BMJ Open

Validation study: evaluation of the metrological quality of French hospital data for perinatal algorithms

Karine Goueslard,[1,2,3,4] Jonathan Cottenet,[1,2,3,4] Eric Benzenine,[1,2,3,4] Pascale Tubert-Bitter,[5] Catherine Quantin [6]

¹Biostatistics and Bioinformatics (DIM), Inserm, CIC 1432, Dijon, France
²University Hospital, Dijon, France
³Bourgogne Franche-Comté University, Dijon, France
⁴Centre Hospitalier Universitaire de Dijon, Dijon, Bourgogne, France
⁵Biostatistics, Biomathematics, Pharmacoepidemiology and Infectious Diseases (B2PHI), Inserm, UVSQ, Institut Pasteur, Université Paris-Saclay, INSERM, Villejuif, Île-de-France, France
⁶Service de Biostatistique et Informatique Médicale, Centre Hospitalier Universitaire, Dijon, France

Correspondence to
Dr Catherine Quantin;
catherine.quantin@chu-dijon.fr

ABSTRACT

Objective The aim of our validation study was to assess the metrological quality of hospital data for perinatal algorithms on a national level.

Design Validation study.

Setting This was a multicentre study of the French medicoadministrative database on perinatal indicators.

Participants In each hospital, we selected 150 discharge abstracts for delivery (after 22 weeks of gestation), in 2014, and their corresponding medical records. Overall, 22 hospitals were included.

Interventions A single investigator performed blind data collection from medical records in order to compare data from discharge abstracts with data from medical records. Finally, 3246 discharge abstracts were studied.

Primary and secondary outcome measures Seventy items, including maternal and delivery characteristics and maternal morbidity, were collected for each delivery stay.

Results The concordance rate of maternal age at delivery was 94.8% (95% CI 93.8 to 95.4). Combining the two forms of pre-existing diabetes, the algorithm presented a PPV of 65.9% and a sensitivity of 75.7%. The concordance rate of gestational age at delivery was 91.8% (90.9 to 92.7). Regarding gestational diabetes, the PPV was 80.8% (79.4 to 82.2) and the sensitivity was 79.5% (78.1 to 80.9). Regardless of the algorithm explored, the PPV for vaginal delivery was over 99%. For the diagnosis codes corresponding to immediate postpartum haemorrhage, the PPV was 77.7% (76.3 to 79.1) and the sensitivity was 75.5% (74.0 to 77.0). The algorithm for stillbirth presented a PPV of 89.4% (88.3 to 90.5) and a sensitivity of 95.4% (94.7 to 96.1).

Conclusions This first national validation study of many perinatal algorithms suggests that the French national hospital database is an appropriate data source for epidemiological studies, except for some indicators which presented low PPV and/or sensitivity.

Strengths and limitations of this study

► The study was conducted on individual data, which made it possible to assess positive predictive value and sensitivity.
► The study included all types of maternity units and all volumes of deliveries.
► The data collection was performed by a single technician.
► There was heterogeneous distribution of maternity units throughout the French territory.
► Data collection was not duplicated by a health professional, and there was no dual data collection by a health professional.

may reflect the health status of individual subjects. In France, the medical information system (Programme de Médicalisation des Systèmes d'Information) covers data relative to all public and private hospitals throughout the territory. These data are particularly interesting for the investigation of perinatal health, seeing as 99.6% of the 800 000 annual births in France take place in hospitals.[7] Even though national perinatal surveys are currently used to follow up on numerous indicators every 5 years, hospital database is also a suitable epidemiological tool because it makes information available on a yearly basis.[8 9] The use of routinely collected health data saves time and money when identifying infrequent and unfavourable delivery outcomes and improves health surveillance of women and their offspring.[10]

Because hospital data were originally collected for administrative or financial purposes, the researchers using these databases are required to ensure data quality using either a single variable or a case-finding algorithm based on several variables.[10–14] It then becomes an evaluation of data reliability: a declared case has to be a confirmed case. A validation study aims to evaluate the gap between the information

INTRODUCTION

Subsequent to the digitisation of hospital data, a great deal of epidemiological data about the hospitalised population have become available; these have been used de facto for scientific research for more than 20 years.[1–6] In this context, health data have focused on identifying morbidity using diagnosis and/or procedure data that

found in the medicoadministrative database and the information found with a gold standard approach. This type of study performed from individual data is extremely time-consuming and expensive because the data from medical records require prior national data collection. Nonetheless, this type of validation study needs to be carried out for each database seeing as variations can occur.[15] For instance, a Canadian study validated an identification algorithm for children with diabetes mellitus, but the identification of children with diabetes using the same algorithm was insufficient in a Colombian database.[16 17] The relevant question here is whether the French hospital database can be used to reliably detect perinatal indicators and therefore to inform clinical studies or investigate quality of care.

The aim of our validation study was to assess the metrological quality of hospital data for perinatal algorithms on a national level.

METHODS

The principle of this transversal multicentre study was to compare data from the French hospital database (named 'Programme de Médicalisation des Systèmes d'Information' or PMSI) with data from medical records, which we considered to be the gold standard. The objective of a validation study for a case-finding algorithm is to estimate the validity indices of the algorithm which quantifies to what extent a variable in the data corresponds to the variable in reality.

French hospital database

The French hospital database, which uses diagnosis-related groups (DRG), is a patient classification system with an objective to describe hospital activity according to the resources consumed. In France, the classification of DRG is based on discharge abstracts recorded for each hospitalisation. These abstracts include some sociodemographic characteristics, the principal and associated diagnoses (coded according to the 10th revision of the International Statistical Classification of Diseases and Related Health Problems), and procedures performed during the stay (coded according to the French Common Classification of Medical Acts) for each inpatient. Several internal and external controls are periodically performed. Although the purpose of this database is the payment of clinical hospital activities, this database has already been the subject of numerous studies in various medical disciplines, notably in a perinatal setting.[7 18–23]

Population

We performed a two-stage sample design. First, 50 health hospitals with a maternity unit were randomly selected in metropolitan France (other than Paris and Paris region), irrespective of the level of the unit. In France, maternity units are classified into levels according to the care they are able to provide (online supplementary

file 1). Two hospitals that had participated in a previous pilot study were automatically included. The heads of the maternity units were contacted by an email where we presented the study and asked if the unit would be interested in taking part. We followed up by telephone hospitals that did not answer.

Second, for each included hospital, 150 delivery discharge abstracts (≥22 weeks of gestation) were randomly selected from all the discharge abstracts from 2014 that contained a Z37 code (the outcome of delivery on the mother's record) and/or a delivery procedure, according to the French Common Classification. We chose to include 150 records per hospital to reach the minimum sample size required,[24] based on the different values of the prevalence of a disease and both the sensitivity and specificity of a screening procedure or diagnostic test. Except for one simulation out of 40, we reached the minimum sample size required for the diagnostic tests.

Particular interest was paid to diabetes mellitus, severe postpartum haemorrhage (PPH), stillbirth and termination of pregnancy for medical reasons. These conditions are relatively rare and could be under-represented in a simple random sample selection. We therefore used the quota method.

The selected discharge abstracts were distributed as follows:
► 80 discharge abstracts selected from all deliveries, whatever the pregnancy outcome.
► 10 discharge abstracts selected from all deliveries, including a diagnosis code for pre-existing type 2 or type 1 diabetes mellitus, in pregnancy, childbirth and the puerperium.
► 10 discharge abstracts selected from all deliveries, including a diagnosis code for gestational diabetes mellitus (GDM).
► 20 discharge abstracts selected from all deliveries, including a diagnosis code for severe PPH.
► 20 discharge abstracts selected from all deliveries, including a diagnosis code for stillbirth.
► 10 discharge abstracts selected from all deliveries, including a diagnosis code for a termination of pregnancy for medical reasons.

For the last two groups, all discharge abstracts were included if the number of cases was not reached in 2014 for a given hospital. To maintain the number of 150 for this hospital, an additional random draw was performed among discharge abstracts selected from all deliveries.

We developed a software program to randomly select 150 discharge abstracts that included a delivery per hospital. Each hospital's medical information department ran this program in its hospital database to extract the data for these 150 stays for delivery and the 150 corresponding medical records.

A list of 25 additional discharge abstracts (12 all deliveries, 3 diabetes mellitus, 3 GDM, 3 severe PPH, 2 stillbirths and 2 terminated pregnancy for medical reasons) was prepared to compensate for inaccessible records if

necessary, thus maintaining the number of 150 discharge abstracts per hospital.

Data collection

A single clinical research associate collected data in each hospital from the hospital perinatal medical records of women who delivered in 2014. These medical records were linked to the previously selected discharge abstracts. The clinical research associate was trained to collect and manage data and was blinded to the data recorded on the discharge abstracts. Data collection took place between September 2016 and December 2017.

The medical record was made up of an electronic or paper document that retraced prenatal care, the delivery and the postdelivery stay, a report of the procedure, and a discharge letter. The data from each record were collected on a standardised form in accordance with national predefined guidelines regarding the collection of hospital data. Guidelines for filling in information and the opportunity to discuss the disputed cases with a supervisor with expertise in obstetrics allowed us to maximise data homogeneity.

The variables studied corresponded to the characteristics of the hospital stay, the pregnancy, the delivery and the newborn: the mode of admission and discharge, the length of stay, the maternal age at delivery, the existence of maternal obesity (body mass index (BMI) $>30\,\mathrm{kg/m^2}$), the weight and gestational age of the newborn, and the parity for vaginal deliveries (accounted after delivery). Maternal diseases included diabetes before the pregnancy, gestational diabetes, hypertensive disorders, premature rupture of membranes (PROM) and premature labour. The characteristics of the labour and delivery included the type of pregnancy (singleton or multiple), the type of presentation, the mode of delivery (spontaneous vaginal, instrumental extraction, caesarean: emergency or not), and PPH, stillbirth and transfer in utero.

Statistical analysis

We compared the data from medical records and the data from discharge abstracts, at an individual level. Several algorithms were explored, including different combinations of codes in the discharge abstracts from delivery stays, pregnancy stays or hospitalisation in the 2 years before delivery (online supplementary file 2).

Means or proportions were calculated for each source of data. To evaluate the metrological quality of the hospital database, the various indicators were calculated for each variable. The medical record was considered the gold standard. Positive predictive value (PPV) and sensitivity were calculated for dichotomous data. PPV corresponded to the probability that the variables recorded in the discharge abstracts were also present in the medical record. Sensitivity corresponded to the probability that variables recorded in the medical record were also present in the discharge abstracts.

Continuous data were assessed by the concordance rate, which corresponded to the number of concordant cases

between the discharge abstracts and the medical records (ie, if the same value was identified in both data sources) over the total number of records examined. However, the variability of a quantitative measure has two main sources: the method itself (analytical variability) and the individual (intersubject or intrasubject variability). We used the Deming regression[25–27] to take these variabilities into account. If the two measures are estimated on the same scale, then the methods are well calibrated when the 95% CI for the intercept includes 0 and the CI for the slope includes 1. Concerning qualitative variables, in order to estimate the concordance between data of discharge abstracts and medical records, we calculated the kappa index,[28] and the interpretation was made using a commonly cited scale.[29] This index is considered as good from 60% (substantial agreement) and excellent from 80% (almost perfect agreement). The rates of false negative and false positive were also calculated in order to select the best algorithms with regard to the likelihood ratio (balancing specificity and sensitivity).

Patient and public involvement

No patients involved.

RESULTS

Twenty-three hospitals agreed to take part in the study (authorised by the head obstetrician and the hospital director): four level 1 maternities, seven level 2 maternities and twelve level 3 maternities.

The distribution of the hospitals was unequal over the territory, but the four major geographical areas were represented.

One hospital did not provide the data from discharge abstracts obtained from the random draw, so only 22 hospitals were finally included.

Fifty-four discharge abstracts could not be linked with the corresponding medical records. For three of them, the data collection was missing in the medical records. Fifty-one discharge abstracts could not be matched with the medical records because a mistake was made in the joint patient identifier for both data sources by the corresponding hospital.

In total, 3246 discharge abstracts were compared with their corresponding medical records.

Maternal indicators

The concordance rate of maternal age at delivery was 94.8% (95% CI 94.0 to 95.6). The concordance rate for postal codes was 97.0% (96.4 to 97.6) and the concordance rate for departments was 100%. The concordance using Deming regression between data from discharge abstracts and data from medical records is presented in figure 1. The data include maternal age at delivery, gestational age, childbirth weight and maternal departments of residence. For all these variables the concordance was almost perfect as the 95% CI for the intercept includes 0 and the CI for the slope includes 1.
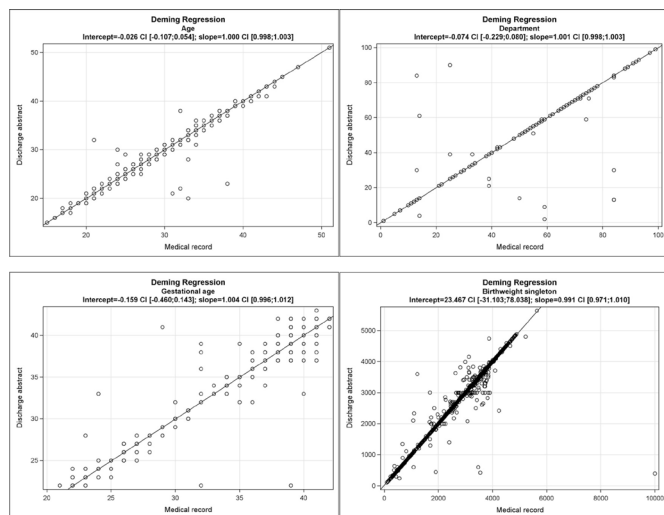
Figure 1 Concordance using Deming regression between discharge abstracts and medical records for maternal age at delivery, gestational age, childbirth weight and maternal departments of residence.

The maternal characteristics are presented in table 1. The maternal weight and/or height were not available in 309 medical records, making it impossible to calculate BMI. The diagnosis code for obesity ($\geq 30 \, \text{kg/m}^2$; E66 codes, except E66.03, E66.13, E66.83 and E66.93) presented a PPV of 91.7% (95% CI 90.7 to 92.7) and a sensitivity of 39.1% (37.3 to 40.9) (kappa index=0.50, moderate agreement).

We explored parity, which was taken into account after vaginal delivery: the PPV and sensitivity of primiparous women were, respectively, 93.3% (92.3 to 94.3) and 92.4% (91.3 to 93.5). For multiparous women, the PPV and sensitivity were of the same order 95.5% (94.7 to 96.3). The kappa index for parity was greater than 0.80 (almost perfect agreement).

Regarding uterine scars, the PPV was 94.8% (94.0 to 95.6) and the sensitivity was 78.5% (77.1 to 79.9) (kappa index=0.84, almost perfect agreement).

Maternal morbidity
Several types of morbidity were explored (table 1).

Diabetes mellitus
First, we focused on pre-existing diabetes mellitus. For type 1 diabetes, the two best algorithms presented a PPV of 50.4% (48.7 to 52.1) and a sensitivity of 73.5% (72.0 to 75.0) (kappa index=0.58, moderate agreement) and were defined as follows:
► Code O24.0 recorded in discharge abstracts established for delivery stay.
► Code O24.0 recorded in discharge abstracts established for delivery stay and/or code of type 1 diabetes mellitus (E10) recorded in discharge abstracts for hospital stay during pregnancy.

For type 2 diabetes, the best algorithm only included the O24.1 code recorded during pregnancy stay or delivery stay, and it presented a PPV of 67.7% (66.1 to 69.3) and a sensitivity of 41.2% (39.5 to 42.9) (kappa index=0.51, moderate agreement).

We explored other algorithms which included the E10 code (or E11 for type 2 diabetes) recorded on pregnancy

Table 1 Metrological quality of discharge abstracts for maternal characteristics and comorbidities

	Medical records		Discharge abstracts		PPV		FP		FN		Sensitivity		Kappa index
	n	%	n	%	%	95%CI	n	%	n	%	%	95%CI	
Parity													
Primiparous women	981	40.2	971	39.8	93.3	92.3 to 94.3	65	2.7	75	3.1	92.4	91.3 to 93.5	0.88 (APA)
Multiparous women	1459	59.8	1458	59.8	95.5	94.7 to 96.3	66	2.7	67	2.7	95.5	94.7 to 96.3	0.89 (APA)
Overweight or obesity $\geq 25 \, \text{kg/m}^2$*	1104	37.6	220	7.5	98.6	98.2 to 99.0	3	0.1	887	30.2	19.7	18.3 to 21.1	0.23 (FA)
Obesity $\geq 30 \, \text{kg/m}^2$*	507	17.3	216	7.4	91.7	90.7 to 92.7	18	0.6	309	10.5	39.1	37.3 to 40.9	0.50 (MA)
Uterus scar	464	14.4	384	11.9	94.8	94.0 to 95.6	20	0.6	100	3.1	78.5	71.1 to 79.9	0.84 (APA)
Diabetes mellitus													
Type 1 diabetes†	98	3.0	143	4.4	50.4	48.7 to 52.1	71	2.2	26	0.8	73.5	72.0 to 75.0	0.58 (MA)
Type 2 diabetes†	51	1.6	31	1.0	67.7	66.1 to 69.3	10	0.3	30	0.9	41.2	39.5 to 42.9	0.51 (MA)
Type 1 or type 2 diabetes	148	4.6	170	5.2	65.9	64.3 to 67.5	58	1.8	36	1.1	75.7	74.2 to 77.2	0.69 (SA)
High blood pressure	29	0.9	34	1.0	32.4	30.8 to 34.0	23	0.7	18	0.6	37.9	36.2 to 39.6	0.34 (FA)

*Missing data=309.
†Missing data <5.
APA, almost perfect agreement; FA, fair agreement; FN, false negative ; FP, false positive; MA, moderate agreement; PPV, positive predictive value; SA, substantial agreement.

Table 2 Metrological quality of discharge abstracts for pregnancy-related disorders

	Medical records		Discharge abstracts		PPV		FP		FN		Sensitivity		Kappa index
	n	%	n	%	%	95% CI	n	%	n	%	%	95% CI	
Gestational diabetes													
Code O24.4	482	14.9	469	14.5	80.6	79.2 to 82.0	91	2.8	104	3.2	78.4	77.0 to 79.8	0.76 (SA)
Codes O24.4–O24.9	482	14.9	474	14.6	80.8	79.4 to 82.2	91	2.8	99	3.1	79.5	78.1 to 80.9	0.77 (SA)
Hypertensive disorders													
Previous or during pregnancy	213	6.6	239	7.4	69.5	67.9 to 71.1	73	2.3	47	1.4	77.9	76.5 to 79.3	0.71 (SA)
Moderate or severe pre-eclampsia	96	3.0	106	3.3	70.8	69.2 to 72.4	31	1.0	21	0.6	78.1	76.7 to 79.5	0.73 (SA)
Eclampsia	2	0.1	8	0.2	12.5	11.4 to 13.6	7	0.2	1	0.0	50.0	48.3 to 51.7	0.20 (SIA)
Premature labour* (O60.0)	141	4.3	87	2.7	57.5	55.8 to 59.2	37	1.1	91	2.8	35.5	33.9 to 37.1	0.42 (MA)
Premature labour*†	141	4.3	175	5.4	48.0	46.3 to 49.7	91	2.8	57	1.8	59.6	57.9 to 61.3	0.51 (MA)
Premature delivery	434	13.4	295	9.1	85.8	84.6 to 87.0	42	1.3	181	5.6	58.3	56.6 to 60.0	0.66 (SA)
Placental abruption	41	1.3	40	1.2	75	73.5 to 76.5	10	0.3	11	0.3	73.2	71.7 to 74.7	0.74 (SA)

*Missing data ≤10.
†Codes O60.0, O60.2 and O47.0.
FN, false negative; FP, false positive; MA, moderate agreement; PPV, positive predictive value; SA, substantial agreement; SIA, slight agreement.

hospitalisation and/or on hospitalisation within 2 years prior to the delivery, but without improvement.

Combining the two forms of pre-existing diabetes, the algorithm which mixed the codes O24.0 or O24.1 in the discharge abstract of delivery stay presented a PPV of 65.9% and a sensitivity of 75.7% (kappa index=0.69, substantial agreement).

High blood pressure
Previous high blood pressure was explored with the O10 diagnosis code. The PPV was 32.4% and the sensitivity was 37.9%, with a fair agreement regarding the kappa index (k=0.34).

Pregnancy-related disorders
The concordance rate of gestational age at delivery was 91.8% (90.9 to 92.7). However, the gestational age was specified in weeks and days of gestation in medical records, while it was specified in weeks of gestation in the hospital data. Rounding up or down to the nearest whole number of weeks of gestation, the concordance rate increased to 98.3%.

The pregnancy-related disorders are presented in table 2.

Gestational diabetes
Regarding gestational diabetes (O24.4 and O24.9 codes in the discharge abstract of delivery stay), the PPV was 80.8% (79.4 to 82.2) and the sensitivity was 79.5% (78.1 to 80.9). The association of these two codes (in comparison with only code O24.4) did not modify the number of false positives but decreased the number of false negatives. The algorithms that included the discharge abstracts from pregnancy hospitalisations decreased the PPV and

increased the sensitivity. The kappa index for gestational diabetes was almost 0.80 (k=0.76, substantial agreement).

Hypertensive disorders
There were 213 cases of hypertensive disorders in the medical records and 239 in the discharge abstracts from the delivery stays (codes O10-O16). The PPV was 69.5% (67.9 to 71.1) and the sensitivity was 77.9% (76.5 to 79.3). When we added the same codes but that were recorded during pregnancy or the P000 code which is recorded in the discharge abstract from a newborn stay, the results were similar. A kappa index of 0.71 indicates a substantial agreement between the two data sources.

Moderate or severe pre-eclampsia was identified in 106 cases in the hospital data (code O14 in the discharge abstracts for delivery stays), while 96 cases were recorded in the medical records; the PPV was 70.8% (69.2 to 72.4) and the sensitivity was 78.1% (76.7 to 79.5) (kappa index=0.73, substantial agreement). Eclampsia is a major and rare event. Nevertheless, we observed many false positives: two cases were recorded in the medical records and eight cases in the hospital data. Thus, the PPV was 12.5% and the sensitivity was 50.0%, with a slight agreement regarding the kappa index (k=0.20).

Premature labour, premature delivery
Premature labour is the motive for a large number of hospitalisations in France. According to the guidelines from the national agency for the management of hospitalisation data (Agence Technique de l'Information Hospitalière (ATIH)), the O60.0 code corresponds to a premature labour. For this code, recorded in discharge abstracts of hospital stays during pregnancy, the PPV was

57.5% (55.8 to 59.2) and the sensitivity was 35.5% (33.9 to 37.1). Adding other codes (eg, O60.0, O60.2 or O47.0, 'false labor before 37 completed weeks of gestation'), the PPV decreased (32.5%) and the sensitivity increased (75.2%). The kappa index for premature labour depending on the algorithm used was, respectively, 0.42 and 0.51 (moderate agreement).

Regarding premature deliveries with or without spontaneous labour (codes O60.1 and O60.3), the PPV was 85.8% and the sensitivity was 58.3% (kappa index=0.66, substantial agreement).

Premature rupture of membranes

Regarding the codes for PROM (O42) or the code for delayed delivery after spontaneous or unspecified rupture of membranes (O75.6) during the delivery stay, the PPV was 54.9% (53.2 to 56.6) and the sensitivity was 57.0% (55.3 to 58.7). When we added to the O42 code the P011 code from the newborn discharge abstract, the PPV was 56.2% (54.5 to 57.9) and the sensitivity was 60.0% (58.3 to 61.7). The kappa index was around 0.50 for both algorithms, indicating a moderate agreement between the two data sources.

For *placental abruption* (code O45), the PPV was 75% (73.5 to 76.5) and the sensitivity was 73.2% (71.7 to 74.7) (kappa index=0.74, substantial agreement).

Delivery

The results of delivery algorithms are presented in table 3. The data for delivery in singleton or twin pregnancies presented, respectively, PPV of 98.3% (97.9 to 98.7) and 94.3% (93.4 to 95.1), sensitivity of 99.7% and 95.2%, and kappa index of 0.64 (substantial agreement) and 0.95 (almost perfect agreement).

Regardless of the algorithm explored, the PPV for vaginal delivery was over 99%. However, the sensitivity increased to 99.6% when we used the algorithm that included diagnosis codes and/or the corresponding codes for delivery procedures. For instrumental deliveries, the PPV decreased by 2% (88.2% (87.1 to 89.3)) and the sensitivity increased to 95.9% when the algorithm included diagnosis codes and/or procedure codes.

As regards caesarean births, the algorithm that included diagnosis codes and/or procedure codes improved the PPV to 98.5% and the sensitivity to 99.5%. The PPV for emergency caesarean sections or planned caesarean sections (diagnosis and procedure codes) was, respectively, 83.0% and 80.4%, and their sensitivity was 95.1% and 84.5%.

For all delivery types, almost perfect agreement was obtained with the kappa index.

The PPV for episiotomy, which was calculated by comparing medical records and the procedure codes, was 90.1% (88.9 to 91.3) and the sensitivity was 86.0% (84.6 to 87.4) (kappa index=0.86, almost perfect agreement). The PPV for perineal tears was 86.2% (84.8 to 87.6) and the sensitivity was 76.4% (74.7 to 78.1) (kappa index=0.65, substantial agreement).

For the diagnosis codes corresponding to immediate PPH (O72.0 or O72.1), the PPV was 77.7% (76.3 to 79.1) and the sensitivity was 75.5% (74.0 to 77.0) (kappa index=0.74, substantial agreement). We explored an algorithm that included these codes and the procedure codes for manual removal of the placenta. In these cases, the PPV was 80.7% and the sensitivity was 13.1%, with a fair agreement regarding the kappa index (k=0.21).

In order to select severe PPH, we explored advanced interventional procedures which indicated a second-line therapy (arterial embolisation, uterine or hypogastric artery ligation, haemostasis hysterectomy). First, when we included the relevant codes of advanced interventional procedures on immediate postpartum, the PPV was 67.8% and the sensitivity was 80.8%. We then explored the performance of an algorithm that included the following:
▶ Relevant codes for advanced interventional procedures during the immediate postpartum period.
▶ Diagnosis codes corresponding to immediate PPH associated with general codes for advanced interventional procedures.

The PPV was 68.4% (66.8 to 70.0) and the sensitivity was 90.0% (89.0 to 91.0).

For both algorithms, we found a substantial agreement with the kappa index (k=0.73 and k=0.77, respectively).

Fetal mortality

The results of fetal mortality algorithms are presented in table 3. As regards medically indicated abortion, we wanted to explore the algorithms proposed by the national agency ATIH for hospitalisation. First, we explored medical abortion from one delivery procedure code and a gestational age greater than or equal to 22 weeks of gestation (WG) and one of the codes specifying stillbirth (Z37.11, Z37.31, Z37.41, Z37.61, Z37.71). The PPV and the sensitivity were 91.5% (90.5 to 92.5) (kappa index=0.91, almost perfect agreement). The other algorithms (only newborn discharge abstract, or the association of maternal and newborn discharge abstracts) were not more powerful.

Regarding medical abortion linked to a fetal condition, the same algorithm was used, but the diagnosis code O35 (as the primary diagnosis) was added. The PPV was equal to 88.9% (87.8 to 90.0) and the sensitivity was equal to 75.9% (74.4 to 77.4) (kappa index=0.81, almost perfect agreement).

According to the ATIH algorithm, among the codes specifying stillbirth (Z37.1, Z37.3, Z37.4, Z37.6, Z37.7 or O36.4 or O31.2), stillbirth presented a PPV of 89.4% (88.3 to 90.5) and a sensitivity of 95.4% (94.7 to 96.1) (kappa index=0.92, almost perfect agreement).

When the P95 code was added, the sensitivity increased to 99.2% but the PPV dropped to 60.6%, with a substantial agreement for the kappa index nonetheless (k=0.73).

The concordance rate between vital status and the diagnosis codes for medical abortion from newborn discharge abstracts was 98.6% (98.2 to 99.0) for singleton

Table 3 Metrological quality of discharge abstracts for delivery Sensitivity

	Medical records		Discharge abstracts		PPV		FP		FN		Sensitivity		Kappa index
	n	%	n	%	%	95% CI	n	%	n	%	%	95% CI	
Type of pregnancy													
Singleton	3134	96.5	3179	97.9	98.3	97.9 to 98.7	54	1.7	9.0	0.3	99.7	99.5 to 99.9	0.64 (SA)
Twin	105	3.2	106	3.3	94.3	93.5 to 95.1	6	0.2	5.0	0.2	95.2	94.5 to 95.9	0.95 (APA)
Triple	5	0.2	8	0.2	50.0	48.3 to 51.7	4	0.1	1.0	0.0	80.0	78.6 to 81.4	0.61 (SA)
Delivery													
Vaginal	2508	77.3	2510	77.3	99.5	99.3 to 99.7	13	0.4	11	0.3	99.6	99.4 to 99.8	0.98 (APA)
Operative	343	10.6	373	11.5	88.2	87.1 to 89.3	44	1.4	14	0.4	95.9	95.2 to 96.6	0.91 (APA)
Caesarean	731	22.5	738	22.7	98.5	98.1 to 98.9	11	0.3	4	0.1	99.5	99.3 to 99.7	0.99 (APA)
Emergency caesarean	450	13.9	516	15.9	83	81.7 to 84.3	88	2.7	22	0.7	95.1	94.4 to 95.8	0.87 (APA)
Planned caesarean	238	7.3	250	7.7	80.4	79.0 to 81.8	49	1.5	37	1.1	84.5	83.3 to 85.7	0.81 (APA)
Episiotomy*	350	14.0	334	13.4	90.1	88.9 to 91.3	33	1.3	49	2.0	86.0	84.6 to 87.4	0.86 (APA)
Perineal tears*	1231	49.3	1092	43.7	86.2	84.8 to 87.6	151	6.0	290	11.6	76.4	74.7 to 78.1	0.65 (SA)
PPH*													
Immediate (O72.0, O72.1)	286	8.8	278	8.6	77.7	76.3 to 79.1	62	1.9	70	2.2	75.5	74.0 to 77.0	0.74 (SA)
Diagnosis codes + manual removal of the placenta*	191	5.9	31	1.0	80.7	79.3 to 82.1	6	0.2	166	5.1	13.1	11.9 to 14.3	0.21 (FA)
Severe PPH													
Relevant advanced interventional procedures	120	3.7	143	4.4	67.8	66.2 to 69.4	46	1.4	23	0.7	80.8	79.4 to 82.2	0.73 (SA)
Relevant or general advanced interventional procedures	120	3.7	158	4.9	68.4	66.8 to 70.0	50	1.5	12	0.4	90.0	89.0 to 91.0	0.77 (SA)
Medical abortion	153	4.7	153	4.7	91.5	90.5 to 92.5	13	0.4	13	0.4	91.5	90.5 to 92.5	0.91 (APA)
Medical abortion, fetal pathology	137	4.2	117	3.6	88.9	87.8 to 90.0	13	0.4	33	1.0	75.9	74.4 to 77.4	0.81 (APA)
Stillbirth													
Relevant Z37 codes† or O36.4 or O31.2	239	7.4	255	7.9	89.4	88.3 to 90.5	27	0.8	11	0.3	95.4	94.7 to 96.1	0.92 (APA)
Relevant Z37 codes + P95	239	7.4	391	12.0	60.6	58.9 to 62.3	154	4.7	2	0.1	99.2	98.9 to 99.5	0.73 (SA)
Transfer in utero	61	1.9	30	0.9	56.7	55.0 to 58.4	13	0.4	44	1.4	27.9	26.4 to 29.4	0.37 (FA)

*Missing data ≤10.
†Z37.1, Z37.3, Z37.4, Z37.6 and Z37.7.
APA, almost perfect agreement; FA, fair agreement; FN, false negative; FP, false positive; PPH, postpartum haemorrhage; PPV, positive predictive value; SA, substantial agreement .

pregnancy. In case of multiple pregnancies, the rate was 100% for all children.

The concordance rate between vital status and the diagnosis codes for stillbirth from newborn discharge abstracts was 95.4% (94.7 to 96.1) for singleton pregnancy. For multiple pregnancies, the rate was 99.7% for the first-born or the second-born child, and 100% for the third-born child.

Newborn indicators

The concordance rate of newborn weight was 91.3% (90.3 to 92.3) in singleton pregnancy. The rate was 79.1% (70.5 to 87.7) for first-born and second-born in cases of multiple pregnancy. The concordance using Deming regression is presented in figure 1. As regards the first-born child, the median gap between the newborn weight mentioned in the medical record and the weight specified in the discharge abstract was 100 g. The same gap was estimated for the first-born or second-born child in case of multiple pregnancies.

Transfer in utero

Sixty-one cases of in utero transfer were identified in the medical records, while only 30 cases were identified in the discharge abstracts. The PPV was 56.7% and the sensitivity was 27.9%, with a fair agreement for the kappa index (k=0.37).

DISCUSSION
Main findings

To our knowledge, this is the first time a national validation study has been done for perinatal algorithms from the French national hospital database. The frequency of data observed in discharge abstracts was sometimes very close to those observed in medical records, particularly for maternal characteristics, pregnancy characteristics (parity, type of pregnancy, type of delivery, stillbirth, termination of pregnancy for medical reasons) and child birth weight. However, we found the algorithms for pregnancy disorders to be insufficient.

The results of our study may allow researchers to target the best performing algorithms. For example, the best case-finding algorithm of premature deliveries presented substantial agreement, while the concordance rate of gestational age at delivery was almost perfect. It appears that gestational age associated with the onset of delivery should be taken into account when exploring prematurity.

It is very important to adjust the design of a study to the quality of the available data. The case-finding algorithms which present lower PPV and/or sensitivity can be used for descriptive purposes if the total number of women is close to what is expected. Regarding premature labour, the concordance between the hospital database and medical records was poor according to the diagnosis code required by the technical agency for information on hospital care (ATIH) that managed the hospital data (O60.0). We explored other diagnosis codes recorded at the end of the hospitalisation for delivery or at the end of at least one hospitalisation during the pregnancy, which improved concordance slightly. This result may be explained by several factors: there is no consensus on the clinical definition, and the definition of diagnosis codes is subject to interpretation. However, the best performing algorithm estimated the prevalence of premature labour to 5.4%. In France, a national perinatal survey is conducted at regular intervals on a representative sample of births. In 2016, this survey estimated that hospitalisation for premature labour occurred in 5.4% of women. These results were close to those of the national perinatal survey conducted in 2010.[9] However, our sample design could significantly increase this rate.

For longitudinal epidemiological studies, caution is required in case of very low sensitivity and/or PPV. For these studies, it is better to use case-finding algorithms that provide substantial or perfect agreement. In a pilot study from 2012 based on 20 cases of gestational diabetes from 300 medical records, our team found a PPV of 88.9% (74.3–100) and a sensitivity of 72.7% (54.1–91.3). The current study found similar results despite a slight improvement in sensitivity. Unfortunately, we are not able to compare the frequency of gestational diabetes with other studies in France because of the artificially increased prevalence in our sample (14.5% in our study vs 10.8% in the 2016 national perinatal survey). The use of a complementary French database for ambulatory care (reimbursement of treatment, biological testing and medical devices) could improve the identification of women with GDM, but the metrological quality of this algorithm has not yet been assessed. In case of stillbirth or termination of pregnancy for medical reasons, it seems possible to study these two indicators not only for descriptive studies, but also for longitudinal studies. Similar to what is already done for maternal mortality,[30] the French hospital database could monitor these two additional indicators.

On the contrary, it would be unwise in the current situation to use the French hospital database to undertake longitudinal studies on certain pregnancy disorders because of their low PPV and sensitivity. Several factors may explain low predictive value or sensitivity seen in these disorders. First, the original and main objective of this standardised database is hospital financing, and a number of procedure or diagnosis codes are not allocated to medical fees. Certain diagnoses that are not financially valuated may remain uncoded.

Furthermore, we explored the identification of severe PPH via obstetrical procedures. One of the two procedures studied (uterine check post natural placental delivery) concerns only a very small proportion of this type of procedures. In most cases, placental delivery occurs after intravenous oxytocin injection (84%), so the placental delivery can no longer considered natural.[9] The procedure code no longer corresponds to clinical reality in these cases, and improvements are needed in terms of information quality.

Strengths and limitations

We conducted a validation study comparing individual data from the hospital discharge abstract with medical records. Some limitations have to be acknowledged. The data collection was performed by a perfectly trained clinical research associate but not by a medical specialist, such as an obstetrician or a midwife. In addition, the maternity units included in our study were not distributed equally throughout the French territory. The results of our study investigated national data and are therefore useful for studies carried out on a national scale, but regional or local studies would need a local evaluation as a result of interhospital variability.

Our study also has some strengths. First, in France almost all deliveries occur in a hospital. Second, we explored a large number of perinatal indicators, whereas many studies focus on a single condition.[22 31–33] Overall, 70 items were collected for each hospital stay by a single clinical research associate. Third, despite the challenges of recruiting centres that would have to dedicate considerable time to the study, our national study included maternity units of all types and with all volumes of deliveries. Finally, we explored more than 3200 pairs of medical records–discharge abstracts.

CONCLUSION

This first national validation study of a large set of perinatal algorithms provides valuable information for researchers about the quality of the French national hospital database. For certain case-finding algorithms, our results suggested that this database may be an appropriate data source for epidemiological studies. For others (with low PPV/sensitivity), we would discourage longitudinal studies. In recent years, a professionalisation of the staff in charge of health information coding may be an important lever for improving the quality of information. Moreover, the joint association of the French hospital database and 'National System of Health Data' (SNDS - *Système National des Données de Santé*) database, which collects individual non-hospital healthcare data, may improve case identification of some conditions that impact the course of pregnancy.

ORCID iD
Catherine Quantin http://orcid.org/0000-0001-5134-9411

REFERENCES

1. Quantin C, Benzenine E, Fassa M, *et al*. Evaluation of the interest of using discharge Abstract databases to estimate breast cancer incidence in two French departments. *Stat J IAOS* 2012;28:73–85.
2. Remontet L, Mitton N, Couris CM, *et al*. Is it possible to estimate the incidence of breast cancer from medico-administrative databases? *Eur J Epidemiol* 2008;23:681–8.
3. Quantin C, Benzenine E, Hägi M, *et al*. Estimation of national colorectal-cancer incidence using claims databases. *J Cancer Epidemiol* 2012;2012:1–7.
4. Abdulmalak C, Cottenet J, Beltramo G, *et al*. Haemoptysis in adults: a 5-year study using the French nationwide Hospital administrative database. *Eur Respir J* 2015;46:503–11.
5. Couris CM, Polazzi S, Olive F, *et al*. Breast cancer incidence using administrative data: correction with sensitivity and specificity. *J Clin Epidemiol* 2009;62:660–6.
6. Lorgis L, Cottenet J, Molins G, *et al*. Outcomes after acute myocardial infarction in HIV-infected patients: analysis of data from a French nationwide hospital medical information database. *Circulation* 2013;127:1767–74.
7. Quantin C, Cottenet J, Vuagnat A, *et al*. [Quality of perinatal statistics from hospital discharge data: comparison with civil registration and the 2010 National Perinatal Survey]. *J Gynecol Obstet Biol Reprod* 2014;43:680–90.
8. Blondel B, Lelong N, Kermarrec M, *et al*. [Trends in perinatal health in France between 1995 and 2010: Results from the National Perinatal Surveys]. *J Gynecol Obstet Biol Reprod* 2012;41:151–66.
9. Blondel B, Coulm B, Bonnet C, *et al*. Trends in perinatal health in metropolitan France from 1995 to 2016: results from the French national perinatal surveys. *J Gynecol Obstet Hum Reprod* 2017;46:701–13.
10. Garrison LP, Neumann PJ, Erickson P, *et al*. Using real-world data for coverage and payment decisions: the ISPOR real-world data Task force report. *Value Health* 2007;10:326–35.
11. James S, Rao SV, Granger CB. Registry-based randomized clinical trials--a new clinical trial paradigm. *Nat Rev Cardiol* 2015;12:312–6.
12. Bossuyt PM, Reitsma JB, Bruns DE, *et al*. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;326:41–4.
13. Benchimol EI, Manuel DG, To T, *et al*. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2011;64:821–9.
14. De Coster C, Quan H, Finlayson A, *et al*. Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: report from an international Consortium. *BMC Health Serv Res* 2006;6:77.
15. Gini R. *A Validation Odyssey: from big data to local intelligence. [Internet]*. Rotterdam: Erasmus University, 2016. https://www.

researchgate.net/profile/Rosa_Gini/publication/309428868_
A_Validation_Odyss ey_From_big_data_to_local_intelligence/
links/5810884a08aea04bbcbd03bd/A-Validation- Odyssey-From-
big-data-to-local-intelligence.pdf

16 Amed S, Vanderloo SE, Metzger D, *et al*. Validation of diabetes case definitions using administrative claims data. *Diabet Med* 2011;28:424–7.

17 Guttmann A, Nakhla M, Henderson M, *et al*. Validation of a health administrative data algorithm for assessing the epidemiology of diabetes in Canadian children. *Pediatr Diabetes* 2010;11:122–8.

18 Sagot P, Mourtialon P, Benzenine E, *et al*. Accuracy of blood transfusion in postpartum hemorrhage to assess maternal morbidity. *Eur J Obstet Gynecol Reprod Biol* 2012;162:160–4.

19 Chantry AA, Deneux-Tharaux C, Bal G, *et al*. [French hospital discharge database: data production, validity, and origins of errors in the field of severe maternal morbidity]. *Rev Epidemiol Sante Publique* 2012;60:177–88.

20 Aboa-Eboulé C, Mengue D, Benzenine E, *et al*. How accurate is the reporting of stroke in hospital discharge data? A pilot validation study using a population-based stroke Registry as control. *J Neurol* 2013;260:605–13.

21 Hanf M, Quantin C, Farrington P, *et al*. Validation of the French National health insurance information system as a tool in vaccine safety assessment: application to febrile convulsions after pediatric measles/mumps/rubella immunization. *Vaccine*. 2013;31:5856–62. 2 déc.

22 Fermaut M, Fauconnier A, Brossard A, *et al*. Detection of complicated ectopic pregnancies in the hospital discharge database: a validation study. *PLoS One* 2019;14:e0217674.

23 p B-levy, Duret S, Gilleron V, *et al*. Étude de validation des codes diagnostics de l'insuffiance cardiaque. *Rev epidemiol sante* 2018;66:S199–200.

24 Bujang MA, Adnan TH. Requirements for minimum sample size for sensitivity and specificity analysis. *J Clin Diagn Res* 2016;10:YE01-YE06.

25 Linnet K. Estimation of the linear relationship between the measurements of two methods with proportional errors. *Stat Med* 1990;9:1463–73.

26 Linnet K. Evaluation of regression procedures for methods comparison studies. *Clin Chem* 1993;39:424–32.

27 Fuhrman C, Chouaïd C. [Concordance between two variables: numerical approaches (qualitative observations - the kappa coefficient-; quantitative measures]. *Rev Mal Respir* 2004;21:123–5.

28 Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360–3.

29 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.

30 Saucedo M, Denaux-Tharaud C. Évaluation de la fiabilité du PMSI pour l'identification des morts maternelles intra hospitalières en France. *Rev epidemiol sante* 2019.

31 Flood M, Pollock W, McDonald SJ, *et al*. Accuracy of postpartum haemorrhage data in the 2011 Victorian perinatal data collection: results of a validation study. *Aust N Z J Obstet Gynaecol* 2018;58:210–6.

32 Bowker SL, Savu A, Lam NK, *et al*. Validation of administrative data case definitions for gestational diabetes mellitus. *Diabetic Medicine* 2017;34:51–5.

33 Korst LM, Gregory KD, Gornbein JA. Elective primary caesarean delivery: accuracy of administrative data. *Paediatr Perinat Epidemiol* 2004;18:112–9.