



Agreement assessment of key maternal and newborn data elements between birth registry and Clinical Administrative Hospital Databases in Ontario, Canada

Qun Miao^{1,2,3} · Deshayne B. Fell^{2,3,4} · Sandra Dunn^{1,2,5} · Ann E. Sprague^{1,2,5}

Received: 6 January 2019 / Accepted: 24 April 2019 / Published online: 20 May 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Purpose Since 2012, BORN Ontario, a maternal-newborn registry, has collected data on every birth in Ontario. To ensure data quality, we assessed the reliability of key elements collected in BORN by comparing these with like data elements in the Canadian Institute for Health Information-Discharge Abstract Database (CIHI-DAD).

Methods We used provincial health card numbers to deterministically link live or stillbirth records and their corresponding mothers' records in the BORN database to the CIHI-DAD in the fiscal years 2012–2013 to 2014–2015. Percentage agreement and Cohen Kappa statistics were used to assess agreement on main elements in both databases.

Results The percentage agreement and Kappa coefficients were 99.98% and 0.740 (95% CI: 0.677–0.803) on live/stillbirth, respectively. The Kappa coefficients for infant sex, gestational age at birth, induction of labour, and caesarean birth were 0.989 (95% CI: 0.988–0.989), 0.920 (95% CI: 0.919–0.920), 0.782 (95% CI: 0.780–0.785), and 0.995 (95% CI: 0.995–0.996), respectively. Kappa agreement for the number of fetuses in a pregnancy was 0.979 (95% CI: 0.977–0.981). Percentage agreement was very high for infants' birthdates (99.9%), infant postal codes (91.8%), infants' birth weight in grams (95.5%), and mothers' dates of birth (99.1%).

Conclusions Overall, the BORN and CIHI-DAD databases had concordance on key birth and maternal data elements; however, additional work is needed to understand discrepancies identified.

Keywords The BORN database · The Canadian Institute for Health Information Discharge Abstract Database (CIHI-DAD) · Data quality · Agreement · Kappa test

Introduction

Ontario's Better Outcomes Registry and Network (BORN) was established in 2009 to collect, interpret, share and rigorously protect critical data about pregnancy, birth and early childhood in the province of Ontario, Canada [1, 2]. In 2012, BORN launched its new information system to bring together data holdings across the continuum of care provided during the prenatal, perinatal, and postnatal periods [2]. Hospitals, midwifery practice groups, screening programs, laboratories and clinics across the province contribute data either by direct entry or via electronic data upload from hospital or clinic-specific health records systems [2]. To date, there are over one million infants' records in the registry capturing data on maternal demographics, health behaviours, pre-existing health problems, prenatal screening results, obstetric complications, intrapartum interventions, fetal anomalies, birth outcomes, early neonatal care, and

✉ Qun Miao
GMiao@bornontario.ca

¹ BORN Ontario, Children's Hospital of Eastern Ontario (CHEO), Ottawa, ON, Canada

² Children's Hospital of Eastern Ontario (CHEO) Research Institute, Ottawa, ON, Canada

³ School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada

⁴ Institute for Clinical Evaluative Sciences (ICES), Ottawa, ON, Canada

⁵ School of Nursing, Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada

newborn screening for all hospital, home, and birth centre births [1]. These data are used to provide near real-time feedback to maternity and newborn care provider groups on their clinical performance, to advise funders on maternal-newborn health policy decisions to help facilitate or improve care, to help adjust screening algorithms for better performance, to monitor system-wide performance, and to support population health surveillance and research [1–3].

The Canadian Institute for Health Information (CIHI) is an organization that provides information on health systems and the health of Canadians [4]. The Discharge Abstract Database (DAD) is a major CIHI data holding, capturing information from hospital discharge records [4]. Related to pregnancy and birth, they collect information on obstetrical deliveries, newborns, and stillbirths from all acute inpatient hospitals in Canada, except Quebec [5]. Among other data elements, the DAD includes codes detailing diagnoses, comorbidities, and procedures, classified according to the Canadian Adaptation of the 10th International Statistical Classification of Diseases and Related Health Problems (ICD-10-CA) and the Canadian Classification of Health Interventions (CCI), respectively [6].

The two systems differ with respect to their intended use and availability to clinical users. While both collect similar data on demographics and outcomes, the BORN database contains more detailed data about maternal-newborn care that can be accessed via a reporting portal for authorized users or through data requests to BORN. Compared to the DAD, there is more in-depth data in the BORN system about maternal obstetrical history, environmental and behaviour exposures, and health behaviours and characteristics that allow contributors and requesters to explore relationships between pregnancy/birth characteristics and outcomes. Conversely, the DAD has more complete inpatient care information including maternal and neonatal transfer and readmission records and captures every hospitalization, not just those related to the labour and birth [5, 6]. It also captures long-term outcomes related to pediatric hospitalizations [5, 6].

High-quality data are essential to provide accurate and valid evidence for perinatal health surveillance, health service planning, and research. BORN continuously strives to enhance data quality in the BORN database [7]. Similar to CIHI, BORN's data quality framework concentrates on five dimensions of quality which include timeliness, accuracy, comparability, usability, and relevance [8].

Since BORN augments its data holdings with the CIHI-DAD data for some areas not routinely captured in the BORN database (e.g., pregnancy termination and maternal mortality) from acute care facilities in Ontario annually, and the DAD has some similar birth and maternal elements, it was considered a good source of comparison for a validation study [4].

The overall objective of this quality assurance project was to evaluate the quality of BORN data by assessing the agreement between key data elements collected in the BORN database and the corresponding data elements in the CIHI-DAD.

Methods

Data linkage

The provincial health card number (Ontario Health Insurance Plan [OHIP]) is a unique identifier, and has been assigned to almost all live newborns and a very small proportion of stillbirth records in both the BORN database and the CIHI-DAD. OHIP numbers are usually not assigned to stillbirths or to infants who die soon following birth, since they do not require provincial health services. We used the infant OHIP number to deterministically link these two data sources.

Figure 1 summarizes the linkage process. Briefly, we first extracted infant birth records with a valid OHIP number from the BORN database for three fiscal years (FY) (April 1 2012–March 31 2013, April 1 2013–March 31 2014, and April 1 2014–March 31 2015). In total, 417,558 infant records with a valid infant OHIP number were identified. A total of 6425 (1.5%) records did not have a valid infant OHIP number (4541 live births and 1884 stillbirths) and were, therefore, removed. We matched all infant records with their corresponding maternal records using an internal unique identifier assigned by BORN. We then similarly extracted newborn and stillbirth hospital abstracts from the CIHI-DAD for Ontario births during the same three fiscal years. Of 410,950 newborn/stillbirth records, we removed 2109 (0.5%) records with no valid OHIP number (1794 of the removed records were stillbirths). In the CIHI-DAD, maternal and live/stillbirth records are obtained from the Ontario hospital maternal delivery and infant live/stillbirth discharge abstracts. Since infants' and mothers' delivery records are completely separate, BORN had previously created a crosswalk file to connect infant records with their maternal records using both deterministic and probabilistic linkage methods. For this study, we made use of this crosswalk file to match infant live/stillbirth records with the respective maternal records to retrieve maternal data elements. The final step was to use the valid infant OHIP number and a deterministic linkage method to link the BORN database and the DAD. We compared type of birth (live birth or stillbirth), infant birth date, sex, postal code, birth weight, and gestational week at birth in the birth records, and compared mother's birth date, number of fetuses, induction of labour, and caesarean delivery in the associated maternal records.

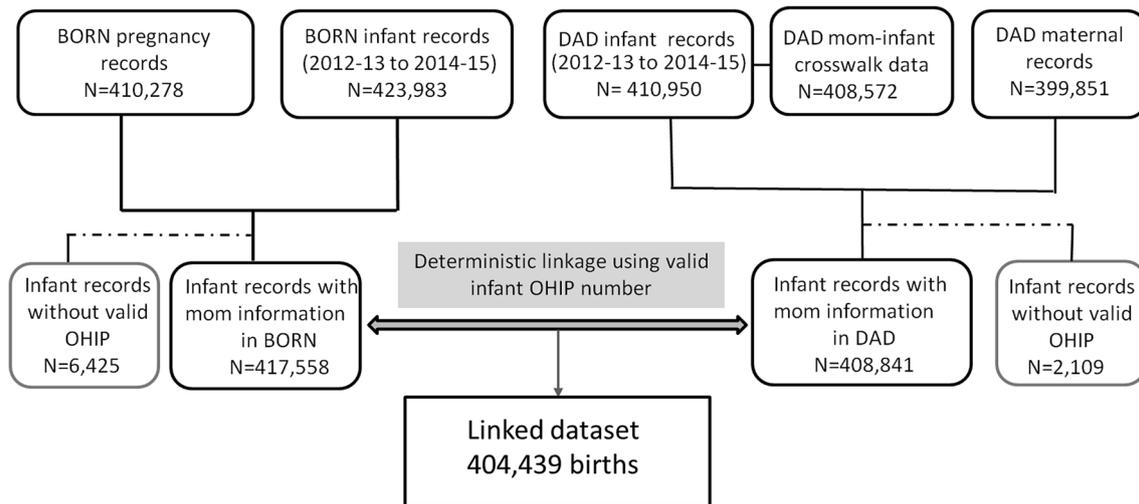


Fig. 1 Flowchart of linkage (fiscal year 2012–2013 to fiscal year 2014–2015). Linkage rate: BORN, 404,439/417,558 = 96.9%; DAD, 404,439/408,841 = 98.9%

CIHI-DAD diagnosis and procedure codes for certain variable definitions

We used maternal intervention codes (CCI codes) “5.AC.30” and “5.MD.60” to identify induction of labour and caesarean delivery, respectively, according to the ICD-10-CA in the DAD.

To obtain additional death information from the DAD for those unmatched records showing live birth in the DAD but stillbirth in the BORN database, we used maternal diagnosis codes “O00” to “O08”, both maternal and infant diagnosis codes “Z37”, “P95” and “P96.4”, and maternal intervention codes “5.CA.89” and “5.PC.91” to determine if there was a spontaneous abortion, pregnancy termination or stillbirth. We used maternal diagnosis codes “O61” and “O60.3” to explore records with failure of induction of labour and preterm induction of labour, respectively, in the DAD.

Statistical analysis

We calculated percentage agreement and Kappa coefficients (simple or weighted) with 95% confidence intervals (CI) for categorical and ordinal variables [9]. We applied the following conventional criteria to judge the strength of the agreement: Kappa coefficient < 0: less than chance agreement; 0.01–0.20: slight agreement; 0.21–0.40: fair agreement; 0.41–0.60: moderate agreement; 0.61–0.80: substantial agreement; 0.81–0.99: almost perfect agreement [9, 10]. For those variables with Kappa coefficients not meeting our criteria for almost perfect agreement, we examined potential reasons for these discrepancies. All data linkages and analysis were performed using SAS 9.4 (SAS Institute Inc., Cary NC).

Ethical considerations

As a quality assurance project, this data quality assessment was exempt from Research Ethics Board review under article 2.5 of the Tri-Council Policy Statement (TCPS2), the overarching ethical framework for research involving human participants in Canada [11].

Results

In total, 404,439 birth records were linked in three fiscal years of integrated maternal-newborn records across the BORN and DAD databases. The linkage rates between the two databases were 96.9% (404,439 out of 417,558) in the BORN database and 98.9% (404,439 out of 408,841) in the DAD (Fig. 1).

Table 1 shows agreement between the BORN and DAD databases among the linked cohort. The percentage agreement and Kappa coefficients were 99.98% and 0.740 (95% CI: 0.677–0.803) for live birth and stillbirth, respectively. The Kappa coefficients for the sex of infants, gestational age (weeks) at birth, induction of labour and caesarean birth were 0.989 (95% CI: 0.988–0.989), 0.920 (95% CI: 0.919–0.920), 0.782 (95% CI: 0.780–0.785), and 0.995 (95% CI: 0.995–0.996), correspondingly. Kappa agreement for the number of fetuses in the pregnancy was 0.979 (95% CI: 0.977–0.981). In this linked dataset, 99.9% of infant birthdates, 91.8% of infant postal codes, 95.5% of birth weights in grams, and 99.1% of maternal date of birth were matched exactly in the two data sources.

We further explored discrepancies in live/stillbirth classification, as well as induction of labour. Related to live

Table 1 Agreement between BORN and CIHI-DAD among the linked cohort

Variables	N	Percent agreement (%)	Kappa test			
			Kappa coefficient	95% CI lower limit	95% CI upper limit	p value
Baby's elements						
Stillbirth or live birth ^a	404,439	99.98	0.740	0.677	0.803	< 0.0001
Baby date of birth	404,439	99.9	NA	NA	NA	NA
Baby sex	404,439	99.4	0.989	0.988	0.989	< 0.0001
Baby weight ^b	403,396	95.5	NA	NA	NA	NA
Gestational age at birth (weeks) ^{b,c,d}	404,005	98.4	0.920	0.919	0.920	< 0.0001
Baby postal code ^b	404,164	91.8	NA	NA	NA	NA
Maternal elements						
Maternal date of birth ^b	403,059	99.1	NA	NA	NA	NA
Induction of labour ^b	402,552	92.1	0.782	0.780	0.785	< 0.0001
Caesarean delivery ^b	403,059	99.8	0.995	0.995	0.996	< 0.0001
Number of fetuses ^{b,d}	403,007	99.9	0.979	0.977	0.981	< 0.0001

^aOut of all stillbirths, 148 records in the BORN database with valid OHIP were linked

^bMissing values (less than 1%) were excluded

^cPercent agreement of gestational age (weeks) at birth calculation, agreement was defined as within 1 week of difference in gestational age

^dWeighted Kappa coefficient was reported

birth and stillbirth, there were 61 stillbirths in the BORN database that were recorded as live births in the DAD. We did not note any live births in the BORN data that were coded as stillbirths in the DAD (Appendix A). After exploring related data elements for these inconsistent live birth/stillbirth records in the DAD, we further observed that 22 of these 61 records had evidence of a “death” event in the DAD

(Fig. 2), including 21 of 22 that were coded as “death” in the discharge disposition DAD data element, 12 of 22 indicating that “death” occurred in a special care unit after birth, or a diagnosis or a procedural code on the DAD record suggesting an abortion, termination, or stillbirth. Table 2 shows the distribution of these 22 records by infant pregnancy outcome in the BORN data, gestational age at birth in weeks in the

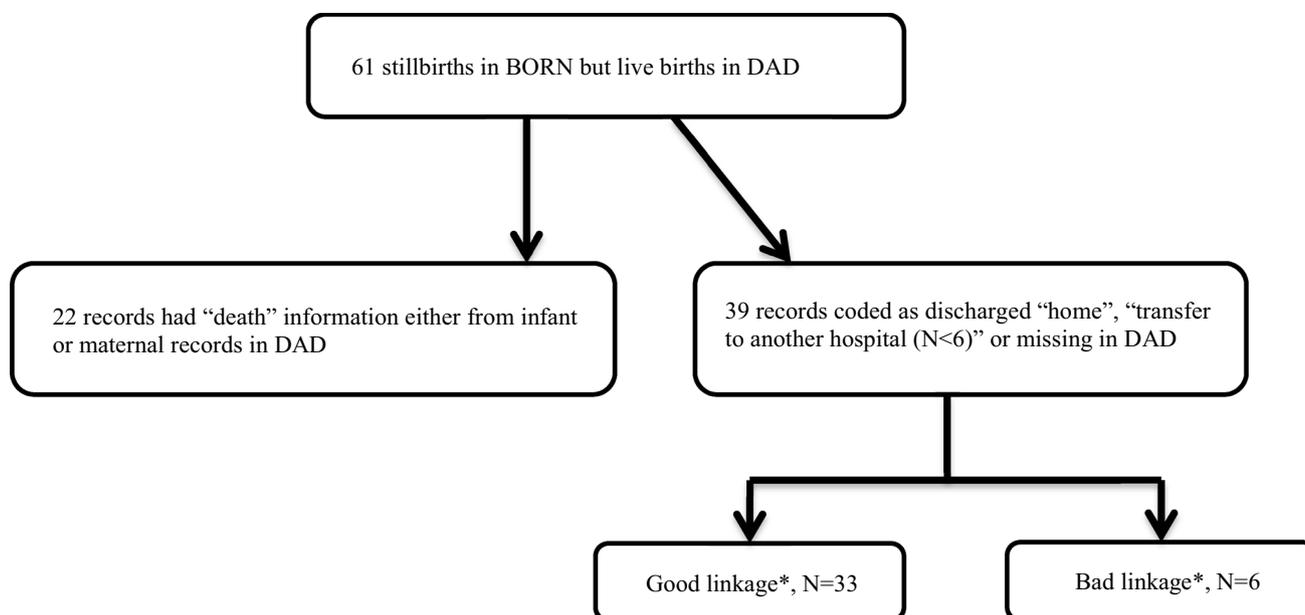


Fig. 2 Further investigation for inconsistent stillbirth/live birth records in two data sources. *Good linkage or bad linkage: if elements of baby birth date, sex, gestational age, postal code are concordant in BORN and DAD, this was called “good linkage”; otherwise “bad linkage”

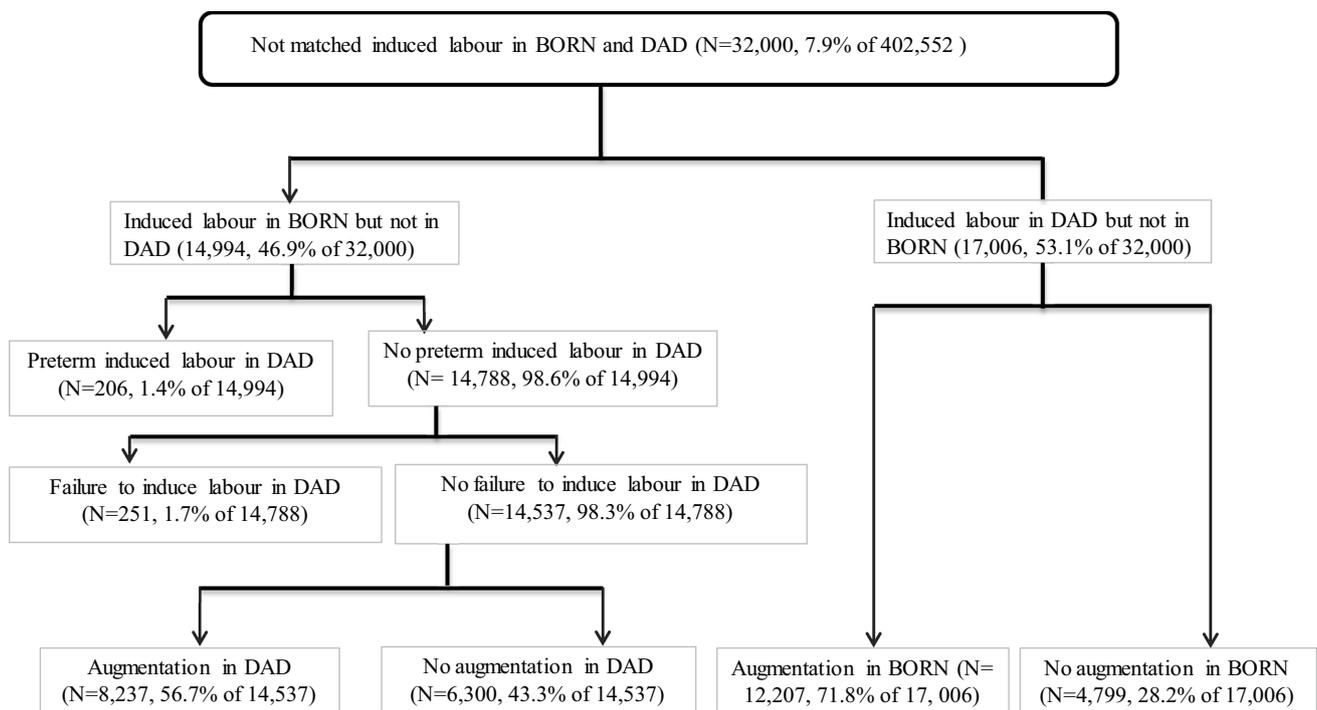
Table 2 Distribution of 22 stillbirths in BORN data but live births with “death” information in DAD

Variables	<i>N</i>	%	Median	Mean	Std
Infant pregnancy outcome in BORN					
Stillbirth-termination	7	31.8%	NA	NA	NA
Stillbirth-spontaneous in ante or intrapartum period	15	68.2%	NA	NA	NA
Gestational age in weeks BORN					
≤28 weeks	16	72.7%	23.0	22.9	2.0
29–41 weeks	6	27.3%	38.5	37.5	3.1
Overall	22	100%	23	26.9	7.0
Gestational age in weeks DAD					
≤28 weeks	16	72.7%	23.0	23.0	1.9
29–41 weeks	6	27.3%	38.5	37.5	3.1
Overall	22	100.0%	23.0	27.0	7.0
Length of stay in special care (hours)					
0	<6	–	–	–	–
No report	≥6	–	NA	NA	NA

“NA” stands for “not applied”. Due to the registry privacy policy, we are not able to report any number less than six individuals in one cell

two databases, and length of stay in special care nursery/neonatal intensive care unit (NICU) (in hours) in the DAD. In the BORN database, 15 (68.2%) of these 22 records were coded as a spontaneous stillbirth in the ante- or intrapartum period. In both the BORN database and the DAD, 72.7% of these 22 records showed a gestational age at birth equal to or less than 28 weeks. Except for a few records ($N < 6$) showing

transfer to another hospital, the DAD indicated that most of the remaining 39 records were either discharged to home or missing (Fig. 2). We suspected that there might be a linkage error due to an incorrect OHIP number, or records incorrectly linked in the data set. However, after we checked the common variables in the two data sources, we found that the baby’s birthdate, sex, gestational age, and postal code were

**Fig. 3** Further investigation for inconsistent induction of labour records in two data sources

concordant in the two databases for 33 of those birth records. There were only 6 records where information collected in the two databases was not fully consistent, suggesting that the linkage was good but there might be coding errors on live birth/stillbirth in either the DAD or BORN database.

For induction of labour, 32,000 (7.9% of 402,552) linked records were discordant (Fig. 3). Among them, 14,994 records (46.9% of 32,000) were coded as induced labour in BORN but not in the DAD, while 17,006 records (53.1% of 32,000) were coded as induced labour in the DAD but not in the BORN database. Among the records coded as induced labour in the BORN database but not in the DAD, 206 records (1.4% of 14,994) were coded as preterm induced labour (ICD-10-CA: “O60.3”) in the DAD. Among the remaining records, 251 (1.7% of 14,788) were coded as failed induction of labour (ICD-10-CA: “O61”) in the DAD. More than half of the remaining records ($N=8237$, 56.7% of 14,537) were coded as labour augmentation in the DAD. Among the inconsistent records with documentation of induced labour in the DAD but not in the BORN database, the majority of records (12,207/17,006; 71.8%) were coded as labour augmentation in the BORN database.

After exploring other related data elements in the DAD and BORN database, there were still discrepancies between the DAD and BORN database related to induction or augmentation of labour. Specifically, 6300 records in the BORN database showed induction of labour, yet were not coded as either induction or augmentation in the DAD. In addition, 4799 records in the DAD showed induction of labour, yet were not coded as induction or augmentation in the BORN database. Thus, we further explored birth-type distribution from the BORN data elements for these records. Detailed results are presented in Table 3. Among the records without information on labour induction or augmentation in the DAD that were coded as induced labour in BORN, only 22 (0.3% of 6300) records showed “No Labour—Caesarean

Section” as the type of birth, which implies that most of records were coded consistently between labour type and birth type in BORN. Among the DAD records coded as induction of labour but without any information on labour induction or augmentation in BORN, 883 (18.4% of 4,799) records showed “No Labour—Caesarean Section”, which indicated there might be a coding discrepancy between BORN and DAD for these elements.

Discussion

We found excellent percentage agreement ($\geq 90\%$), and substantial to almost perfect agreement on Kappa tests for all assessed key maternal and infant birth elements between the BORN dataset and the CIHI-DAD. However, there were discrepancies for the live/stillbirth and induction of labour variables. Overall, our findings of good concordance on assessed elements are consistent with other perinatal database validation studies conducted in the provinces of British Columbia and Nova Scotia, Canada and in Ontario using the legacy data from the Niday Perinatal Database [7, 12, 13]. Regarding induction of labour, another BORN data quality study evaluated “labour type” as one of the data elements re-abstracted and found 98% agreement with a kappa of 0.61 demonstrating substantial agreement, which is consistent with our finding (personal communication—Sandra Dunn).

Since 22 of 61 discrepant stillbirth cases had other evidence in the DAD indicating a “death” had occurred (i.e., abortion, termination, stillbirth-related elements or discharge disposition as death), this may imply some coding error or data entry variation in the DAD for these records. Additionally, when we compared the definitions of live/stillbirth in the two databases, we noted slight differences for stillbirth definitions. For instance, BORN adopted the Ontario Vital Statistics definition of stillbirth and live birth [14], according

Table 3 Distributions of birth type from BORN among not matched induction of labour records among linked cohort

Birth type distribution from the BORN data element	Not coded as induced labour or augmentation in DAD but was coded as induced labour in BORN data ^a ($N=6300$)		Not coded as induced labour or augmentation in BORN data but was coded as induced labour in DAD ($N=4799$)	
	<i>N</i>	%	<i>N</i>	%
Spontaneous vaginal	3774	59.9	2974	62.0
Assisted vaginal	756	12.0	359	7.5
Induced or spontaneous labour caesarean section	1748	27.7	583	12.1
No labour—caesarean section	22	0.3	883	18.4
Total	6300	100%	4799	100%

Include records not matched on induction of labour and no indication of augmentation in BORN and DAD

^aExclude records that were coded as failure to induction of labour or preterm induction of labour

to which a stillbirth is defined as a fetus who is either 500 g or more in birth weight or 20 weeks or more in gestational age that does not show any signs of life following birth, and a live birth is a newborn infant with any signs of life following birth, regardless of the gestational age or birth weight [14]. Importantly, the latter definition also includes situations in which a live born infant dies within the first few minutes or hours after birth. In contrast, CIHI's definition does not include a birth weight restriction and is based on an ICD10-CA code which defines a live birth as "the birth of a fetus greater than or equal to 20 weeks of gestation showing any signs of life at delivery" and a stillbirth as "any intra-uterine fetal demise or termination of pregnancy occurring at or after 20 completed weeks of gestation in which the fetus shows no signs of life" [15]. This slight variation in definitions between the two databases is one possible reason for the discrepancy, and coding errors in either database may also have contributed. Nevertheless, stillbirth is a rare event and the observed discrepancy only occurred in one direction; there were 61 stillbirths in the BORN database that were recorded as live births in the DAD, but no stillbirths in the DAD that were coded as live births in the BORN database. It is likely that other unrecognized factors may account for this discrepancy and further investigation of this discrepancy is needed.

With respect to induction of labour, BORN identifies this event with one data element (labour type), whereas in the DAD, induction of labour is identified with a procedure code ("5.AC.30") [15, 16]. In this linked cohort, more cases of induction of labour were identified in the DAD (95,717 in BORN vs. 97,729 in the DAD) (Appendix B). Among 14,994 (46.9% of 32,000) records defined as induction of labour in BORN but not in DAD, we found that a small proportion of records (457 of 14,994) were diagnosed as either failure of induction of labour (O.61) or preterm induction of labour (O.603) in the DAD (Fig. 3). In this study, we found that a major contributor to discrepancies on induction of labour between the two databases was coding variation between labour induction and labour augmentation. The key difference between "induced" labour and "augmented" labour is the timing of when the intervention is delivered, which is dependent on the judgement of onset of labour (or active labour) [17, 18]. In our study, among 32,000 not-matched records, 54.9% of records (8237 of 14,994) were coded as labour induction in BORN data but augmentation of labour in the DAD, and 71.8% of records (12,207 of 17,006) were coded as labour induction in the DAD but augmentation of labour in BORN (Fig. 2 and Appendix B). Another validation study conducted in Nova Scotia also compared the detection of labour induction between the DAD and the provincial perinatal database, and identified a variation of induction of labour coding between the two databases [13]. They found the rates of sensitivity and specificity for induction of labour in the DAD were 89.2%

and 96.9%, respectively, based on the assumption that the Nova Scotia perinatal database was a gold standard database. The authors also noted that one major area of discrepancy between the two databases was ascertaining if the labour was induced or not. For example, they found 45 out of 191 cases were coded as artificial rupture of membranes in both databases, but were coded as induction of labour in the DAD but not in the Nova Scotia database [13]. Difficulty judging the onset of labour, non-consensus among care providers, and unclear and varying definitions on different guidelines could possibly explain why it is not straight forward to differentiate induction of labour and augmentation of labour [17–19].

Furthermore, during an internal audit of BORN data, we found instances where cervical ripening with prostaglandins leading to labour was recorded as a spontaneous labour (personal communication—Jill Wiley). Since May 2015, a clear definition of induction of labour has been added to the BORN data dictionary tool to help data entry personnel avoid this coding issue [20]. This relatively recent definition of induction in the BORN database was based on the Society of Obstetricians and Gynaecologists of Canada (SOGC) guideline, which defines it as "the artificial initiation of labour by any intervention/s before the spontaneous onset of the latent phase of labour to help achieve a vaginal birth in 24–48 h" [19].

There are several limitations to this study. First, a large number of stillbirths ($N=1,884$) were excluded from the BORN file prior to data linkage with the DAD due to invalid OHIP numbers [5]. It is unclear whether the agreement on live/stillbirth status between the two databases would have been better or worse among those records with no OHIP number assigned. Considering the great importance of accurate classification of live/stillbirth status in a birth registry, further investigations are planned once we can resolve the data linkage method to include more complete stillbirth records. Second, our assessment did not include a hospital chart audit. We plan to conduct a data abstraction study to explore potential reasons for discrepancies based on a hospital chart review method in the next phase of our work. Third, the discrepancies between the two databases could be due to a potential linkage error, which may occur when the data entry personnel manually enter infants' OHIP numbers [21]. However, we expect the impact of this type of linkage error on our study would be minor because both BORN Ontario and CIHI have applied strict rules and algorithms to identify correct infant OHIP numbers [4, 8]. Due to the broader ICD-10-CA diagnosis categories on O60.3 (Preterm delivery without spontaneous labour), we might misclassify some cases with preterm delivery without induction of labour into preterm induction of labour. However, the potential impact on this misclassification is minor since there were only 1.4% records with ICD-10-CA O60.3 among inconsistent induction of labour records (Fig. 3). Finally, although both BORN and CIHI have standard protocols and guidelines for data

entry, errors are possible for both manual data entry and upload from electronic medical records.

In this study we conclude that the BORN database and the CIHI-DAD had a high concordance on key birth/maternal data elements. The criteria and methods for data collection are similar in the two systems. Our findings indicate that, as compared with the DAD, BORN data are accurate for maternal-newborn health care surveillance and health service planning. In addition, neither system as they are currently designed meets the full spectrum of needs of clinicians, hospitals, health system planners and funders, and researchers. The ability to link these two data sources facilitates both in-depth exploration of maternal and child health issues, as well as important longitudinal research [22]. Through the BORN and DAD data linkage, we are able to build more powerful and complete data for some elements (e.g., fetal and newborn congenital anomalies) that are limited in both databases to support surveillance and decision making, and answer broader research questions. Finally, we should be aware that the two systems differ with respect to their intended use and availability to clinical users. Although both systems collect similar data on demographics and outcomes, it is possible that different reporting strategies on potential confounders, such as pregnancy-related hypertensive disorders may have influenced the comparison of the maternal and neonatal outcomes in clinical research [23–25].

In summary, although agreement between the two databases was high, some discrepancies exist. Potential explanations include slight differences in data element definitions, different time points of data entry, different data element sources, and potential errors from manual or upload data entry. In the near future, we plan to conduct a data abstraction study to explore the potential reasons for discrepancies based on a hospital chart review.

Acknowledgements We thank all BORN staff, especially Ms. Jessica Reszel, Ms. Catherine Riddell, Ms. Farzana Yasmin, Ms. Vivian Holmberg, Dr. Jill Wiley and Dr. Mary (Yanfang) Guo for their excellent comments and editing on this manuscript. We also thank CIHI for providing the CIHI-DAD data to BORN.

Author contributions Q Miao: project development, data management and analysis, manuscript writing and editing. DB Fell: project development, manuscript writing and editing. S Dunn: manuscript writing and editing. AE Sprague: project development, manuscript writing and editing.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

Disclaimer Parts of this material are based on data and information compiled and provided by CIHI. However, the analyses, conclusions, opinions and statements expressed herein are those of the author(s), and not necessarily CIHI.

Ethical approval As a quality assurance project, this data quality assessment was exempt from Research Ethics Board review in Canada.

Appendix A: Distribution of type of birth in BORN and CIHI-DAD in the linked dataset

		BORN database		
		Live birth N (%)	Stillbirth N (%)	Total N (%)
CIHI-DAD	Live birth, N (%)	404,291 (100)	61 (41.2)	404,352 (99.98)
	Stillbirth, N (%)	0	87 (58.8)	87 (0.02)
	Total, N (%)	404,291 (100)	148 (100)	404,439 (100%)

There are more stillbirth records, but 148 with valid OHIP in linked dataset

Appendix B: Distribution of induction of labour (yes vs no) in BORN and CIHI-DAD in the linked dataset

		BORN database		
		Yes, N (%)	No, N (%)	Total, N (%)
CIHI-DAD	Yes, N (%)	80,723 (84.3)	17,006 (5.5)	97,729 (24.3)
	No, N (%)	14,994 (15.7)	289,829 (94.5)	304,823 (75.7)
	Total, N (%)	95,717 (100)	306,835 (100)	402,552 (100)

In the CIHI-DAD, induction of labour was identified by an ICD-10-CA procedure code: “5.AC.30”

References

1. Sprague AE, Sidney D, Darling EK, Wagner V, Soderstrom B, Rogers J, Graves E, Coyle D, Sumner A, Holmberg V, Khan B, Walker MC (2018) Outcomes for the first year of Ontario’s Birth Center demonstration project. *J Midwifery Womens Health* 63(5):532–540. <https://doi.org/10.1111/jmwh.12884>
2. Dunn S, Sprague AE, Grimshaw JM, Graham ID, Taljaard M, Fell D, Peterson WE, Darling E, Harrold J, Smith GN, Reszel J, Lanes A, Truskoski C, Wilding J, Weiss D, Walker M (2016) A mixed methods evaluation of the maternal-newborn dashboard in Ontario: dashboard attributes, contextual factors, and facilitators and barriers to use: a study protocol. *Implement Sci* 11:59. <https://doi.org/10.1186/s13012-016-0427-13>
3. Dunn S, Sprague A, Fell D, Dy J, Harrold J, Lamontagne B, Walker M (2013) The use of a quality indicator to reduce

- elective repeat caesarean section for low-risk women before 39 weeks' gestation: the Eastern Ontario Experience. *J Obstet Gynaecol Can* 35(4):306–316. [https://doi.org/10.1016/S1701-2163\(15\)30957-9](https://doi.org/10.1016/S1701-2163(15)30957-9)
4. Canadian Institute for Health Information (CIHI) (2016) Data quality study of the 2015–2016 Discharge Abstract Database: a focus on hospital harm 2016. https://secure.cihi.ca/free_products/DAD_15_16_Reab_Report_EN.pdf. Accessed 22 Oct 2018.
 5. Canadian Institute for Health Information (CIHI) (1996–2018) Discharge Abstract Database Metadata (DAD). <https://www.cihi.ca/en/discharge-abstract-database-metadata>. Accessed 22 Oct 2018.
 6. Canadian Institute for Health Information (CIHI) (1996–2019). DAD Data Elements 2015–2016. p. 2015–6. https://secure.cihi.ca/estore/productSeries.htm?pc=PCC189&_ga=2.107224642.115153162.1540238008-2019196711.1518113758. Accessed 22 Oct 2018.
 7. Dunn S, Bottomley J, Ali A, Walker M (2011) 2008 Niday perinatal database quality audit: report of a quality assurance project. *Chronic Dis Inj Can* 32(1):32–42
 8. BORN Ontario (2013) BORN Data Quality Framework. <https://datadictionary.bornontario.ca/data-quality/>. Accessed 22 Oct 2018.
 9. Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:37–46
 10. Viera AJ, Garrett JM (2015) Understanding interobserver agreement: the kappa statistic. *Fam Med* 37(5):360–363
 11. Government of Canada (2014) Ethical Conduct for Research Involving Humans. <https://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/default/>. Accessed 22 Oct 2018.
 12. Frosst G, Hutcheon J, Joseph KS, Kinniburgh B, Johnson C, Lee L (2015) Validating the British Columbia Perinatal Data Registry: a chart re-abstraction study. *BMC Pregnancy Childbirth* 15(1):1–11. <https://doi.org/10.1186/s12884-015-0563-7>
 13. Joseph KS, Fahey J (2009) Validation of perinatal data in the Discharge Abstract Database of the Canadian Institute for Health Information. *Chronic Dis Can* 29(3):96–100
 14. Statistics Canada (2005) Data quality, concepts and methodology, Database, Vital Statistics - Stillbirth. 2002. <https://www150.statcan.gc.ca/n1/pub/84f0211x/2005000/4153052-eng.htm>. Accessed 22 Oct 2018.
 15. Newfoundland and Labrador Centre for Health Information (2014) Provincial standard related to stillborn abstracting. <https://www.nlchi.nl.ca/images/PDFs/Provincial%20Standard%20for%20Stillborn%20Abstracting%202014.pdf>. Accessed 22 Oct 2018.
 16. Canadian Institute for Health Information (CIHI) (2015) International Statistical Classification of Diseases and Related Health Problems, Tenth Revision. https://www.cihi.ca/en/icd_volum_e_one_2015_en.pdf. Accessed 22 Oct 2018.
 17. Hanley GE, Munro S, Greyson D, Gross MM, Hundley V, Spiby H, Janssen PA (2016) Diagnosing onset of labor: a systematic review of definitions in the research literature. *BMC Pregnancy Childbirth*. 16:71. <https://www.ncbi.nlm.nih.gov/pubmed/27039302>
 18. Lange I, Ab C, Lee L, Bc V, Muise S, On ST (2013) Induction of labour. 296:1–18. https://sogc.org/wp-content/uploads/2013/08/September2013-CPG296-ENG-Online_REV-D.pdf. Accessed 22 Oct 2018.
 19. Crane J (2001) Induction of labour at term. *J SOGC* 23(8):717–728. [https://doi.org/10.1016/S0849-5831\(16\)31465-320](https://doi.org/10.1016/S0849-5831(16)31465-320)
 20. BORN Ontario (2018) BORN Data Dictionary. <https://datadictionary.bornontario.ca/>. Accessed 22 Oct 2018.
 21. Harron K, Dibben C, Boyd J, Hjern A, Azimae M, Barreto ML, Goldstein H (2017) Challenges in administrative data linkage for research. *Big Data Soc* 4(2):1–12. <https://doi.org/10.1177/2053951717745678>
 22. Dusetzina SB, Tyree S, Meyer AM, Meyer A, Green L, Carpenter WR (2014) Linking data for health services research: a framework and instructional guide. Agency for Healthcare Research and Quality, Maryland
 23. Laganà AS, Vitale SG, Sapia F, Valenti G, Corrado F, Padula F, Rapisarda AMC, D'Anna R (2018) miRNA expression for early diagnosis of preeclampsia onset: hope or hype? *J Matern Fetal Neonatal Med* 31(6):817–821. <https://doi.org/10.1080/14767058.2017.1296426.24>
 24. Laganà AS, Giordano D, Loddo S, Zoccali G, Vitale SG, Santamaria A, Buemi M, D'Anna R (2017) Decreased Endothelial Progenitor Cells (EPCs) and increased Natural Killer (NK) cells in peripheral blood as possible early markers of preeclampsia: a case-control analysis. *Arch Gynecol Obstet* 295:867–872. <https://doi.org/10.1007/s00404-017-4296-x>
 25. Schimmel MS, Bromiker R, Hammerman C, Chertman L, Ioscovich A, Granovsky-Grisaru S, Samueloff A, Elstein D (2015) The effects of maternal age and parity on maternal and neonatal outcome. *Arch Gynecol Obstet* 291:793–798. <https://doi.org/10.1007/s00404-014-3469-0>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.