

Defining a low-risk birth cohort: a cohort study comparing two perinatal data sets in Ontario, Canada

 Elizabeth Kathleen Darling^{1,2,*}, Olivia Marquez¹, and Alison L. Park³

Submission History

Submitted:	29/09/2023
Accepted:	13/02/2024
Published:	18/04/2024

¹McMaster Midwifery Research Centre, 1280 Main Street West, HSC 4H24, Hamilton, ON, L8S 4K1 Canada

²ICES McMaster University, 1280 Main Street West, HSC 4N43, Hamilton, ON, L8S 4K1 Canada

³ICES University of Toronto, 155 College Street, Suite 424, Toronto, ON, M5T 3M6, Canada

Abstract

Introduction

There are two main data sources for perinatal data in Ontario, Canada: the BORN BIS and CIHI-DAD. Such databases are used for perinatal health surveillance and research, and to guide health care related decisions.

Objectives

Our primary objective was to examine the level of agreement between the BIS and CIHI-DAD. Our secondary objectives were to identify the differences between the data sources when identifying a low-risk birth (LRB) cohort and to understand their implications.

Methods

We conducted a population-based cohort study comparing characteristics and clinical outcomes of all linkable births in BIS and CIHI-DAD between 1st April 2012 and 31st March 2018. We excluded out-of-hospital births, those with invalid healthcare numbers, non-Ontario residents and gestational age <20 weeks. We compared the portion of the cohort that met the criteria of a provincial definition of LRB based on each data source and compared clinical outcomes between the groups.

Results

During the study period, 779,979 eligible births were linkable between the two data sources. After applying the LRB exclusions, there were 129,908 cases in the BIS and 136,184 cases in CIHI-DAD. Most exclusion criteria had almost perfect, substantial or moderate agreement. The agreement for non-cephalic presentation and BMI ≥ 40 kg/m² (kappa coefficients 0.409 and 0.256, respectively) was fair. Comparison between the two LRB cohorts identified differences in the prevalence of cesarean (14.3% BIS versus 12.0% CIHI-DAD) and NICU admission (8.7% BIS versus 7.5% CIHI-DAD) and only 0.01% difference in the prevalence of ICU admission.

Conclusions

Overall, we found high levels of agreement between the BIS and CIHI-DAD. Identifying a LRB cohort in either database may be appropriate, with the caveat of appropriate understanding of the collection, coding and definition of certain outcomes. The decision for selecting a database may depend on which variables are most important in a particular analysis.

Keywords

perinatal; database; birth registry; data accuracy; administrative data; cohort study; population data; linkage; low-risk birth

*Corresponding Author:

Email Address: darlinek@mcmaster.ca (Elizabeth Kathleen Darling)

Introduction

The use of high-quality big data is important for accurate perinatal surveillance. Databases that collect perinatal information allow researchers, analysts and policy makers to monitor maternal and neonatal health trends over time and among specific sub populations. In Ontario, Canada, both a provincial perinatal database and a national hospital discharge database are used to examine perinatal outcomes and therefore the quality of these databases is of great importance.

Ontario has two main sources of perinatal health data: BORN (Better Outcomes Registry and Network) [1] and CIHI-DAD (Canadian Institute for Health Information Discharge Abstract Database) [2]. The BIS (BORN Information System) was first established in 2012 to collect, interpret and share critical data about pregnancy, birth and the early childhood period to facilitate and improve the provision of healthcare [1, 3, 4]. The registry collects a rich set of data from a range of sources including hospitals, laboratories, midwifery practice groups and clinical programs [1, 3, 5]. The quality of data housed in the BIS has been a priority and is guided by BORN's data quality framework which supports quality in the collection, analysis and disclosure of information [1, 6, 7]. The quality of BIS data has been verified in previous studies, and a recent chart re-abstraction study found the accuracy of most data elements was very good [1, 3, 5].

The CIHI-DAD is a hospital discharge database that captures administrative, demographic and clinical data on all hospitalizations [3, 8]. The database collects information related to pregnancy, livebirths, stillbirths and newborns from all acute inpatient hospitals in Canada [3]. Among other data elements, the CIHI-DAD includes codes detailing diagnoses and comorbidities, classified according to the Canadian Adaptation of the 10th International Statistical Classification of Diseases and Related Health Problems (ICD-10-CA), and procedures, classified according to the Canadian Classification of Health Interventions (CCI) [3]. The CIHI-DAD similarly prioritizes data quality and has a data and information quality program whose goal is to continuously improve existing data quality and ensure that new data and information meet CIHI's rigorous standards [9, 10].

The two databases are similar yet differ with respect to their data holdings based on their intended use [3]. Both data sources emphasize data quality and utilize data quality frameworks to guide their use and activities. Compared to the CIHI-DAD, the BIS collects more detailed information regarding maternal-newborn care, including information about health history, risk factors, exposures and indications for interventions; however, it does not include hospital level data that are typically used for costing and funding purposes [1, 3]. The BIS is the unique source of provincial data on births occurring at home and in birth centres, prenatal and newborn screening, cytogenetic laboratories, fertility clinics and billings for midwifery services. The CIHI-DAD houses broader diagnostic and intervention data regarding inpatient care and longitudinal outcomes pertaining to hospitalizations. This includes prior and subsequent hospitalizations of the pregnant person as well as pediatric hospitalizations of the newborn, thereby capturing information beyond that related to the labour and birth [3]. While one other published study has compared ten elements in the two data sets [3], agreement

between many variables remained unassessed prior to our work and the implications of variation between the two sources had not been documented.

Within perinatal databases, there is a need to be able to accurately identify and select populations in order to monitor perinatal outcomes in specific cohorts. Such a need arose in 2017, when Ontario's Ministry of Health and Long Term Care (MOHLTC) developed a set of "Quality-Based Procedures" (QBPs), which were designed to incentivise appropriate care and reduce inappropriate variation [11]. A Low Risk Birth (LRB) QBP was developed with a goal of reducing the variation in cesarean section rate across Ontario by adopting evidence-based guidelines that promote vaginal birth [11, 12]. The QBP defined a LRB population using variables collected in the BIS, with the goal of identifying a homogeneous group of pregnant people with characteristics likely to promote vaginal birth [11, 12]. Given variation between stakeholders in their access to either BIS or CIHI-DAD data, the LRB QBP created an incentive to compare the two different sources of routinely collected perinatal health data in Ontario and to explore the implications of their application in monitoring health system performance [11].

Objectives

Our primary objective for this study was to examine the level of agreement between the BIS and CIHI-DAD data sources. Our secondary objective was to understand the implications of differences between these two data sources when identifying a LRB cohort and measuring clinical outcomes.

Methods

Design, setting, and population

We conducted a retrospective population-based cohort study of people who gave birth in Ontario between 1st April 2012 and 31st March 2018, and whose births were included in both the BORN BIS and CIHI-DAD datasets and were linkable. We excluded out-of-hospital births, those with invalid health card numbers, non-female sex of parturient (linked on healthcare to RPDB), duplicate delivery (same parturient and date of birth but different delivery identification), delivery 1–140 days following another delivery (with the same health card number), non-Ontario residents, births prior to 20 weeks gestational age and stillbirth outcomes to create a cohort of linked births with data available from both sources.

Data sources and ethics statement

The research was conducted at ICES, an independent, non-profit research institute whose legal status under Ontario's health information privacy law allows it to collect and analyze health care and demographic data, without consent, for health system evaluation and improvement. We used the following data sources held at ICES: BORN BIS, CIHI-DAD, the Registered Persons Database (RPDB) and the Institution Information System (INST), as well as the following ICES-derived cohorts: MOMBABY, the Ontario Hypertension Database (HYPER) and the Ontario Diabetes Database

(ODD). These datasets were linked using unique encoded identifiers and analyzed at ICES. Details about each data source are available in Appendix A in Appendix Table 1.

Variable definition

We described the demographic characteristics of the full linked cohort, and of the LRB cohorts, using data from each data source regarding age of the person at the time they gave birth, and gestational age at birth. We used postal code data from the RPDB and census data to derive residential income quintile and to identify rural residency. We used hospital level data from INST to describe the type of hospital (academic, community, small) and the local health integration network (i.e. health region) where births occurred.

Exclusion criteria used to create the LRB QBP target population were a list of maternal and neonatal health conditions developed by a clinical expert advisory group using an evidence-based consensus process [11]. The LRB exclusion criteria were: stillbirths, maternal age <10 or >35 years of age at the time of delivery, pre pregnancy BMI ≥ 40 , multiparity, multiple gestation, non-cephalic fetal presentation, gestational age <37 weeks, induction of labour or no labour and any significant maternal or neonatal health condition (e.g. autoimmune, cancer, cardiovascular, diabetes, gastrointestinal, genitourinary, haematology, hypertension, musculoskeletal, neurology, pulmonary, placental, fetal complications).

To operationalize these variables using BORN BIS data, we used programming code provided by BORN which was developed during the creation of the original LRB QBP definition. This code was modified to be run within the ICES analytic environment but was otherwise untouched from the original algorithm. To create the CIHI-DAD version of the LRB QBP target population, two researchers (EG, EKD) used the published LRB QBP definition and identified equivalent variables available in the CIHI-DAD. We used the ICES-derived MOMBABY cohort to identify records of births in CIHI-DAD. Details of the exact definitions used in both BORN BIS & CIHI-DAD are provided in Appendix A in Appendix Table 2.

Prior to applying exclusions to create the LRB QBP, we compared the ascertainment of six variables (extreme preterm birth (i.e., <32 weeks' gestation), parity, gestational age in weeks, birthweight (for liveborn singletons only)), small-for-gestational-age <10th percentile (SGA) and large-for-gestational-age >90th percentile (LGA) between BORN BIS and CIHI-DAD. To explore options within available ICES data to ascertain pre-existing maternal hypertension and pre-existing maternal diabetes, we also identified the portion of cases in the linked cohort prior to the LRB exclusions who were in a) the Ontario Hypertension Database (HYPER) and b) the Ontario Diabetes Database (ODD). HYPER and ODD, two ICES-derived cohorts, are known to be more sensitive in identifying individuals with hypertension and diabetes, respectively, than CIHI-DAD data alone.

Once the LRB cohorts had been created based on the two data sources, we examined five outcomes: cesarean birth, intensive care unit (ICU) admission, neonatal intensive care unit (NICU) admission, SGA and LGA. Details of the definitions for each of these outcomes are also provided in Appendix A in Appendix Table 2.

Statistical analysis

We compared the count and frequency within the overall population for each component of the QBP LRB definition between the two perinatal data sources, as well as the agreement for each definition component. We measured overall agreement, overall probability adjusted kappa, hospital-specific stratified population counts and agreement for each exclusion criteria included in the QBP definition. We interpreted the kappa coefficient using Landis and Koch's classification of Cohen's kappa, which rate the strength of agreement as poor, slight, fair, moderate, substantial, and almost perfect using divisions which aim to provide useful benchmarks for interpretation [13]. We also compared agreement for parity, gestational age and birthweight between the two sources, and compared pre-existing hypertension and pre-existing diabetes mellitus between the BIS, and HYPER and ODD, respectively. Lastly, we compared agreement of the outcomes (cesarean section, ICU admission, NICU admission, SGA and LGA) between the LRB cohorts identified through the two perinatal data sources. All analyses were conducted in SAS 9.4 (SAS Institute, Inc, Cary, NC).

Results

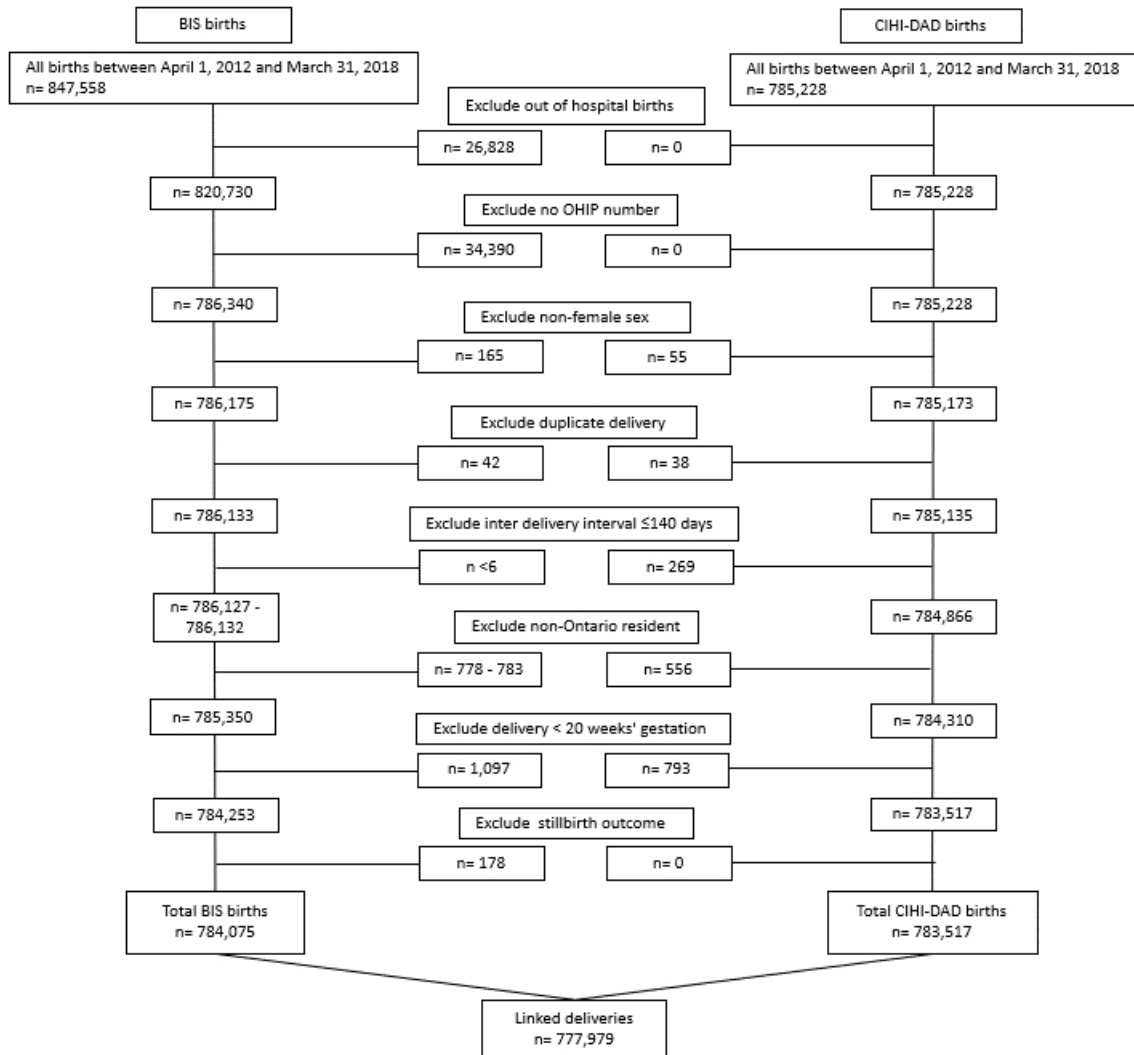
A total of 777,979 people who gave birth during the study period were linkable between the two data sources after applying initial exclusions (Figure 1). Table 1 shows the characteristics, based on data from the BIS and from CIHI-DAD, of the individuals included in the full linked data set, as well as the characteristics of LRB populations identified through each source.

Table 2 describes the prevalence, overall percent agreement and positive percent agreement for each exclusion criterion in the LRB definition. Most criteria in the LRB definition had almost perfect, substantial, or moderate agreement between BIS and CIHI-DAD. Indicators that had fair or slight agreement were primarily 'maternal conditions' from the LRB definition. Non-cephalic presentation and obesity (BMI ≥ 40 kg/m²) also had fair agreement (kappa coefficient (95% CI) 0.409 (0.406–0.412) and 0.256 (0.250–0.262), respectively). While agreement for non-spontaneous labour was moderate, this outcome had the lowest overall percent agreement between the BIS and CIHI-DAD (77.7% agreement).

We found fair agreement and slight agreement, respectively, when looking at pre-existing diabetes and pre-existing hypertension, which compared the ODD and HYPER databases respectively to BIS among the linked deliveries, and almost perfect agreement on SGA, LGA and extreme preterm birth between the BIS and CIHI-DAD (see Table 2).

Table 3 displays agreement in the full linked cohort (before applying LRB exclusions) between CIHI-DAD and BIS for parity, gestational age at birth and birthweight (for liveborn singletons). There was almost perfect agreement between the databases for all three of these variables. We found 89.3% exact agreement between CIHI-DAD and BIS for gestational age (in weeks) at birth and note that a remaining 9.4% agreement was within ± 1 week of the exact date.

Figure 1: Cohort creation from BORN-BIS and CIHI-DAD of total linked deliveries



The LRB cohorts derived from the BIS and the CIHI-DAD were similar in size (BIS N = 129,908 and CIHI-DAD N = 136,184), with 101,835 individuals included in both LRB cohorts, 28,073 only included in the BIS LRB cohort and 34,349 only included in the CIHI-DAD LRB cohort. Overall % agreement between the two sources for any exclusion was 92.0%. CIHI-DAD identifies 94.70% of LRB cases identified in the BIS, and the BIS identifies 95.63% of LRB cases in CIHI-DAD. The Kappa statistic was 0.717 (0.715–0.719) indicating substantial agreement.

After creation of the LRB cohorts, we compared cesarean, ICU admission and NICU admission between the BIS LRB cohort and the CIHI-DAD LRB cohort, and identified differences in the prevalence of cesarean (14.3% in the BIS LRB cohort versus 12.0% in the CIHI-DAD LRB cohort) and NICU admission (8.7% in the BIS LRB cohort versus 7.5% in the CIHI-DAD LRB cohort) and only a slight difference in the prevalence of ICU admission (0.05% in the BIS LRB cohort versus 0.06% in the CIHI-DAD LRB cohort)). Within each of the two LRB cohorts, agreement was almost perfect for cesarean and NICU admission when BIS and CIHI-DAD were compared as the source of data to measure the outcome, and there was moderate agreement

for ICU admission (see Table 4). Rates of SGA and LGA were very similar between the BIS LRB cohort and the CIHI-DAD LRB cohort (5.8% in the BIS LRB cohort versus 5.5% in the CIHI-DAD LRB cohort and 10.8% in the BIS LRB cohort versus 10.7% in the CIHI-DAD LRB cohort, respectively).

Discussion

Our findings confirm that overall, there are high levels of agreement between the BORN BIS and CIHI-DAD, the two main sources of perinatal data in Ontario. At the same time, we identified three key nuances in how the two sources differ. First, in general, the BIS identifies a higher proportion of cases with pre-existing health conditions, including obesity (BMI ≥ 40 kg/m²), diabetes and hypertension. Two exceptions to this were genitourinary conditions and fetal complications. Second, the BIS captures some variables pertaining to labour and birth that are not directly captured in CIHI-DAD, including labour type and fetal presentation. Third, observed differences between the two data sources will contribute to substantial differences in who's record is

Table 1: Baseline characteristics of linked dataset before and after applying Low Risk Birth cohort exclusions in the BIS and of CIHI-DAD datasets






Characteristic	Value	BIS values before exclusions	BIS Low Risk Birth cohort (after exclusions)	CIHI-DAD values before exclusions	CIHI-DAD Low Risk Birth cohort (after exclusions)
		N = 777,979	N = 129,908	N = 777,979	N = 136,184
<i>Of the parturient at delivery</i>					
Age	<10 years	352 (0.0)	0 (0.0)	1–5 (0.0%) ¹	0 (0.0)
	10–19 years	17,792 (2.3)	7,369 (5.7)	17,781–17,785 (2.3%) ²	7,755 (5.7)
	20–24 years	83,851 (10.8)	23,537 (18.1)	83,937 (10.8)	24,761 (18.2)
	25–29 years	209,910 (27.0)	48,563 (37.4)	210,024 (27.0)	50,672 (37.2)
	30–35 years	328,021 (42.2)	50,439 (38.8)	328,185 (42.2)	52,996 (38.9)
	36+ years	138,053 (17.7)	0 (0.0)	138,051 (17.7)	0 (0.0)
Gestational weeks	20–36	58,285 (7.5)	0 (0.0)	57,897 (7.4)	0 (0.0)
	37	57,356 (7.4)	6,619 (5.1)	57,447 (7.4)	7,108 (5.2)
	38	149,967 (19.3)	18,763 (14.4)	152,495 (19.6)	20,472 (15.0)
	39	225,352 (29.0)	40,785 (31.4)	227,128 (29.2)	43,415 (31.9)
	40	191,620 (24.6)	48,200 (37.1)	190,533 (24.5)	50,118 (36.8)
	41	91,969 (11.8)	15,034 (11.6)	90,463 (11.6)	14,778 (10.9)
	42+	3,430 (0.4)	507 (0.4)	2,016 (0.3)	293 (0.2)
Residential income quintile (Q)	Unknown	2,236 (0.3)	405 (0.3)	2,236 (0.3)	402 (0.3)
	Q1 (lowest)	172,506 (22.2)	27,850 (21.4)	172,506 (22.2)	28,937 (21.2)
	Q2	155,671 (20.0)	27,205 (20.9)	155,671 (20.0)	28,478 (20.9)
	Q3	159,526 (20.5)	27,226 (21.0)	159,526 (20.5)	28,459 (20.9)
	Q4	160,789 (20.7)	26,886 (20.7)	160,789 (20.7)	28,589 (21.0)
	Q5 (highest)	127,251 (16.4)	20,336 (15.7)	127,251 (16.4)	21,319 (15.7)
Rural residence	Unknown	769 (0.1)	129 (0.1)	769 (0.1)	134 (0.1)
	No	700,832 (90.1)	117,559 (90.5)	700,832 (90.1)	123,527 (90.7)
	Yes	76,378 (9.8)	12,220 (9.4)	76,378 (9.8)	12,523 (9.2)
<i>Of the hospital at delivery</i>					
Type	Unknown	353 (0.0)	62 (0.0)	353 (0.0)	53 (0.0)
	Academic	192,527 (24.7)	28,652 (22.1)	192,527 (24.7)	28,554 (21.0)
	Community	573,670 (73.7)	99,104 (76.3)	573,670 (73.7)	105,467 (77.4)
	Small	11,429 (1.5)	2,090 (1.6)	11,429 (1.5)	2,110 (1.5)
Local Health Integration Network (LHIN)	Unknown	444 (0.1)	89 (0.1)	1–5 (0.0%) ¹	0 (0.0)
	Erie St. Clair	33,446 (4.3)	5,068 (3.9)	33,450 (4.3)	4,596 (3.4)
	South West	56,366 (7.2)	8,679 (6.7)	56,399 (7.2)	8,351 (6.1)
	Waterloo	44,213 (5.7)	7,799 (6.0)	44,232 (5.7)	8,126 (6.0)
	Wellington				
	Hamilton	75,973 (9.8)	13,260 (10.2)	76,044 (9.8)	12,579 (9.2)
	Niagara				
	Haldimand Brant				
	Central West	47,435 (6.1)	8,713 (6.7)	47,437 (6.1)	8,333 (6.1)
	Mississauga	69,108 (8.9)	9,381 (7.2)	69,117 (8.9)	13,177 (9.7)
	Halton				
	Toronto Central	112,588 (14.5)	17,971 (13.8)	112,622 (14.5)	18,799 (13.8)
	Central	99,086 (12.7)	18,856 (14.5)	99,216 (12.8)	19,258 (14.1)
	Central East	76,063 (9.8)	13,058 (10.1)	76,094 (9.8)	14,468 (10.6)
	South East	23,096 (3.0)	3,760 (2.9)	23,103 (3.0)	3,706 (2.7)
	Champlain	73,812 (9.5)	11,697 (9.0)	73,821 (9.5)	12,709 (9.3)
North Simcoe	22,983 (3.0)	4,232 (3.3)	23,024 (3.0)	4,428 (3.3)	
Muskoka					
North East	29,212 (3.8)	5,234 (4.0)	29,258 (3.8)	5,303 (3.9)	
North West	14,154 (1.8)	2,111 (1.6)	14,161–14,165 (1.8%) ²	2,351 (1.7)	

¹Suppressed due to small cells (1–5).²Suppressed to avoid recalculation of small cells.

Table 2: Prevalence, overall percent agreement, positive percent agreement, and Kappa statistic for agreement of variables between CIHI-DAD and BIS data sources in full linked dataset

Condition (N = 779,979)	Prevalence, %		Difference DAD-BIS	Overall % Agreement	Positive % Agreement		Kappa coefficient (95% CI)
	DAD	BIS			DAD cases in BIS	BIS cases in DAD	
Exclusions for the QBP Low Risk Birth definition							
Age <10 or >35 years	17.74%	17.79%	-0.05%	99.9%	99.7%	99.9%	0.998 (0.997–0.998)
Multifetal birth	1.86%	1.78%	0.08%	99.9%	99.3%	95.2%	0.971 (0.969–0.973)
Parous	56.45%	57.29%	-1.00%	97.7%	97.3%	98.7%	0.953 (0.953–0.954)
Preterm birth <37 weeks' gestation	7.44%	7.49%	-0.05%	99.4%	95.3%	96.0%	0.953 (0.952–0.954)
Not a livebirth	0.50%	0.47%	0.03%	99.9%	94.8%	89.1%	0.919 (0.912–0.925)
Diabetes (pre-existing or gestational)	7.99%	7.50%	0.49%	98.0%	88.4%	82.9%	0.844 (0.841–0.846)
Cesarean section without labour	14.39%	15.05%	-0.66%	96.0%	84.6%	88.5%	0.842 (0.840–0.843)
Hypertensive disorder	5.67%	4.86%	0.81%	97.1%	78.1%	66.9%	0.706 (0.702–0.709)
Non-spontaneous labour	26.31%	40.36%	-14.26%	77.7%	55.0%	84.3%	0.509 (0.507–0.511)
Placental disorder	2.30%	1.57%	0.73%	97.9%	55.6%	38.0%	0.441 (0.434–0.448)
Fetal complication	7.75%	5.72%	2.03%	93.0%	56.8%	41.9%	0.446 (0.442–0.450)
Non-cephalic presentation	9.60%	11.06%	-1.46%	89.0%	43.87%	50.55%	0.409 (0.406–0.412)
Autoimmune condition	0.13%	0.48%	-0.35%	99.5%	14.8%	53.5%	0.231 (0.215–0.246)
Haematologic disorder	1.17%	1.75%	-0.58%	97.8%	20.6%	30.8%	0.236 (0.229–0.244)
Cancer	0.14%	0.23%	-0.09%	99.7%	20.5%	34.1%	0.255 (0.234–0.276)
Obesity (BMI \geq 40 kg/m ²)	1.70%	2.88%	-1.18%	97.0%	21.7%	36.6%	0.256 (0.250–0.262)
Gastrointestinal condition	0.80%	0.83%	-0.03%	98.7%	20.1%	20.8%	0.198 (0.188–0.207)
Cardiovascular disease	1.16%	2.08%	-0.92%	97.1%	9.0%	16.1%	0.102 (0.096–0.107)
Neurologic condition	0.16%	1.50%	-1.34%	98.5%	3.9%	37.6%	0.068 (0.062–0.075)
Genitourinary condition	4.54%	1.37%	3.17%	94.4%	11.9%	3.6%	0.035 (0.032–0.038)
Musculoskeletal condition	0.02%	0.5%	-0.48%	99.5%	0.85%	28.21%	0.016 (0.011–0.022)
Pulmonary condition	0.10%	4.02%	-3.92%	95.9%	0.6%	25.6%	0.010 (0.009–0.012)
Any exclusion	82.50%	83.30%	-0.80%	92.0%	94.7%	95.6%	0.717 (0.715–0.719)
Small-for-gestational age <10 th percentile	9.38%	9.48%	-0.11%	98.8%	93.3%	94.4%	0.932 (0.931–0.933)
Large-for-gestational age >90 th percentile	10.06%	10.05%	0.01%	98.9%	94.5%	94.4%	0.938 (0.937–0.940)
Extreme preterm birth <32 weeks' gestation	1.25%	1.26%	-0.01%	99.9%	96.7%	97.5%	0.970 (0.968–0.973)
Pre-existing diabetes (ODD vs BIS)	0.76%	1.00%	-0.24%	98.7%	20.68%	27.3%	0.229 (0.219–0.238)
Pre-existing hypertension (HYPER vs BIS)	0.66%	0.95%	-0.29%	98.7%	16.3%	23.4%	0.186 (0.177–0.195)

Legend: Based on Landis and Koch's classification of Cohen's kappa.¹³

Almost perfect	
Substantial	
Moderate	
Fair	
Slight	

included or excluded in a LRB cohort when operationalizing the LRB QBP definition, and subsequently will lead to differences in the prevalence of key outcomes such as cesarean birth. The latter two points are discussed further below.

The two LRB exclusion criteria contributing the greatest absolute difference between the two data sources were

non-spontaneous labour and non-cephalic presentation, with both being more prevalent in the BIS. As noted above, CIHI-DAD does not contain variables specifically intended to capture labour type or fetal presentation. In both cases, identifying these variables with CIHI-DAD relied on the use of intervention codes and ICD-10 codes that do not ideally capture the phenomenon of interest. For example,

Table 3: Comparison of outcomes in full linked cohort: Agreement between CIHI-DAD and BIS for parity, gestational age at birth (in weeks), and infant birthweight (among singleton livebirth deliveries)

Condition	Mean (SD)		Mean (SD) difference DAD - BIS	Percent agreement (%)				Weighted Kappa coefficient (95% CI)
	DAD	BIS		Exact	±1	±2	≥ ±3	
Parity	0.89 (1.1) ^a	0.91 (1.1) ^b	-0.02 (0.32) ^c	94.7% ^c	4.6% ^c	0.5% ^c	0.02% ^c	0.940 (0.940 to 0.941) ^c
Gestational age at birth (in weeks)	38.7 (2.0)	38.8 (2.1)	-0.01 (0.50)	89.3%	9.4%	0.8%	0.5%	0.929 (0.929 to 0.930)
				Difference in grams (%)				Intra-class coefficient
				Exact	±1 to 500	±501 to 1000	±1001+	
Birthweight (in grams)	3366.0 (552.2) ^d	3365.8 (564.6) ^e	1.0 (149.2) ^f	95.8% ^f	3.8% ^f	0.3% ^f	0.1% ^f	0.031

^aLimited to 777,933 non-missing parity in CIHI-DAD.

^bLimited to 768,620 non-missing parity in BIS.

^cLimited to 768,575 non-missing parity on both CIHI-DAD and BIS linked records.

^dLimited to 760,034 liveborn deliveries with recorded birthweight in CIHI-DAD.

^eLimited to 760,534 liveborn deliveries with recorded birthweight in BIS.

^fFurther limited to 756,119 liveborn deliveries with recorded birthweight in both CIHI-DAD and BIS.

Table 4: Comparison of outcomes between Low Risk Birth cohorts: Prevalence, overall percent agreement, positive percent agreement (with respect to CIHI-DAD data and with respect to BIS data), and Kappa statistic for agreement between CIHI-DAD and BIS data sources

Condition (N = 779,979)	Prevalence, %		Difference DAD-BIS	Overall % Agreement	Positive % Agreement		Kappa coefficient (95% CI)
	DAD	BIS			DAD cases in BIS	BIS cases in DAD	
Cesarean, ICU and NICU agreement after BIS exclusions N = 129,617							
Cesarean	14.32%	14.30%	0.02%	99.9%	99.6%	99.5%	0.994 (0.993–0.995)
ICU ^a	0.09%	0.05%	0.04%	99.9%	77.8%	39.8%	0.527 (0.440–0.614)
NICU ^a	7.78%	8.73%	-0.95%	98.2%	84.2%	94.5%	0.881 (0.876–0.886)
Small-for-gestational age <10 th percentile ^b	10.61%	10.80%	-0.19%	98.7%	93.2%	94.9%	0.933 (0.930–0.937)
Large-for-gestational age >90 th percentile ^b	5.96%	5.83%	0.13%	99.3%	94.7%	92.7%	0.933 (0.929–0.937)
Cesarean, ICU and NICU agreement after DAD exclusions N = 135,869							
Cesarean	12.03%	12.06%	-0.03%	99.9%	99.4%	99.6%	0.994 (0.993–0.995)
ICU	0.06%	0.04%	0.02%	100.0%	73.6%	45.4%	0.561 (0.462–0.660)
NICU	7.47%	8.37%	-0.90%	98.3%	84.4%	94.5%	0.883 (0.878–0.888)
Small-for-gestational age <10 th percentile ^c	10.69%	10.77%	-0.08%	98.8%	94.0%	94.7%	0.937 (0.934–0.940)
Large-for-gestational age >90 th percentile ^c	5.47%	5.48%	-0.01%	99.2%	92.8%	92.8%	0.924 (0.920–0.929)

^aTotal included in analysis N = 129,587.

^bTotal included in analysis N = 129,358 – limited to births between 22- and 43-weeks' gestation with non-missing birthweight and sex in both CIHI-DAD and BIS.

^cTotal included in analysis N = 136,725 – limited to births between 22- and 43-weeks' gestation with non-missing birthweight and sex in both CIHI-DAD and BIS.

CIHI-DAD captures non-cephalic presentations associated with obstructed labour or when interventions to treat malpresentation are required, so spontaneous vaginal births with face or compound presentations may not be captured in

the CIHI-DAD and thus contribute to the observed discrepancy in prevalence between the two data sources. Further, this difference could also contribute to the observed difference in the prevalence of outcomes.

While the absolute difference of 2% in cesarean prevalence between the BIS and CIHI-DAD LRB cohorts is small, it represents a 14-17% difference relative to the estimated cesarean prevalences of 14.3% and 12.0%, respectively. This magnitude of discrepancy would be of significance to potential recipients of pay for performance incentives. Given the almost perfect levels of agreement *within* each LRB cohort when comparing cesarean and NICU admissions based on the two data sources, the observed difference *between* the BIS and CIHI-DAD LRB cohorts is mostly attributable to differences in who is included in each cohort (i.e., differences between the two sources in ascertainment of the LRB exclusion criteria). Our results showed that >20% of the individuals in each LRB cohort are not included in the LRB cohort that is based on the other data source. In addition to differences already described, we note that it is possible that the codes included in the CIHI-DAD genitourinary conditions and fetal complications variables captured conditions not captured in the BIS definitions of these variables. Our comparison of the frequency of SGA and LGA between the two LRB cohorts revealed minimal differences in these two outcomes, suggesting that differences in who is excluded in each data set create minimal bias with respect to the distribution of birth weight for gestational age.

One interesting finding was that after applying LRB exclusions, the proportion of births that were LGA was notably lower (~5.5% versus ~10%) but the proportion of births that were SGA rose slightly (~10.7% versus ~9.4%). The shift to a slightly higher proportion of SGA births in the LRB cohort was likely due to two factors: 1) the LRB cohort is restricted to nulliparous births, and nulliparity is associated with a higher rate of SGA [14], and 2) the exclusion of pregnancies with greater risk of LGA (diabetes, obesity) led to many LGA births being excluded and contributed to an overall shift towards a lower birth weights after the exclusions. This observation raises the question of whether the LRB cohort as defined has truly removed all pregnancies at risk. The answer to such a question depends on the adverse outcome or outcomes being considered and will vary based on the intended purpose of examining a LRB cohort. Case by case considerations would be needed to explore which data source might best be used to capture individuals who are truly 'low risk'; however, our findings highlight that while either source is a reasonable choice, consistency in methods and data source is important for any comparison of outcomes between organizations or monitoring of trends over time.

One previous study assessing quality and comparing data in the BIS and CIHI-DAD found similar results and conclusions. In 2019, Miao *et al* compared key perinatal data elements from the BIS and CIHI-DAD and found concordance on key birth and maternal data elements and excellent percentage agreement ($\geq 90\%$) [3]. There was limited overlap in our study with respect to which variables were examined, and our research also differed by examining some variables using a categorical rather than a continuous approach (e.g., maternal age category vs. maternal date of birth, multifetal birth vs. number of fetuses). Our findings of high levels of agreement are also consistent with other perinatal database validation studies conducted in the provinces of British Columbia and Nova Scotia, and in Ontario using legacy data from the Niday Perinatal Database and the current BORN database

[1, 15, 16]. Previous authors have also noted that differences between provincial perinatal registry data and the CIHI-DAD may be attributable to variability in coding and the use of combined codes to yield similar data elements between the systems [8].

While neither database can meet the full spectrum of needs of clinicians, hospitals, health system planners and researchers, linking the two data sources as we have done in our study seems to be the most robust method to obtain complete data for elements that would otherwise not be as robust (e.g. fetal and newborn congenital anomalies) [3, 17, 18]. However, a limitation of this approach would be the exclusion of cases which cannot be linked between the two databases (e.g., out of hospital births or non-insured cases captured in BORN BIS) [19]. Findings from our study also support previous recommendations to strengthen data element definitions, data entry guidelines and training support for those who contribute data to perinatal information sources as this may improve the quality of data in both systems [5, 16, 17].

The existence of two publicly funded databases with overlapping content raises the question of whether this creates inefficiencies in cost and effort. It is important to note that while our study examined variables that overlapped between the two data sources, many of the variables captured in the BIS are not captured in the CIHI-DAD. The additional data in the BIS facilitate research on a wide range of important perinatal health topics that could not be examined using CIHI-DAD data. Furthermore, data unique to the BIS, such as indications for interventions and health history and risk factor data, are key to obtaining health care provider buy in to enhance the success of quality improvement endeavors addressing variations in clinical care. The BIS also adds value through its use of collected data to facilitate direct improvements in care, e.g., ensuring that newborn screening is not missed. The potential for improved efficiency is likely limited. While CIHI-DAD data is abstracted from hospital records by a data abstractor (who typically has no clinical expertise in perinatal care), data in the BORN-BIS is either uploaded directly from electronic medical records or manually entered by clinicians. Although in theory, it would be possible for the BIS to pull in data from CIHI-DAD for overlapping variables, this approach would likely offer minimal gain in cost or effort (given the widespread use of upload from electronic medical records) and likely would come with a trade off in terms of data accuracy, given a lack of relevant clinical expertise among DAD abstractors [20].

A main strength of our study is that it provides the first comparison between many variables in the two datasets. This will be of great value for researchers when making decisions regarding which data source to use. While the data we examined were collected between 2012 and 2018, there have been limited changes in the variable definitions and data collection processes since then, so our findings are likely similar to what they would be if the analysis were repeated with more recent data. Furthermore, many perinatal cohort studies will continue to use the historical data collected from 2012 onward, so documentation of the agreement of data during this timeframe continues to be relevant. Our study has a couple of minor limitations, mainly regarding the inherent differences between the two data sources. Our main limitation is that we compared two data sources without a clear gold standard. For this reason, we cannot determine if the differences observed

between the data sources were a result of accuracy issues in one data source and/or the other; however, our results still accurately reflect levels of agreement between the two sources. Second, since BORN's data elements were created by clinical experts and the CIHI-DAD uses the Canadian Classification of Health Interventions (CCI) and International Classification of Diseases (ICD) coding system, there are challenges in aligning the two different coding systems. The LRB definition was originally developed based on the BIS codes and definitions [11], and some of the LRB exclusion criteria were not captured in identical ways in the CIHI-DAD. We mapped the BIS codes onto the CIHI-DAD codes carefully; however, inherent differences in codes and their respective definitions likely explain some of the discrepancies seen between the LRB cohort in each database. Miao *et al* similarly discussed this difference in their study comparing congenital anomalies captured in the BIS and CIHI-DAD [18]. Again, we believe that our results accurately reflect levels of agreement between the two sources. However, as a caveat, researchers planning analyses focused on a specific pre-existing or pregnancy-related health condition that is used as an LRB exclusion criterion could consider conducting chart validation studies to determine the best CIHI-DAD codes to ascertain that condition and to ensure adequate specificity and sensitivity prior to conducting such an analysis using CIHI-DAD data.

Researchers using the two data sources should also be aware that differences in how data is entered, time points of data entry, and data element sources may lead to additional discrepancies between the two sources. Finally, given the different aims and purposes of the two data sources, there may be differing incentives regarding which variables are well captured.

Conclusion

In summary, our study compared agreement between linked births in the CIHI-DAD and the BIS, and between LRB cohorts identified from each source, showing that overall, there are high levels of agreement between the two sources. Indicators with lower agreement between the two sources might be improved by using more consistent coding and variable definitions. This study supports the use of either the CIHI-DAD or the BIS for basic perinatal health surveillance and for the identification of a LRB cohort, with the caveats that use of either data source requires appropriate understanding of the collection, coding and definition of outcomes, and that comparisons across the sources are problematic. Overall, our findings suggest that the BIS is a better data source if an analysis relies on more detailed medical history information. The optimal choice of data source, if both are available, should involve consideration of which variables are most important to the planned analysis.

Ethics statement

The research was conducted at ICES, an independent, non-profit research institute whose legal status under Ontario's health information privacy law allows it to collect and analyze health care and demographic data, without consent, for health

system evaluation and improvement. The use of the data in this project is authorized under section 45 of Ontario's Personal Health Information Protection Act (PHIPA) and does not require review by a Research Ethics Board. The project was approved through an internal ICES privacy impact assessment.

Conflicts of interest

The authors have no conflicts of interest to declare.

Acknowledgement and disclaimer

The authors wish to acknowledge the contributions of Erin Graves to earlier versions of the analysis leading to this publication. This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health (MOH) and the Ministry of Long-Term Care (MLTC). This document used data adapted from the Statistics Canada Postal Code^{OM} Conversion File, which is based on data licensed from Canada Post Corporation, and/or data adapted from the Ontario Ministry of Health Postal Code Conversion File, which contains data copied under license from ©Canada Post Corporation and Statistics Canada. Parts of this material are based on data and information compiled and provided by MOH and CIHI. However, the analyses, conclusions, opinions and statements expressed herein are solely those of the authors and do not reflect those of the funding or data sources; no endorsement is intended or should be inferred. This Study is also based in part on data provided by Better Outcomes Registry and Network ("BORN"), part of the Children's Hospital of Eastern Ontario. The interpretation and conclusions contained herein do not necessarily represent those of BORN Ontario.

Data availability statement

The dataset from this study is held securely in coded form at ICES. While legal data sharing agreements between ICES and data providers (e.g., healthcare organizations and government) prohibit ICES from making the dataset publicly available, access may be granted to those who meet pre-specified criteria for confidential access, available at www.ices.on.ca/DAS (email: das@ices.on.ca). The full dataset creation plan and underlying analytic code are available from the authors upon request, understanding that the computer programs may rely upon coding templates or macros that are unique to ICES and are therefore either inaccessible or may require modification.

References

1. Murphy MS, Fell DB, Sprague AE, Corsi DJ, Dougan S, Dunn SI, *et al*. Data Resource Profile Data Resource Profile: Better Outcomes Registry & Network (BORN) Ontario. *International Journal of Epidemiology* 2021;50(5):1416–25. <https://doi.org/10.1093/ije/dyab033>
2. Canadian Institute for Health Information. Data Quality Documentation, Discharge Abstract Database

- Current-Year Information, 2019–2020. Ottawa; 2020. [cited 2023 Sept 29]. Available from: <https://www.cihi.ca/sites/default/files/document/dad-data-quality-current-year-information-2019-2020-en.pdf>.
3. Miao Q, Fell DB, Dunn S, Sprague AE. Agreement assessment of key maternal and newborn data elements between birth registry and Clinical Administrative Hospital Databases in Ontario, Canada. *Arch Gynecol Obstet*. 2019;300:135–143. <https://doi.org/10.1007/s00404-019-05177-x>
 4. About BORN - BORN Ontario [Internet]. 2022 [cited 2023 Sept 29]. Available from: <https://www.bornontario.ca/en/about-born/about-born.aspx>.
 5. Dunn S, Lanes A, Sprague AE, Fell DB, Weiss D, Reszel J, et al. Data accuracy in the Ontario birth Registry: A chart re-abstraction study. *BMC Health Serv Res*. 2019;19:1001. <https://doi.org/10.1186/s12913-019-4825-3>
 6. BORN Ontario. BORN Ontario's Data Quality Framework. [cited 2023 Sept 29]. Available from: <https://www.prenatalscreeningontario.ca/en/data/resources/Documents/Nov-2020-BORN-Ontario-Data-Quality-Framework-Summary.pdf>.
 7. Better Outcomes Registry & Network. BORN Data Quality Report 2012-2014 – Executive Summary. 2016. [cited 2023 Sept 29]. Available from: <https://www.bornontario.ca/en/data/resources/Documents/Nov-2020-BORN-Data-Quality-Report-2012-2014—Executive-Summary.pdf>.
 8. Joseph KS, J. Fahey. Validation of perinatal data in the Discharge Abstract Database of the Canadian Institute for Health Information. *Chronic Dis Can*. 2009;29(3):96–100. Available from: <https://www.canada.ca/content/dam/phac-aspc/migration/phac-aspc/publicat/hpcdp-pspmc/29-3/pdf/cdic29-3-1-eng.pdf>.
 9. Canadian Institute for Health Information. CIHI's Information Quality Plan. Ottawa; 2017. [cited 2023 Sept 29]. Available from: <https://www.cihi.ca/sites/default/files/document/corporate-information-quality-plan-2016-en.pdf>.
 10. Canadian Institute for Health Information. CIHI's Information Quality Framework. Ottawa; 2017. [cited 2023 Sept 29]. Available from: https://www.cihi.ca/sites/default/files/document/iqf-summary-july-26-2017-en-web_0.pdf.
 11. The Provincial Council for Maternal and Child Health & Ministry of Health and Long-Term Care. Quality-Based Procedures Clinical Handbook for Low-Risk Birth. Toronto; 2017. [cited 2023 Sept 29]. Available from: <https://www.pcmch.on.ca/wp-content/uploads/2022/02/Low-Risk-Birth-QBP-Clinical-Handbook-2017-08-30.pdf>.
 12. The Provincial Council for Maternal and Child Health & Ministry of Health and Long-Term Care. Quality-Based Procedure Low-Risk Birth Toolkit. Toronto; 2018. [cited 2023 Sept 29]. Available from: <https://www.pcmch.on.ca/wp-content/uploads/2022/02/Low-Risk-Birth-QBP-Toolkit-18-05-15.pdf>
 13. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Source: Biometrics*. 1977;33(1):159–74. <https://doi.org/10.2307/2529310>
 14. El Adam S, Hutcheon JA, Mcleod C, McGrail K. Why are babies in Canada getting smaller? *Health Reports*. 2022;33(1):1–15. Available from: <https://www150.statcan.gc.ca/n1/en/pub/82-003-x/2022001/article/00001-eng.pdf?st=npEA7wqi>
 15. Dunn S, Bottomley J, Ali A, Walker M. 2008 Niday Perinatal Database quality audit: report of a quality assurance project. *Chronic Dis Inj Can*. 2011;32(1):32–42. <https://doi.org/10.24095/hpcdp.32.1.05>
 16. Frosst G, Hutcheon J, Joseph KS, Kinniburgh B, Johnson C, Lee L. Validating the British Columbia Perinatal Data Registry: A chart re-abstraction study. *BMC Pregnancy Childbirth*. 2015; 15:Article 123. <https://doi.org/10.1186/s12884-015-0563-7>
 17. Metcalfe A, Lyon AW, Johnson JA, Bernier F, Currie G, Lix LM, Tough SC. Improving completeness of ascertainment and quality of information for pregnancies through linkage of administrative and clinical data records. *Ann Epidemiol*. 2013;23(7):444–7. <https://doi.org/10.1016/j.annepidem.2013.05.002>
 18. Miao Q, Moore AM, Dougan SD. Data Quality Assessment on Congenital Anomalies in Ontario, Canada. *Front Pediatr*. 2020;8:573090. <https://doi.org/10.3389/fped.2020.573090>
 19. Harron K, Dibben C, Boyd J, Hjern A, Azimae M, Barreto ML, Goldstein H. Challenges in administrative data linkage for research. *Big Data & Society*, 2017;4(2). <https://doi.org/10.1177/2053951717745678>
 20. Zozus MN, Pieper C, Johnson CM, Johnson TR, Franklin A, Smith J, Zhang J. Factors Affecting Accuracy of Data Abstracted from Medical Records. *PLoS One*. 2015 Oct 20;10(10):e0138649. <https://doi.org/10.1371/journal.pone.0138649>

Appendix A

Appendix Table 1: Data sources

Data source	Full name	Description
BORN BIS	Better Outcomes Registry and Network Information System	A provincial perinatal registry established in 2012 to collect data about every birth in Ontario, as well as data about pregnancy and the early childhood period.
CIHI-DAD	Canadian Institute for Health Information Discharge Abstract Database	A national hospital discharge database that captures administrative, demographic, and clinical data on all hospitalizations.
RPDB	Registered Persons Database	An ICES derived dataset containing demographic information about anyone who has ever had an Ontario health insurance plan number. Based on data from the Ministry of Health and enhanced with other ICES data holdings.
INST	Institution Information System	A set of linkable datasets containing information about Ontario health care institutions funded by the Ministry of Health.
MOMBABY	MOMBABY	An ICES derived cohort based on linked DAD inpatient admission records of mothers (people giving birth) and their newborns.
HYPER	Ontario Hypertension Database	An ICES derived cohort containing all Ontario hypertension patients identified since 1991. Based on hospital admission and same day surgery diagnosis codes and physician billing codes.
ODD	Ontario Diabetes Database	An ICES derived cohort containing all individuals in Ontario with non-gestational diabetes identified since 1991. Based on hospital admission codes and physician billing codes.



Appendix Table 2: Data used to define low risk birth (LRB) cohort entry, exclusions, baseline characteristics and outcomes

Assessment	Disease or procedure or condition	BORN BIS	ICD-10-CA or CCI codes in CIHI-DAD, or in MOMBABY
Inclusion criteria	All deliveries in Ontario between April 1, 2012, and March 31, 2018	All records with a birthdate (B_BDATE) between 01APR2012 and 31MAR2018	All records with a birthdate (B_BDATE) between 01APR2012 and 31MAR2018 (see MOMBABY)
Exclusion criteria applied to all included records	Out of hospital births	BIRTH_LOCATION_TYPE_ID = 1012730 (hospital)	N/A
	Invalid healthcare number	M_VALIKN not 'V'	M_VALIKN not 'V'
	Invalid sex	SEX in RPDB not Female	SEX in RPDB not Female
	Non-Ontario residents	Province code in RPDB not Ontario	Province code in RPDB not Ontario
	Gestational age < 20 weeks	GA_AT_BIRTH_WEEKS < 20	B_GESTWKS_DEL < 20 (if missing use M_GESTWKS_DEL)
	Overlapping/duplicate deliveries	Same M_IKN and B_BDATE but different PREGNANCY_ID_ENC, or a delivery < 20 weeks following the previous delivery	Same M_IKN and B_BDATE but different M_KEY, or a delivery < 20 weeks following the previous delivery
	Non-birth outcome	AGG_PREGNANCY_OUTCOME_ID is not live birth (1021030) or stillbirth (1021060, 1021070, 1021080, 1021090)	N/A
Link BIS & CIHI-DAD deliveries	Deliveries from 01APR2012 to 31MAR2018	M_IKN in MOMBABY = M_IKN in BIS B_BDATE in MOMBABY +/- 7 days of B_BDATE in BIS	M_IKN in MOMBABY = M_IKN in BIS B_BDATE in MOMBABY +/- 7 days of B_BDATE in BIS
Exclusions applied to linked BIS + CIHI-DAD delivery records:			
<i>Recorded at delivery, or during pregnancy (BMI)</i>	Stillbirth	AGG_PREGNANCY_OUTCOME_ID includes 1021060, 1021070, 1021080, or 1021090	M_STILLBIRTH in MOMBABY = 'T'
<i>Same as above</i>	<10 or >35 years of age	MATAGEATSTILLORLIVEBIRTHYEARS < 10 or > 35	AGE in RPDB < 10 or > 35
<i>Same as above</i>	Pre-pregnancy BMI > = 40.0 kg/m ²	MATERNAL_BMI >= 40.0 (if BMI is missing, then include in dataset)	E66.(Obesity). Any DXTYPE.
<i>Same as above</i>	Multiparous	PARITY is not 0	The sum of PREVTERM (Number of previous term deliveries) and PREVPRETERM (Number of previous pre-term deliveries) > 0
<i>Same as above</i>	Multifetal gestation	NUMBER_OF_FETUSES is not 1	O30 (same as M_MULTIBIRTH = 'T' in MOMBABY)
<i>Same as above</i>	Non-Cephalic presentation	PRESENTATION_TYPE_ID is not 1021150 (Cephalic), 3000088 (Cephalic \ Cephalic type unknown), 1021190 (Cephalic \ Vertex)	5MD56^^ (Breech delivery) OR O32^^ (malpresentation) OR O64^^ (obstructed labour)
<i>Same as above</i>	Gestation < 37 weeks	GA_AT_BIRTH_WEEKS < 37	B_GESTWKS_DEL in MOMBABY < 37 (if missing use M_GESTWKS_DEL in MOMBABY)
<i>Same as above</i>	Non-Spontaneous labour	LABOUR_TYPE_ID is not 1014640 (Spontaneous)	5.AC.24 or 5.AC.30 (inductions)
<i>Same as above</i>	No labour C-section	BIRTH_TYPE_ID = 1012900 (No Labour - Cesarean Section) OR (BIRTH_TYPE_ID = 1012890 [Induced or Spontaneous Labour Cesarean Section] and LABOUR_TYPE_ID = 1014645 [No Labour])	5.MD.60^^ AND Intervention Status Attribute code: N2 (Elective primary c/s with indications), N3 (Elective primary c/s without indication – e.g. mother's request), or N4 (Elective repeat c/s), PA (Primary, Indicated, Planned), PC (Primary, Not-Indicated, Planned), RA (Repeat, Indicated, Planned), or RC (Repeated, Not-Indicated, Planned)
<i>BIS: Pre-existing or diagnosed / active during the index pregnancy; CIHI-DAD: Diagnosed ≤ 2 y pre-conception & before delivery discharge</i>	Maternal autoimmune condition	MAT_PRE_EXIST_HEALTH_COND_ID IN (1016530, 1016540, 1016560, 1016550)	L93, M32, A18.4, M05, M06, M45, M35.9

Continued

Appendix Table 2: Continued

Assessment	Disease or procedure or condition	BORN BIS	ICD-10-CA or CCI codes in CIHI-DAD, or in MOMBABY
Same as above	Maternal cancer	MAT_PRE_EXIST_HEALTH_COND_ID IN (1016570, 1016580, 1016590) or EXPOSURE_MEDICATION_ID = 1020640	C00-C97, Z51.1, Z54.2, Z92.6
Same as above	Maternal cardiovascular condition	MAT_PRE_EXIST_HEALTH_COND_ID IN (1016600, 1016610, 1016620, 1016630, 1016640, 1016650, 1016680, 1016660, 1016670)	O10.1, O10.3, I51.6, Q20, Q21, Q22, Q23, Q24, O10.4, O10.9, I10, I11, I13, O10.2, I12, O26.8, N00-N07; N13.0-N13.5; N13.7-N13.9; N17-N19
Same as above	Maternal diabetes	MAT_PRE_EXIST_HEALTH_COND_ID IN (1016730, 1016740, 1016750, 1016760, 1016770, 1016780, 3000004) or DIABETES_AND_PREGNANCY_ID IN (1013400, 1013410, 1013420, 1013430, 1013440, 1013450, 1013460, 1013465, 1013470, 1013480, 1013490, 1013500, 1013510, 1013520, 1013530, 1013540, 1013545, 1013550, 1013560, 1013570, 3000001, 3000002)	O24, E10, E11, E13, E14
Same as above	Maternal gastrointestinal condition	MAT_PRE_EXIST_HEALTH_COND_ID IN (1016900, 1016910, 1016920, 1016930, 1016905, 1017383)	O26.2, K80.0, K80.1, K80.2, K80.4, K81, K51, K52, K50, O98.4, B15-B19
Same as above	Maternal genitourinary condition	MAT_PRE_EXIST_HEALTH_COND_ID IN (1016960, 1016970, 1016980, 1017010, 1016990, 1017000)	O99.2, O11, O12.1, N17, N18, N19, O34.0, O34.1, O34.29, N99
Same as above	Maternal haematologic disorder	MAT_PRE_EXIST_HEALTH_COND_ID IN (1017020, 1017030, 1017035, 1017040, 1017050, 1017090, 1017060, 1017070, 1017080)	O28.0, O99.1, D65-D89, D66, D67, D68.0, D69, D57, D56, D68.5
Same as above	Maternal musculoskeletal condition	MAT_PRE_EXIST_HEALTH_COND_ID IN (1017100, 1017150, 1017110, 1017120, 3000186, 1017130, 1017140, 1017160, 1017165)	Q77.4, G71.0, G71.2, G71.1, Q78.0, G71.8, G71.9
Same as above	Maternal neurologic condition	MAT_PRE_EXIST_HEALTH_COND_ID IN (1017170, 1017180, 1017190, 1017189, 1017200, 1017210, 1017230, 1017220)	G80, G35, G70.0, Q05, G90-G99
Same as above	Maternal pulmonary condition	MAT_PRE_EXIST_HEALTH_COND_ID IN (1017320, 1017192, 1017330, 1017340, 1017370, 1017350, 1017360)	O28.5, E84, I27.0, I27.1, I27.2, I28
During index pregnancy or delivery hospitalization	Maternal hypertensive disorder of pregnancy	PREG_HYPERTENSION_DISORDER_ID IN (1020800, 1020810, 1020820, 1020850, 1020854, 1020840)	O11, O13, O14, O15, O16
Same as above	Maternal placental disorder in the index pregnancy	COMPLICATION_ID IN (1020245, 1020365, 1020310, 1020320, 1020330, 1020340, 1020300)	O43.2, O44, O45, O43.0, O43.1, O43.80, O43.81, O43.88, O43.9
Same as above	Fetal complication in the index pregnancy	COMPLICATION_ID IN (1020190, 1020195, 1020200, 1020210, 1020215, 1020220, 1020235, 1020230)	O35, O36.0, O36.1, O36.5, O41.01, O41.02, O41.03, O41.09, O36.7, O36.8, O36.9
Additional comparisons in full linked cohort	Maternal pre-existing diabetes	MAT_PRE_EXIST_HEALTH_COND_ID IN (1016730, 1016740, 1016750, 1016760, 1016770, 1016780, 3000004) OR DIABETES_AND_PREGNANCY_ID IN (1013500, 1013510, 1013520, 1013530, 1013545, 1013540, 1013550, 1013560, 1013570)	ODD: ICES-derived cohort created using algorithms applied to DAD, SDS & OHIP physician claims databases. Definition: 2 OHIP records with ICD-9 250, or 1 DAD / SDS record with ICD-9 [ICD-10] 250 [E10, E11, E13, E14] in 2-y. Excludes claims & admissions with diabetes Dx in 120-d before or 180-d after a gestational hospitalization record
	Maternal pre-existing hypertension	MAT_PRE_EXIST_HEALTH_COND_ID IN (1016660, 1016620)	HYPERT: ICES-derived cohort created using algorithms applied to DAD, SDS & OHIP physician claims databases. Definition: 2 OHIP claims with ICD-9 401-405, or 1 DAD/SDS with ICD-9 [ICD-10] 401-405 [I10-I13, I15]. Excludes claims & admissions with hypertension Dx in 120-d before or 180-d after a gestational hospitalization record

Continued

Appendix Table 2: Continued

Assessment	Disease or procedure or condition	BORN BIS	ICD-10-CA or CCI codes in CIHI-DAD, or in MOMBABY
	Extreme preterm birth (<32 weeks gestational age)	GA_AT_BIRTH_WEEKS < 32	B_GESTWKS_DEL < 32 (if missing use M_GESTWKS_DEL)
	Gestational age	GA_AT_BIRTH_WEEKS	B_GESTWKS_DEL (if missing use M_GESTWKS_DEL) in MOMBABY
	Birthweight	BIRTH_WEIGHT_GRAMS (link to AGG_INFANT on PREGNANCY_ID_ENC)	B_WEIGHT in MOMBABY
	Small-for-gestational-age ¹	BIRTH_WEIGHT_GRAMS, GA_AT_BIRTH_WEEKS, B_SEX (link to AGG_INFANT on PREGNANCY_ID_ENC)	B_SEX in MOMBABY, Gestational age (defined above), Birthweight (defined above)
	Large-for-gestational-age ¹	BIRTH_WEIGHT_GRAMS, GA_AT_BIRTH_WEEKS, B_SEX (link to AGG_INFANT on PREGNANCY_ID_ENC)	B_SEX in MOMBABY, Gestational age (defined above), Birthweight (defined above)
	Parity	PARITY	Sum of PREVTERM (Number of previous term deliveries) and PREVPRETERM (Number of previous pre-term deliveries)
Baseline characteristics of LRB cohort	Maternal age	MATAGEATSTILLORLIVEBIRTHYEARS	AGE in RPDB
	Gestational age	GA_AT_BIRTH_WEEKS	B_GESTWKS_DEL (if missing use M_GESTWKS_DEL)
	LHIN of hospital Hospital type	INDEX_HOSPITAL_LHIN_CODE_NUM MATERNAL_INDEX_HOSPITAL_ID	INSTLHIN from the LHIN file INST from the institution file provided by the Ministry of Health
	Rural residence	Postal code in RPDB outside the commuting zone of centres with population 10,000+ (Statistics Canada, PCCF+)	Postal code in RPDB outside the commuting zone of centres with population 10,000+ (Statistics Canada, PCCF+)
	Neighbourhood income quintile	Postal code in RPDB linked with Census area data on median household income (Statistics Canada, PCCF+)	Postal code in RPDB linked with Census area data on median household income (Statistics Canada, PCCF+)
Outcomes compared between BIS LRB cohort and CIHI-DAD LRB cohort	Cesarean section	BIRTH_TYPE_ID = '1012890' (Induced or Spontaneous Labour Cesarean Section) or '1012900' (No Labour - Cesarean Section)	5.MD.60. AND Intervention Status Attribute code NOT 'A' (Abandoned)
	Maternal ICU admission	MATERNAL_OUTCOME_ID = '1016500' (Transfer to ICU/CCU) at labour/birth encounter (LBM) or post-partum (PPM) or AGG_PREGNANCY (link on PREGNANCY_ID_ENC)	DAD (SCU1-SCU6) 10 (Medical) 20 (Surgical) 25 (Trauma) 30 (Combined Medical-Surgical) 35 (Burn) 40 (Cardiac) 45 (Coronary) 60 (Neurosurgery) 80 (Respirology)
	Neonatal ICU admission	NICU_ADMISSION_FLAG (link to AGG_INFANT on PREGNANCY_ID_ENC)	DAD (SCU1-SCU6) 50, 51, 52, 53

BIS: Better Outcomes Registry & Network (BORN) Information System; CCI: Canadian Classification of Health Interventions; CIHI: Canadian Institute for Health Information; DAD: Discharge Abstract Database; HYPERT: Ontario Hypertension Database; ICD-9: International Classification of Diseases, 9th Revision; ICD-10-CA: International Classification of Diseases, 10th Revision, Canada; LHIN: Local Health Integration Network; ODD: Ontario Diabetes Dataset; OHIP: Ontario Health Insurance Plan; PCCF+: Statistics Canada's Postal Code^{OM} Conversion File Plus; RPDB: Registered Persons Database; SDS: Same Day Surgery Database.

¹Defined according to Kramer *et al*'s population-based Canadian reference for birth weight for gestational age: Kramer MS, Platt RW, Wen SW, Joseph KS, Allen A, Abrahamowicz M, Blondel B, Bréart G, Fetal/Infant Health Study Group of the Canadian Perinatal Surveillance System. A new and improved population-based Canadian reference for birth weight for gestational age. *Pediatrics*. 2001 Aug;108(2):E35. <https://doi.org/10.1542/peds.108.2.e35>. PMID: 11483845.