A Data Validation Exercise between the CIHI-DAD and the BORN Information System

Grace Miao, Deshayne Fell, Ann Sprague

BORN Ontario

April 24, 2017
BORN Information System (BIS)—‘Data In’ Process

BIS System

- Cycles from all IVF clinics (2013)
- Births from all 97 birthing hospitals in Ontario
- Births from all 87 midwifery practice groups in Ontario
- All prenatal screening results from 4 labs (NIPT in dev't)
- All newborn screening results from NSO
- All Level 2 NICU stays - 50% of Level 3
- Prenatal and Newborn Screening follow-up results from clinics
- Primary Care: 10 sites provide OAR, RBR, ndds, clinic visits, height/weight

Automatic linking and matching regardless of order of entry

2 million individuals in the BIS
## BORN data quality framework

### TIMELINESS
- **Data currency**
- **Documentation currency**

### ACCURACY
- **Comprehensiveness**
  - Coverage
  - Capture & collection
- **Completeness**
  - Stakeholder non-response
  - Item non-response
- **Precision**
  - Measurement Error
  - Editing and processing
  - Technical specifications

### COMPARABILITY
- **Data collection**
- **Data dictionary**
- **Linkage**
- **Equivalency**
- **Historical comparability**

### USABILITY
- **Interpretability**
- **Accessibility**
- **Documentation**

### RELEVANCE
- **Adaptability**
- **Value**
BORN data quality framework

TIMELINESS
- Data currency

ACCURACY
- Comprehensiveness
  - Coverage
  - Capture & collection
- Completeness
  - Stakeholder non-response
  - Item non-response
- Precision
  - Measurement Error
  - Editing and processing
  - Technical specifications

COMPARABILITY
- Data collection
- Data dictionary
- Linkage
- Equivalency
- Historical comparability

USABILITY
- Interpretability
- Accessibility
- Documentation

RELEVANCE
- Adaptability
- Value
Aim

• To monitor and provide evidence of data quality in BIS data

• To provide rationale that a linkage between CIHI-DAD and BIS datasets could increase data completeness
Objective

• To assess the reliability of key elements collected in the BIS by comparing these with data in the Canadian Institute for Health Information Discharge Abstract Database (CIHI-DAD)
CIHI-DAD

- **Canadian Institute for Health Information (CIHI)**
  - “CIHI is an independent, not-for-profit organization that provides essential information on Canada’s health systems and the health of Canadians”.

- **Discharge Abstract Database (DAD)**
  - “DAD captures administrative, clinical and demographic information on hospital discharges”.
  - BORN receives an annual dataset which includes all maternal and newborn records from acute care facilities in Ontario.

References:
Maternal-newborn linkage in BIS

• Based on aggregated infant data with infant birth date in fiscal years 2012-2013, 2013-2014 and 2014-2015
  – 5,615 records from hospital birth and Ontario residence without valid OHIP (3,784 live births and 1,831 stillbirths) were removed.
  – N = 417,558

• Aggregated infant data is linked to aggregated pregnancy data in the BIS to match infant's maternal information
Maternal-newborn linkage CIHI-DAD

  - Only 2,109 baby records (0.5% of 410,950 birth records showing Ontario health records) without valid OHIP (0=321, 9=1,788). ‘0’= HCN not available, ‘9’=Stillbirth.
  - Total records: 408,841

- Link the cross-walk data file (baby-mom in CIHI-DAD) developed by BORN and then patch infant’s maternal information in DAD
Flowchart of linkage
(FY 2012-2013 to 2014-2015)

BIS Aggregated pregnancy

BIS Aggregated infant (FY 2012-13 to 2014-15)

DAD birth records (FY 2012-13 to 2014-15)

DAD maternal records

Deterministic linkage using valid baby OHIP number

Infant records with mom information in BIS

Infant records with mom information in DAD

Linked dataset 404,439 births

Linkage rate: BIS: 404,558 / 417,549 = 96.8%; DAD: 404,439 / 408,841 = 98.9%
Methods

• Comparison variables
  – Infant records
    • Live birth or stillbirth (pregnancy outcome), birth date, sex, postal code, birth weight, gestational week at birth
  – Maternal records
    • Mother’s birth date
    • Number of fetuses
    • Induction of labour: intervention code (“5.AC.30”) in ICD-10-CA
    • Cesarean delivery: the intervention code (“5.MD.60”) in ICD-10-CA

• Analysis methods
  – Percentage of agreement and Kappa coefficients (simple or weighted)
Methods

**Kappa agreement assessment criteria**

- < 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

References:
Results

Table 1. Distribution of live birth or stillbirth in BIS and CIHI-DAD in linked dataset

<table>
<thead>
<tr>
<th>CIHI-DAD</th>
<th>Live birth</th>
<th>Stillbirth</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live birth</td>
<td>404,291</td>
<td>61</td>
<td>404,352</td>
</tr>
<tr>
<td>Stillbirth</td>
<td>0</td>
<td>87</td>
<td>87</td>
</tr>
<tr>
<td>Total</td>
<td>404,291</td>
<td>148</td>
<td>404,439</td>
</tr>
</tbody>
</table>

Percent agreement between CIHI-DAD and BIS: 99.98%
Kappa coefficient agreement between CIHI-DAD and BIS: 0.74, 95% CI (0.68-0.80)

*Note: Out of all stillbirths, 148 records in the BIS with valid OHIP were linked*
# Results

## Table 2. Agreement between BIS and DAD among linked birth records

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>Percent agreement (%)</th>
<th>Kappa coefficient</th>
<th>95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Kappa test</td>
<td>Lower limit</td>
<td>Upper limit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baby's date of birth</td>
<td>404,439</td>
<td>99.9%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Baby's sex</td>
<td>404,439</td>
<td>99.4%</td>
<td>0.9885</td>
<td>0.9880</td>
<td>0.9889</td>
</tr>
<tr>
<td>Baby's weight*</td>
<td>403,396</td>
<td>95.5%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Gestational age at birth (weeks) * ,§,£</td>
<td>404,005</td>
<td>98.4%</td>
<td>0.9195</td>
<td>0.9186</td>
<td>0.9204</td>
</tr>
<tr>
<td>Baby's postal code*</td>
<td>404,164</td>
<td>91.8%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

*: missing values (less than 1%) were excluded  
§: percent agreement of gestational age (weeks) at birth calculation, agreement was defined as within 1 week of difference in gestational age  
£: weighted Kappa coefficient was reported
### Table 3. Agreement between BIS and DAD among linked maternal records

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>Percent agreement (%)</th>
<th>Kappa coefficient</th>
<th>95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower limit</td>
<td>Upper limit</td>
</tr>
<tr>
<td>Mom's date of birth*</td>
<td>403,059</td>
<td>99.1%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Induction of labour*</td>
<td>402,552</td>
<td>92.1%</td>
<td>0.7823</td>
<td>0.7800</td>
<td>0.7845</td>
</tr>
<tr>
<td>Cesarean delivery *</td>
<td>403,059</td>
<td>99.8%</td>
<td>0.9952</td>
<td>0.9949</td>
<td>0.9956</td>
</tr>
<tr>
<td>Number of fetuses*, §</td>
<td>403,007</td>
<td>99.9%</td>
<td>0.9787</td>
<td>0.9770</td>
<td>0.9805</td>
</tr>
</tbody>
</table>

*: missing values (less than 1%) were excluded

§: weighted Kappa coefficient was reported
Results

- Excellent percent agreement (≥ 90%) for all assessed key elements
- Substantial to almost perfect agreement on Kappa test
  - Substantial agreement (0.61–0.80): *stillbirth or live birth, induction of labour*
  - Almost perfect agreement (0.81–0.99): *cesarean delivery, number of fetuses, baby’s sex*

References:
Discussion

• Potential reasons for discrepancies could include
  – Slight difference in some elements’ definitions
  – Different time points of data entry
  – Different data element sources
  – Errors from manual data entry
Next steps...

• Data abstraction study to explore the potential reasons for discrepancies based on hospital chart review

• Report to BORN’s data quality committee and discuss strategies

• Future work on data quality
  – A review of the data quality framework
  – Implement strategies to improve data quality and perform random audits
Conclusion

• Overall, the BIS and CIHI-DAD had a high concordance on main birth/maternal elements

• The criteria and methods for data collection are consistent in two systems

• More work is needed to understand discrepancies
Acknowledgements

• BORN staff, especially:
  - Catherine Riddell
  - Deborah Weiss
  - Nicole Roberts

• CIHI
Thank you!
Questions?

www.BORNOntario.ca

Follow us on Twitter @BORNOntario