Author’s response to reviews

Title: Infant feeding practices within a large electronic medical record database

Authors:

Emily Bartsch (emily.bartsch@mail.utoronto.ca)
Alison Park (alison.park@ices.on.ca)
Jacqueline Young (jacqueline.young@ices.on.ca)
Joel Ray (rayj@smh.ca)
Karen Tu (karen.tu@ices.on.ca)

Version: 2 Date: 28 Jun 2017

Author’s response to reviews:

Reviewer 1:

1) p. 5 line 18 change "indicated" to "indicate"

The text now reads “indicate”.

2) p. 5 line 19 "Small for gestational age birthweight" is not a typical term. Delete birthweight after “Small for gestational age"

“Birthweight” was removed.

3) In Table 1. Specifying the amounts in each income quintile would add clarity. Could add to legend.

This is not straightforward because the neighbourhood income quintiles calculated by Statistics Canada are community-specific; for instance, the absolute value of Q1 in the GTA is not the same as the value of Q1 in Thunder Bay. As stated within the Postal CodeOM Conversion File User’s Guide (Postal CodeOM Conversion File (PCCF), Reference Guide, 2013. Statistics Canada Catalogue no. 92-154-G.), “The quintiles were defined within each area in order to better reflect the relative nature of this measure, to minimize the effect on household welfare of large differences in housing costs, and to ensure that each census metropolitan area or census agglomeration would have about an equal percentage of the population in each income quintile.”

Reviewer 2:
1) My main issue with this paper is that I am not sure what the main, overarching point of it was. Was it to demonstrate that EMR data held in the EMERALD database is good enough to be used for observational research in this field or was it to determine the rates of breastfeeding in this cohort and associated factors? These are quite different research questions and are not entirely compatible with each other. The background section suggests the latter aim, but the conclusion suggests the former aim.

If the aim is the latter (to describe breastfeeding), this kind of work has been done many times before and in similar many cohorts, so is not particularly ground-breaking - although it will be of local interest. If however, this paper is more about validating the use of EMR data for such research, this is potentially more interesting to me as work validating and describing such routinely-extracted data sources is lacking. I would like to see the authors focus more on one of these aims explicitly.

If the paper does seek to validate the EMRALD database for observation research in relation to infant feeding, it would be good to more clearly define a "gold-standard" against which the data are to be compared (or to discuss why there is not one available). The authors could then focus on demonstrating the data quality (data completeness, accuracy, etc.) for the EMRALD dataset in relation to this particular question, and then examine the generalizability of EMRALD vs Ontario more widely and test for systematic differences.

The intent of this paper was to look at breastfeeding rates amongst mothers in Ontario and to examine characteristics associated with breastfeeding drawn from a relatively unbiased data source. Our review of the literature did not find this kind of reporting drawn from primary care records. Previous studies of infant feeding in Ontario and across Canada have been limited by small sample size or limited longitudinal follow-up. One Canadian study analysed infant feeding up to 16 weeks postpartum, through self-reported questionnaires but included only 1184 mothers (1). Two other studies evaluating infant feeding in Ontario and Canada, respectively, had sample sizes of 2560 and 6421 (2, 3). One group included over 90,000 births; however, these data were limited to feeding practices only at discharge, with no subsequent follow-up (4, 5). Our study using the EMRALD™ cohort is robust because it captures the feeding practices of nearly 9000 infants born in Ontario, with follow-up to 6 months after birth. It may be more representative of the general Ontario population, as it draws upon the secondary use of family physician records as recorded in routine clinical practice and is not subject to sampling bias and reporting bias such as with a self-report survey, which may be over-representative of only English-speaking mothers or those having particular interest in this topic. We are unaware of any comprehensive registry capturing this type of information in Ontario or in other provinces in Canada.


2) The numbers are quite small - only 8,815 infants in the EMRALD dataset vs 1.2 million births. And only about 88% of the 8,815 infants had data on breastfeeding.

□ The numbers are relatively small because the data are only from GP practices that have enrolled in EMRALD™, and therefore do not include all births. However, the practices represent a diversity of geographical areas across Ontario. As data on infant feeding practices is often unavailable, we believe that having feeding information on 88% of infants in a sample, as in our study, is quite high.

□ Additionally, our study is unique in that it comprises a large population and includes follow-up to 6 months after birth (see response #1).

3) The authors state that the EMRALD population "were highly similar to those across Ontario" - I disagree with this statement; there are potentially very important difference between the EMRALD cohort and Ontario births in most of the characteristics presented in table 1 (rural residence, maternal region of birth etc.). The way the authors have presented the differences (Standardized difference) seems misleading to me and is not described in detail - I would like to see a more meaningful description of the difference such as absolute difference and CIs.

□ To describe some of the differences of mothers in the EMRALD™, we have now changed this sentence to read: “The infants and mothers in EMRALD™ were similar to those across Ontario in terms of characteristics such as age and parity, however mothers in EMRALD™ were somewhat more likely to be a rural resident and Canadian-born, and less likely to be South or East Asian born and in lower income quintiles (Table 1).”

□ A standardized difference of > 10% (0.1) is commonly used to define differences in the distribution of covariates between groups (see for example: Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. Communications in Statistics - Simulations and Computation 2009; 38:1228–1234).

4) Given the differences between the EMRALD population and the Ontario population it would be useful to see the authors apply an approach to account for these differences in their results.

□ Our approach was to describe relative rates of breastfeeding by demographic characteristics as captured in the primary care setting from practices throughout Ontario. Although EMRALD™ is
not exactly representative of the Ontario population, it is likely as close to representative as possible given that it is not from a specialty practice and it is not single center – rather, it comprises multiple clinics and family physicians distributed throughout Ontario.

☐ We have added this sentence to the Discussion, paragraph 3: “Despite its differences from the Ontario population, EMRALD™ may be the most closely representative dataset available, as the data do not come from a specialty practice, and comprises multiple clinics and physicians across Ontario.”

5) The EMRALD dataset are limited to those mother-infant dyads that visited a GP within 750 days - this is clearly not a representative group of all dyads in Ontario - this should be explored more

☐ In Ontario, it is recommended that children have a well-baby/child visit within the 1st week of life, and at months 1, 2, 4, 6, 12-13, 18 and at 2-3 years, and have feeding, growth, immunizations and other health indicators recorded. This is routine practice, therefore, the majority of dyads in EMRALD™ will have visited a GP at least once within the first 750 days of life.

6) Can the authors please be more specific about when each infant had data available and the proportions of the total population that had data at each point (2, 4, 6 months). How many infants had a "complete dataset" and how did they compare to the wider Ontario population?

☐ We performed a subgroup analysis that included only infants who had visits at 2-3 months, 4-5 months and 6 months or later. This analysis included 3959 infants, and we found that, compared to our main analysis, the exclusive breastfeeding rates were the same at 4 months, and ~1% higher at 2 and 6 months.

☐ We have added this sentence to the Methods, paragraph 4: “In addition, we performed a subgroup analysis including only infants who had visits at all three of these time points.”

☐ We have added this sentence to the Results, paragraph 3: “Among those infants that had feeding visits at 2, 4, and 6 months, the rates of exclusive breastfeeding were similar at 4 months, and approximately 1% higher at 2 and 6 months, compared to the main analysis.”

7) Where more than one data point was available for an infant at one time (e.g. there had been more than 1 visit between 2 and 4 months) how did the authors decide which data point to include?

☐ We chose the data point closest to but not preceding the target time point. For example, for 2 month rates, we did not include infants with a visit at 1 month only in the denominator.

8) I am still unclear how the authors extracted data for each point. From the description in the methods section it seems as though any visit before 60 days was used as data for "2 months" - does this mean that a visit on day 3 would be coded as 2 months? Similarly, any visit after 60
days (e.g. 61 days) seems to be coded as 4 months! The authors at the very least need to provide data on the mean and SD age of the data used to describe infants at 2, 4 and 6 months.

We stated in the methods, “The denominators for the rates at 2, 4 and 6 months included infants with any visit at ≥ 60 days, ≥ 122 days, and ≥ 182 days of age, respectively”, meaning that the rate at 2 months was based on infants with data at 60 or more days.

9) The data were extracted in 2 very different ways: either automatically (4955 infants) or by hand searching (2816 infants). It is important to know how different the characteristics and results are in these two groups - there are likely to be systematic differences in both of these.

When we compared the characteristics of patients that were manually extracted to those for whom the information was extracted automatically, the characteristics were similar though the rates of breastfeeding were lower in the manually extracted data. The higher rate of breastfeeding with the automated algorithm was expected, given that all charts were put through an automated data extraction as an initial pass, followed by manual abstraction of those charts where no clear result was found – primarily charts without structured documentation or incomplete documentation. This paper was not intended to be a validation study, but to rather to report on rates of breastfeeding and associated characteristics.

We have added this sentence to the Results, paragraph 2: “The rates of breastfeeding were higher among infants in the automatically abstracted group than those of the manually abstracted group at 2, 4, and 6 months.”

We have revised the Discussion, paragraph 3, such that it now reads: “A few limitations herein need be acknowledged. Even after completing the manual chart abstraction, no feeding information could be obtained for over 1000 records. This suggests that there may be inconsistencies in the way that EMRs, and standardized forms like the Rourke Baby Record, are completed. This highlights some challenges in using EMR data for secondary purposes. A second limitation of using EMR data was the resultant discrepancy in rates of breastfeeding between the infants in the automatic abstraction group compared to the manual abstraction group; however, the higher rate of breastfeeding with the automated algorithm was expected, given that all charts were put through an automated data extraction as an initial pass. Third, although the patients in EMRALD™ appeared to be ethnically diverse, mothers from South Asia and East Asia were under-represented, and rural dwelling women were over-represented in EMRALD™, compared to the entire population of Ontarians. Despite its differences from the Ontario population, EMRALD™ may be the most closely representative dataset available, as the data do not come from a specialty practice, and comprises multiple clinics and physicians across Ontario. Last, we were unable to assess the impact of maternal education or employment on exclusive breast-feeding rates, as this information was not available in our data.”

10) In the results (Table 2) there seem to be many difference factors that have been tested for significance in relation to breastfeeding at multiple different points. The authors should explain how they have dealt with multiple testing here.

We did not account for multiple testing.
11) The conclusion states that "primary care EMR data can be used to assess breast-feeding practices in large populations" - I do not feel that this conclusion is warranted from the data presented. To convince me of this statement the authors need to more clearly define the comparator against which they have examined the EMRALD database and clearly explain the data quality of EMRALD as well as considering the generalisability of EMRALD data to the larger population, qualify the systematic differences between EMRALD and the wider Ontario population and consider how these would need to be accounted for.

EMRALD™ currently contains data on over 500,000 patients and at the time of the study was ~300,000 patients. Although only containing data on a small percentage of the 13 million people in the Ontario population, a data set with the level of detail as contained in EMRALD™ is large. The Canadian Community Health Survey, considered a “large” survey, contains data on 130,000 Canadians across the country so we imagine the Ontario participants are only in the 10’s of thousands. We believe that EMRALD™ is the largest provincial database of its kind in Ontario. Although in Canada there is one other larger EMR database (CPCSSN), it does not contain full chart data as EMRALD™ and this type of study would not be possible in that database. The quality and comprehensiveness of EMRALD™ data has been assessed in the past (1, 2). Although EMRALD™ differs on some characteristics compared to the Ontario population, disease prevalence is similar (1). We believe the characteristic differences of EMRALD™ compared to Ontario is more reflective of the type of people that seek medical care (more females, higher SES) and not something in particular of EMRALD™ participating physicians and patients.


We have changed our Conclusion, such that it now reads: “Over 85% of newborns registered within a large primary care EMR received at least some breastfeeding at age two months. Exclusive breastfeeding was more likely among mothers who were older and residing in an urban and higher income quintile neighbourhood, and among immigrant women living in a lower income quintile neighborhoods. Our study suggests that primary care EMR data can be used to assess breastfeeding practices in large populations.”