Introduction
Public Health Ontario (PHO) undertook a data quality assessment of the BORN Ontario public health data cube, examining accuracy, timeliness, usefulness, and relevancy to public health. The main objective of this assessment was to contribute to the understanding of the data quality of the BORN information system for routine public health monitoring.

High levels of missingness in the BORN data cube have been observed for a number of reproductive health indicators key to public health (1). This project sought to determine how missing data is distributed across public health units, peer groups, and across a neighbourhood measure of socioeconomic status (SES), (i.e., the dimensions of the Ontario Marginalization Index (ON-Marg)).

High levels of missing data can result in inaccurate values for important public health indicators, particularly if missingness is not random but differs by important factors such as geography and SES.

Methods
Data from the BORN public health data cube representing all live births in 2013 was used to analyse current and proposed reproductive and child health indicators developed by the Association of Public Health Epidemiologists in Ontario (APHEO) (2).

For indicators with more than 5% missing data, results were stratified by public health unit (PHU), geographic peer group, and by a neighbourhood measure of SES - the Ontario Marginalization Index (ON-Marg) which is calculated based on data from the 2006 census and includes four dimensions of marginalisation: material deprivation, ethnic concentration, residential instability, and dependency (3).

ON-Marg quintiles were assigned throughout linking with census dissemination area. Per cent missing data by ON-Marg quintiles was measured and differences were examined. The slope index of inequality (SII) was calculated for Ontario and selected PHUs to determine whether there were significant differences across quintiles of the ON-Marg dimensions. PHUs were selected to be representative of each peer group. The SII was used to report the absolute difference in missing data between the highest and lowest SES in the population. Negative values for the SII indicate higher levels of missing data among the least marginalized neighbourhoods and a positive value represents higher levels of missing data among the most marginalized neighbourhoods.

Results

Missingness in BORN
Missingness varied considerably across the APHEO core indicators, ranging from 0% missing data for stillbirths, perinatal mortality and multiple birth rate, to 31.4% missing data for maternal weight gain group (See Table 1 for indicators with missingness greater than 5%).

PHU and Peer Group Ranks
Across indicators, a number of PHUs consistently showed high missing data. High missingness was more frequently observed in the following peer groups: urban centres (peer group B), metro centres (G), and rural northern regions (H). Lower missingness was consistently observed in urban/rural mix regions (A), sparsely populated urban/rural mix (C), and mainly rural regions (J) (see Figure 1).

Missingness across Neighbourhood SES
In general, patterns of socioeconomic disparities in data missingness were consistent across indicators. In Ontario, across all indicators, higher ethnic concentration was strongly associated with higher missingness, higher residential instability was often associated with higher missingness, higher material deprivation was associated with lower missingness, and higher dependency was associated with lower missingness.

Inequalities in missingness across ON-Marg dimensions are likely influenced by geographic factors, for example, PHUs with high missingness (e.g., metro and urban centres) also often have higher scores for ethnic concentration. Results were examined at the local level for selected PHUs representing each peer group. When examining results at the PHU level disparities in missingness across ON-Marg dimensions were sometimes significant but across different PHUs no consistent pattern in missingness by ON-Marg dimension was observed, suggesting that SES alone does not directly influence data missingness (see Figure 2).

Discussion
For a number of core child and reproductive health indicators, data missingness is substantial. There were large differences in data missingness by geography with higher missingness observed in metro centres and cities.

Reasons for higher missingness in certain urban/metro centres may in part be influenced by low levels of reporting in a few high birth hospitals. For all reproductive indicators in Ontario, disparities in data missingness were found across ON-Marg dimensions. However, a causal relationship has not been established and patterns of inequalities in Ontario were not observed at the PHU level, suggesting that neighbourhood SES does not directly influence missingness.

When using data from BORN it is important to understand which populations may be under-represented due to high missingness. In Ontario overall it is clear that certain neighbourhoods (e.g., those with high ethnic concentration, low dependency) as well as urban and metro centres, generally have high missing data. Understanding these differences at the provincial and PHU level will be important for properly interpreting information from BORN.

References