



Data Quality and Confidentiality Standards and Guidelines (Public)

2011 National Household Survey

Table of contents

	Page
1. Introduction.....	3
1.1 Background.....	3
2. Confidentiality (non-disclosure) rules.....	3
2.1 Area suppression for standard ¹ and non-standard geographic areas.....	3
2.1.1 Population universes used for suppression routines.....	4
2.2 Postal Code minimum aggregation rules.....	5
2.3 Random rounding.....	5
2.4 Disclosure avoidance for statistics.....	6
2.4.1 Statistic suppression.....	6
2.4.2 Special statistic calculations.....	7
2.5 Suppression of NHS estimates for confidentiality protection.....	8
3. Confidentiality practices.....	10
3.1 Area suppression for income characteristic data.....	10
3.2 Confidentiality adjustment for place of work estimates.....	10
3.3 Confidentiality adjustment for daytime population estimates.....	10
3.4 Preventing disclosure.....	11
3.5 Census of Agriculture tabulations.....	11
3.6 Public Use Microdata Files (PUMFs).....	11
4. Data quality practices.....	12
4.1 Data quality measures.....	12
4.1.1 Data quality indicators for tabulations based on place of residence geographies.....	12
4.1.1.1 Data quality indicators.....	12
4.1.1.1.1 Incompletely enumerated areas.....	13
4.1.1.1.2 Partially enumerated areas.....	13
4.1.1.1.3 Global non-response rates.....	13
4.1.2 Data quality indicators for tabulations based on place of work geographies.....	14
4.2 Other methods of data quality suppression.....	15
4.2.1 Distribution suppression.....	15
4.3 Calculation of order statistics.....	15
4.4 Data quality rule for disseminating data for population aged 100 and over.....	16
4.5 Data quality rule for disseminating data on same-sex and opposite-sex couples.....	16
5. Suppression – Indian reserves.....	17
5.1 Indian reserve N2 suppression.....	17
6. Data suppression – Other.....	18
6.1 Incidence reporting.....	18
6.2 Zero suppress.....	18
6.3 Place of work flow suppression.....	18
6.4 Data quality and confidentiality table symbols.....	19

1. Introduction

1.1 Background

Information previously collected by the long-form census questionnaire was collected as part of the new voluntary National Household Survey (NHS). Approximately 4.5 million households received the NHS questionnaire.

Data disseminated by the NHS are subjected to a variety of automated and manual processes to determine whether the data needs to be suppressed. This is done primarily for two reasons, (1) to ensure non-disclosure of individual respondent identity and characteristics (which will subsequently be referred to as **confidentiality**) and (2), to limit the dissemination of data of unacceptable quality (which will subsequently be referred to as **data quality**).

Additionally, suppression of data may be applied for product-specific reasons due, typically, to formatting issues. The term **product** refers, primarily, to tabular output. Data may either be modified in the product or removed from the product altogether to reflect the suppression rules required.

2. Confidentiality (non-disclosure) rules

The following describes the various rules used to ensure confidentiality (or non-disclosure) of individual respondent identity and characteristics. All NHS data are subject to confidentiality (non-disclosure) rules.

2.1 Area suppression for standard¹ and non-standard geographic areas

Area suppression is used to remove all characteristic data for geographic areas below a specified population size.

The specified population size for all standard¹ areas or aggregations of standard areas is 40, except for blocks, block-faces or postal codes. Consequently, no characteristics or tabulated data are to be released for areas below a population size of 40.

1. For more information on standard areas, refer to the 2011 *Census Dictionary*.

The specified population size for six-character postal codes (forward sortation area – local distribution unit [FSA-LDU]), geocoded areas and custom areas built from the block, block-face or LDU levels is 100. Consequently, no characteristics or tabulated data are to be released if the total population of the area is less than 100. Generally, blocks and individual urban block-faces (one side of the street between two intersections) will be too small to meet the above-specified population size thresholds. Where an aggregation of blocks or block-faces fall above the threshold specified by the population size, data can be retrieved through a custom tabulation.

Please refer to section 2.2 Postal code minimum aggregation rules for additional rules applicable to postal code data.

2.1.1 Population universes used for suppression routines

The population under consideration for all data tabulations is the NHS estimate of population in private households.

For place of work data, the population under consideration is the employed labour force having a usual place of work or worked at home.

Population universes used for suppression routines

NHS	POW geographic areas
Estimate of population in private households	Employed labour force having a usual place of work or worked at home

For NHS tabulations that are based on place of work geographies or areas, all criteria are to be based on estimates of the employed labour force having a usual place of work or worked at home. That is, the 40 population, 100 population and 250 population thresholds are estimates of the employed labour force having a usual place of work or worked at home, rather than the population of the areas. Tabulations containing both places of residence and places of work as geographic areas have the 40, 100 and 250 size limits applied to both place of residence (population) and place of work (employed labour force having a usual place of work or worked at home).

2.2 Postal Code minimum aggregation rules

In addition to the confidentiality rules on disseminating National Household Survey data with the postal codes, the following rules are applied to postal codes. These rules fall under clause 03.01 (n) of the Commercial Non-Mailing licence between Statistics Canada and Canada Post Corporation.

- All requests must include batches of two or more postal codes; the only exception being for postal codes which have a zero as the second digit (rural postal codes);
- Groups of postal codes are to be assigned a unique classification/number (e.g. K1A 0T6, 0T7, 0T8 = Custom Area 1); under the terms of the contract listed above, clients cannot be provided with lists of postal codes, only the name specified in the client's request can be used.
- All other confidentiality rules for custom extractions still apply as per section 2.1.

Also, the following disclaimer is applicable to all postal code custom requests:

Postal code validation disclaimer: Statistics Canada makes no representation or warranty as to, or validation of the accuracy of any postal code^{OM} data submitted to Statistics Canada.

Please note these rules are applicable to historical postal code requests as well.

2.3 Random rounding

All estimates in NHS tabulations are subjected to a process called random rounding. Random rounding transforms all raw estimates to random rounded estimates. This reduces the possibility of identifying individuals within the tabulations.

All estimates greater than 10 are rounded to base 5, estimates less than 10 are rounded to base 10. This means that any estimates less than 10 will always be changed to 0 or 10. The table below shows the effect of rounding on estimates with a value less than 10.

Table 2.1 Random rounding frequency

Estimate of	Will round to 0	Will round to 10
1	9 times out of 10	1 time out of 10
2	8 times out of 10	2 times out of 10
3	7 times out of 10	3 times out of 10
4	6 times out of 10	4 times out of 10

Estimate of	Will round to 0	Will round to 10
5	5 times out of 10	5 times out of 10
6	4 times out of 10	6 times out of 10
7	3 times out of 10	7 times out of 10
8	2 times out of 10	8 times out of 10
9	1 time out of 10	9 times out of 10
0	Always	Never

The random rounding algorithm uses a random seed value to initiate the rounding pattern for tables. In these routines, the method used to seed the pattern can result in the same estimate in the same table being rounded up in one execution and rounded down in the next.

2.4 Disclosure avoidance for statistics

Statistics (such as mean, sum, median, percentile, ratio or percentage) are not subject to random rounding. However, when shown in tabulations accompanying the estimates used to calculate the statistic, their presence can result in disclosure of individuals. To prevent this, we use statistic suppression methods or special statistic calculations.

2.4.1 Statistic suppression

The following three situations will result in the suppression of statistics:

- (1) Statistics for a cell will be suppressed if the range of data (i.e., the maximal dollar amount of the cell minus the minimal dollar amount) over the maximum of absolute values is below a threshold parameter. This method of suppression is applied only to statistics calculated from quantitative values measured in dollar (\$) units such as income or value of dwellings.
- (2) For all quantitative variables, a statistic is suppressed if the number of actual records used in the calculation (not rounded or weighted) is less than 4. For quantile statistics, an alternate minimum number of records apply: for quartiles, quintiles and deciles, 20 records are required, and for percentiles, 400 records are required.
- (3) Statistics for a cell will be suppressed if it contains an outlier. A cell is considered to contain an outlier if the largest absolute value divided by the sum of the absolute values is above a threshold parameter.

Note: The number of records used in the calculation is not necessarily the number of records in the cell but, rather, the number of records that are applicable or available to the calculation of the statistic in the cell.

Example:

Consider a cell containing the following records:

The eight records in the cell represent 47.6 persons (the sum of the weights). Since for the variable 'Wages' only non-zero values are used in the calculation, the average \$22,727.27 will be suppressed because only three records are used in the calculation.

Example of eight records showing the weight applied and wages of each respondent

Record number	Weight	Wages (\$)
1	5.5	16,500
2	2.9	345,600
3	8.1	12,900
4	6.2	0
5	6.6	0
6	5.9	0
7	5.4	0
8	6.9	0

(4) For all quantitative variables, all statistics are suppressed if the sum of the weights is less than 10.

2.4.2 Special statistic calculations

- (1) The statistic value is never rounded, except for frequencies.
- (2) All statistics based on ranks (medians, percentiles) are calculated the usual way, that is, never rounded.
- (3) When a sum is specified, if the program sums a dollar value, a number of weeks, a number of hours, or an age, then the program multiplies the unrounded average of the group in question by the rounded, weighted frequency. Otherwise, the program rounds the actual weighted sum.

When a division is specified (averages, percentages, ratios, etc.), the program must apply the point (3) to both numerator and denominator before it proceeds with the division.

Note: Statistics based on ranks like median and percentiles are always calculated via linear interpolations. That means that, for cells with low estimates, these statistics are not reliable. That is the reason why no additional confidentiality measures are applied to them.

Note: The average of dollar value, a number of weeks, a number of hours or an age is not altered by the rounding because the numerator is the product of the true average by the rounded frequencies and the denominator is the rounded frequencies. The two frequencies cancel each other leaving the true average untouched.

2.5 Suppression of NHS estimates for confidentiality protection

Section 2.3 discussed random rounding for estimates in NHS tabulations. Random rounding is used as a means of protecting confidentiality in estimates. Analysis of NHS data revealed that even with random rounding in place, in some cases, data with elevated risks of disclosure could be released.

These elevated risks arise because non-response adjustment in the NHS required a relatively wide range of weights. High weights may enable individuals with rare characteristics to be more easily identified in a table, particularly if their characteristics are publicly known.

To minimize these risks, a rule was instituted for estimates that are similar to the rule for quantitative variables described in Section 2.3. A cell estimate will be suppressed if the number of records with the attribute or combination of attributes represented by the cell (unrounded and unweighted) is less than 4. In these cases, the cell will show the number 0 instead of the suppressed value, and thus will be indistinguishable from a genuinely empty cell.

Example:

Suppose we have the following records for a given geography:

Example of 15 records showing the weight applied and the age of each respondent

Record number	Weight	Age
1	6.5	20
2	4.9	22
3	8	25
4	6.8	26
5	5.4	27
6	6.1	27
7	4.7	27
8	5.7	29
9	2.8	32
10	6.8	36
11	41.1	39
12	5	39
13	81.4	40
14	5.1	50
15	3.2	54

Applying only random rounding, the NHS estimates would be published (randomly rounded) as illustrated in the following table. The number of records would never be published, but is only for illustrating the effect of the rule.

Example of estimates that would be published without applying cell suppression

	Age range				
Values	20 to 29	30 to 39	40 to 49	50 to 59	Total
NHS estimate	50	55	80	10	195
Number of records	8	4	1	2	15

Suppression based on cell estimates suppresses both the 40 to 49 and 50 to 59 age ranges, as fewer than four (4) records have the attribute in question, and result in the following table. The total remains unchanged as the total cell represents at least four individuals.

Example of estimates that would be published with cell suppression applied

	Age range				
Values	20 to 29	30 to 39	40 to 49	50 to 59	Total
NHS estimate	50	55	0	0	195
Number of records	8	4	1	2	15

The primary intent of the rule is to prevent disclosure of personal information related to certain individuals. Within the example above, if there is only one individual in the area in question who is in the 40 to 49 age range, it is not divulged that they are an NHS respondent, and thereby it minimizes the risk that further information from the NHS on this individual is disclosed.

3. Confidentiality practices

3.1 Area suppression for income characteristic data

Area suppression is used to replace all income characteristic data with an 'x' for geographic areas with populations and/or number of households below a specific threshold.

If an NHS tabulation contains quantitative income data (e.g., total income, wages), qualitative data based on income concepts (e.g., low income before tax status) or derived data based on quantitative income variables (e.g., indexes) for individuals, families or households, then the following rule applies: income characteristic data are replaced with an 'x' for areas where the population is less than 250 or where the number of private households is less than 40. The private household threshold does not apply for tabulations based on place of work geographies.

3.2 Confidentiality adjustment for place of work estimates

The place of work estimates for dissemination blocks are available on a custom basis. These estimates will be adjusted to reinforce the confidential nature of the data. In fact, all dissemination block estimates for employed labour force having a usual place of work or worked at home will be rounded to a base of 5. This adjustment, however, will be controlled. That is, aggregates (totals) of the adjusted population estimates for dissemination areas will always be within 5 of the actual values.

3.3 Confidentiality adjustment for daytime population estimates

Daytime population estimates will be determined by taking the population living in a specific area, adding in the workers who live elsewhere and commute into the area, and subtracting the workers who live in the area and commute out of the area. The number of workers will be based on persons in the employed labour force having a usual place of work or worked at home. Daytime population estimates will be adjusted to reinforce the confidential nature of the data by controlled rounding of the estimates to a base of 5.

3.4 Preventing disclosure

Prevention of direct or residual disclosure must also be addressed when determining product content. When assessing the potential for disclosure, a number of factors must be considered. The detail of individual variables, cross-classification of variables and the geographic level of the data will all contribute to the risk. For example, there may be no risk in producing households by number of rooms in the dwelling and detailed groupings of dwelling value showing various characteristics of the household members for large geographic areas. However, the risk of disclosure would increase for the lower levels of geography.

The most common method used for preventing disclosure is defining content that is appropriate for a given geographic level. Increasing population thresholds or applying manual suppression as needed are other methods that can be employed. Since these are typically product-specific requirements, they are not part of the automated suppression systems.

3.5 Census of Agriculture tabulations

Census of Agriculture and National Household Survey (NHS) data are matched using geographic information and the characteristics of farm operators (i.e., age and sex). Match rates are about 95% and weighting is performed to account for non-matches. Data are available for all members of households where a farm operator resides.

Census of Agriculture data include farm type, farm sales, area of crops and numbers of livestock while the NHS provides socioeconomic data, including education, income and occupation of families and household members. Pre-planned standard products are produced at the province level only.

Custom products are available at provincial levels. The data are random-rounded and low-bounded to ensure confidentiality. Suppressions are done manually if cells are below a specified size.

All verification of tables is done internally by Census of Agriculture staff.

3.6 Public Use Microdata Files (PUMFs)

The 2011 NHS PUMF products will consist of two microdata files: the individual file and the hierarchical file. The individual file will contain records from approximately 3% of the Canadian population and the hierarchical file will contain records from 1% of the population in private households.

Microdata files are unique among NHS products in that they give users access to non-aggregated data. This makes PUMF a powerful research tool. The files contain a large number of variables. Users can

group and manipulate these variables to suit their own data and research requirements. Tabulations not included in other NHS products can be created, or relationships between variables can be analyzed using various analytical tools.

The NHS Public Use Microdata Files (PUMFs) provide quick access to a comprehensive social and economic database about Canada and its people. They consist of samples of anonymous responses to the NHS questionnaire (Forms N1, N2). The PUMFs contain statistical information about Canadians, the families and households to which they belong and the dwellings in which they live and allow researchers to study the relationships between these universes.

Statistics Canada has to protect the confidential information that it collects. Owing to the very nature of a microdata file, various measures are taken to fulfil this commitment. The Microdata Release Committee reviews all requests for release of microdata.

Data for small geographic areas will not be available in these files. The user will find information only for selected census metropolitan areas, the provinces and the territories. Variable data is aggregated to preserve confidentiality while providing as much detail as possible to maintain the analytical value of the file. Some of the values of sensitive variables are suppressed because their combination could be used to identify a person, a family or a household. Also, all quantitative dollar value variables are subjected to random rounding and are top and bottom coded.

4. Data quality practices

The following section describes the methods used to restrict the dissemination of NHS data of unacceptable quality.

4.1 Data quality measures

4.1.1 Data quality indicators for tabulations based on place of residence geographies

4.1.1.1 Data quality indicators

Data quality indicators (commonly referred to as data quality flags) are attached to each place of residence standard geographic area disseminated. In the NHS database environments, the data quality indicators consist of a five-digit numeric field. On the database and in electronic products browsed via Beyond 20/20, these flags are displayed as a five-digit numeric code (example: 1 0 0 1 0). On the NHS website, flagging of partially enumerated areas to end users is done through the use of symbols. Specific symbols in use for the 2011 NHS are documented in Section 6.4.

4.1.1.1.1 Incompletely enumerated areas

In 2011, there were a total of 36 Indian reserves and Indian settlements that were 'incompletely enumerated' in the NHS. For these reserves or settlements, NHS enumeration was either not permitted or was interrupted before it could be completed, or was not possible because of natural events (specifically forest fires in Northern Ontario).

There are no data for incompletely enumerated Indian reserves and settlements on the NHS database. Higher-level geographic areas containing these areas are identified in the NHS products.

Although NHS data are not available for incompletely enumerated Indian reserves and settlements, the areas themselves are included as part of the standard geographic hierarchies on the NHS databases. Retrieval and tabulation software will retrieve these areas but with no data. For place of work geographies, these areas will be suppressed.

4.1.1.1.2 Partially enumerated areas

Any geographic area that contains an incompletely enumerated area is considered a partially enumerated area. Partially enumerated areas are flagged to end users as containing incompletely enumerated areas.

4.1.1.1.3 Global non-response rates

The global non-response rate (GNR) is an indicator of data quality which combines complete non-response and partial non-response to the survey. A smaller GNR indicates a lower risk of non-response bias, i.e., a lower risk of lack of accuracy. Global non-response rates are determined for each of the NHS geographic areas. These areas are flagged on the database according to the non-response rate. Geographic areas with a global non-response rate higher than or equal to 50% are suppressed from standard data products but will be available as a custom request. Geographic areas with a global non-response rate lower than 50% are identified in tabulations, but not suppressed. In electronic products, a numeric flag, as well as the actual global non-response rate is provided.

Table 4.1 Data quality indicators for place of residence – 2011 NHS

Digit	Description	Flag	Flag description
1st (0XXXX)	Incomplete enumeration flag	0	Default
		1	Incompletely enumerated Indian reserve or Indian settlement (suppressed)
		2	Excludes National Household Survey data for one or more incompletely enumerated Indian reserves or Indian settlements
2nd (X0XXX)	Not applicable	0	Default
3rd (XX0XX)	Not applicable	0	Default
4th (XXX0X)	Data quality flag	0	Data quality index showing a global non-response rate lower than 50%
		1	Data quality index showing, a global non-response rate higher than or equal to 50% (suppressed)
5th (XXXX0)	Not applicable	0	Default

4.1.2 Data quality indicators for tabulations based on place of work geographies

As indicated in Section 4.1.1.1.3, global non-response rates (GNRs) are determined for each of the NHS geographic areas. Therefore, place of work geographic areas (POW) have their own global non-response rates. POW GNRs are based on the population aged 15 years and over who worked at any given time between January 2010 and May 2011 at a usual place of work or at home, located in the specific place of work geographic area whereas place of residence geographic areas (POR) have global non-response rates based on the population residing in the area. Consequently, place of work geographic areas might have different global non-response rate values when compared to their equivalent place of residence geographic area. For example, the global non-response rate for the place of work census subdivision of Toronto might not be the same as the global non-response rate for the place of residence census subdivision of Toronto.

POW GNRs like POR GNRs are an estimate, not an absolute metric, and both GNR values are variable. However, POW GNRs are more variable than POR GNRs, in precisely the same way that POW population estimates, not being calibrated to a known POW population enumerated through the census, will be more variable than POR population estimates, which are calibrated to known populations enumerated through the census.

As is the case for place of residence geographic areas, data for place of work geographic areas with a global non-response rate of 50% or above will be suppressed in standard products, but will be available as a custom request. However, it is important to note that in standard products, data might be available

for some place of residence geographic areas (if their global non-response rate is below 50%) but not for the equivalent place of work geographic area (if the equivalent place of work geographic area has a global non-response rate of 50% or above), and vice-versa.

The data quality indicator for place of work uses the 4th digit of the five-digit numeric code.

Table 4.2 Data quality indicator for place of work, 2011 NHS

Digit	Description	Flag	Flag description
4th (XXX0X)	Data quality flag	0	Data quality index showing a global non-response rate lower than 50%
		1	Data quality index showing a global non-response rate higher than or equal to 50% (suppressed)

4.2 Other methods of data quality suppression

The methods of suppression mentioned to this point provide sufficient data quality suppression and identification for most NHS data products. However, in some products, the specifying area or production area may require that additional data quality suppression be performed. Examples of additional suppression could include increasing population thresholds or applying distribution or cell suppression. These are typically product-specific requirements and therefore are not part of the automated suppression systems. In all cases, some form of manual process is required.

4.2.1 Distribution suppression

The most common example of other methods of data quality suppression is distribution suppression. This occurs in selected **standard income products** where income distributions are suppressed when the **total number of units** (persons, families, households) within the income distribution is less than **250**. A variation of this procedure is applied to standard income products that feature only number, median and average statistics for employment or total income only.

4.3 Calculation of order statistics

Medians and more generally quantiles are calculated using linear interpolations. The quantile interval (that is the interval where the value of the quantile is located) is determined using two methods based on the kind of values of the statistical variables:

- (1) Variables that take values with decimals and any variables with dollar values

The quantile interval is constructed to ensure that relative errors made by using the linear interpolation are less than 0.78%. For example, if the true quantile is \$30,000.00, the error made by using the built-in algorithm is less than \$234.00.

- (2) Variables that take integer values that are not dollars

For these variables, the quantile interval is always of size 1. For example, if the true quantile is 23.46, the interpolation is applied to the interval [23, 24].

4.4 Data quality rule for disseminating data for population aged 100 and over

Data for the population aged 100 years and over cannot be disseminated in single years of age. For custom requests that require a more detailed breakdown than provided in standard data products, in which the population aged 100 years and over is grouped together, the most detailed age breakdown which can be provided is as follows, and it can only be provided for 'Canada':

Total population 100 years and over

100 years to 104 years

105 years to 109 years

110 years and over

4.5 Data quality rule for disseminating data on same-sex and opposite-sex couples

The questionnaires of the 2011 Census of Population and the 2011 National Household Survey introduced for the first time a specific response on household relationships to determine the number of same-sex married couples. Analysis of the data on same-sex married couples has shown that there may be an overestimation of this family type and marital status. The 2011 National Household Survey shows a total of 63,920 same-sex couples in Canada, of which 20,280 are married couples. The range of overestimation of both these estimates, at the national level, is between 0 and 3,800.

For levels of geography such as Canada, provinces, territories and census metropolitan areas (CMAs), estimates are generally higher, so the potential overestimation is expected to be small in relative terms; however, the data should still be interpreted with caution.

At lower levels of geography, the same potential overestimation could be relatively large, and not only should the data be interpreted with caution, but certain suppression rules restrict their publication. These rules apply to both the 2011 Census and the 2011 National Household Survey.

First, the breakdown of same-sex couples or opposite-sex couples by conjugal status, that is, whether they are married or living common law, cannot be disseminated for geographic areas other than Canada, provinces, territories and CMAs.

Second, data cannot be disseminated that identify either same-sex or opposite-sex couples (in total, married or living common law) of any area with a population of less than 5,000 (as measured in the 2011 NHS).

In summary,

- All data may be disseminated for same-sex or opposite-sex couples for Canada, provinces, territories, census metropolitan areas (CMAs), although they should still be interpreted with caution.
- Data on same-sex couples and opposite-sex couples may be disseminated for other geographic areas if they have a population of 5,000 or more, provided that the breakdown by conjugal status (married, living common law) is not included.
- No data may be disseminated that identify any same-sex or opposite-sex couples for areas of population less than 5,000.

5. Suppression – Indian reserves

5.1 Indian reserve N2 suppression

Suppression of data also occurs when certain questions are not asked of all respondents. Persons living on Indian reserves and Indian settlements who were enumerated with the 2011 NHS N2 questionnaire were not asked the questions on citizenship (Question 10), landed immigrant status (Question 11) and year of immigration (Question 12). However, it was possible that a census subdivision (CSD) or lower geographic area was enumerated using both the N2 questionnaire (for the on-reserve population) and the N1 questionnaire (for the off-reserve population). In this case, the following rules were used to determine if suppression had to be applied to all citizenship and immigration data for that CSD (or lower geographic area):

- (1) If the population estimate from N1 questionnaires is higher than the population estimate from N2 questionnaires (based on weighted results), then include citizenship and immigration estimates.

If the population estimate from N2 questionnaires is higher than or equal to the population estimate from N1 questionnaires (based on weighted results), then exclude citizenship and immigration estimates.

Consequently, citizenship, landed immigrant status and period of immigration data are suppressed for Indian reserves and Indian settlements at census subdivision and lower levels of geography where the majority of the population was enumerated with the N2 form. These data are, however, included in the totals for larger geographic areas, such as census divisions and provinces.

For a complete list of Indian reserves and Indian settlements for which citizenship, landed immigrant status and period of immigration data are suppressed, please refer to:

http://www12.statcan.gc.ca/nhs-enm/2011/ref/sup_N2-eng.cfm.

6. Data suppression – Other

As indicated in Section 5, suppression of data occurs when certain questions are not asked of all respondents. Additionally, suppression of data may be applied for product-specific reasons due, typically, to the size of the product and/or the constraints of the media on which the product is being disseminated.

6.1 Incidence reporting

Incidence reporting is a process used to order or rank characteristic data by size within products. It can be used as a method to select only the 'n' highest categories of a characteristic for inclusion in a product.

6.2 Zero suppress

Zero suppress refers to the removal of records in which all of the estimates are equal to zero. This method is used to reduce the size of an output product by removing any rows of the output matrix where all data are equal to zero.

6.3 Place of work flow suppression

The following is a guideline only therefore its implementation in all flow tables is not mandatory. The primary purpose is for its application to standard data products displayed on the Internet.

Flow suppression is a process used to remove records with very low estimates from place of work flow tables. It can be used as a method to select only those commuting flows with an estimate greater than a threshold value for inclusion in a product. The default value is 20. These are typically product-specific requirements and therefore are not part of the automated suppression systems.

6.4 Data quality and confidentiality table symbols

NHS standard products will contain two Statistics Canada [standard table symbols](#). The table below shows each symbol and its description.

Data quality and confidentiality table symbols

Symbol	Description
...	not applicable
x	suppressed to meet the confidentiality requirements of the <i>Statistics Act</i>

Note: When statistic suppression is applied (e.g., MMM, record count, estimate, outlier), a zero is outputted instead of the 'x' symbol. This is done to ensure there is no risk of residual disclosure.