

Applying confidentiality rules to 2018 Census data and summary of changes since 2013





Crown copyright ©

[See Copyright and terms of use](#) for our copyright, attribution, and liability statements.

Citation

Stats NZ (2019). *Applying confidentiality rules to 2018 Census data and summary of changes since 2013*.

Retrieved from www.stats.govt.nz.

ISBN 978-1-98-858360-0 (online)

Published in September 2019 by

Stats NZ Tatauranga Aotearoa

Wellington, New Zealand

Contact

Stats NZ Information Centre: info@stats.govt.nz

Phone toll-free 0508 525 525

Phone international +64 4 931 4600

www.stats.govt.nz

Contents

Purpose and summary	4
Purpose	4
Summary of key points	4
About the 2018 Census confidentiality rules	5
Applying the rules	5
2018 Census confidentiality rules	6
Rule 1: Geography variable rule	6
Rule 2: Mean cell size rule	7
Rule 3: Threshold rule.....	7
Rule 4: Perturbation rule	7
Rule 5: Proportions from counts rule	8
Rule 6: Suppression for measures rule	8
Rule 7: Noise for measures rule.....	8
How we apply confidentiality rules.....	9
Summary of main changes from 2013 to 2018.....	11
Detail of main changes from 2013 to 2018	12
Geographic variables rule changes	12
Mean cell size rule changes	13
Threshold rule changes.....	13
Perturbation rule changes	13
Proportions from counts rule changes	14
Suppression for measures rule changes	14
Noise for measures rule changes.....	15
Rounding of measures rule changes.....	15
Release under license changes	15
How we reviewed the confidentiality rules	16
More information	16

Purpose and summary

Purpose

Applying confidentiality rules to 2018 Census data and summary of changes since 2013 explains the confidentiality rules and procedures for 2018 Census, and summarises changes since 2013 Census.

Summary of key points

For 2018 Census, we reviewed and updated the 2013 Census confidentiality rules based on internal and external feedback. We developed seven rules with the aim of ensuring the safety of outputs while maximising data utility to our customers.

Rules 1 and 2 are decision rules that determine whether a table is sensitive, and rule 3 suppresses small cells in sensitive tables. Rule 4 adds noise (the amount that is added or subtracted from the true value in order to confidentialise it) to the cells and is always applied. Rule 5 is applied to proportions, ratios, and percentages. Rule 6 decides when to suppress a measure. Rule 7 applies noise to unsuppressed measures.

The overall effect of these rules is that all tables of counts can be released, but some of them (those deemed 'sensitive' by rule 1 or rule 2) will have counts below six suppressed. All remaining counts for release are randomly rounded to base 3 to add noise.

The primary changes since 2013 Census are:

- The meshblock rule is no longer applicable, as Statistical Area 1 geographies (SA1) are now the smallest standard output geography.
- Random rounding is now fixed and made consistent across different tables, to ensure consistency and enable publicly available configurable tables.
- Measures are now calculated from unrounded counts, with noise added afterwards; previously measures were calculated from rounded counts.
- Rounding of measures has been removed as a confidentiality rule, because it is not done for confidentiality reasons.
- Release under license has been discontinued.

About the 2018 Census confidentiality rules

The Statistics Act of 1975 requires Stats NZ and its employees to protect the identity of individuals and organisations contained within statistical outputs. To achieve this, we use a variety of confidentiality techniques to protect individuals and organisations from harm and to maintain public trust in our work.

For 2018 Census, we redeveloped the 2013 Census confidentiality rules based on internal and external review and consultation. We developed seven rules for 2018 Census with the aim of ensuring the safety of outputs while maximizing data utility to our customers.

Our goals for confidentiality are:

- Utility – fulfilling Stats NZ's aim to maximise the use of data.
- Safety – maintaining sufficient levels of confidentiality protection.
- Simplicity – providing procedures that can be efficiently described, understood, automated, and implemented.
- Consistency – providing consistent procedures and methodology across collections and modes of output.

Applying the rules

There are seven main confidentiality rules for the 2018 Census, as well as some supporting rules for special types of data. Rules 1 and 2 are decision rules that determine whether a table is sensitive, and rule 3 suppresses small cells in sensitive tables. Rule 4 adds noise to the cells and is always applied. Rule 5 is applied to proportions, ratios and percentages. Rule 6 decides when to suppress a measure. Rule 7 applies noise to unsuppressed measures.

The overall effect of applying these rules is that all tables of counts can be released, but some of them (those deemed 'sensitive' by rule 1 or rule 2) will have counts below six suppressed. All remaining counts for release are randomly rounded to base 3 to add noise. When we refer to a 'table' for rules 1–3, we mean a very specific type of table – a dimensional table for a single geographic area. If a customer requests a table with two variables (for example, age by sex), then this table can be considered to contain several different dimensional tables (for example, a one-dimensional age table, a one-dimensional sex table, and a two-dimensional age by sex table).

We apply these rules separately to the tables for each geographic area. 'Geographic area' is a generic term used to describe the various standard Stats NZ output geographies, such as statistical area 1 geographies (SA1) and statistical area 2 geographies (SA2), territorial authority areas, and regional council areas.

The geographies covered by the standard 2018 Census confidentiality rules are:

- single standard output geographies at SA1 level or larger
- standard and non-standard groupings of standard output geographies at SA2 level or larger
- standard groupings of meshblocks and SA1 geographies, which are limited to:
 - SA2 geographies, urban rural geographies
 - territorial authorities, regional councils
 - electoral boundaries:

- ward, community board (includes local board), territorial authority subdivision, regional council constituency, or Māori constituency
- general electoral district or Māori electoral district
- police station, police area, or police district
- district health board
- statistical areas or other legal or 'government' boundaries as approved by the Census general manager
- subject population totals of meshblocks.

The rules are applied to all census data we release, and have been automated in the data output systems where possible. Any exemptions to the rules are documented and approved by the Census general manager and/or the Chief Methodologist.

These 2018 Census confidentiality rules are used for the production of tables of counts from 2018 Census data, and the proportions and measures derived from them. They do not fully cover the requirements for protecting unit record data (or 'microdata'). Access to the Integrated Data Infrastructure (IDI) and 2018 Census microdata is granted to approved researchers through the Stats NZ Data Lab. Data Lab processes protect the microdata and regulate outputs in ways that are equivalent to these rules.

[Access our microdata](#) has more information.

2018 Census confidentiality rules

Rule 1: Geography variable rule

The purpose of this rule is to identify sensitive-skewed tables where small groups or individuals can be identified. This rule asks whether two or more geographic variables from the following list are present in a table:

- census night address
- usual residence
- usual residence 1 year ago
- usual residence 5 years ago
- workplace address
- educational institute address.

If two more of these geographic variables are present, the table is deemed sensitive and requires suppression of small cells by rule 3.

Download the Excel file on the webpage for:

- Table 1: Example of a table passing rule 1 (table using one geographic variable and one geographic area)
- Table 2: Example of a table passing rule 1 (table using one geographic variable and two geographic areas)
- Table 3: Example of a table failing rule 1 (table using two geographical variables)

Rule 2: Mean cell size rule

The purpose of this rule is to identify sensitive sparse tables where small groups or individuals can be identified, by looking at the *mean cell size* of each table.

The mean cell size is applied separately at each geographic level and is calculated like this:

$$\text{mean cell size} = \frac{\left(\begin{array}{c} \text{total unrounded subject} \\ \text{population for the geographic area} \end{array} \right)}{\left(\begin{array}{c} \text{number of cells in the} \\ \text{table for one geographic unit} \end{array} \right)}$$

A table is sensitive if the mean cell size is less than or equal to 2. We then apply rule 3 to each sensitive table for each geographic unit.

Exceptions to rule 2

A table may contain the total count for its subject population by geographic area only, and not use any other variables. Such a table is not subject to rule 2, and is treated as not sensitive. Rule 3 is therefore not required, and only rule 4 needs to be applied.

Download the Excel file on the webpage for:

- Table 4: Example of a table passing and failing rule 2 (table using one variable)
- Table 5: Example of a table passing and failing rule 2 (table using two variables)

Rule 3: Threshold rule

When the table for a geographic unit is deemed sensitive by either rule 1 or rule 2, all counts of less than six are suppressed.

The purpose of this rule is to enhance both utility and safety. The larger and more useful counts of six or more can always be released. The smaller and more disclosive counts are protected, where tables are likely to be sparse or related to small subject populations.

Download the Excel file on the webpage for:

- Tables 3 to 5: Examples of suppressed sensitive cells that have failed rule 1 or rule 2.

Rule 4: Perturbation rule

Perturbation, the altering of cells, is performed by applying fixed random rounding to base 3 (FRR3). 'Fixed' means that random rounding is applied consistently across different tables. 'Random rounding to base 3' involves randomly rounding every unsuppressed count in a table to a number divisible by three.

We apply fixed random rounding to base 3 to census tables, by rounding values to:

- the nearest multiple of three with a probability of two-thirds (applied approximately two-thirds of the time)
- the second closest multiple of three with a probability of one-third (applied approximately one-third of the time).

We don't change values that are already multiples of three.

We apply fixed random rounding of all counts to base 3, large and small, to:

- disguise small counts of 0, 1, and 2, which have the highest disclosure risk
- protect against the recalculation of small counts from differencing large counts
- retain almost all of the statistical properties of the original table by adding only a little noise to the larger counts
- allow for **configurable tables**, which are dynamic tables where variables are selected and filtered by customers and built automatically.

Each value in a table is rounded independently, including the totals. This means that the marginal totals can differ slightly from the corresponding sum of the rows or columns. That is, if the columns or rows in a table are added, they will not always equal the total given. Almost all the statistical properties of the table are retained, as the values will never be more than two higher or lower than the original number.

Rule 5: Proportions from counts rule

The purpose of this rule is to prevent the reconstruction of the original unrounded counts.

All proportions, percentages, and ratios (not totals) that come from counts must be calculated using fully confidentialised data. A proportion will only be available for release if both counts in the proportion have not been suppressed.

Rule 6: Suppression for measures rule

A measure is suppressed if the total unrounded count of individuals contributing to the measure is less than the threshold value for that type of measure.

For each measure there is a threshold value:

- for means and medians this threshold value is 6, so any table with less than 6 individuals/households/families etc contributing to it will not have a median or mean shown
- for quartiles the threshold is 12
- for quintiles it is 15
- for deciles it is 30.

Rule 7: Noise for measures rule

The purpose of this rule is to prevent the reconstruction of the original unrounded counts. All measures are calculated from raw, unperturbed counts and have a fixed amount of random noise applied to them. 'Fixed' means that the noise is applied consistently across different tables. 'Noise' is a value that is added or subtracted from the true measure in order to confidentialise it. The noise is calculated differently depending on the type of measure.

How we apply confidentiality rules

Figures 1 and 2 show the process for deciding how to confidentialise 2018 Census data.

Figure 1: The decision process for confidentialising tables of counts (rules 1 to 4)

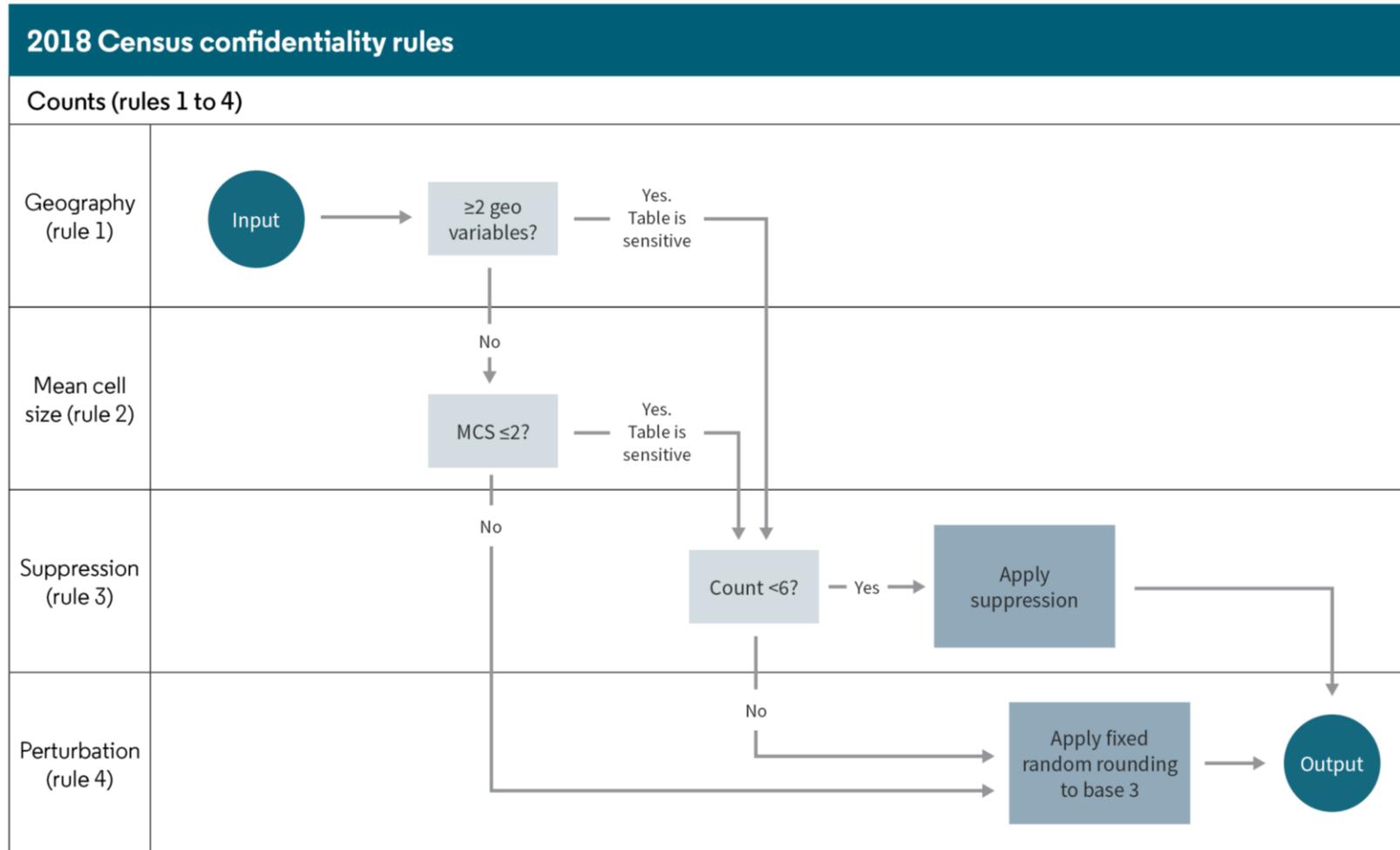
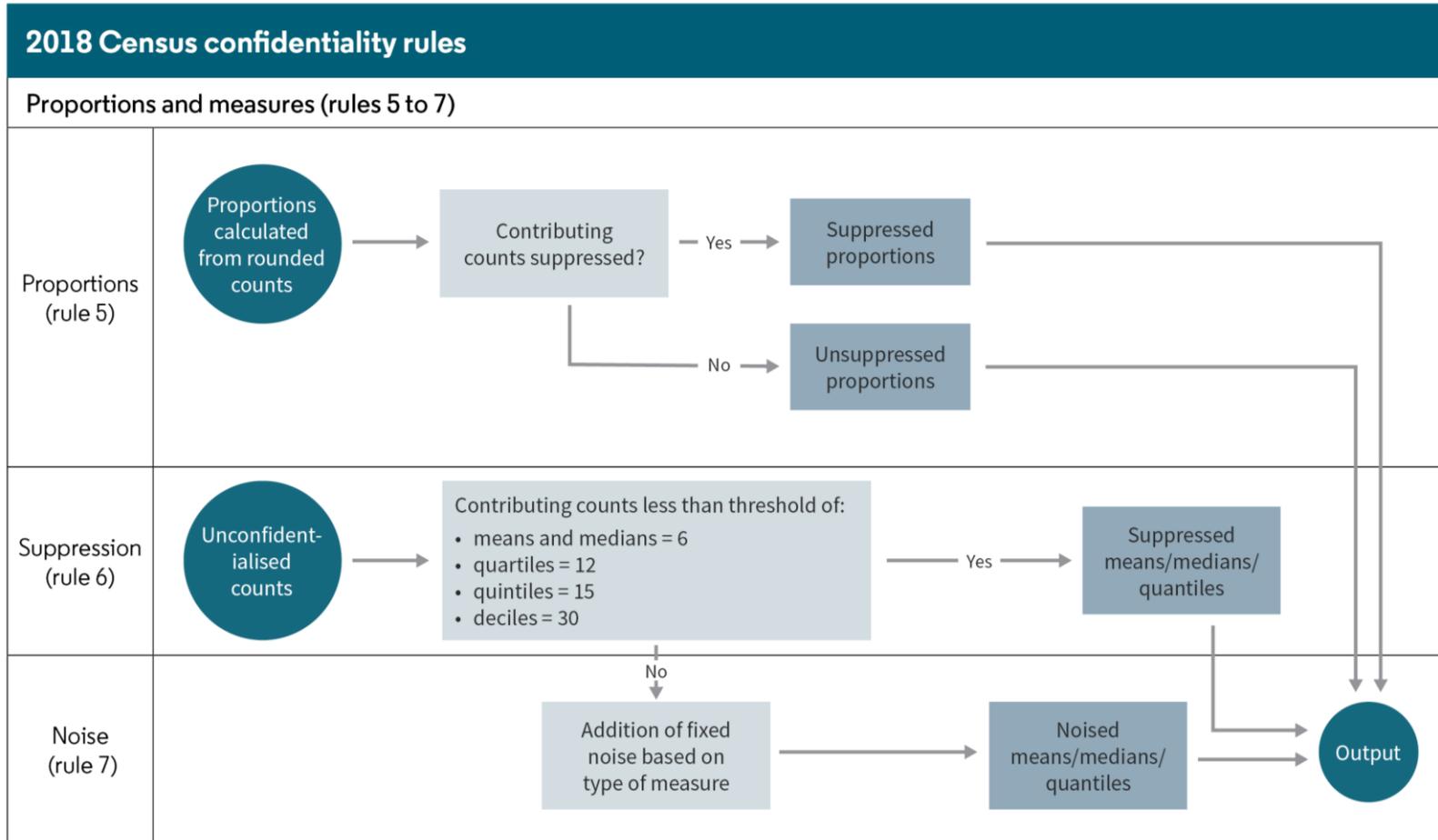


Figure 2: The decision process for confidentialising output containing proportions or measures (rules 5 to 7).



Summary of main changes from 2013 to 2018

The following table summarises the main changes between the rules for 2013 and 2018 censuses.

Changes in confidentiality rules from 2013 to 2018 censuses		
2013	2018	Changes
Rule 1: Meshblock rule	Rule 1: Geographic variables rule	Meshblock rule no longer required after SA1s replaced meshblocks as the smallest standard output geography Geographic variable rules made explicit (previously this was part of mean cell size rule)
Rule 2: Mean cell size rule	Rule 2: Mean cell size rule	None
Rule 3: Threshold rule	Rule 3: Threshold rule	None
Rule 4: Random rounding rule	Rule 4: Perturbation rule	Random rounding is fixed (made consistent across different tables)
Rule 5: Proportions from counts rule	Rule 5: Proportions from counts rule	None
Rule 6: Suppression for measures rule	Rule 6: Suppression for measures rule	None
Rule 7: Use of Rounded counts rule	Rule 7: Noise for measures rule	Measures now calculated from unrounded counts with fixed noise added
Rule 8: Rounding of Measures		Discontinued because this rounding is not required for confidentiality
Release under license		Discontinued based on internal review

Detail of main changes from 2013 to 2018

Geographic variables rule changes

2013 – Rule 1: Meshblock rule

If the geographic unit for the table is a meshblock, and if the table uses a variable that is at a level more detailed than its highest level, or one that uses two or more variables, then the table is deemed to be sensitive. Then rule 3, the threshold rule, will apply to the table.

2018 – Rule 1: Geographic variables rules

If a table contains two or more geographic variables, then the table is deemed to be sensitive. Then rule 3, the threshold rule, is applied to sensitive tables.

Geographic variables for 2018 Census are:

- census night address
- usual residence
- usual residence 1 year ago
- usual residence 5 years ago
- workplace address
- educational institute address.

Explanation of changes

For 2018 Census, we dropped the 2013 Census rule 1, the meshblock rule. This was made possible because census is now using SA1s as the smallest output geography, replacing meshblocks. The SA1 geography groups together the smallest meshblocks, providing a more consistently-sized and sensibly-grouped output geography than the meshblock geography.

Dropping rule 1 means that SA1 tables with a variable more detailed than the highest level or with more than one variable are not automatically deemed as sensitive. Single meshblock outputs are no longer covered by the standard census confidentiality rules, with the exception of standard subject population totals.

The 2013 meshblock rule has been removed. The 2018 geographic variables rule was previously part of rule 2, the 2013 Census mean cell size rule. This has been made explicit and its own rule in 2018 (see below). This has the effect of identifying sensitive skewed tables, where potentially small groups or individuals could be identified.

Mean cell size rule changes

2013 – Rule 2: Mean cell size rule

A table is defined as sensitive if the mean cell size is equal to or less than 2. The mean cell size is calculated as the total unrounded subject population of a geographic area, divided by the number of categories for that same geographic area.

2018 – Rule 2: Mean cell size rule

A table is defined as sensitive if the mean cell size is equal to or less than 2. The mean cell size is calculated as the total unrounded subject population of a geographic area, divided by the number of categories for that same geographic area.

Explanation of changes

No changes.

Threshold rule changes

2013 – Rule 3: Threshold rule

When the table for a geographic unit is deemed sensitive by either rule 1 or rule 2, all counts less than six are suppressed.

2018 – Rule 3: Threshold rule

When the table for a geographic unit is deemed sensitive by either rule 1 or rule 2, all counts less than six are suppressed.

Explanation of changes

No changes.

Perturbation rule changes

2013 – Rule 4: Random rounding rule

All counts for individuals, families, households, and dwellings must be randomly rounded to base 3 (RR3).

2018 – Rule 4: Perturbation rule

All counts for individuals, families, households, and dwellings must have fixed random rounding to base 3 (FRR3) applied.

Explanation of changes

The perturbation method has changed from RR3 in 2013 Census to FRR3 in 2018 Census. This means that identical table cell outputs will be rounded in the same consistent manner, regardless of source or output tool. This is to ensure that outputs are safe for configurable tables, which are dynamic tables where the variables are selected and filtered by customers and built automatically. This rule makes different outputs consistent, increases data utility, and makes data easier for customers to understand.

Proportions from counts rule changes

2013 – Rule 5: Proportions from counts rule

Proportions are calculated from perturbed counts and suppressed if any contributing count is suppressed.

2018 – Rule 5: Proportions from counts rule

Proportions are calculated from perturbed counts and suppressed if any contributing count is suppressed.

Explanation of changes

No changes.

Suppression for measures rule changes

2013 – Rule 6: Suppression for measures rule

Suppress a measure if the total unrounded count of individuals contributing to that measure are below a specified threshold for that measure.

- For medians and means the threshold is 6.
- For quartiles it is 12.
- For quintiles it is 15.
- For deciles it is 30.

2018 – Rule 6: Suppression for measures rule

Suppress a measure if the total unrounded count of individuals contributing to that measure are below a specified threshold for that measure.

- For medians and means the threshold is 6.
- For quartiles it is 12.
- For quintiles it is 15.
- For deciles it is 30.

Explanation of changes

No changes.

Noise for measures rule changes

2013 – Rule 7: Use of rounded counts for calculating measures rule

All measures need to be calculated from randomly rounded counts.

2018 – Rule 7: Noise for measures rule

Measures are calculated from unperturbed counts, and random noise is applied.

Explanation of changes

2013 Census rule 7 compounded and exaggerated RR3 noise, producing some measures that were far from their true value. This has been changed for 2018 by calculating measures directly from unrounded counts, then applying fixed random noise directly to those unperturbed measures. This has the effect of making noise more consistent for measures, while being fixed and maintaining confidentiality.

Rounding of measures rule changes

2013 – Rule 8: Rounding of measures rule

All measures have simple conventional rounding applied.

2018 – Removed

Explanation of changes

Rounding is not done because it is not required.

Release under license changes

2013 – Release under license

Stats NZ may offer a licensed release of less well confidentialised counts to institutions under strict conditions and only for internal institutional use.

2018 – Removed

Explanation of changes

We reviewed our historic practices regarding the release of data under license and decided not to release data under license for 2018.

How we reviewed the confidentiality rules

In 2016 and 2017, we consulted with internal and external customers about their thoughts on the 2013 Census confidentiality methods, and potential methods we might use in the future.

We learned that external customers:

- want to create their own outputs online via an easy-to-use tool
- would like outputs to be more detailed, multivariate, and customisable
- want less suppression of small area and population data
- would prefer noise to come in the form of Random Rounding to base 3 (RR3) rather than the uncertainty of adding noise in a less consistent manner
- consider additivity desirable but not a high priority, particularly if it comes at the expense of more suppression.

Internal customers generally agreed with the points above, and wanted more automation when applying confidentiality methods.

We used this feedback as we reviewed the existing confidentiality rules, and developed new rules for 2018 Census.

More information

[2013 Census confidentiality rules and how they are applied](#)

[2013 Census confidentiality rules – summary of changes from 2006](#)

[Geographic hierarchy](#)