

Methodological standard for
confidentiality in the 2023 Census
2023 Census | Tatauranga 2023





Crown copyright ©

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

Citation

Stats NZ (2024). *Methodological standard for confidentiality in the 2023 Census*. Retrieved from www.stats.govt.nz.

ISBN 978-1-99-104976-6 (online)

Published in April 2024 by

Stats NZ Tatauranga Aotearoa
Wellington, New Zealand

Updated April 2025

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	4
Summary	4
About this standard	5
Structure of this standard	5
Rationale.....	5
Legislation, protocols, and policies	6
Census confidentiality principles	7
Scope	10
Inclusions	10
Exclusions.....	10
Responsibilities	10
Implementation	11
Requirements of the standard	12
Contexts and purposes	12
Approaches to table construction.....	12
Random numbers	16
Count data in static tables.....	16
Medians and other measures from a distribution in static tables.....	20
Other geographies	21
Configurable tables	21
User-defined derived variables	22
Sensitive variable rule.....	22
Data from past censuses	23
Access to microdata.....	23
Resolving issues with this standard	24
Safety and utility.....	25
Adherence to the standard.....	26
Appendix 1: Confidentiality methods and tools for census data in the Data Lab	27
Appendix 2: List of sensitive variables and subject populations	28
References	28

Purpose

Methodological standard for confidentiality in the 2023 Census has been developed by Stats NZ to explain the principles, policies, and processes for managing people's information in the course of producing census outputs (data and statistical releases, products, services, and research).

This standard is part of how Stats NZ meets its obligation to keep data safe and confidential, protect people's privacy, and ensure appropriate use of the data.

Summary

This standard applies to the production of any output from census data, including from census form responses and administrative (admin) data produced by Stats NZ or external researchers who are granted access to census microdata. It applies to 2023 Census data as well as data from past censuses. This standard is largely based on the 2018 Census confidentiality standard, with some implemented improvements.

In August 2021, it was assessed that the 2018 Census confidentiality rules were substantively fit for purpose for use for 2023 Census products and services.

[Applying confidentiality rules to 2018 Census data and summary of changes since 2013](#) has more information on the 2018 Census confidentiality rules.

The assessment included a review of public consultation feedback on 2018 Census products and services.

[Developing 2023 Census products and services – feedback](#) has more information.

Implemented improvements for the 2023 Census include minor method variations, ease of use and scale improvements, and improvement to the Automated Confidentiality Service (ACS) that applies the standard.

Training

Stats NZ provides 'Art of confidentiality' training courses for confidentiality practitioners (technical experts and managers who supply and use data in New Zealand to implement, evaluate, and verify techniques for protecting confidentiality). This training includes an understanding of confidentiality in the New Zealand context, the confidentiality standards Stats NZ uses, and practical statistical disclosure control exercises.

If you are interested in attending this course, please email:

smsconfidentialitynetwork@stats.govt.nz.

About this standard

Structure of this standard

This section includes the rationale, background to, and structure of the standard, as well as census confidentiality principles.

[Scope](#) covers inclusions and exclusions, responsibilities, and implementation of the standard using ACS.

[Requirements of the standard](#) contains the instructions for data providers that ensure that we are in accordance with census confidentiality principles.

[Adherence to the standard](#) contains a flow diagram outlining the main steps for using confidentiality rules.

[References](#) has links to Stats NZ documents referenced in this standard.

Rationale

Stats NZ relies on the cooperation and goodwill of the New Zealand public to collect high-quality information in a census. For our part, we must ensure that private identifiable data is held securely, is used only for statistical purposes, and is not revealed in published outputs. This is fundamental to maintaining the trust and confidence of respondents.

Confidentiality protection is designed to keep individual information safe, and is achieved in two main ways:

- A range of statistical disclosure control methods that modify the data or the design of the published statistics may be used.
- Access may be given to more detailed data under strictly controlled conditions (for example access to microdata in the data laboratory).

Note that 'Individual', when discussing census confidentiality, refers to an individual person, dwelling, household, family, or extended family.

At the same time, we want to make best use of the information provided, releasing as much relevant data as possible.

Most people in New Zealand respond to the census, and information can be produced for over 100 variables. Data is available not just for individuals, but for dwellings, households, families, and extended families. Millions of tables can potentially be produced with census data, many of them interrelated and

overlapping. These features make the census a uniquely valuable source of information for small areas and sub-population groups with cross-tabulations by multiple variables.

These same features of the data also mean that the risk of disclosure of individual information can be high in some situations. No table can be assessed for risk in isolation but must be considered as part of a large group of related tables. If there is no disclosure control, statistical data derived directly from one or two respondents may be traced back to the individuals. The risk of disclosure is high for tables with cells that are derived from one or two respondents, and for microdata records that are unique in the population. The disclosure risk is higher where contributors to the table are more likely to know others personally, for example people in small communities. A count of zero can also create a disclosure risk, as information can be deduced from knowing that an individual does not fall within a category.

When releasing census data, we consider the types of census data produced, the main uses of this data, and the main risks to confidentiality for this data. The ability to efficiently implement confidentiality methods is also an important consideration.

Legislation, protocols, and policies

[Data and Statistics Act 2022](#) (for example, sections 39–42) makes confidentiality related provisions that the production of census data, as with all Stats NZ data, is subject to.

Further, production of census data is guided by the United Nations' (UN) fundamental principles for the production of official statistics, of which Principle 6 states:

“Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.”

[Fundamental Principles of National Official Statistics](#) provides more information.

Within the New Zealand official statistics system, principles and protocols are in place for producers of tier 1 statistics. They embody key aspects of the former Statistics Act 1975, the Privacy Act 1993, the Official Information Act 1982, Public Records Act 2005, and the UN Fundamental Principles of National Official Statistics.

Protocol 4 addresses respondents' rights to confidentiality, privacy, and security. While this protocol acknowledges that there is always an element of risk, it requires Stats NZ (and others) to:

“Take all reasonable steps to ensure that particulars relating to an individual business, household or person are neither divulged in, nor able to be deduced from, published statistics or microdata.”

[Principles and protocols for producers of tier 1 statistics](#) has more information.

Stats NZ staff are also required to adhere to privacy and confidentiality guidelines.

[Information, privacy, security, and confidentiality policy](#) has more information.

[Ngā Tikanga Paihere](#) framework guides ethical and culturally appropriate data use in this standard.

Census confidentiality principles

General approach to census confidentiality

1. Statistical disclosure control (confidentiality) methods are designed to provide a level of confidentiality protection acceptable to the Government Statistician, while aiming to preserve as much data utility as possible, within operational constraints. Access to census microdata is an important way of providing greater benefits from the data collected to produce official statistics.

Census will consider providing microdata within the context of other products and services to be produced from the 2023 Census and previous censuses, and in the context of corporate initiatives and guidelines. Census microdata products will be developed in accordance with the confidentiality standard for microdata access applied by Stats NZ, and the Stats NZ [Microdata output guide](#).

[Appendix 1: Confidentiality methods and tools for census data in the Data Lab has more information.](#)

2. Confidentiality protections are in place for all census products; tables, graphs and maps, microdata, and analytic models will be treated as a coherent whole.
3. Censuses have historically designed statistical disclosure control methods for release of static tables only. Dynamic user-defined tables for external users, also known as ‘configurable tables’, which demonstrably introduce a new dimension of risk, were considered for 2018 Census, but rejected on grounds of residual risk by Stats NZ’s information, privacy, security, and confidentiality governance group. This residual risk remains in the substantively unchanged methods planned for use for 2023 Census. However, configurable tables are out of scope for 2023 Census outputs.

4. Classifications for geographic areas and other variables have a key impact on the design of tables, on the application of confidentiality methods, and on the census data that can be published. The 2023 Census is allowing for tables with user-defined geographies and [other geographies](#).

Tables with user-defined geographies and other geographies have more risk, so require additional consideration when designing statistical disclosure control methods.

5. Any data integration carried out under Stats NZ data integration guidelines will take this standard into account for the release of data.

[How to apply to integrate new data](#) has more information.

6. In response to some requests for data or special products, Stats NZ may consider and approve variations to this standard.
7. Development costs and operational efficiency are important considerations when developing statistical disclosure control methods, and it may result in a more conservative approach than is needed for confidentiality reasons.
8. Stats NZ's approach aligns with the Australian Bureau of Statistics and Statistics Canada.

Confidentiality for published census data

1. In tables, cells with zero, one, or two contributors have a high risk of disclosing individual information. Cells with zero, one, or two contributors will be suppressed or subject to modification to introduce uncertainty about the true value. This requirement applies equally to tables for individuals, families, households, and dwellings, and for all variables (including source variables). It applies to tables of cell counts and proportions calculated from those counts. Means, medians, and other measures are suppressed if there are less than six contributors to the total. This is to protect disclosure of counts of zero, one, or two contributors.
2. Statistical disclosure control methods will treat each census table as part of a large group of related tables, and it will consider the potential for disclosure of information about an individual through comparing different tables.
3. There is a higher risk of disclosing individual information in tables related to small geographic areas or small communities. Additional restrictions may be applied to protect confidentiality for these situations.

[Other geographies](#) has more information.

4. Statistical disclosure control methods may not provide good protection for sparse/skewed tables (those with a high proportion of zeros and small counts). Additional restrictions may be applied to protect confidentiality for these tables.
5. In general, information relating to a group of people (for example, all those with a particular country of birth or religion) is not considered confidential. The Government Statistician may restrict the release of some information in exceptional circumstances.
6. Where publishable information is not fit for purpose, arrangements for controlled access to microdata are available, where data can be selected and analysed before confidentiality methods are applied.
7. Census will publish information about the confidentiality protection methods used (without compromising their effectiveness).

Scope

Inclusions

This standard applies to the production of any output from census data (including census data that has had gaps filled by admin data). Such output may include static and configurable:

- tables
- analytical output
- graphs and maps
- microdata.

It applies to 2023 Census data as well as data from past censuses.

[Data from past censuses](#) has more information.

This standard applies to the production of any output from census data by Stats NZ or by any external researchers who are granted access to census microdata.

Exclusions

This standard does not cover specific security and privacy issues.

This standard does not contain a method for dealing with breaches of confidentiality.

In the course of their work, Stats NZ staff may need to work with non-confidentialised data at times. All employees must sign a 'Certificate of confidentiality' consistent with the Data and Statistics Act 2022 and/or a 'Declaration of secrecy' consistent with the previous Statistics Act 1975 before accessing confidential data.

This standard only applies to the New Zealand Census of Population and Dwellings, and not to other Stats NZ collections that may be full coverage (for example, the agricultural production census).

Responsibilities

Any business unit that produces outputs from census data is responsible for ensuring this standard is adhered to, including the Data Lab (whether this is where researchers have access to standalone census microdata datasets or Integrated Data Infrastructure integrated microdata, including census data) and the customised data requests service.

Data Lab is responsible for ensuring external researchers who have access to unit-record census data are aware of their obligation to produce outputs in accordance with this standard, or where this standard is not fully implemented in Data Lab with the [Microdata output guide](#).

Stats NZ will review this standard at the completion of production of products and services from the 2023 Census. Stats NZ is responsible for maintaining staff capability to support implementation of this standard. And Stats NZ is responsible for decisions about interpretations of rules in the standard, and for management of issues related to implementation of it.

Implementation

The implementation of this standard requires that all relevant parts of the 2023 Census confidentiality processes will be automated in all software tools that output data, by connecting to the Automated Confidentiality Service (ACS).

The ACS carries out most of the confidentiality rules consistently across all software tools. The software output tools, and Stats NZ users, fill any remaining gaps to complete confidentiality methods (for example, calculating proportions from confidentialised counts). The ACS is the key mechanism used to apply the rules in this standard.

[2013 Census confidentiality rules and how they are applied](#) describes the confidentiality methods in use before 2018, for 2023, and earlier data.

[2013 Census confidentiality rules – summary of changes from 2006](#) also has more information.

Stats NZ took an agile approach when developing and implementing these rules allowing for changes and improvement in methodology along the way.

Requirements of the standard

Contexts and purposes

This section describes the confidentiality system in use for the 2023 Census.

These requirements consider census data in different contexts, such as:

- our approaches to tables
- how and why random numbers are used
- counts data
- medians and other measures
- how other geographies are treated
- variables derived from standard variables, data from past censuses, access to microdata, and ongoing development of our processes.

The standard set of confidentiality rules for counts and measures (for public release, in static tables, and using the geographic units specified in [Geographic units](#)) are described in:

- [Count data in static tables](#)
- [Medians and other measures from a distribution in static tables.](#)

Separate data products (for example, count data, medians and other measures, configurable tables, and microdata) are described in separate sections.

Approaches to table construction

The confidentiality system is based on the ways in which we approach or regard the census dataset. These approaches are noted here. The specific rules described later use the approaches noted below.

Geographic units

All published census data relates to some geographic unit (for example, one regional council is equivalent to one geographic unit). The confidentiality processes are applied to each geographical unit separately. For example, when we make a table of counts for people by regional council and gender, the processes treat this as a set of separate tables, one for each regional council. The rules below convey this by referring to ‘the table for a geographic unit’.

The purposes behind this approach are the importance of geographical location for both data utility and safety, and the desire to target the confidentiality processes precisely. Geographic units with large populations receive minor modification, and those with small populations have more modification. This approach was implemented for previous censuses.

Note that when grouping geographic units, the first step is to add the geographic units together, and the second step is to apply confidentiality methods.

The standard 2023 Census confidentiality rules are applied to tables with the following geographies:

- single standard output geographies at SA1 level and above, which are:
 - Statistical area 1 (SA1)
 - Statistical area 2 (SA2)
 - Statistical area 3 (SA3)
 - urban rural
 - territorial authority/local board (TALB)
 - regional council
 - New Zealand
- standard and non-standard groupings of standard output geographies at SA2 level and above
- other standard groupings of meshblocks and SA1s, which are limited to:
 - electoral boundaries
 - ward, community board (includes local board), territorial authority/local board (TALB), subdivision, regional council constituency, Māori constituency
 - general electoral district and Māori electoral district
 - police station, police area, police district
 - district health boards, and/or their national health agency equivalents
- meshblocks (for subject populations).

If a table does not state a geographical classification, then its geographical unit is New Zealand as a whole.

[Other geographies](#) have additional confidentiality methods applied.

Tables and their marginal tables of totals

Most published data is in tables that are constructed using a geographical classification and one, two, or more categorical variables. For example, we can have a table for the geographic unit 'Christchurch city', with counts of people by the two categorical variables 'Age' and 'Gender'. A two-dimensional table like this has two associated 'marginal tables'. In this example, these are: 'Age totalled across gender for Christchurch city', and 'Gender totalled across age for Christchurch city'.

Each of these tables is assessed against the rules as a stand-alone table. Sometimes a high-dimension table will fail a rule, while its lower-dimension marginal tables will pass, as they will be less sparse.

When a table contains counts, the cells are randomly rounded independently of each other.

[Rule 4: Perturbation rule](#) has more information.

The table's marginal tables also have their cells randomly rounded independently of each other. This means that:

- a minimal amount of random rounding noise is added to any cell
- the total in a marginal table may be different from (and more precise than) the total from summing the cells in the more detailed table
- recalculation of suppressed cells is difficult.

A similar situation arises where a variable has two or more levels of detail, in a hierarchical relationship. For example, a table of counts for 'Age in single years' may have totals for five-year groups. These totals form a separate table, and the rules apply to it separately.

The purpose behind the approach is the need for consistency. For example, the table for 'Age' can occur on its own, as a marginal from (subtotal in) the two-dimensional table for 'Age' and 'Gender', or as a marginal from (subtotal in) the two-dimensional table for 'Age' and 'Occupation'. Each time it appears, a table needs to have consistent treatments applied.

Contents of tables

A table is a collection of cells, and these cells may contain:

- counts (that is, frequencies) of variables about individuals (for example, people, households)
- proportions (that is, percentages) derived from these counts

- medians or means for a numerical variable, or a variable that can be treated as such
- other measures calculated for a numerical variable.

Each of these types of content has its own set of treatments, and these are explained below.

Variables

There are several standard output classifications available from the census. Many have different groupings for the same variable, and these often form a hierarchy of classifications. For example, 'Age' exists as single years, and also as five-year groups, ten-year groups, and other groupings.

[DataInfo+](#) lists and defines these output classifications.

The 'highest level' in the hierarchy of classifications for a variable is normally the least detailed one with the broadest groups.

A user may request a table with a 'custom classification' for a variable – this is any classification (or grouping or subtotal) other than a standard approved one. A custom classification must be built in one of two ways:

- Either, it can be built from a standard classification, by first applying the confidentiality rules to the standard classification, then summing to create the custom classification.
- Or, where the customised data requests service views a custom classification as a 'user-defined derived variable' (UDDV), the customised data requests team create a UDDV using non-confidentialised data, and all counts outputs are treated as 'sensitive', so the following rules are always applied:
 - [Rule 3: Threshold rule](#)
 - [Count data in static tables](#) rules 4 and 5
 - [Medians and other measures from a distribution in static tables](#) rules 6 and 7.

It is best practice to choose the highest-level standard classification that can be summed to create the custom classification.

[User-defined derived variables](#) has more information.

For example, consider the custom classification for 'Age' for the four-year span from 16 to 19 years. The confidentiality rules must first be applied to single-year-of-age,

the standard approved classification. Then the confidentialised counts of single-year-of-age (ages 16, 17, 18, and 19) would be summed to create the new four-year span.

The purpose of this is consistency. In the example, the four-year span is an aggregated subset of the single-year classification. Any request for a subset or custom classification like this needs to produce tables that are consistent with those from the standard approved classification itself.

Note that the total stated grouping is treated as a standard classification, even though it is not included in all classifications. In other words, all total stated groupings are dealt with in a specific way, which is that they are always treated as a standard classification. This is because this grouping is output on a regular basis and the utility added by treating these as standard classifications outweighs the added risk.

Some variables, for example 'Income', have a subject population that is a subset of the whole census population. In this example, the subject population is the census usually resident population count aged 15 and over.

Random numbers

To add fixed noise to counts and measures, we first must assign permanent random numbers to each unit record of our input datasets, also known as 'record keys'.

When a table of counts is created, the record keys are then summed for each cell to create a corresponding table of 'cell keys'. The cell keys are used in the calculation of noise for both counts and measures to ensure the same amount of noise is added consistently in every output tool, with every application of the confidentiality rules.

More information on the use of cell keys in the calculation of noise can be found under:

- [Rule 4: Perturbation rule](#)
- [Rule 7: Noise for measures rule.](#)

Count data in static tables

Much of the data produced from census consists of tables of counts of individual people, as well as of dwellings, households, families, and extended families, often disaggregated by one or more demographic, categorical, and geographic variable(s).

This refers to unit records within relevant census data sets, where the unit records in these data sets are representative of individual people, dwellings, households, families, or extended families. For the Data and Statistics Act 2022, all but natural persons are treated as de facto business entities.

Stats NZ produces static tables for public release on its website and elsewhere. It also produces them in response to client requests, and releases these as confidentialised tables so that they too can be publicly available. This section states the procedures and rules for public release of static tables of counts.

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

More information is available under:

- [Rule 1: Geographic variables rule](#)
- [Rule 2: Mean cell size rule](#)
- [Rule 3: Threshold rule](#)

Special exceptions can also be made that deem a table as 'sensitive' and have their unrounded counts below six suppressed by rule 3.

More information is available under:

- [Other geographies](#)
- [User-defined derived variables](#)
- [Sensitive variable rule](#)

The rules apply separately to the table for each geographic unit.

[Geographic units](#) has a list of these.

Rules 1, 2, and 3 may be applied in that order. Rule 4 and rule 5 follow these in sequence. The order in which rules 1, 2, and 3 are applied can depend on computing efficiency.

Rule 1: Geographic variables rule

A table is identified as 'sensitive' if it contains two or more 'geographic variables'. Then rule 3 (the threshold rule) will apply to the table.

A 'geographic variable' is a variable in a dataset that allows you to select one or more geographic units, and it gives context for how the geographic unit relates to the individual or unit record. Geographic variables are: 'Dwelling address', 'Census night address', 'Usual residence', 'Usual residence 1 year ago', 'Usual residence 5 years ago', 'Workplace address', and 'Educational institute address'. Note that 2018

Census had two new geographic variables – ‘Usual residence 1 year ago’ and ‘Educational institute address’.

A ‘geographic unit’ is a defined area on a map, such as a single SA1, a regional council, or New Zealand. A ‘geographic classification’ is a standard collection of geographic units, such as the SA1 or regional council classifications.

[Geographic hierarchy](#) outlines how different geographic classifications in New Zealand relate to each other.

For example, a table that contains SA1 geographic units for ‘Usual residence’ and SA2 geographic units for ‘Usual residence’ is not sensitive by rule 1. However, a table that contains SA1 geographic units for ‘Census night address’ and SA1 geographic units for ‘Usual residence’ is sensitive by rule 1. Before 2018 Census, this rule was part of the Rule 2: Mean Cell Size rule.

Rule 2: Mean cell size rule

This rule is applied separately for each geographic unit. For any table for a geographic unit, the mean cell size is the total unrounded subject population for that geographic unit, divided by the number of cells in the table for that geographic unit.

If the mean cell size is less than or equal to two, then the table for the geographic unit is deemed to be sensitive. Then rule 3, the threshold rule, will apply to the table.

A table may contain the total count for its subject population by geographic unit only, and not use any other variables. Such a table is not subject to the geographic variables rule or the mean cell size rule, and it is treated as not sensitive unless another rule deems it sensitive.

The purpose of this rule is to provide targeted protection for the small counts in tables that are likely to have sparse areas in them, while allowing access to counts in tables that are less likely to have sparse areas. The rule was first applied in this form for the 2006 Census, and was also applied for the 2013 Census, and the 2018 Census. It is similar to one used for the 2001 Census.

Rule 3: Threshold rule

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

The purpose of this rule is the enhancement of 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.

A threshold rule was originally introduced for the 2006 Census, and this rule for 2023 Census is substantively identical to the threshold rule used for the 2018 Census.

Rule 4: Perturbation rule

All tables of counts have rule 4 applied. Counts that have not been suppressed are randomly rounded to base 3.

The purposes for random rounding all counts, large and small, to base 3 include:

- disguising the small disclosive counts of zero, one, two, and three
- protecting against the recalculation of small counts from differencing large counts
- preserving utility by adding a relatively small amount of noise to the larger and more useful counts.

This process was introduced for the 1981 Census. Fixed random rounding to base 3 (FRR3) was newly introduced for 2018 Census, which ensures the random rounding occurs consistently in every output tool, with every application of the confidentiality rules. Sets of partly non-unique random numbers added to each data record set are used to accomplish this.

Rule 5: Proportions from counts rule

All proportions, percentages, and ratios that come from counts must use data that is fully confidentialised, using rules 1 to 4 above. This means that:

- where either of the two counts in the proportion is suppressed, then the ratio is suppressed
- where both counts are available for release, then the proportion is calculated from these rounded counts.

The purpose of this rule is to prevent the reconstruction of the original unrounded counts. It corresponds with part of the derivations rule from the 2001 and 2006 censuses, and was the approach taken for 2013 Census.

The overall effect of these rules

The overall effect is that all tables can be released, and the only cells suppressed are those that are both small and in a sensitive table.

The table provider needs to ask: Is the table sensitive under any combination of rule 1, rule 2, or another rule?

- If the answer is no, the table is released with 0s, 3s, 6s, 9s, etc.
- If the answer is yes, the table is released with Cs (or some other symbol denoting confidentialised data), 6s, 9s, etc.

Medians and other measures from a distribution in static tables

Census datasets contain variables where the categories have numerical values or can be assigned to them (for example, 'Age in single years' and 'Annual income'). Hence the cells in a table can contain medians or other measures from the distribution of the numerical variable. Stats NZ releases these measures in static tables on its website and elsewhere, as well as in response to client requests. This section states the procedures and rules for public release of static tables of measures.

The purpose of the following rules is to maximise the quantity of values released, while protecting particulars about individuals.

Rule 6: Suppression for measures rule

For each measure, there is a threshold value, which is calculated as the number of times you are splitting the data (total) multiplied by 3:

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

Stats NZ can decide whether it is appropriate to release other measures, on request.

If the total unrounded count of individuals related to a measure (including those in residual categories) is less than the threshold value for the measure, then that measure is suppressed.

As an example, median income of males for an SA2 requires at least six males for release.

Rule 7: Noise for measures rule

Means, medians, and other quantiles (but not proportions, which are covered by rule 5) are calculated based on raw, unperturbed counts plus unbiased noise/perturbation.

Sets of partly non-unique random numbers added to each data record set are used to accomplish this.

All measures are to have a degree of rounding (that is, rounding to a number of decimal places). Rounding is done using Stats NZ's rounding of numbers policy.

If a measure has a higher-level classification requested with it (for example, five-year age groups), the measure should still be calculated from the lowest-level classification (that is, one-year age groups).

Other geographies

Additional protection is required for other geographies not covered by the standard 2023 Census confidentiality rules. These other geographies are:

- non-standard geographies defined by xy coordinates (for example, iwi boundaries, water catchment areas)
- non-standard groupings of SA1s and meshblocks (for example, a non-standard grouping of SA1s along a street)
- single meshblock tables that include variables.

Tables with these geographies are riskier, so they will always be treated as 'sensitive', and therefore always apply the threshold rule (rule 3). The remaining standard confidentiality rules for counts and measures are also applied (rules 4 and 5, and rules 6 and 7, respectively).

This extra protection is designed for static tables only. This includes any pre-confidentialised data tables, whether or not they are subsequently wrapped in a dynamic user interface. Customised data requests will maintain a register of requests with these geographies and check this against any new requests to ensure there are no problematic overlaps. Other teams that produce similar types of outputs will also be encouraged to adopt a register-of-requests approach to non-standard geography disclosure control.

Configurable tables

Configurable tables would allow users to create their own tables from a selection of variables in a publicly available manner. These types of tables are inherently riskier than tables with pre-defined variables, because of residual remaining method risk, and how these tables are proposed to be supplied to external users 'on the fly'. They do not necessarily give Stats NZ an opportunity to spot risky tables and intervene to not release them.

Developing an appropriate confidentiality approach for releasing these types of tables safely remains a work in progress. Configurable tables are not currently produced.

User-defined derived variables

A client may request novel data (detailed, non-standard output data), which is derived from standard census output variables. For example, 'Iwi household', where iwi have asked for comparisons of their iwi's households, against non-Māori descent households or total households. All tables with user-defined derived variables are treated as sensitive, so the threshold rule (rule 3) is always applied, as well as the remaining standard confidentiality rules (rules 4 and 5, and rules 6 and 7, respectively).

When a derived variable is calculated by dividing or multiplying counts, these counts should be confidentialised first. If an input count is suppressed, then the output should also be suppressed. The new variable may also need to be rounded to a number of decimal places (in the example, to the nearest whole number of vehicles) in keeping with quality and safety.

Sensitive variable rule

There will always be some risk of disclosure when releasing data. Confidentiality rules are designed to balance that risk with utility, to ensure the safety of outputs as much as possible while ensuring the released data is useful.

While these rules have been determined by Stats NZ to be adequate in protecting individual privacy, some variables, due to the nature of their distribution, or the perception of their risk within communities they are about, may warrant increased protection over the standard rules. This may happen where, for example, the variables' results are highly skewed, or where there are particular concerns about harm due to disclosure.

In such cases, variables may be classified as sensitive. Any table containing a sensitive variable is also deemed sensitive, so the threshold rule (rule 3) is always applied, as well as the remaining standard confidentiality rules (rules 4 and 5, and rules 6 and 7, respectively).

A list of all the variables that have been classified as sensitive can be found in [Appendix 2](#).

Data from past censuses

Where possible, the confidentiality processes for the 2023 Census are applied to data from previous censuses.

The details are:

- 2006 and 2013 and 2018: the 2023 process
- 2001 and before: FRR3 only for tables of counts, proportions from rounded counts, and the process in part 5 for medians and other measures, if possible – if there are technical or resourcing difficulties, then non-fixed random rounding to base 3 only for tables of counts, and calculating proportions, medians, and other measures from rounded counts as was done in the 2013 Census process. See [2013 Census confidentiality rules and how they are applied](#).

While the purpose of this is to have consistent methodology across all census years, to have consistent output from here on out for all census years, to seek to enable all 2006 and later census data to be publishable in configurable tables, and to have efficiency in table production because for older datasets disclosure risk is less acute, in practice, there are difficulties relating to applying these rules to the 2001 and earlier censuses. Hence, the 2013 Census rules are applied instead. As a result, these older censuses should not be published in a configurable table.

Access to microdata

It is very important, in terms of data utility, for census data to be made accessible as unit-record microdata, for researchers and other technical and professional users, in situations that meet user needs and maintain safety.

Access to census microdata is provided via Stats NZ's Data Lab. Researchers there may have access to the standardised census datasets or census data integrated into Stats NZ's Integrated Data Infrastructure (IDI).

[Microdata output guide](#) describes the purposes and processes for these. This standard applies to all outputs of census microdata.

In these modes of access, respondent confidentiality is maintained by:

- [Five safes framework](#) – safe data, safe projects, safe uses, safe environment, and safe outputs
- [Ngā Tikanga Paihere guidelines](#) – a framework guiding ethical and culturally appropriate data use.

In terms of the rules used in the Data Lab, the standard for applying confidentiality to Data Lab microdata can be found via the microdata output guide, linked above.

Appendix 1 later in this report discusses the risks with having multiple relevant methodologies, and refers to a decision during 2018 Census that confidentiality methods for census microdata:

- are clearly laid out in the microdata output guide
- remain largely the same as 2013 Census methods
- don't encompass the updates to the Stats NZ Census Products and Services and Customised Data Requests rules, as outlined here for 2018 and 2023
- are different to these rules, but are still considered to protect data to the required levels of safety
- need future updating, to align them better with the standards for confidentiality being applied outside the Data Lab, but this is still a work in progress.

Resolving issues with this standard

This standard is designed to cover all standard outputs and almost all customer requests. It incorporates previous decisions about requests for tabular or aggregate output. However, we expect that customers will continue to make stronger demands on census data. Customers will make requests that may not be covered sufficiently by the requirements stated in this standard. Also, requests may be for output that passes the requirements, but could still lead to disclosure about individuals. Stats NZ will resolve issues about such requests.

Codefile requests

Codefiles are made up of actual and likely synonyms of survey responses from respondents. Codefiles can include alternative terms, misspellings, abbreviations, and known survey responses.

Codefiles are sometimes referred to as coding indexes, line codes, or synonyms. Coded responses in census outputs are linked to a codefile entry.

For example, the finest level of detail for the occupation classification, has an entry for 'management consultant' (code 224711). However, a range of other descriptions provided by respondents also get coded to this code 224711.

Codefile requests are where customers request tables, which disaggregate counts of the finest level of detail within our classifications, into the synonyms from the codefile. Although these are not based on standard classifications, codefile requests are tables, and as such must adhere to the confidentiality guidelines as outlined in this document.

Table 1. Codefile example

224711	Budgets and planning officer
224711	Business administrator
224711	Business adviser
224711	Business analyst
224711	Business architect

Safety and utility

[Data and Statistics Act \(2022\)](#) outlines that Stats NZ is required to withhold any output that may identify the characteristics of a particular person or business entity. If, despite these requirements of the standard being met, there is any reason to suspect that an output may identify the characteristics of a particular person or business entity, the data should not be released.

This overarching confidentiality requirement reflects the equivalent requirement for previous censuses.

The New Zealand Census of Population and Dwellings is an extremely rich source of information. Its use holds great benefits for New Zealand. This standard aims for intensive use, careful risk management, and efficiency in achieving the benefits.

The census confidentiality flowchart below provides a brief and introductory overview of the confidentiality processes for production of count and measures data.

For public release in static tables, see:

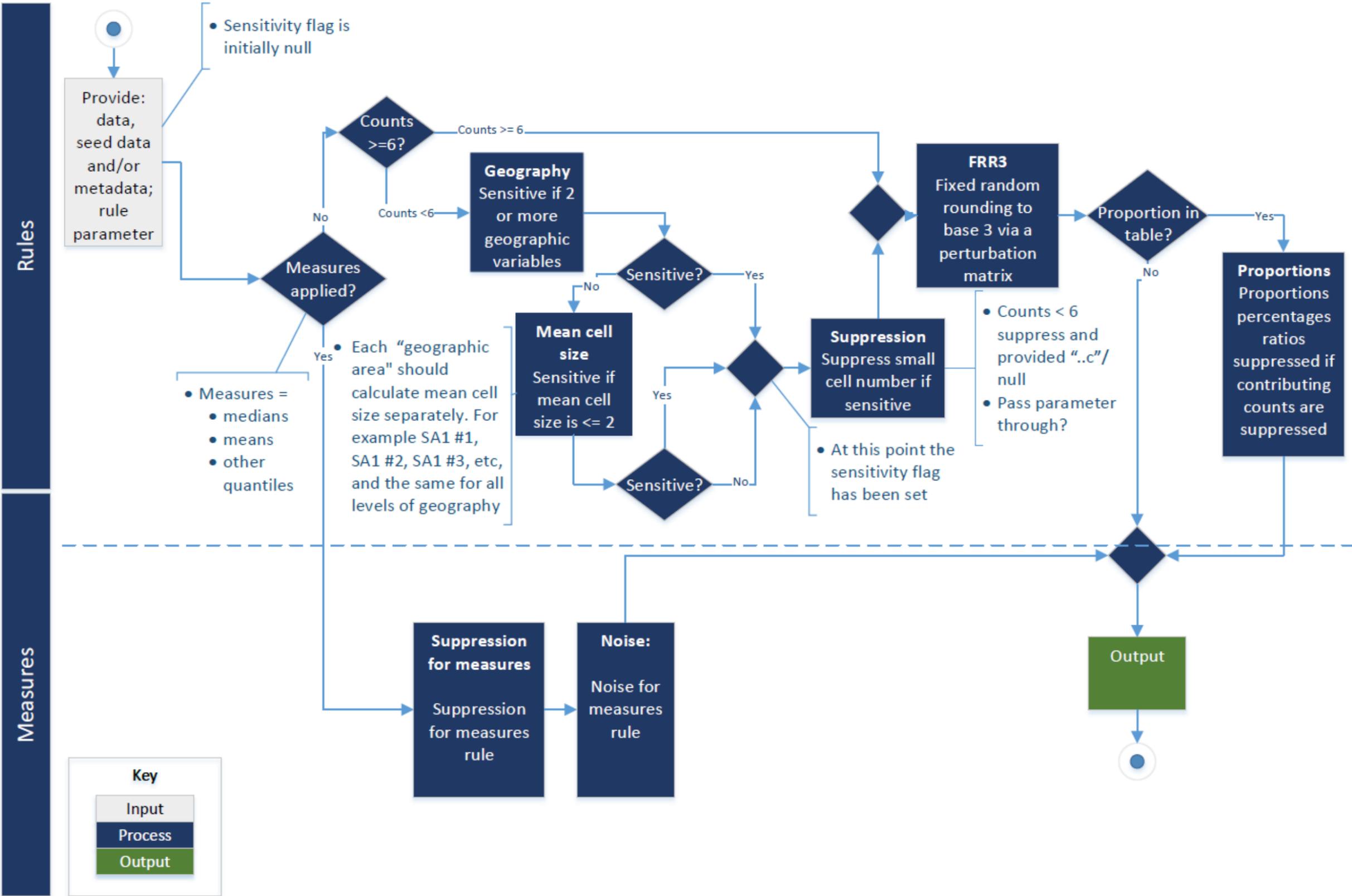
- [Count data in static tables](#)
- [Medians and other measures from a distribution in static tables.](#)

This flowchart is not intended to be a complete statement about these processes.

Other orderings for rules may be possible. Rules may be improved and refined as needed, providing they remain consistent with Stats NZ's [Information, privacy, security, and confidentiality policy](#).

Adherence to the standard

Figure 1: Census confidentiality flowchart



Appendix 1: Confidentiality methods and tools for census data in the Data Lab

Sometimes our confidentiality methods differ across output channels. For example, it is possible for Stats NZ to produce public release outputs based on 2023 Census data, and for Data Lab researchers to produce similar outputs from similar data and release them independently:

- The confidentiality methods used across these two release channels are different because the Automated Confidentiality Service (ACS), which Stats NZ 2023 Census outputs rely upon, has not been implemented inside the Data Lab. The Data Lab currently uses a different, simpler confidentiality method, which is similar to the method used for 2013 Census. Across these example outputs, it is possible for output table cell values and other measure values to exhibit minor differences in value and therefore impacts output data quality. Therefore, these potential release outputs potentially create diminished data safety and utility, due to lack of consistency.
- Data Lab researchers using census data are required to be familiar with the rules and processes around applying Data Lab confidentiality methods. If release outputs are conducted with appropriate diligence, then safety and utility can be managed, mitigated, and minimised to a low level. If not, then the level of diminished safety and utility could be greater than, up to, and including the ability for attackers to potentially unpick the confidentiality of release outputs.

Appendix 2: List of sensitive variables and subject populations

Variables whose inclusion deems a table sensitive by the sensitive variable rule are:

- Sexual identity
- Cisgender and transgender status
- Variations of sex characteristics
- Rainbow/LGBTIQ+ indicator
- Type of couple by gender (and related variables for same-gender couples and another gender partners).

Subject populations relating to the variables above are also considered sensitive, except for male-female couples which do not need extra suppression under the sensitive variable rule (though may under other rules).

References

Stats NZ

Stats NZ (2023). [Apply to use microdata for research](#). Retrieved from www.stats.govt.nz.

Stats NZ (2023). [DataInfo+](#). Retrieved from www.stats.govt.nz.

Stats NZ (2022). [Geographic hierarchy](#). Retrieved from www.stats.govt.nz.

Stats NZ (2022). [How to apply to integrate new data](#). Retrieved from www.stats.govt.nz.

Stats NZ (2022). [Microdata output guide](#). Retrieved from www.stats.govt.nz.

Stats NZ (2022). [Information, privacy, security, and confidentiality policy](#). Retrieved from www.stats.govt.nz.

Stats NZ (2022). [Privacy, security, and confidentiality of survey data](#). Retrieved from www.stats.govt.nz.

Stats NZ (2020). [Ngā Tikanga Paihere: A framework guiding ethical and culturally appropriate data use](#). Retrieved from www.stats.govt.nz.

Stats NZ (2019). [Applying confidentiality rules to 2018 Census data and summary of changes since 2013](#). Retrieved from www.stats.govt.nz.

Stats NZ (2018). [Partnering with Māori on real-world issues: Improving data access and building capability](#). Retrieved from www.stats.govt.nz.

Stats NZ (2013). [2013 Census confidentiality rules and how they are applied](#). Retrieved from www.web.archive.org.

Stats NZ (2013). [2013 Census confidentiality rules – summary of changes from 2006](#). Retrieved from www.web.archive.org.

Stats NZ (2013). [2013 Census Data Dictionary](#). Retrieved from www.web.archive.org.

Stats NZ (2007). [Principles and Protocols for Producers of Tier 1 Statistics](#). Retrieved from www.web.archive.org.

Other sources

International Statistical Institute (2023). [Declaration on Professional Ethics](#). Retrieved from www.isi-web.org.

Parliamentary Counsel Office (2022). [Data and Statistics Act \(2022\)](#). Retrieved from www.legislation.govt.nz.

Parliamentary Counsel Office (2021). [Statistics Act \(1975\)](#). Retrieved from www.legislation.govt.nz.

United Nations Economic Commission for Europe (2019). [General Statistical Business Processing Model \(GSBPM\)](#). Retrieved from www.statswikiunece.org.

United Nations (2014). [Fundamental Principles of Official Statistics \(FPOS\)](#). Retrieved from www.unstats.un.org.