Stats NZ
Tatauranga Aotearoa

Editing, data sources, and imputation in the 2023 Census

2023 Census | Tatauranga 2023

Te Kāwanatanga o Aotearoa
New Zealand Government

Citation

Stats NZ (2024). *Editing, data sources, and imputation in the 2023 Census.* Retrieved from www.stats.govt.nz.

ISBN 978-1-99-104995-7

Published in May 2024 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

List of tables

List of figures

Purpose and summary

Purpose

Editing, data sources, and imputation in the 2023 Census summarises the approach to filling in gaps in census attributes when valid information has not been provided on 2023 Census forms. This includes questions for which no information is provided, as well as those where the information provided is inconsistent or not usable for other reasons.

This paper first provides a general overview of the different data sources and methods used for filling gaps. These methods are:

- editing
- deterministic derivation, based on information from other census variables
- historical census responses
- administrative (admin) data
- statistical imputation.

It then provides more detailed information about the sources and methods used for each 2023 Census variable. Finally, the paper describes the methods used to assess the quality of each source.

This paper is one of a collection of documents summarising the methodology used to combine 2023 Census responses with admin data.

Using a combined census model for the 2023 Census provides links to the other papers.

Summary of key points

For the 2023 Census, we use a range of methods to ensure valid values for individual and dwelling attributes where possible. This includes editing to detect and resolve errors, as well as the use of alternative data sources to fill gaps.

The 2023 Census builds on the combined model first implemented in the 2018 Census. Like in 2018, the 2023 Census uses historical census data, admin data, and statistical imputation to supplement 2023 Census responses. Using these data sources makes more information available for data users and improves the overall quality of census outputs.

Compared with the 2018 Census, we have expanded the use of alternative data sources for the 2023 Census. Some variables, such as post-school field of study and usual residence one year ago, use admin data for the first time. Other variables, such as Māori descent and total personal income, have incorporated new sources,

or improved methods, to increase the quality of data provided. For the new sex at birth and gender variables, we use both admin data and statistical imputation to provide the highest quality outputs.

However, despite extensive research, some variables still lack viable alternative data sources. These variables will have higher amounts of missing data and will be affected more by differential response rates across subpopulations.

Introduction

The New Zealand Census of Population and Dwellings is the official count of how many people and dwellings there are in New Zealand. It provides a snapshot of our society at a point in time.

While the census aims to collect information directly from all people in New Zealand on census night, some people will not complete a census form, and some will submit forms without answering every question. Even when questions are answered, the information provided may not be usable.

Therefore, while processing census data, Stats NZ applies rules called edits to detect and resolve errors and uses statistical methods to deal with missing or otherwise invalid data. These processes are used to improve the overall quality of census outputs.

The 2018 Census was the first in New Zealand to use historical census data (2013 Census) and admin data to fill gaps in census attributes for individuals and dwellings (Stats NZ, 2019). These sources were primarily accessed via the Integrated Data Infrastructure (IDI).

The 2023 Census has further extended this model by design. Historical census data, including both the 2013 and 2018 Censuses, is used for concepts that are unlikely to change over time, while admin data is used when we have evidence the concepts, and measurement of these concepts, are consistent with census responses. Statistical imputation is also used to fill remaining gaps for some variables, which maintains distributions across census variables. Many variables use all these methods, and the Data sources and methods by variable section explains which sources are used for each variable.

This chapter introduces the guiding principles, process, and methods used to fill gaps and resolve inconsistencies in this data.

Guiding principles

The general approach to editing, data sourcing, and imputation was aimed at producing high-quality census outputs while keeping the amount of clerical review low. The design followed three main guidelines:

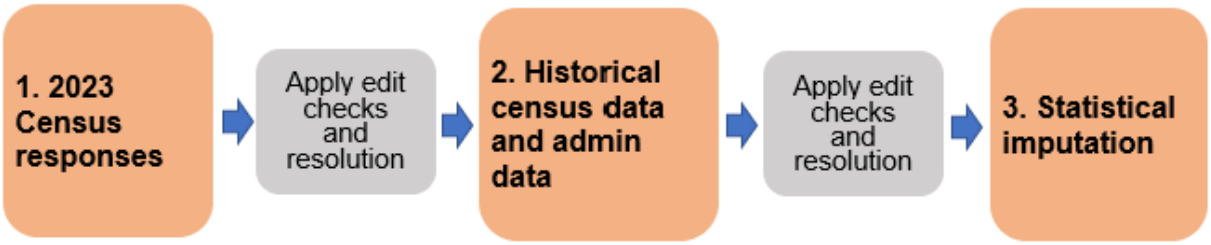
- Prioritising the use of 2023 Census responses and respecting respondent intentions where possible.
- Automating data checks, and resolution of issues and inconsistencies, as far as possible, to minimise the need for manually classifying the data (manual intervention). In specific cases, however, manual intervention is preferred over automated solutions, for quality reasons.
- Using admin and historical census data, as well as statistical imputation, to minimise missing attributes in the census file, to achieve accurate aggregated counts and reduce possible bias from non-response.

In addition, it was important to ensure the methods used were appropriate and would improve the quality of the data, and that we are able to provide information for users about the sources, methods, and quality of each attribute.

Process for filling gaps in 2023 Census variables

Figure 1 shows an overview of the general hierarchy of the steps for filling gaps in data from alternative data sources for most attributes. The standard attribute data source classification, as shown in Appendix 1, for the 2023 Census can be found at Census attribute data source indicator output.

Figure 1: General hierarchy for filling data from other sources



In our approach, three broad sources of information are considered for a given variable:

1. **2023 Census responses** – information provided directly from a 2023 Census individual form, dwelling form, or household set-up form.
2. **Historical census data and admin data** – information about an individual or dwelling from the following sources:
   - Historical (2013 or 2018) census responses – information provided about an individual from a previous census form. Only information from an actual response is included. Values alternatively sourced within the 2018 Census are not considered to be a 2018 Census response.

- Admin data – information about an individual taken from an admin data source, such as birth registrations or school enrolments.

3. **Statistical imputation** – the replacement of missing information with values from a statistical process, in contrast to our methods of sourcing real values for the individual or dwelling in the previous steps. Three types of statistical imputation are used:
   - Probabilistic imputation – missing information is randomly allocated a given value.
   - Within-household donor imputation – information is copied from the person closest in age within an individual's usual residence household.
   - Nearest-neighbour-donor imputation – information is copied from a similar record in the census file, with donors found using nearest-neighbour donor imputation methodology.

In addition, **deterministic derivation** happens at multiple stages in this process. This is where a value is assigned for a variable based on values for one or more other variables to provide consistency across them. See the Deterministic derivation in the 2023 Census section below.

For most variables, sources will be used in the order described above. Valid 2023 Census responses are prioritised first. Historical census and admin are preferred over statistical imputation, as they represent information about the specific person or dwelling. Statistical imputation is only used when no other sources are available – or have been considered as acceptable – for a given record.

Data quality ratings for 2023 Census variables provides more information for each variable.

The Data sources and methods by variable section provides more details about the ordering of sources for each variable.

Editing methods

The first step in the process of filling gaps in the data is identifying cases where data may be erroneous or is unusable. This process is called editing. Examples include multiple boxes being ticked for a single-response question, illegible responses, or contradictory information such as indicating having no income sources but providing an income.

The main types of edits are:

- **Capture edits** identify respondent errors, such as marking multiple responses to a single response question. Where possible these are corrected to a single response. For example, the study participation question has three answers with

tick-boxes. If both ‘full-time’ and ‘part-time’ study tick-boxes are marked but not ‘neither of these’, the response is coded as ‘full-time’. All other combinations of multiple tick-boxes are coded to ‘response unidentifiable’.

- **Validity edits** identify responses that are not valid. For example, providing a year of arrival in New Zealand that is outside the range of possible years, such as a year in the future. In these cases, values can be set to a residual category, or replaced in a later step (using another data source or imputation).
- **Consistency edits** involve comparing responses to different questions. Consistency edits identify responses which do not correctly follow routing instructions, responses which in combination are logically impossible, or combinations of responses which are unacceptable for more complex subject matter and classification reasons. These edits require investigation to identify which response is likely to be erroneous, and fixes are case-specific. Consistency edits can also include between-form or within-household checking, such as checking that household members’ ages are consistent with the relationships they have provided. Three examples of consistency edits are:
  - People under the age of 15 years cannot be legally married in New Zealand.
  - People that reported one or more children born must have a female sex at birth.
  - Someone that is under the age of 16 cannot be driving themselves to work/study and is instead more likely to be a passenger.

Data sources overview

This section describes census response, and historical census and admin data sources used for 2023 Census attributes. Although we consider statistical imputation and deterministic derivation alternative data sources, these are broken out as later steps in this process.

Table 3 shows which sources were used for a specific variable.

2023 Census responses

Responses to the 2023 Census were captured both online and on paper. An online household set-up form and paper dwelling form collected information on all people present at the dwelling on census night. Each person had to be listed by name and had to provide their age, gender, and relationship to the reference person (the person who filled out the household set-up or dwelling form). A separate individual form was to be completed for each person present on census night. However, this does not always occur, and some people listed as being present did not submit an individual form.

Where possible, information from a 2023 Census form was prioritised. If a person did not provide information on their individual form, information from their household set-up or dwelling form was used, where available.

There are also some **derived variables**. Instead of directly asking about them on census forms, we determine them based on combinations of responses to other variables. For instance, the total number of rooms in a house is derived from the counts for each room type.

Integrated Data Infrastructure (IDI)

The Integrated Data Infrastructure (IDI) is a research database which holds microdata about people and households, linked and de-identified at an individual level. This is the database in which admin and historical census data for the 2023 Census is held. Data is gathered from a range of government agencies, Stats NZ surveys, the 2013 Census and the 2018 Census, and non-government organisations. The data are integrated – linked together – to form the IDI.

All admin and historical census data for the 2023 Census was accessed through a secure admin production environment specific to 2023 Census. Access to this environment was strictly controlled, with records de-identified, similar to the measures used for the Integrated Data Infrastructure (IDI). Privacy impact assessment for the use of admin data in the 2023 Census describes these measures.

2023 Census responses are linked to the IDI spine. Dwellings can also be linked to the IDI using shared address identifiers. These links for individuals and dwellings allow us to source information from admin and historical census data. Note that there is an estimated false positive linking rate between census responses and the IDI Spine of about 1 percent. For records that are incorrectly linked, values from the IDI are much more likely to be incorrect. Linking 2023 Census responses to the Integrated Data Infrastructure provides more information about the linking process.

For each variable present in the IDI for a person or dwelling, a single value – or a combination of values, for multi-response variables – is returned. If there are multiple sources reporting a value for a specific variable, we prioritise the available sources and take the information only from the highest priority source.

Historical census data

Historical responses can come from the 2013 Census and the 2018 Census. Older census responses are not currently available in the IDI and were not included as alternative sources for 2023.

Historical census data was primarily used for variables that should not change over time, such as birthplace or year of arrival in New Zealand.

A person's answer about concepts such as ethnicity and religious affiliation may change over time. After careful consideration, we used historical census data for a limited set of variables that are subject to change. These are variables where a response provided in the 2018 Census or even the 2013 Census remained a strong indicator of how an individual would have responded in the 2023 Census.

Historical census data was limited to responses provided directly by the respondents in the relevant census. Information from a historical census that came from an alternative source was not used to fill in values for the 2023 Census.

Admin data

Admin data in general refers to data that has been collected by a range of government and other organisations to operate, such as the collection of tax; to deliver health and education services; or to record events such as births, deaths, and marriages.

Admin data from the IDI that was used in the 2023 Census for census attributes comes from the following sources:

- Accident Compensation Corporation (ACC).
- Department of Internal Affairs (DIA).
- Inland Revenue (IR).
- Kāinga Ora (KO).
- Ministry of Business, Innovation and Employment (MBIE).
- Ministry of Education (MOE).
- Ministry of Health (MOH).
- Ministry of Social Development (MSD).
- New Zealand Customs Service (CUS).
- New Zealand Transport Authority (NZTA).

Other admin data sources may also contribute to the central IDI tables, which were used for attributes such as age and sex at birth. Electoral roll data was also sourced from the Electoral Commission (EC). This is used as a potential source for census attributes, although it is not formally part of the IDI. Department of Corrections data was used to identify those in prison on census night, but it was not used for any attributes. Methodology for using admin data to count people in the 2023 Census describes the use of Corrections data.

Parental data

Parental data refers to information about an individual's parents, where parent-child relationships are identified through birth registrations. Data for the parent could be from a 2023 Census response or from an alternative data source. The 2023 Census has introduced the use of parental information to fill gaps in the Māori descent and iwi affiliation variables. DIA birth records are used to identify parent-child relationships, with information for the parents used to fill gaps for the child. See the Alternative data sources in the 2023 Census: Parental data section below.

Imputation methods

This section describes statistical imputation. Unlike the other alternative data sources discussed to this point, statistical imputation is not intended to capture precise information about specific individuals. Instead, we aim to find similar individuals, dwellings, or households, and use their information to fill in missing data.

This process helps to ensure realistic distributions in the data, representative of those who did not respond to the 2023 Census, with the purpose of improving data quality for users. Statistical imputation is split into two broad sub-categories: Probabilistic imputation and CANCEIS nearest-neighbour donor imputation.

Probabilistic imputation

Probabilistic imputation is a method of randomly allocating information to a given category based on known distributions. Two different processes were labelled as probabilistic imputation in the standard attribute data source classification:

- Probabilistic imputation of age and 'years since arrival to New Zealand'. Where these cannot be calculated due to missing or invalid elements of date of birth, the missing or invalid elements are imputed at random.
- Within-household donor imputation. A gap is filled by choosing a valid response from an individual in the same household.

CANCEIS nearest-neighbour donor imputation

CANCEIS is the CANadian Census Editing and Imputation System, developed by Statistics Canada. CANCEIS is Stats NZ's standard tool for statistical imputation, and it is also used by many other statistical agencies around the world. It implements nearest-neighbour donor imputation (Statistics Canada, 2015).

When an individual (the donee) does not have a valid value for a variable, even after we have tried the earlier data sourcing steps appropriate to that variable, CANCEIS finds respondents similar to that person. It uses 'matching variables', for example age and gender, to define a 'distance function' and searches for records close to the

donee in the file. The closest match is selected as the donor, and the required information is copied from the donor to the donee. Records for which the matching variables themselves were missing were excluded from being prospective donors.

Matching variables can be weighted, giving some matching variables more influence in determining the distance from donees to prospective donor. Matching variables are chosen to be informative of the missing value. Two things inform the choices of matching variables:

- Subject matter expert recommendations.
- Investigations into the predictive power of prospective matching variables for the imputation variable using machine learning methods.

Consistency edits are included in CANCEIS so that donors selected do not create implausible combinations of data for donees. For example, a person aged 15 years or younger with a missing value for main means of travel to education cannot be given a donor who drove to their place of education, as the donee cannot legally drive in New Zealand.

Deterministic derivation methods

Deterministic derivation assigns a value for a variable based on values for one or more other variables to provide consistency across them. For example, ‘usual residence one year ago’ is set to ‘not born one year ago’, for those less than one year of age.

Editing in the 2023 Census

This section describes how different kinds of edits were applied in the 2023 Census.

Editing

The 2023 Census largely reused the approach for the 2018 Census, with automated processes to both check for and resolve these problems where appropriate. Edits are applied at different phases in the 2023 Census processing system and can be resolved in different ways.

For gaps or inconsistencies without a fit-for-purpose automatic resolution, responses were reviewed by census personnel (manual intervention) to improve the quality of the data. Responses that could not be clearly assigned a valid value were coded to residual categories such as ‘Response unidentifiable’ or ‘Response outside scope’. If a residual category has been coded, these records continue through the later steps to see if a valid value can be provided from another data source.

The online form applied some edits at point-of-capture, which avoided or reduced the prevalence of errors compared with those from paper forms. The online form had some basic checks to reduce invalid responses, suggested matching responses using as-you-type lists, and automatically routed the respondent through the form's questions.

Having online edits reduced the amount of editing required later, although it did not eliminate all errors. The increased number of paper responses in the 2023 Census also meant that all types of edits listed under Editing methods were still required.

Changes to the editing method since 2018

The 2023 Census followed the same editing strategy as the 2018 Census, with only minor differences. We introduced more automation. It also became possible to use available admin data to help resolve failed edits and to target effort and optimise efficiency during manual intervention, as it directed personnel toward prioritising certain tasks. Several variables were sent directly to manual intervention, to avoid automated code introducing errors in complex cases.

Alternative data sources in the 2023 Census

Historical census and admin data

When choosing to fill in gaps in data using alternative sources from historical census and admin data, each different source is prioritised separately. This prioritisation considers which source aligns most closely with the census concept, as well as which source has higher levels of agreement at an individual record level.

For example, for birthplace, DIA births data is prioritised highest as DIA is responsible for birth registrations in New Zealand, and thus has high-quality data for everyone born in New Zealand.

Parental data

Parental data sourcing was only used for the iwi affiliation and two Māori descent variables. We used DIA birth records to identify individuals' parents, where possible. Values from parents could then be used to determine values for their children, if necessary. Parental values used in this step include those sourced from 2023 Census responses and admin and historical census data.

Where we could not identify values from parents, values from grandparents or great-grandparents were sometimes used to determine values for their grandchildren or great-grandchildren, respectively. Only DIA birth records were used to identify parents, as these records are likely to represent biological relationships, which is

important for both Māori descent and iwi affiliation concepts. Information about parents, grandparents, and great-grandparents who have passed away (as identified using DIA death records) is considered tapu and was excluded from this process.

Note that there are two Māori descent variables. As the name suggests, Māori descent (output) is used in census output and has three categories: ‘Māori descent’, ‘No Māori descent’, and ‘Don't know’, as some people do not know whether they have Māori descent. Māori descent (electoral) is used to calculate Māori electoral populations and determine the number of Māori electorates. Māori descent (electoral) has two categories: ‘Māori descent’ and ‘No Māori descent’, and therefore requires all ‘Don't know’ answers to be imputed to be either ‘Māori descent’ or ‘No Māori descent’.

See Data sources and imputation for Māori descent in the 2023 Census for more information about the use of parental data for the Māori descent (output) variable. Information about parental data used for iwi affiliation will also be published in late 2024.

Statistical imputation in the 2023 Census

Probabilistic imputation

Two types of probabilistic imputation are used in the 2023 Census. The following subsections explain how they were implemented.

Probabilistic imputation of dates and age

Probabilistic imputation was used when only partial age or date of birth information was available. This includes instances when age could not be determined from the available information – for example, when year of birth was available, but month of birth was missing – as well as instances where age was known, but date of birth was inconsistent or incomplete.

Probabilistic imputation was only used for year of birth when age was available; otherwise, age was filled using nearest-neighbour donor imputation, explained below. When the missing elements of the date of birth were imputed, all valid values had an equal probability of being selected. For example, if someone was born in January, but no day of birth was provided, a value between 1 and 31 was randomly selected.

Similarly, years since arrival is determined by the date a person first arrived to live in New Zealand, and partial information can be completed with random data. For instance, if the record shows only that someone arrived in May 2022, then probabilistic imputation can randomly select a value for day of arrival, between 1 and 31.

Within-household donor imputation

This method was used for ethnicity, languages spoken, and religious affiliation. The person in the same household who was closest in age to the person missing valid responses to these variables was selected and their values copied. People within the same household are more likely to have the same response to these questions than other types of donors.

In the 2018 Census this approach was also used for the two Māori descent variables. However, in the 2023 Census within-household donor imputation was replaced with the parental step described above.

CANCEIS nearest-neighbour donor imputation

For most variables in the 2023 Census, statistical imputation was applied through CANCEIS. The 2023 Census implementation of CANCEIS used largely the same approach as in the 2018 Census. Data sources, editing, and imputation in the 2018 Census provides more information about the 2018 approach.

Statistical imputation was run in a series of modules, where each module has different imputation variables – variables that will be imputed in that module – and different matching variables. Variables grouped together within a CANCEIS module were imputed together (copied from the same donor) to preserve any relationship between those variables. Table 1 describes the CANCEIS modules.

Age and gender are heavily weighted matching variables for all individual modules because they are important for all individual variables being imputed. Other key matching variables are the detailed geographic variables, derived from the usual residence address and census night address. This helps create representative cross-tabular distributions of variables (after imputation) when they are broken down by age, gender, and geography.

The key benefits of our implementation of CANCEIS are that it:

- leads to more accurate aggregated counts
- maintains relationships between variables
- can reduce some potential non-response bias.

Table 1: CANCEIS imputation modules

Module and subject population	Variables imputed	Matching variables
Individual module 1: Age, gender, sex at birth Census night population	Age, gender, sex at birth	Age, gender, sex at birth, usual residence address variables, census night address variables, ethnicity, dwelling type, study participation, highest qualification, home ownership status, income, employment, smoking
Individual module 2: Usual residence address location (down to X,Y coordinates) Census night population	Detailed usual residence address geographies down to X,Y coordinates {meshblock, SA1, SA2, SA3, usual residence TALB, region code}	Age, gender, usual residence address variables, census night address variables, ethnicity
Individual module 3: Census night address location (down to X,Y coordinates) Census night population	Detailed census night address geographies down to X,Y coordinates {meshblock, SA1, SA2, SA3, usual residence TALB, region code}	Age, gender, usual residence address variables, census night address variables, ethnicity
Individual module 4: Cultural and education variables Usually resident population	Ethnicity, language, religion-related and study-related variables	Age, gender, usual residence address variables, census night address variables, ethnicity, language, religion-related and study-related variables, Māori descent (output)

Module and subject population	**Variables imputed**	**Matching variables**
Individual module 4b: Māori descent (output) Usually resident population	Māori descent (output)	Age, gender, usual residence address variables, census night address variables, ethnicity, language, usual residence address 1 year ago
Individual module 4c: Māori descent (electoral) Usually resident population	Māori descent (electoral)	Age, gender, usual residence address variables, census night address variables, ethnicity, Māori descent (output), language, usual residence address 1 year ago
Individual module 5: Income, work-related variables, smoking Usually resident population aged over 15	Income, smoking-related variables, employment-related variables	Age, gender, usual residence address variables, census night address variables, ethnicity, total income, smoking-related variables, employment-related variables, highest qualification
Dwelling module: Type, number of rooms, rent, tenure, sector All dwellings	Dwelling record type, dwelling type, number of rooms and bedrooms (for private dwellings), sector of landlord, tenure of household, weekly rent paid by household and rent period (for occupied private dwellings)	Location variables, dwelling record type, dwelling type, number of rooms and bedrooms, sector of landlord, tenure of household, weekly rent paid by household

Module and subject population	Variables imputed	Matching variables
Household modules: Relationships between all household members Households with two or more usual residents	Relationships for each member of the household. Each household size has a separate module (for household sizes 2 to 8)	Relationship variables and living arrangements
Note: SA1 – statistical area 1. SA2 – statistical area 2. SA3 – statistical area 3. TALB – territorial authorities and Auckland local boards. A full discussion of the family coding system is beyond the scope of this paper. More information will be available in a paper covering families and households in the 2023 Census, to be published by Stats NZ in November 2024. **Source:** Stats NZ		

Changes from the 2018 Census approach

The 2023 Census methodology for statistical imputation fixed some minor technical issues from 2018, improved accuracy and computational performance, and adjusted the details around the imputation of several variables. Adjusting the imputation details for those variables increased the consistency of imputed values and reduced the likelihood of unrealistic combinations being imputed.

The 2023 Census explicitly asked a question on gender for the first time. It also asked a question on sex at birth, replacing the previous question, ‘Are you: Male/Female’. Both gender and sex at birth had to be complete variables with no missing values. Although some alternative data sources were available, some values for both gender and sex at birth had to be imputed through CANCEIS. The data sources and imputation used are described in Methodologies for filling gaps in gender and sex at birth concepts for the 2023 Census.

We also improved the methodology for imputing the Māori descent (output) and Māori descent (electoral) variables. In the 2018 Census, the Māori descent variables were imputed together, alongside ethnicity and other variables. In addition, CANCEIS only imputed people into the ‘Māori descent’ and ‘No Māori descent’ categories, even though ‘Don’t know’ is a valid response to the question in the census and for Māori descent (output). For the 2023 Census, based on consultation about the concept and appropriate use of the data, alongside statistical analyses, we

decided that Māori descent (output) and Māori descent (electoral) would each be imputed in separate modules.

The first (Module 4b) filled in missing data in the Māori descent (output) variable, imputing values in three categories (including ‘Don’t know’). Statistical investigations found that this dedicated module imputed values more accurately than a combined module. A separate module let us choose and weight matching variables specifically for this variable, resulting in higher quality imputation for the Māori descent (output) variable.

The second module (Module 4c) filled in missing data in the Māori descent (electoral) variable, ensuring that everyone with a ‘Don’t know’ response was imputed into either ‘Māori descent’ or ‘No Māori descent’. Both Māori descent variables had different matching variables than in 2018. Language and usual residence 1 year ago were both added to help impute the Māori descent variables, and birthplace was no longer used.

See Data sources and imputation for Māori descent in the 2023 Census for more information. There will also be a specific description for the Māori descent (electoral) variable, expected to be published in October 2024.

Imputation for smoking-related variables has changed compared with the 2018 Census. In the 2023 Census, for current cigarette-smoking behaviour, imputation has been chosen over historical census responses, due to expected changes in smoking rates since previous censuses.

Historical census data was only used to determine whether someone had previously smoked – that is, if an individual reported either being a regular smoker or having previously being a regular smoker in the 2018 or 2013 Census. Historical census data was not used for people that have never smoked as it cannot be assumed that this is still true.

Historical census data was instead added as a matching variable for CANCEIS imputation. This allowed donees reporting they smoked in the 2018 Census to get their donor-imputed variable from other records that reported smoking in the 2018 Census, resulting in better predictive accuracy for regular smoking behaviour. As a result, more records in the 2023 Census went through imputation for these variables than in the 2018 Census.

Deterministic derivation in the 2023 Census

The 2023 Census introduced the term **deterministic derivation** for instances where a value for a given variable is assigned based on values for one or more other variables. This is only used when the census response from the form was missing or coded to a residual category.

This approach can only be used where variables are directly related to each other. In this case, a certain value for one variable implies or precludes certain values for the other variable. For example, the name given in the business name question can be used to code workplace address. This follows the assumption that the address of an individual's employer is likely to be the address of their workplace.

The rules for applying deterministic derivation in the 2023 Census are similar to those in the 2018 Census. However, because the variables used to derive the final response can come from a variety of alternative data sources, a new data source category is created to capture responses sourced through this method. Table 2 lists all the deterministic derivation scenarios in the 2023 Census.

Table 2: Deterministic derivation use cases

Variable	Use case
Age	For those without a valid age, a manual operator determines age from the relationships to other members of the household and the ages of the related people.
Languages spoken	Set to ‘Too young to speak’ when individual is aged 0.
Māori descent	Set to ‘Māori descent’ when the individual has a valid iwi value.
Usual residence one year ago	Set to ‘Not born one year ago’ when individual is aged 0.
Usual residence five years ago	Set to ‘Not born five years ago’ when individual is aged 0–4.
Years at usual residence	Set to ‘Less than one year’ when individual is aged 0.
Individual home ownership	For those without a valid response for individual home ownership: 1. Set to ‘Do not own and do not hold in a family trust’ when individual is aged under 18. 2. Set to ‘Do not own and do not hold in a family trust’ when individual’s usual residence is listed as rented. 3. Set to ‘Do not own and do not hold in a family trust’ when individual’s usual residence is a non-private dwelling type that cannot be owned or held in a family trust.

Variable	Use case
	4. Align with tenure of household when individual is identified as living alone.
Main means of travel to work	Set to ‘Work from home’ when usual residence address and workplace address are the same.
Workplace address	1. Coded to a specific workplace address when they have provided the name of their employer in an earlier question, and this information can be linked to the Business Register. 2. Set to an individual’s usual residence address when the ‘travelling to work’ variable is imputed to ‘work from home’. 3. Set to the territorial authority of usual residence when individual has indicated they do not work at home, and when they have not provided a valid workplace address.
Job indicator	1. Set to ‘Employed’ when there is evidence of employment. 2. Set to ‘Not employed’ when there is no evidence of employment, and when there is some information about searching for work or being available for work.
Total personal income	Set total personal income to ‘Zero income’ when individual has indicated that they had ‘No sources of income’ and have not provided a total personal income.
Educational institution address	Set to the territorial authority of usual residence when individual has indicated they do not study at home, and when they have not provided a valid educational institution address.
Tenure of Household	Derived for dwellings which have been flagged as private dwellings in a registered retirement village.

Data sources and methods by variable

This section describes the data sources and methods used for each variable in the 2023 Census.

The criteria used to determine which sources were suitable for each variable are described in *Editing, data sourcing, and imputation: Planned approach for the 2023 Census* found on Using a combined census model for the 2023 Census. The 2023 Census was intended to be a minimal change census, so the default position was to validate methods from the 2018 Census, and to reuse these where they were deemed suitable. Improvements were focused on areas where there were changes to concepts, and where limitations or ethical concerns had been identified in the 2018 Census.

In particular, this included changes to the approach when using other data sources or methods for certain variables, including Māori descent, iwi affiliation, gender, and sex at birth. These needed to be designed in partnership with Māori, stakeholders, and community groups.

Table 3 describes the data sources and methods for individual and dwelling variables.

This includes:

- whether each variable had information from each of the 2023 Census responses, historical census responses, admin data, deterministic derivation, or statistical imputation
    - Summary of admin data used in the 2023 Census provides more details on admin data sources that were used
- changes implemented for the 2023 Census.

For most variables, sources are prioritised in the order listed in the table. However, there are some instances where the rules are more complex, or where sources are prioritised differently. Table 4 provides more detail about these examples.

Table 3: 2023 Census data sources, by census attribute

Census attribute	Use of data source					Detailed admin data sources[1]	Changes for 2023 Census
	2023 Census response	Historical census response	Admin data	Statistical imputation	Deterministic derivation		
Date of birth / Age	Y	Y	Y	Y	Y	Multiple[2]	--
Sex at birth[3]	Y	Y	Y	Y	N	DIA Multiple[2]	Concept change. Use of admin data and statistical imputation guided by admin data.
Gender[3]	Y	N	Y	Y	N	MSD	New concept. Use of admin data and statistical imputation guided by admin data.
Census night address	Y	N	Y	Y	N	Multiple[4]	--

Census attribute	Use of data source					Detailed admin data sources[1]	Changes for 2023 Census
	2023 Census response	Historical census response	Admin data	Statistical imputation	Deterministic derivation		
Usual residence address[3]	Y	N	Y	Y	N	Multiple[4]	Improved method for determining usual residence address from admin data.
Usual residence one year ago	Y	N	Y	N	Y	Multiple[4]	Use of admin data.
Usual residence five years ago	N	Y	Y	N	Y	Multiple[4]	Use of admin data.
Years at usual residence	Y	Y	Y	N	Y	Multiple[4]	Use of admin and historical census data.
Number of children born	Y	Y	Y	N	N	DIA	Extended use of admin data for those not in historical census data.

Census attribute	Use of data source					Detailed admin data sources[1]	Changes for 2023 Census
	2023 Census response	Historical census response	Admin data	Statistical imputation	Deterministic derivation		
Ethnicity	Y	Y	Y	Y	N	DIA MOE MOH MSD	Use of additional admin data sources such as parents on birth records.
Sexual identity	Y	N	N	N	N	…	--
Variations of sex characteristics	Y	N	N	N	N	…	--
Māori descent (output), Māori descent (electoral)[3]	Y	Y	Y	Y	Y	DIA EC	Extended use of admin data. Updated statistical imputation methodology. Inclusion of parental data.

Census attribute	Use of data source					Detailed admin data sources[1]	Changes for 2023 Census
	2023 Census response	Historical census response	Admin data	Statistical imputation	Deterministic derivation		
Iwi affiliation[3]	Y	Y	Y	N	N	MOE MSD	Use of admin and historical census data. Inclusion of parental data.
Birthplace[3]	Y	Y	Y	N	N	DIA MBIE	--
Date of arrival / years since arrival in New Zealand	Y	Y	Y	Y	N	DIA MBIE	--
Languages spoken	Y	Y	N	Y	Y	…	--
Religious affiliation	Y	Y	N	Y	N	…	--
Highest secondary school qualification[3]	Y	Y	Y	N	N	MOE	--
Level of post-school qualification[3]	Y	Y	Y	N	N	MOE	--

Census attribute	Use of data source					Detailed admin data sources[1]	Changes for 2023 Census
	2023 Census response	Historical census response	Admin data	Statistical imputation	Deterministic derivation		
Post-school field of study[3]	Y	Y	Y	N	N	MOE	Use of admin data.
Study participation	Y	N	Y	Y	N	MOE	--
Main means of travel to education	Y	N	N	Y	N	…	--
Educational institution address	Y	N	N	N	Y	…	--
Total personal income	Y	N	Y	Y	Y	IR MSD WFF	Improved method for extracting personal income from admin data.
Sources of personal income	Y	N	Y	Y	N	IR MSD WFF	Improved method for extracting sources of income from admin data

Census attribute	Use of data source					Detailed admin data sources[1]	Changes for 2023 Census
	2023 Census response	Historical census response	Admin data	Statistical imputation	Deterministic derivation		
Job search methods	Y	N	N	Y	N	…	--
Employment status	Y	N	N	Y	N	…	--
Hours worked in employment per week	Y	N	N	Y	N	…	--
Industry	Y	N	Y	Y	N	IR	--
Occupation	Y	N	N	Y	N	…	--
Sector of ownership	Y	N	Y	Y	N	IR	--
Status in employment	Y	N	N	Y	Y	…	--
Unpaid activities	Y	N	N	N	N	…	--
Main means of travel to work	Y	N	Y	Y	Y	…	--

Census attribute	Use of data source					Detailed admin data sources[1]	Changes for 2023 Census
	2023 Census response	Historical census response	Admin data	Statistical imputation	Deterministic derivation		
Workplace address	Y	N	Y	N	Y	IR	--
Regular smoker indicator	Y	N	N	Y	N	…	Use of historical census data as a predictor for statistical imputation. Removal of historic data as a source

Census attribute	Use of data source					Detailed admin data sources[1]	Changes for 2023 Census
	2023 Census response	Historical census response	Admin data	Statistical imputation	Deterministic derivation		
Ever smoked indicator	Y	Y	N	Y	N	…	Use of historical census data as a predictor for statistical imputation. Change to historic data so that we only use a positive indication of past smoking behaviour.
Disability / activity limitations	Y	N	N	N	N	…	--
Relationship status - legally registered	Y	Y	Y	N	N	DIA MBIE	Extended method for deriving admin family relationships.

Census attribute	Use of data source					Detailed admin data sources[1]	Changes for 2023 Census
	2023 Census response	Historical census response	Admin data	Statistical imputation	Deterministic derivation		
Relationship status - partnership status	Y	Y	Y	N	N	DIA MBIE MSD	Extended method for deriving admin family relationships.
Relationship to reference person	Y	Y	Y	N	N	DIA MBIE MSD	Extended method for deriving admin family relationships.
Living arrangements	Y	Y	Y	N	N	DIA MBIE MSD	Extended method for deriving admin family relationships.

Census attribute	Use of data source					Detailed admin data sources[1]	Changes for 2023 Census
	2023 Census response	Historical census response	Admin data	Statistical imputation	Deterministic derivation		
Individual home ownership	Y	Y	N	N	Y	…	Use of 2018 Census when individuals are identified as being at the same address as five years ago.
Dwelling type	Y	Y	Y	Y	N	MBIE	--
Number of rooms	Y	Y	N	Y	N	…	--
Number of bedrooms	Y	Y	Y	Y	N	KO MBIE	--
Tenure of household	Y	Y	Y	Y	Y	KO MBIE	--

Census attribute	Use of data source					Detailed admin data sources[1]	Changes for 2023 Census
	2023 Census response	Historical census response	Admin data	Statistical imputation	Deterministic derivation		
Sector of landlord	Y	N	Y	Y	N	KO MBIE	Kāinga Ora prioritised as highest quality source.
Weekly rent paid by households	Y	N	Y	Y	N	KO MBIE	--
Access to basic amenities	Y	N	N	N	N	…	--
Access to telecommunication systems	Y	N	N	N	N	…	--
Dwelling dampness indicator	Y	N	N	N	N	…	--
Dwelling mould indicator	Y	N	N	N	N	…	--
Main types of heating	Y	N	N	N	N	…	--

Census attribute	Use of data source					Detailed admin data sources[1]	Changes for 2023 Census
	2023 Census response	Historical census response	Admin data	Statistical imputation	Deterministic derivation		
Number of motor vehicles	Y	N	N	N	N	…	--

1. ACC – Accident Compensation Corporation; COR – Department of Corrections; CUS – New Zealand Customs Service; DIA – Department of Internal Affairs; EC – Electoral Commission; IR – Inland Revenue; KO – Kāinga Ora; MBIE – Ministry of Business, Innovation, and Employment; MOE – Ministry of Education; MOH – Ministry of Health; MSD – Ministry of Social Development; NZTA – New Zealand Transport Agency; WFF – Working for Families.

2. Sources include: ACC, DIA, IR, MBIE, MOE, MOH, MSD, NZTA.

3. Table 4 on the next page has additional notes on this variable.

4. Sources include: ACC, COR, DIA, KO, MOE, MOH, MSD, NZTA.

Symbols: ... Not applicable, -- No notable changes for 2023 Census

Source: Stats NZ

Table 4: Additional notes on 2023 Census data sources, by census attribute

Census attributes	Notes
Gender	See Methodologies for filling gaps in gender and sex at birth concepts in the 2023 Census for more information.
Sex at birth	See Methodologies for filling gaps in gender and sex at birth concepts for the 2023 Census for more information.
Usual residence address	Usual residence address is determined by selecting the best address from across the various sources within the IDI. Predicting usual residence address from admin data in the 2023 Census describes the model used.
Usual residence one year ago	Admin data is used to identify those who were not born, or were overseas, one year ago. For those in New Zealand one year ago, usual residence one year ago is determined by selecting the best address from across the various sources within the IDI. Predicting usual residence address from admin data in the 2023 Census describes the model used.
Usual residence five years ago	Usual residence five years ago was not asked on 2023 Census forms, so all information comes from an alternative data source. Admin data is used to identify those who were not born, or were overseas, five years ago. For those in New Zealand one year ago, usual residence one year ago is determined by selecting the best address from across the various sources within the IDI. Predicting usual residence address from admin data in the 2023 Census describes the model used.

Census attributes	Notes
Māori descent (output), Māori descent (electoral)	Parental data and deterministic derivation were also used to fill in missing data. These steps happened after the use of admin data and before the use of CANCEIS. See Data sources and imputation for Māori descent in the 2023 Census for more information on Māori descent (output). More information on Māori descent (electoral) will be available in October 2024.
Iwi affiliation	Parental data was also used to fill in missing data. More information about the data sources and methods for iwi affiliation will be published in late 2024.
Birthplace	DIA birth registrations are prioritised over historical census data.
Highest secondary school qualification, Level of post-school qualification, Post-school field of study	The highest qualification from across all sources is selected, rather than one source being prioritised.
Total personal income, Sources of income	Income is derived by combining information from across all sources. In general, MSD information is prioritised over IR data when information is recorded in both sources. See Estimating income from linked admin data: Impact of new sources for more information.
Relationship status, Living arrangements	Relationship information is taken by combining all sources, rather than prioritising a single source. For legally registered relationship status and partnership status in a current relationship, DIA data has also been prioritised over 2023 Census responses for civil unions.

Census attributes	Notes
Sector of landlord	Kāinga Ora data is prioritised over 2023 Census response, as the admin source is considered to have complete coverage of Kāinga Ora dwellings.
Source: Stats NZ	

Measuring the quality of data sources

As with previous censuses, the 2023 Census assesses and reports on the data quality of final output variables. The quality rating scale consists of three metrics that contribute to the overall rating for each variable.

Data quality assurance in the 2023 Census and Data quality ratings for 2023 Census variables provide more information.

Metric 1: Data sources and coverage consists of a derived score rating the overall quality of the data sources used for a census output variable. To calculate a score for a variable, each source that contributes to the output for that variable is rated and multiplied by the proportion it contributes to the total output. The rating for a valid census response is defined as 1.00, and missing values are given scores of 0.00. Ratings for other sources are the best estimates available of their quality relative to a census response. This section describes the process for assessing the quality of alternative data sources.

Quality of admin and historical census data, and parental data

Ratings for admin and historical census data, and parental data, are derived by comparing values from these alternative data sources with 2023 Census responses for linked individuals or dwellings. Different ratings are calculated for each of these sources based on these comparisons.

For these comparisons, only valid responses received on 2023 Census forms were included. These comparisons were made based on the assumption that the quality of alternative sources for the records that were linked to the IDI can provide a good indication of the quality of the alternative sources for those whose census data are missing. Census records that were linked to the IDI were used because these are the ones for which both response values and alternatively sourced values are available, allowing us to make a comparison.

Where a value is determined to fill a gap in a variable, a priority ranking method is used for which alternative data sources is used first. The same priority ranking is applied in these quality comparisons. For example, when deriving ethnicity, historical 2018 Census records are prioritised over historical 2013 Census records, and both previous censuses are prioritised over admin sources.

To derive the 2018 Census rating for ethnicity, 2018 Census responses are compared with 2023 Census responses. To derive the 2013 Census rating for ethnicity, 2013 Census responses are compared with 2023 Census responses, but only for individuals who did not also have a 2018 Census response. To derive the admin data rating, admin data are compared with 2023 Census responses, but only for individuals who did not also have a 2018 Census response or a 2013 Census response.

This approach is the same as the approach used for determining quality ratings in the 2018 Census. All ratings were redetermined for the 2023 Census, to accurately reflect the quality of the current data sources, and adjustments to the methods used. Data quality ratings for 2018 Census variables provides the ratings for 2018 Census variables.

For some variables, the admin value is given a rating of 1.00, where it is known to be as good as, or better quality than, census responses – for example, for age, industry, and sector of ownership. For other variables, different methods are used:

- Where reasonable, we assessed the level of exact agreement – how many of the alternative source values were the same as the 2023 Census responses – for example, for country of birth, and Māori descent.
- For some variables, we considered needing an exact match to be unnecessarily precise, and scores were calculated from agreements within one band or one year – for example, for total personal income, and weekly rent paid.
- For variables with detailed hierarchical classifications, we made decisions about which levels of these classifications to use – for example, religious affiliation was compared at the least detailed level of the classification, while languages spoken were compared at the most detailed level of the classification.
- For some variables, multiple responses needed to be allowed for. For these variables – for example, iwi affiliation and sources of income – we assessed whether each value listed in the 2023 Census responses was present in the alternative sources. Consistency is then summarised to have a weighting of one for each person. For example, where an individual listed two sources of income, and admin data only reported one of these, the rating for that person would be 0.50).

Quality of statistical imputation

For the 2023 Census we used a similar approach to measure the quality of statistical imputation as in the 2018 Census, by comparing 2023 Census responses against values that potentially could have been assigned through imputation.

Assessing the quality of within-household donor imputation

For within-household donor imputation, source ratings were based on the following method. We remove valid responses and fill the gaps with imputation, so that the original values can be compared with the imputed ones. We assigned source ratings based on the proportion of records with consistent values.

Assessing the quality of nearest-neighbour donor imputation

We tested the quality of CANCEIS imputation in several ways, particularly when we developed changes to the 2018 Census CANCEIS methodology, for the 2023 Census CANCEIS methodology.

This testing was done by first sampling from data for which the 2023 Census response was known, then randomly removing variables so that the file contained missing values. CANCEIS was run on this test dataset, and imputed values were then compared to the real, known values. Two key metrics were used for this comparison:

- Consistency, or accuracy – how often the imputed values agreed with the real, known values for individuals.
- Distributional consistency, or aggregate consistency – how much the distribution of responses shifted after running CANCEIS and adding imputed values.

A shift in post-imputation distributions is expected sometimes, due to CANCEIS reducing some of the non-response bias in the input data. The results of the imputation tests are checked to verify that any skews that are introduced are expected.

As statistical imputation is not expected or intended to accurately reflect a person, consistency of individual values is not as relevant as it is for other alternative data sources. However, assessments of the similarity of imputed values, at both an individual level and at an aggregated level, were still important to inform our decisions about where and how imputation was used for each variable, to prevent impacting on the association between variables, or introducing skew into the data.

For the 2023 Census, we conducted many analyses into the quality and consistency of statistical imputation, including counterfactual assessment of scenarios where data were not missing at random. This informed several adjustments to the

CANCEIS modules, as well as our decision to not impute the addresses of workplaces and educational institutions.

When determining the quality level of CANCEIS imputation, instead of using the exact consistency that was measured, variables were grouped into categories of either high, medium, or low consistency. With the exception of sex at birth and gender, which get a score of 0.90, a variable in the high consistency group gets a score of 0.80 for CANCEIS imputation, medium consistency gets 0.60, and low consistency 0.40, and all variables in a group are assigned the same indicative consistency.

This is because the goal of statistical imputation is more about aggregate consistency than individual level consistency, and there is uncertainty about exact consistency, because of the nature of this method of imputation. For example, we found in our analyses, that the consistency CANCEIS imputation for ‘travel to work” was around 75 percent, thus this variable was put in the medium category and gets a score of 0.60.

Quality of deterministic derivation

Deterministic derivation ratings were decided on a variable-by-variable basis. In general, ratings accounted for the specific logic and quality of the variables used for deterministic derivation. Where information was taken directly from a different variable, we used the rating for that variable.

Conclusion

The use of alternative sources, including historical census data, admin data, and statistical imputation, is integral to the design of the 2023 Census as a combined model. Using these data sources in combination with census responses makes more information available for data users and improves the overall quality of 2023 Census outputs.

The use of point-of-capture edits built into the online forms avoided many errors. These methods, which supported respondents in supplying valid responses, reduced the need for automatic edits and manual intervention. The online edits therefore improved efficiency and data quality.

The 2023 Census design built on the 2018 Census design, and we also made improvements to many elements of this design. These include updated methodologies for extracting admin and historical census data, as well as adaptations required given a change in variable concepts, such as sex at birth, and gender. As a result, the number of variables using alternative data sources increased, and the level of missing information is reduced.

Another change from the 2018 Census was enhancing the statistical imputation process in the 2023 Census. We did additional testing to better understand the impacts of statistical imputation. For example, we tested how biases in the input data impacted the effectiveness of statistical imputation. The results of these tests were used to improve the implementation of our statistical imputation methodology. There is potential for further improvement in future censuses, to implement the powerful ‘minimum change’ editing features offered by the CANCEIS software, and to use the household as a unit for editing and imputation. There is also potential to investigate other statistical techniques for imputation.

The nearest-neighbour donor imputation methodology, implemented with CANCEIS software for the 2023 Census, doesn't introduce bias. In addition, it is designed to preserve distributions, rather than optimised for finding the true value for any given individual. However, the use of imputation increases the uncertainty in 2023 Census data, and caution should be exercised with this data where imputation rates are high.

These methods and the scale of their use are a significant part of the 2023 Census combined model, which was designed with alternative data sourcing in mind. It is also important to acknowledge that demographics with lower response rates in the 2023 Census field collection thus required greater use of admin enumerations. This included Māori, Pacific peoples, and young adults. For these groups, for the 2023 Census, we were more reliant on alternative sources and imputation than we were for the population overall. It is important that these methodologies are developed within a good, professional understanding of and respect for this context.

References and further reading

References

Stats NZ (2019). Data sources, editing, and imputation in the 2018 Census. Retrieved from www.stats.govt.nz.

Statistics Canada (2015). CANCEIS user's guide version 5.2. Ottawa: CANCEIS Development Team, Social Survey Methods Division, Statistics Canada

Further reading

Stats NZ (2019). Data quality ratings for 2018 Census variables. Retrieved from www.stats.govt.nz.

Stats NZ (2022a). Editing, data sourcing, and imputation: Planned approach for the 2023 Census found on Using a combined census model for the 2023 Census. Retrieved from www.stats.govt.nz.

Stats NZ (2022b). Integrated Data Infrastructure. Retrieved from www.stats.govt.nz.

Stats NZ (2022c). Using a combined census model for the 2023 Census. Retrieved from www.stats.govt.nz.

Stats NZ (2023a). Estimating income from linked admin data: Impact of new sources. Retrieved from www.stats.govt.nz.

Stats NZ (2023b). Census attribute data source indicator V4.0.0 Retrieved from www.aria-prod.stats.govt.nz.

Stats NZ (2023c). Privacy impact assessment for the use of admin data in the 2023 Census. Retrieved from www.stats.govt.nz.

Stats NZ (2024a). Data sources and imputation for Māori descent in the 2023 Census. Retrieved from www.stats.govt.nz.

Stats NZ (2024b). Data quality assurance in the 2023 Census. Retrieved from www.stats.govt.nz.

Stats NZ (2024c). Data quality ratings for 2023 Census variables. Retrieved from www.stats.govt.nz.

Stats NZ (2024d). Linking 2023 Census responses to the Integrated Data Infrastructure. Retrieved from www.stats.govt.nz.

Stats NZ (2024e). Methodologies for filling gaps in gender and sex at birth concepts for the 2023 Census. Retrieved from www.stats.govt.nz.

Stats NZ (2024f). Methodology for using admin data to count people in the 2023 Census. Retrieved from www.stats.govt.nz.

Stats NZ (2024g). Summary of admin data used in the 2023 Census. Retrieved from www.stats.govt.nz.

Appendix 1: Standard attribute data source classification

Table 5: Standard attribute data source classification

Code	Description
11	2023 Census form
12	2023 Census individual variable sourced from dwelling/household set -up form
21	2018 Census
22	2013 Census
31	Admin data
41	Probabilistic imputation
51	Deterministic derivation from another variable
61	CANCEIS donor's response sourced from 2023 Census form
62	CANCEIS donor's response sourced from 2023 Census dwelling/household set-up form
63	CANCEIS donor's response sourced from admin data
64	CANCEIS donor's response sourced from probabilistic imputation from members of their usual residence (UR) household
65	CANCEIS donor's response sourced from deterministic derivation from another variable
66	CANCEIS donor's response sourced from 2018 Census
67	CANCEIS donor's response sourced from 2013 Census
71	No information
Source: Stats NZ	

The above table is a standard data source classification.

Māori descent and iwi affiliation data source classification for 2023 provides an example of a non-standard classification.