

Data sources and imputation for
Māori descent 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). *Data sources and imputation for Māori descent in the 2023 Census*. Retrieved from www.stats.govt.nz.

ISBN 978-1-99-104981-0 (online)

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

Purpose and summary	6
Purpose	6
Summary of key points	6
Background to the Māori descent concept	7
Definition used by Stats NZ	7
Māori descent question in the 2023 Census.....	7
‘Māori descent – output’ variable in the 2023 Census	9
How this data is used by Stats NZ.....	9
2023 combined census model	9
How the 2023 combined census model works.....	9
Integrated Data Infrastructure (IDI).....	10
Linking census respondents to the IDI.....	11
Limitations of alternative sources	11
Historical census and admin sources for ‘Māori descent – output’	11
Historical census data.....	12
Admin data	13
Methods for using data sources for ‘Māori descent – output’	16
Changes in methods from 2018 Census.....	16
Overview of methods	17
Decisions on alternative sources	18
Details of method.....	19
Proportion of data sources used in the 2023 Census	22
Confidentiality	23

Sources used for 'Māori descent – output'	23
Assessing quality	25
How consistency is measured	25
Consistency of 2018 Census, 2013 Census, and admin data sources	26
Consistency of deterministic derivation.....	27
Consistency of CANCEIS statistical imputation	29
Overall assessment	30
Glossary	31
References	32

List of tables and figures

List of tables

1 How responses to the Māori descent question are coded to categories in the 'Māori descent – output' variable	9
2 'Māori descent – output' data sources in the 2023 Census	23
3 Consistency of values sourced from 2023 Census responses and comparison sources	26
4 Consistency of people coded to 'Māori descent' using deterministic derivation	28
5 Consistency of values in the 2018 Census and the 2023 CANCEIS statistical imputation method	29

List of figures

1 Māori descent question and responses on the 2023 Census online form	8
2 Māori descent question and responses on the 2023 Census paper forms	8
3 Admin enumeration of records and data sourcing of variables	10
4 Māori descent question and responses in the 2013 and 2018 Censuses	13
5 Māori descent question and responses when registering a newborn with DIA	14
6 Māori descent question and responses on the electoral roll enrolment form	16
7 How we determined values for 'Māori descent – output' in the 2023 Census	18

Purpose and summary

Purpose

Data sources and imputation for Māori descent in the 2023 Census outlines the data sources and methods used to determine values for the 'Māori descent – output' variable in the 2023 Census, including the use of alternative data sources to mitigate missing census responses.

It also discusses changes since the 2018 Census, and how census responses and alternative data are linked. A description of the data sources and method used for the 'Māori descent – electoral' variable will be published with electoral data in October 2024.

[Editing, data sources, and imputation in the 2023 Census](#) provides more information on the process used for other census variables.

Summary of key points

The 'Māori descent – output' variable is one of the highest-priority variables that we release in the census. It is used to provide insights into the needs and outcomes of people of Māori descent and to determine the subject population for the 'Iwi affiliation' variable.

It is crucial that quality data for and about iwi and Māori can be produced from the census, to enable iwi and Māori to use census data to support their aspirations, to inform decisions that affect iwi and Māori, and to meet the Crown's partnership obligations under Te Tiriti o Waitangi.

However, non-response in the 2023 Census means that some data is missing. The interim collection response rates show that people coded to 'Māori descent' have a lower response rate (77.0 percent) than the overall population (88.3 percent). This means the proportion of missing data is higher for people of Māori descent, which can lead to bias in our outputs.

To produce the Māori descent values for all usual residents counted in the census, we combine 2023 Census responses with historical census data (2013 and 2018 Censuses) and administrative (admin) data (data collected by government agencies and non-government organisations). We also use statistical imputation (replacing an individual's missing values with those copied from another individual) where no other data is available.

We have extended the method used in the 2018 Census by incorporating additional data sources and improving our statistical imputation process. The increased use of historical census and admin data has led to a decrease in the use of statistical imputation and an increase in data quality.

Improving the quality of this variable helps to support fair representation in the iwi affiliation subject population, in the allocation of funding and resources, in research activities, and in the monitoring of outcomes for people of Māori descent.

Stats NZ has worked in partnership with the Data Iwi Leaders Group on the 2023 Census. Working with Te Kāhui Raraunga Charitable Trust, we have developed and tested the methods (including the sources used and their order of use) and ensured the methods are appropriate within te ao Māori context.

[Mana Ōrite Relationship Agreement](#) provides more information on the relationship between Stats NZ and the Data Iwi Leaders Group.

[Te Kāhui Raraunga](#) provides more information on the trust.

Background to the Māori descent concept

This section describes Stats NZ's definition of Māori descent, how the Māori descent question was asked in the 2023 Census, and what we use Māori descent information for.

Definition used by Stats NZ

A person is of Māori descent if they are descended from Māori. Māori descent is based on a genealogical or biological concept. For the purposes of the Māori descent classification, having Cook Island Māori ancestry is not classified as being of Māori descent.

The Stats NZ definition of Māori descent is used throughout this paper.

Māori descent question in the 2023 Census

The 2023 Census asked a single question about Māori descent, with options for 'Yes', 'No', and 'Don't know' ('Āe', 'Kāore', and 'Aua', shown in figures 1 and 2).

Figure 1: Māori descent question and responses on the 2023 Census online form

2023 Census English online form

Are you descended from Māori (that is, did you have a Māori birth parent, grandparent or great-grandparent, etc)? ? Help

yes

no

don't know

2023 Census te reo Māori online form

He tūpuna Māori ōu? ? Āwhina

āe

kāore

aua

Figure 2: Māori descent question and responses on the 2023 Census paper forms

2023 Census English paper form

12 Are you descended from Māori (that is, did you have a Māori birth parent, grandparent or great-grandparent, etc)?

yes → go to **13**

don't know

no → go to **14**

2023 Census te reo Māori and English bilingual paper form

12 He tūpuna Māori ōu?

haere ki **13** ← āe
aua
 haere ki **14** ← kāore

‘Māori descent – output’ variable in the 2023 Census

‘Māori descent – output’ has three categories. We code responses from the Māori descent question into these categories, as shown in table 1.

Table 1: How responses to the Māori descent question are coded to categories in the ‘Māori descent – output’ variable

English response	Te reo response	Output category
Yes	Āe	Māori descent
No	Kāore	No Māori descent
Don’t know	Aua	Don’t know

‘Māori descent – output’ is considered a priority one variable. Priority one variables are those that make up the core reason for conducting a census and have the highest output quality need.

[2023 Census: Final content report](#) has more information on the 2023 Census priority levels.

How this data is used by Stats NZ

The ‘Māori descent – output’ variable is used to report on the Māori descent census usually resident population in all census outputs.

Additionally, individuals coded to ‘Māori descent’ are the subject population for iwi affiliation in census outputs. Further information about the data sources and method used for iwi affiliation in the 2023 Census will be published in late 2024.

For information about how data is used outside Stats NZ, please see [Māori descent – 2023 Census: Information by concept](#).

2023 combined census model

This section describes how 2023 Census responses are combined with alternative sources to reduce the amount of missing data in the census dataset.

How the 2023 combined census model works

The 2023 Census is a combined model by design, with alternative data sources used to supplement census responses. When an individual has not responded to the

census and we are confident in the quality of the admin record, we use admin data to add records (referred to as admin enumerations).

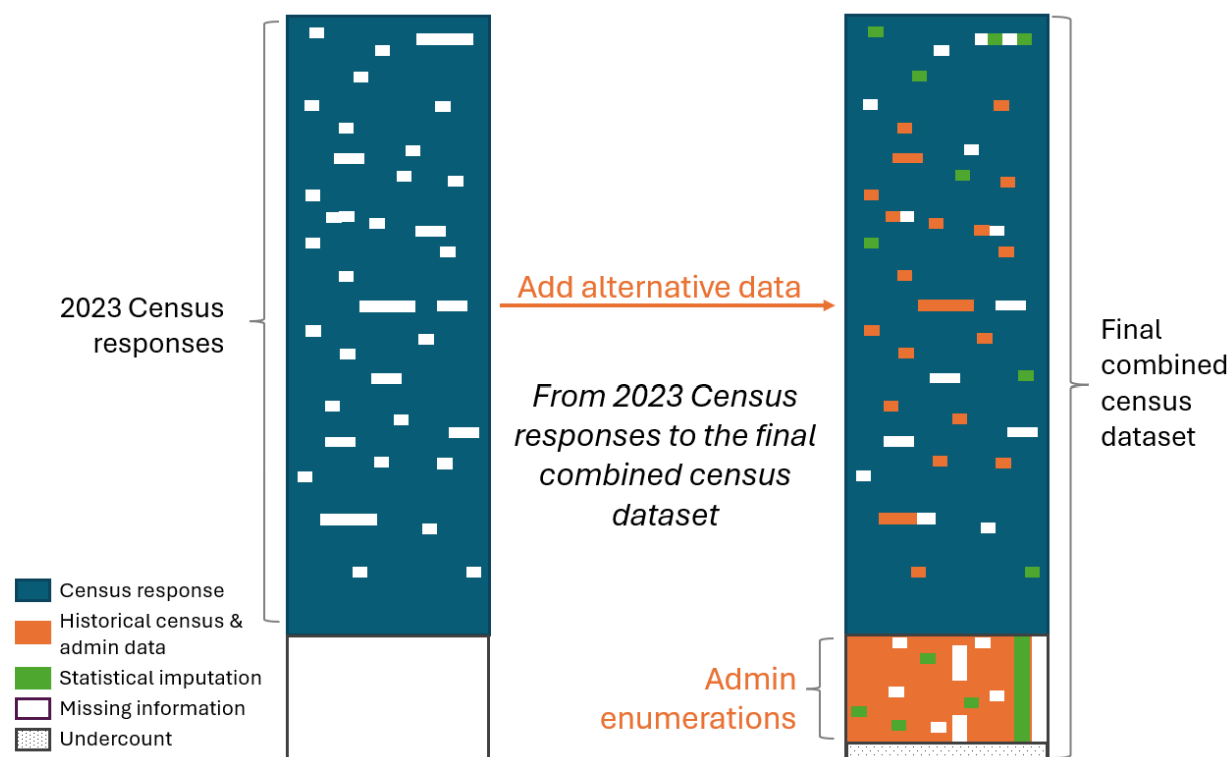
[Methodology for using admin data to count people in the 2023 Census](#) provides more information about the admin enumeration process.

When we are missing responses within a census record, we use historical census data, admin data, and statistical imputation to fill in the gaps.

[Editing, data sources, and imputation in the 2023 Census](#) provides more information.

Figure 3 shows the difference between filling in variables and admin enumeration.

Figure 3: Admin enumeration of records and data sourcing of variables



Integrated Data Infrastructure (IDI)

[Integrated Data Infrastructure \(IDI\)](#) is a large research database, which holds microdata (unit-level data, such as at the individual or business level) about people and households. Data are gathered from a range of government agencies and non-government organisations (admin data) and from Stats NZ surveys (including the 2013 and 2018 Censuses).

[Integrated data](#) provides more information about how data is linked together, or integrated, to form the IDI.

Unlike most access to the IDI, which uses Stats NZ's Data Lab, development of the methodology for filling in missing values in census variables was conducted in a separate secure environment.

[Privacy impact assessment for the use of admin data in the 2023 Census](#) provides more information.

Linking census respondents to the IDI

2023 Census responses are linked to the IDI using a probabilistic approach, which confirms a link when there is a high enough likelihood that sets of records match to the same individual.

[Linking 2023 Census responses to the Integrated Data Infrastructure](#) describes this process in more detail. Overall, 97.9 percent of all census responses were linked.

These links enable us to identify where we are missing census responses for an individual, including for the Māori descent question. We can then use admin and historical census data about that individual to fill in these gaps.

Limitations of alternative sources

The alternative sources we use to fill in missing responses may not reflect how an individual would have responded to the 2023 Census. This could be due to changes in individual knowledge or perception over time, differences in collection method, based on the agency, or for other unknown reasons. This can lead to inconsistencies between data sources.

We provide an indication of the quality of each alternative data source through metric 1 of the quality rating scale.

[Data quality assurance in the 2023 Census](#) provides more information on quality ratings.

Historical census and admin sources for 'Māori descent – output'

This section provides background information about the historical census and admin data sources used to replace missing values in the 'Māori descent – output' variable for the 2023 Census.

Historical census data

Māori descent information was collected in a similar way in the 2013, 2018, and 2023 Censuses. A key difference is that between 2013 and 2018, the main collection mode shifted from paper forms to online forms.

The wording of the Māori descent question on the English form has also changed from 'Are you descended from a Māori?' in the 2013 and 2018 Censuses (figure 4) to 'Are you descended from Māori?' in the 2023 Census (shown in figures 1 and 2). The choice of answers has remained the same.

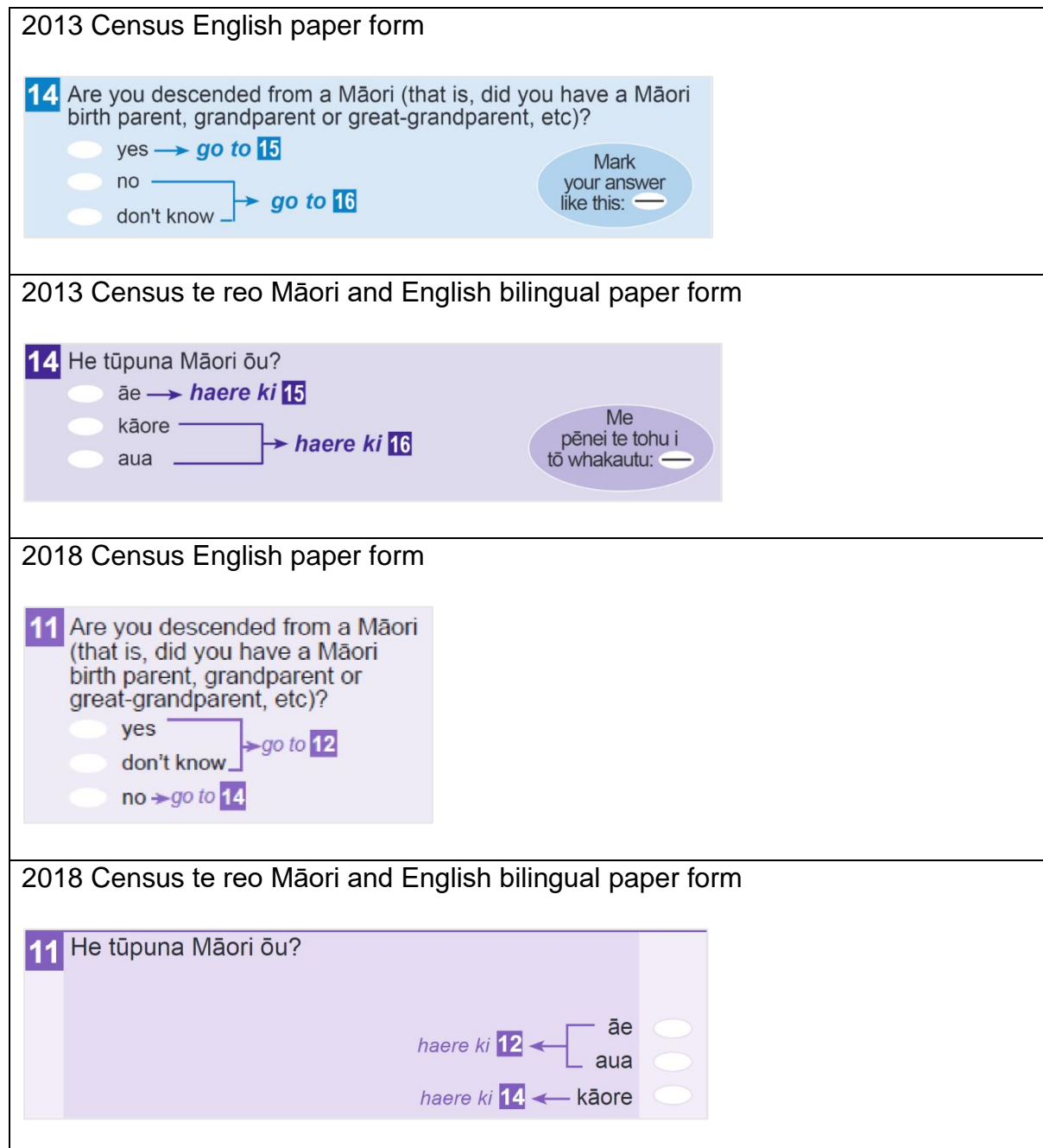
The reo Māori wording of the Māori descent question on the online form and the bilingual paper form has remained the same.

[Māori descent](#) classification has remained the same across the 2013, 2018 and 2023 Censuses, as with the choice of answers.

The aim of all censuses is to measure the entire population at a given point in time. However, there is always a certain level of non-response. Also, while the Māori descent question was mandatory in the online form in 2023 and 2018, paper-form respondents have always been able to skip the question or tick multiple boxes. This means that, across all censuses, paper-form responses may have a missing or unclear response to the Māori descent question.

While there are historical census datasets from before 2013 that contain Māori descent information, these are not currently available in the IDI and are not included as an alternative source for 2023.

Figure 4: Māori descent question and responses in the 2013 and 2018 Censuses



Admin data

Department of Internal Affairs birth records

When babies are born in New Zealand, parents are required to register their birth. The Department of Internal Affairs (DIA) is responsible for maintaining the birth register in New Zealand.

[Registering a new baby and getting a birth certificate](#) provides more information.

Birth registration asks for Māori descent information about the child, the person who gave birth to the child (the child's mother), and the child's second parent. If the second parent is not known, the question is skipped.

The wording of the question for the child's second parent may use the title of father, mother, or parent (all shown in figure 5). Responses to the Māori descent question align with those used in census – 'Yes', 'No', and 'Not sure' (the latter being equivalent to 'Don't know').

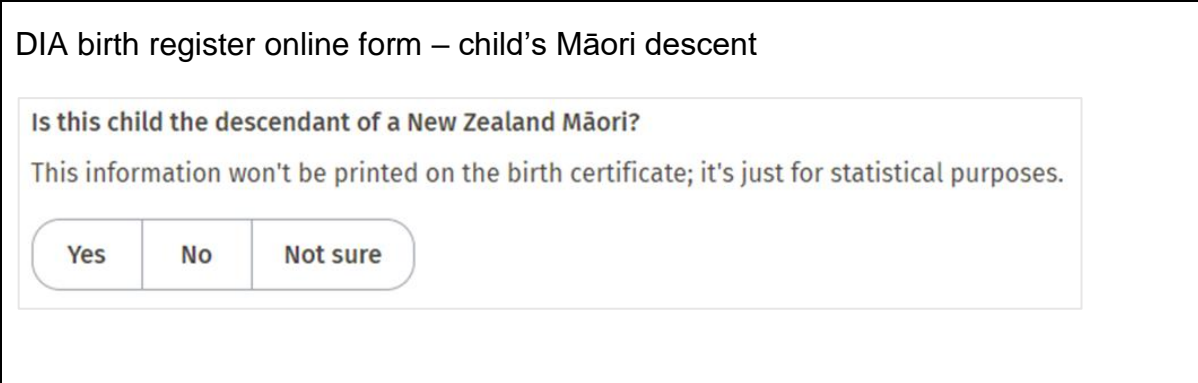
The collection of Māori descent on birth records began in 1995 and is available for all birth records from 1998.

On each birth record, the person born and their parent(s) are separately linked to the IDI. Because birth records from 1998 and later have been fully digitised (with most records since 1985 being fully digitised), they are much more likely to link to the IDI than older records, which are only partially digitised. This means for children born during or after 1985, we have a higher chance of linking them to their parents than for children born before 1985.

While birth records generally indicate a biological relationship, there are some exceptions. Children who are adopted, or whose parents used a sperm or egg donor, may not have their biological parents listed on their birth certificate.

When registering a birth online, individuals must answer the Māori descent question in order to submit the application. However, when registering a birth using a paper form, individuals can skip the Māori descent question. This means that the DIA data is missing Māori descent information for some people.

Figure 5: Māori descent question and responses when registering a newborn with DIA



DIA birth register online form – child's Māori descent

Is this child the descendant of a New Zealand Māori?

This information won't be printed on the birth certificate; it's just for statistical purposes.

Yes No Not sure

DIA birth register online form – mother’s Māori descent

Is the mother the descendant of a New Zealand Māori?

This information won't appear on the birth certificate, it's just for statistical purposes.

Yes

No

Not sure

DIA birth register online form – father’s Māori descent

Is the father the descendant of a New Zealand Māori?

This information won't appear on the birth certificate, it's just for statistical purposes.

Yes

No

Not sure

DIA birth register online form – parent’s Māori descent

Is the parent the descendant of a New Zealand Māori?

This information won't appear on the birth certificate, it's just for statistical purposes.

Yes

No

Not sure

Electoral Commission electoral roll

Approximately 92 percent of the eligible population (aged 18 or older) is enrolled to vote. The Electoral Commission (EC) maintains the electoral roll, which includes everyone who is currently enrolled to vote within New Zealand.

[Enrolment statistics](#) provides more information.


Everybody enrolling for the electoral roll is asked whether they are of Māori descent or not (see figure 6). ‘Don’t know’ is not listed as an option.

For the 2023 Census, we have the active electoral roll data as at 7 March 2023, including information about the usual residence address, Māori descent, and occupation of those enrolled.

While people can register or update their information at any time, it is likely there could be differences in the coverage and quality of electoral roll data at different points throughout the election cycle.

Unlike the census and DIA datasets, everyone enrolled to vote must be assigned a value for Māori descent. This means that the electoral roll data is not missing any Māori descent information for people enrolled to vote.

Figure 6: Māori descent question and responses on the electoral roll enrolment form

EC electoral roll online form	
<p>Please choose one of the following statements about <u>Māori descent</u>:</p> <p><input type="radio"/> I am of Māori descent and want to choose which roll to go on</p> <p><input type="radio"/> I am not of Māori descent (you will go on the general roll)</p>	
EC electoral roll paper form	
<p>Step 4 Your roll</p> <p>This is an important choice. To learn about Māori descent and roll choice, turn over to the QUESTIONS section </p>	<p>Please tick ONE statement that applies to you.</p> <p><input type="radio"/> I am of Māori descent. Please enrol me on the Māori roll.</p> <p><input type="radio"/> I am of Māori descent. Please enrol me on the general roll.</p> <p><input type="radio"/> I am not of Māori descent. (You will be enrolled on the general roll.)</p>

Methods for using data sources for ‘Māori descent – output’

Changes in methods from 2018 Census

The data sources used for ‘Māori descent – output’ in the 2023 Census build on the 2018 Census design with seven key changes:

1. Use of historical census data has been expanded to include two previous censuses (2013 and 2018), rather than one.
2. Parent’s Māori descent value on DIA birth records has been added as an admin source (for that parent).
3. EC electoral roll data has been added as an admin source.
4. Parents’ Māori descent values (from 2023, 2018 or 2013 Census responses, DIA, or EC) have been used, with parents being identified through birth records.
5. The within-household donor step has been removed.
6. Iwi affiliation has been used to determine Māori descent (deterministic derivation).

7. Statistical imputation has been improved to include 'Don't know' values and to achieve better accuracy and alignment with other census variables.

For someone missing a Māori descent response in 2018, the within-household donor step found the person living in the same household of the closest age and copied their Māori descent response. This step has been removed in 2023.

The increased availability and use of historical census and admin data in the 2023 Census has led to a decrease in the proportion of records requiring statistical imputation.

[Proportion of data sources used in the 2023 Census](#) has more information below.

The addition of new sources has also led to a new data source classification for the 'Māori descent – output', 'Māori descent – electoral', and 'Iwi affiliation' variables. This is different to the classification used for other census variables.

[Māori Descent and Iwi Affiliation Data Source Classification for 2023 Census](#) has more information.

Overview of methods

We use the following steps (in the order listed) to search for a valid value for Māori descent for each record (including both 2023 Census responses and admin enumerations). When a valid value is obtained, the search is halted.

1. Start with individual's response to the Māori descent question in the 2023 Census.
2. If no valid response, use their Māori descent values from:
 - a. 2018 Census
 - b. 2013 Census.
3. If still no valid value, then use Māori descent values from:
 - a. DIA birth records
 - b. EC electoral roll.
4. If still no valid value, then use Māori descent values from:
 - a. parents
 - b. grandparents
 - c. great-grandparents.
5. If still no valid value, then look for valid iwi affiliation data from:
 - a. 2023 Census

b. historical census (2013 and 2018), admin data.

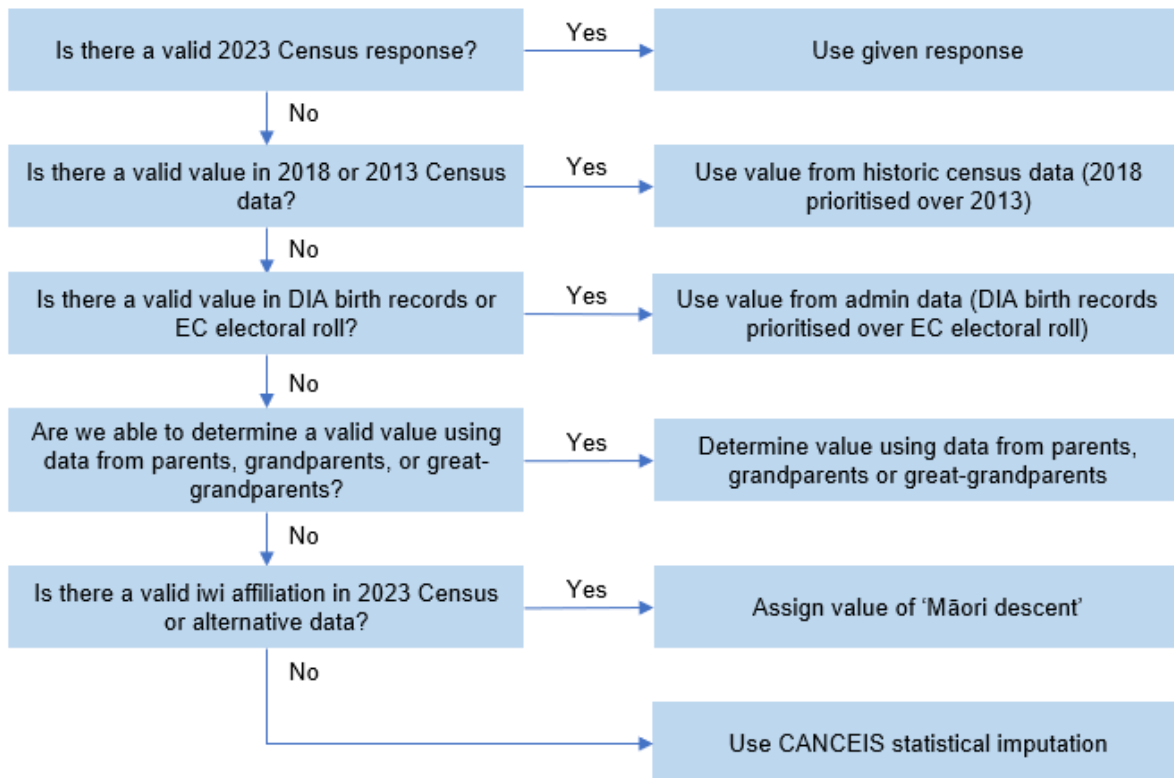
6. If still no valid value, then use statistical imputation.

Note that a ‘valid’ value for the ‘Māori descent – output’ variable is ‘Māori descent’, ‘No Māori descent’, or ‘Don’t know’.

‘Valid’ iwi information excludes iwi affiliation responses coded to ‘Don’t know’, ‘Refused to answer’, ‘Response unidentifiable’, ‘Response outside scope’, or ‘Not stated’.

Figure 7 illustrates the flow of this process.

Figure 7: How we determined values for ‘Māori descent – output’ in the 2023 Census



Decisions on alternative sources

The order and use of alternative data sources in this process have been worked on together by Stats NZ and Data Iwi Leaders Group technicians under the Mana Ōrite relationship agreement. The final decisions align with recommendations made by Stats NZ and the Data Iwi Leaders Group.

Stats NZ has summarised the factors contributing to these decisions as:

- how closely the classification used aligns with the 2023 Census

- how closely the collection method aligns with the 2023 Census
- the recency of the collection
- the quality (coverage and consistency) of the source.

We aim to reflect the response an individual may have given if they responded to the Māori descent question in the 2023 Census.

A note on the iwi affiliation variable

The ordering of sources used for the 'Māori descent – output' variable is different to those used for the 'Iwi affiliation' variable.

The classification and method used to code iwi affiliation responses is significantly more varied between sources than for Māori descent responses. Using the most up-to-date classification is important so that iwi added more recently can be better represented in the 2023 Census data.

The 'Māori descent – output' variable is also different from the 'Iwi affiliation' variable in that the former is single-response, while the latter allows for multiple responses.

For these reasons, the iwi affiliation variable prioritises values from the individual's parents (or grandparents/great-grandparents) over the individual's own values, if these values are closer to a classification recognised in the 2023 Census.

For the 'Māori descent – output' variable, an individual's own values are prioritised over values from others (parents, grandparents, and great-grandparents, or from a donor in statistical imputation). This is in line with the approach used for other 2023 Census variables.

Details of method

1. 2023 Census responses

The first step in producing 'Māori descent – output' is to check if we have a valid 2023 Census response. If a response is missing or unidentifiable (for example, if the respondent has ticked multiple boxes), the process will move to the next step.

2. Historical census responses

If we do not have a valid Māori descent response from the 2023 Census, we check for valid values in historical census data. Values from the 2018 Census are prioritised over values from the 2013 Census.

Only census form responses to the 2013 and 2018 Censuses are used in the 2023 Census. We exclude values that were alternatively sourced or imputed for those previous censuses.

Note that using historical census and admin data depends on linking 2023 Census respondents to the IDI and linking individuals across sources within the IDI.

3. Admin data

If we cannot find a valid value for 'Māori descent – output' in historical census data, we look for values from DIA birth records and EC electoral roll data. Data from birth records is prioritised over data from the electoral roll.

Using birth records

As discussed in the Department of Internal Affairs birth records section, DIA birth records collect three pieces of Māori descent information:

1. Māori descent of the child
2. Māori descent of parent 1
3. Māori descent of parent 2.

In this step, Māori descent information collected about a child (#1) is only used to fill in values for that child, and information collected about parents (#2 and #3) are only used for those parents.

Where an individual is present on multiple birth records (for example, if they have multiple children), we take the Māori descent information from the most recent record. This also means that if we have Māori descent information from an individual's own birth record (#1), and a record where they are a parent (#2 or #3), we prioritise the record where they are the parent.

This is distinct from the use of birth records in the parental data step, where Māori descent values from any of the previous sources can be passed to a second generation (described below).

4. Parental data through census and admin data sources

If an individual is still missing Māori descent values after the previous steps, we can use their birth record to identify their parents. Māori descent values from the parents can then be used to determine Māori descent values for the individual.

Note that parents' Māori descent values used in this step can be sourced from 2023 Census responses, historical (2013 and 2018) censuses, birth records, and electoral roll data.

To determine Māori descent using the parents' values, we apply these three rules:

1. If at least one parent is coded to 'Māori descent', the child is coded to 'Māori descent'.
2. Else, if one parent is coded to 'Don't know' and the other parent is coded to 'Don't know' or 'No Māori descent', the child is coded to 'Don't know'.
3. Else, if both parents are coded to 'No Māori descent', the child is coded to 'No Māori descent'.

Rules 2 and 3 require two parents to be linked to the child, who both have valid Māori descent values.

Additionally, values of the 'Māori descent – output' variable for parents who have passed away (as identified using DIA death records) are considered tapu and excluded from this process.

Where Māori descent values cannot be obtained from parents, we look for the individual's grandparents or great-grandparents.

The same rules are applied as for parents – only one grandparent or great-grandparent coded to 'Māori descent' is required to code a grandchild to 'Māori descent' (rule 1), but rules 2 and 3 require information from all four grandparents or all eight great-grandparents.

While a decision was made to include great-grandparents as a source, it was not used in the production of the 'Māori descent – output' variable. There was no great-grandparent data available for individuals who reached this step, which is likely because their great-grandparents have passed away or have non-digitised birth records, meaning they could not be linked.

5. Deterministic derivation (using iwi responses)

If a Māori descent value is still missing, but the individual has a valid iwi affiliation response from the 2023 Census or historical census, or admin data from the Ministry of Education or Ministry of Social Development or parental data, then we code them to 'Māori descent'.

In this context, a valid iwi affiliation response excludes responses coded to 'Don't know', 'Refused to answer', 'Response unidentifiable', 'Response outside scope', or 'Not stated'.

Note that the Māori descent question is mandatory in the 2023 Census online form, and the online form will only ask for iwi affiliation from respondents who answered 'Yes' or 'Don't know' to the Māori descent question. Therefore, only respondents using a paper form can have a missing value for Māori descent and still provide an iwi affiliation response.

6. Statistical imputation

Statistical imputation is used to fill any missing Māori descent values after the previous steps have been applied. This is performed using the Canadian Census Editing and Imputation System (CANCEIS).

CANCEIS works by matching a person with a missing response (donee) to a person with valid response (donor). Matching is done based on the similarity of other information available (matching variables) for the two people. After matching is completed, the donee's missing response is filled in by copying the donor's response.

In the 2018 Census, 'Māori descent – output' was imputed at the same time as its matching variables. In the 2023 Census, 'Māori descent – output' is imputed after its matching variables have been imputed, providing more opportunity for correlation between the values.

The matching variables used for 'Māori descent – output' in the 2023 Census include age, gender, ethnicity, languages spoken, and several geographic variables. These have been chosen using 'random forest' machine learning models to estimate variable importance. These models helped us to understand how much including or removing a variable affected the accuracy of statistical imputation.

This method helps us to maximise the chance that imputed Māori descent values are realistic and consistent with the donee's other census variables. While imputed values may not always reflect the true value at an individual level, imputation lets us achieve representative data at an aggregated level.

[Editing, data sources, and imputation in the 2023 Census](#) provides more information.

Proportion of data sources used in the 2023 Census

This section describes the proportion of each source used to produce the 'Māori descent – output' variable for the census usually resident population.

Confidentiality

The 2023 Census confidentiality rules have been applied to the following data. Counts are calculated using fixed random rounding to base 3 (FRR3) and suppression of 'sensitive' counts less than six, where tables report small populations. Individual figures may not always sum to stated totals.

[Applying confidentiality rules to 2023 Census data and summary of changes since 2018 and 2013 Censuses](#) contains more information.

Sources used for 'Māori descent – output'

For the 2023 Census, the number of people coded to 'Māori descent' is 978,246, which is 19.6 percent of the total census usually resident population. The number of people coded to 'No Māori descent' is 3,873,726 (77.6 percent), while there are 141,951 people coded to 'Don't know' (2.8 percent).

People coded to 'Māori descent' have a lower percentage of data from the 2023 Census form (73.9 percent), compared with people coded to 'No Māori descent' (88.3 percent) and 'Don't know' (80.6 percent).

In the past, the census has had a lower response rate for people of Māori descent. That trend has continued in 2023 (noting that the final response rates will not be known until published on 9 December 2024 with the results of the Post-enumeration Survey).

Table 2: 'Māori descent – output' data sources in the 2023 Census

Source	'Māori descent – output' value			Total count	Total percent
	Māori descent	No Māori descent	Don't know		
2023 Census response					
2023 Census form	723,051	3,420,153	114,357	4,257,561	85.3
Historical census					
2018 Census	96,183	217,584	12,594	326,361	6.5

Source	‘Māori descent – output’ value			Total count	Total percent
	Māori descent	No Māori descent	Don’t know		
2013 Census	53,910	69,132	4,899	127,941	2.6
Admin data					
DIA birth records	75,150	62,841	7,161	145,152	2.9
EC electoral roll	8,055	19,791	--	27,846	0.6
Parental data					
Parents’ data	3,096	1,470	144	4,710	0.1
Grandparents’ data	21	--	--	21	0.0
Deterministic derivation					
Iwi affiliation from 2023 Census response	504	--	--	504	0.0
Iwi affiliation from an alternative data source	3,558	--	--	3,558	0.1
Statistical imputation					
CANCEIS nearest-neighbour imputation	14,715	82,758	2,796	100,269	2.0
Total	978,246	3,873,726	141,951	4,993,923	100.0

Assessing quality

The tables below highlight the consistency of 'Māori descent – output' values between individual census responses and alternative sources. This is equivalent to the consistency measurement used in metric 1 of the quality rating scale, but without source weighting applied.

[Data quality assurance in the 2023 Census](#) provides more information.

[Data quality ratings for 2023 Census variables](#) provides the final quality ratings for 2023 Census variables.

How consistency is measured

To determine a consistency score, we take people who have responded to the Māori descent question in the 2023 Census and who also have a Māori descent value in an alternative data source.

We then check how often their Māori descent value in the alternative data source is the same as the response they gave in the 2023 Census (using the proportion of matches). This is expressed as a score between 0.00 and 1.00.

By definition, data from the 2023 Census has a consistency score of 1.00.

In 2023, the consistency score for 'Māori descent – output' values sourced from 2018 Census data is 0.95. This means that for the people we looked at, the response they have in the 2018 Census data is the same as the response they gave in the 2023 Census, 95 percent of the time.

This implies that if we use this data source to determine a Māori descent value for someone who did not answer the question in 2023, we would code that person to the correct category 95 percent of the time. Five percent of the time we would code the person to an incorrect category (for example, we code them to 'Māori descent', when this is not true).

While consistency provides insight into the quality of alternative sources, it can only measure performance against the people who responded to this census question. It is possible that results would be different for the people who did not respond to this census question – for instance, if they tend to have more or less alternative data available, or the data tends to be more or less accurate. Still, this consistency score is the best measurement available of how well alternative data sourcing can replace census responses.

Consistency of 2018 Census, 2013 Census, and admin data sources

This comparison only includes respondents who had a valid response for the Māori descent question on a 2023 Census form and who also had data in at least one alternative data source. The result is shown in table 3, with each cell showing the number of matches, total number of people in the category (in brackets), and the consistency score.

Table 3: Consistency of values sourced from 2023 Census responses and comparison sources

Source	'Māori descent – output' value			Total
	Māori descent	No Māori descent	Don't know	
Historical census				
2018 Census	438,177 (481,305) 0.91	2,474,451 (2,526,150) 0.98	19,899 (83,142) 0.24	2,932,527 (3,090,597) 0.95
2013 Census	83,901 (91,800) 0.91	247,770 (256,023) 0.97	2,031 (12,354) 0.16	333,702 (360,177) 0.93
Admin data				
DIA birth records	100,464 (112,473) 0.89	254,697 (262,968) 0.97	1,887 (10,743) 0.18	357,048 (386,184) 0.92
EC electoral roll	11,772 (13,320) 0.88	79,068 (80,256) 0.99	0 (2,415) 0.00	90,840 (95,991) 0.95
Parental data				
Parents' data	2,397 (2,502) 0.96	1,974 (2,298) 0.86	39 (231) 0.17	4,410 (5,031) 0.88

Source	'Māori descent – output' value			Total
	Māori descent	No Māori descent	Don't know	
Grandparents' data	15	0	0	15
	(15)	--	--	24
	1.00	0.00	0.00	0.62
Total	636,720	3,057,963	23,859	3,718,542
	(701,406)	(3,127,695)	(108,888)	(3,937,989)
	0.91	0.98	0.22	0.94

Comparing the consistency of each source highlights there are key differences between each source, but perhaps more significantly, across each value in the 'Māori descent – output' variable.

For people coded to 'No Māori descent', the consistency score is 0.97 or higher in historic census and admin data sources. Because this category makes up the majority of the overall population, it also has the biggest impact on the total consistency scores, which are 0.92 or higher for historical and admin sources.

Consistency scores are lower for people coded to 'Māori descent' in all sources except for 'Parent' and 'Grandparent'. However, these sources only make up a very small proportion of the alternative sources used.

For people coded to 'Don't know', consistency scores are low across all sources. Further investigation is necessary to understand the reasons for this result.

This means that people coded to 'Māori descent' and 'Don't know' based on a 2023 Census response are more likely to have a different value in alternative sources, compared with those coded to 'No Māori descent'.

Consistency of deterministic derivation

Deterministic derivation is only used for people who did not answer the Māori descent question in the 2023 Census, so we needed a different way to measure consistency.

The first group was people who had given an iwi affiliation in 2023 Census but had not answered the Māori descent question. To test this step, we asked: "Of the people who did not answer the Māori descent question in 2023 Census, but provided a valid

iwi in the iwi affiliation question, how many are of Māori descent according to an alternative data source?”

The second group was people we had no Māori descent values for, but we had found a valid iwi affiliation for them in an alternative data source. To test this step, we asked: “Of the people who do not have an alternative data source for Māori descent, but have a valid iwi affiliation in an alternative data source, and who answered the Māori descent question in the 2023 Census, how many said they are of Māori descent?”

For this deterministic derivation step, we determined a final consistency score of 0.90 by multiplying the proportion of people with the consistency score from each step, and then adding the weighted scores together. Table 4 shows these calculations.

Table 4: Consistency of people coded to ‘Māori descent’ using deterministic derivation

Source	Count of each source	Proportion of each source	Consistency score	Weighted score
Deterministic derivation				
Iwi affiliation from 2023 Census response	504	0.12	0.65	0.08
Iwi affiliation from an alternative data source	3,558	0.88	0.93	0.82
Total	4,062	100		0.90

The consistency score of 0.65 for the first step was lower than expected. This means that some people who skipped the Māori descent question but gave a valid iwi affiliation in 2023 Census were coded to ‘No Māori descent’ or ‘Don’t know’ in a different source.

However, this only applied to a small number of people, and the overall score suggests we were able to code people to ‘Māori descent’ correctly 90 percent of the time.

Consistency of CANCEIS statistical imputation

To evaluate how well imputation performs, we run CANCEIS on a test dataset where the missing values are known beforehand. We then compare how often CANCEIS imputes the correct values.

We are aware of non-response bias in the census dataset, where people who do not fill out the census have different characteristics than people who complete the census. In order to reduce this bias, the test dataset needs to include values from current census responses as well as alternative sources (to avoid circularity, statistical imputation is excluded).

Because CANCEIS testing was done in parallel with the development of 2023 Census data, we were unable to use the final Māori descent values from the 2023 Census to complete this testing. For this reason, we used 2018 Census as a proxy for the 2023 Census.

Table 5 shows the individual-level consistency of using the 2023 CANCEIS method to impute values in the 2018 Census.

Table 5: Consistency of values in the 2018 Census and the 2023 CANCEIS statistical imputation method

Source	'Māori descent – output' values			Total
	Māori descent	No Māori descent	Don't know	
Statistical imputation				
CANCEIS nearest-neighbour imputation	0.86	0.95	0.36	0.91

The overall consistency of statistical imputation is slightly lower than for most of the previous sources. We expect statistical imputation to have somewhat lower consistency for individuals because it does not use true values associated with the individual, but rather fills in with values from similar individuals.

The consistency of imputed values in the 'Māori descent – output' variable follows a similar pattern to the historical census and admin-based sources. Statistical imputation performs best for people coded to 'No Māori descent', while those coded to 'Māori descent' have a lower consistency.

People coded to 'Don't know' have a very low consistency. This suggests there are no strong associations between this response type and values found in the matching variables.

Final consistency scores (metric 1 rating) using CANCEIS

Unlike other sources, the consistency scores for CANCEIS are not based solely on the individual consistency.

For the official CANCEIS metric 1 rating, we also do the following:

- We consider the aggregated consistency, because the purpose of statistical imputation is not necessarily to assign every record the correct value, but to achieve representative aggregated data.
- We account for the fact that there is uncertainty around the true consistency, because of non-response bias.

Variables are assigned to groups of either high, medium, or low consistency, and assigned a consistency score of 0.80, 0.60, and 0.40, respectively.

For this reason, the consistency scores for 'Māori descent – output' values imputed using CANCEIS is 0.80.

[Data quality assurance in the 2023 Census](#) explains the quality rating process for the 2023 Census.

Overall assessment

The Māori descent information produced in the census is used to provide important insights, to inform decision-making activities, and to enable the data aspirations of iwi and Māori as treaty partners.

However, people who are coded to 'Māori descent' have lower response rates than the overall population and require a higher proportion of alternative data sources to fill in missing responses.

This highlights the value of using reliable alternative data sources when filling in missing information and invalid responses for the Māori descent variables. These alternative data sources enable us to reach a more accurate representation at an individual level than would be achieved using statistical imputation alone.

Statistical imputation is the last possible step for filling in a variable, used if other sources did not lead to a valid value connected to the respondent. The increased

number of alternative sources available in the 2023 Census means that the use of statistical imputation has decreased from 2018. The quality rating and consistency scores highlighted in this paper suggest that the overall quality of the Māori descent – output variable has improved from 2018 to 2023.

However, people coded to ‘No Māori descent’ make up a greater proportion of the population, and therefore have a larger impact on the final quality scores.

The consistency between the 2023 Census responses and alternative data scores is lower for people coded to ‘Māori descent’, compared with people coded to ‘No Māori descent’. Additionally, people who did not answer the Māori descent question, and therefore required alternative data, were more likely to be coded to ‘Māori descent’.

For these reasons, it is imperative that we continue to analyse and improve on our method, build a deeper understanding of the impacts of missing Māori descent data, and work with iwi and Māori to develop accurate and viable long-term solutions.

Glossary

[Census glossary](#) provides definitions of terms used in the 2023 Census.

References

Stats NZ (2019). [Deriving the 2018 Māori descent electoral population](#). Retrieved from www.stats.govt.nz.

Stats NZ (2021). [2023 Census: Final content report](#). Retrieved from www.stats.govt.nz.

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

Stats NZ (2024a). [Data quality assurance in the 2023 Census](#). Retrieved from www.stats.govt.nz.

Stats NZ (2024b). [Data quality ratings for 2023 Census variables](#). Retrieved from www.stats.govt.nz.

Stats NZ (2024c). [Editing, data sources, and imputation in the 2023 Census](#). 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). [Methodology for using admin data to count people in the 2023 Census](#). Retrieved from www.stats.govt.nz.