Ethnicity Data Protocols

HISO 10001:2017

September 2017

MINISTRY OF HEALTH
MANATŪ HAUORA

Keeping standards up to date

HISO standards are regularly updated to reflect advances in health information science and technology. See our website (**health.govt.nz**) for information about the standards' development process. We welcome your ideas for improving this standard. Email **standards@health.govt.nz** or write to Health Information Standards, Ministry of Health, PO Box 5013, Wellington 6140.

Acknowledgements

This document is available at **health.govt.nz**

The Ethnicity Data Protocols were originally published in 2004 by the Ministry of Health, with the Ethnicity Data Protocols Supplementary Notes published in 2009. This edition replaces the first edition and the supplementary notes.

Contents

List of Tables

List of Figures

Requirement for change

New Zealand is recognised as a world leader in our ability to analyse health data by ethnicity.

The process of collecting and reporting ethnicity data in New Zealand has evolved significantly over time. We have moved from non-standardised collection of data to a process of collecting, recording and output of ethnicity data governed by **Stats NZ's *Statistical Standard, Ethnicity V1.0.0*** (statistical standard) and the *Ethnicity Data Protocols for the Health and Disability Sector* (the protocols).

In the last decade, there has been a rapid shift in primary care from a system that did not record ethnicity routinely to one with near complete ethnicity data recording. This transition has been an important step in our evolving ability to monitor and report key health indicators by ethnicity. This updated document supports the next step in our ongoing efforts to capture high-quality ethnicity data. As we shift the health systems to providing more care closer to home and our data systems towards greater integration, these protocols align with data quality improvement efforts across the sector. This will assist in ensuring we can monitor both the health status of our populations and our interventions to improve health outcomes.

The protocols have been updated to address the move in the health and disability sector to electronic collection and storage of data. The protocols define appropriate processes for confirmation or correction of ethnicity where existing data is held for a respondent and an appropriate frequency for collecting ethnicity data.

They have been updated alongside other key strategic documents.[1] This review allows us to fully integrate the health and disability sector protocols and the statistical standard. The updated protocols support a transition from the previous minimum requirements of recording up to three ethnicities at level 2 classification to recording up to six ethnicities at level 4 classification. This reflects the requirement for information systems to capture the greater population diversity and improved granularity of information to plan, fund and monitor health services. These changes represent a significant move forward in terms of ethnicity data collection and will make a valuable contribution for health.

---

[1] For example, the *New Zealand Health Strategy: Future direction*, *The Guide to He Korowai Oranga – Māori Health Strategy*, Wellington: Ministry of Health, **http://www.health.govt.nz/publication/guide-he-korowai-oranga-maori-health-strategy**

Summary of protocol requirements

The *Ethnicity Data Protocols for the Health and Disability Sector* (the protocols) describes procedures for the standardised collection, recording and output of ethnicity data for the New Zealand health and disability sector.

Collecting and confirming ethnicity

The standard ethnicity question for the health and disability sector is the Stats NZ 2018 Census ethnicity question. The format is to remain the same and the font size and dimensions must not be reduced.

Where a respondent may not be able to fill in a form or questionnaire themselves due to disability, incapacity, being deceased or being a newborn or child, the approach should be adjusted.

The respondent must identify their own ethnicity (called self-identification) regardless of collection method – for example, face-to-face contact, use of a form, electronic collection or telephone.

The collector must not guess ethnicity on behalf of the respondent or limit the number of ethnicities given.

Ethnicity may be collected at any time but must be collected at least every three years.

Ethnicity may be confirmed at any time when other personal details such as gender and contact details are also confirmed.

Classifying, recording and storing ethnicity

The latest version of the Stats NZ's *Ethnicity New Zealand Standard Classification 2005* structure must be used to code ethnicity data.

Ethnicity data must be recorded at level 4 (the most detailed level of the classification).

Information systems must be capable of recording up to six responses.

Where a respondent reports more than six ethnicities, the Stats NZ standard process for reducing multiple ethnic responses must be followed.

Where a respondent reports multiple ethnicities, a 'principal ethnicity' must not be recorded.

'New Zealander' and like responses should be coded to the 'New Zealander' code at level 4.

Protocol requirements for output

The same output method and aggregate categories must be used for both numerator and denominator datasets.

'New Zealander' type responses should continue to be aggregated to 'New Zealand European' in the process of output to maintain time series in the sector. It may also be output to the 'Other' category where this will enable comparability or maintain time series with a dataset.

Where there are multiple ethnicities for a respondent, one of the following three methods of output must be used:

- total response (overlapping)
- prioritised (the most common form of output in the health and disability sector)
- sole/combination.

The method used must be described or noted along with any analysis. Standard descriptions are available in the statistical standard.

Up to six ethnicities at level 4 must be reported to Ministry of Health National Systems.

1   Background

The *Ethnicity Data Protocols for the Health and Disability Sector* (the protocols) describe procedures for the standardised collection, recording and output of ethnicity data for the New Zealand health and disability sector.

This second protocol revision is consistent with Stats NZ's *Statistical Standard, Ethnicity V1.0.0* (statistical standard) and outlines a new set of minimum standards that apply across the health and disability sector to all organisations that collect, record and use ethnicity data. Integration of the revised protocols into the health and disability sector will assist in the ongoing process of improving the quality, accuracy and consistency of ethnicity data.

1.1   Applicability of the protocols

The individuals and groups in the health and disability sector to which this set of protocols applies are:

- patients and/or consumers – for example, health and disability service users
- employees of health and disability organisations and agencies – for example, for health workforce statistics
- populations.

The users of these protocols in the health and disability sector include:

- collectors and processers of ethnicity data including health and disability administrators, clerks and health professionals
- users of ethnicity data including all those who use health and disability ethnicity data for activities such as research, service planning or quality control, monitoring performance and targeting resources or for specific activities like deriving funding formulae
- health information software developers.

The person giving their ethnicity is referred to as the respondent.

1.2   Purposes for collecting ethnicity data

High-quality ethnicity data in the health and disability sector are important for the following reasons:

- Ethnicity data are part of a set of routinely collected administrative data used by health sector planners, funders and providers to design and deliver better policies, services and programmes. Better information can help improve every New

Zealander's health by providing a sound basis for decision-making. Some individually-focused care planning includes the offer of services that include ethnicity as a variable (for example, cardiovascular risk assessment) where targeted interventions are designed to address population-level inequalities in access and/or health outcomes.

- In New Zealand, there are significant ethnic health inequities. The impact of these inequities is particularly evident for Māori and Pacific Peoples, whose health status is lower on average than that of other New Zealanders.

- Key health sector documents, including the *New Zealand Health Strategy*,[2] highlight quality information systems as an important element of improving system integration and health outcomes.

- The health and disability sector has a role in providing quality ethnicity information that enables wider state sector analysis of economic, social and cultural experiences of particular ethnic groups within the New Zealand population.

1.3   Treaty of Waitangi obligations

The Crown recognises the Treaty of Waitangi as the founding document of New Zealand and is committed to fulfilling its obligations as a Treaty partner. The *New Zealand Health Strategy* similarly acknowledges this status of the Treaty, along with the Government's commitment to it.[3]

To date, the relationship between Māori and the Crown in the health and disability sector has been based on three key principles.

a. **Partnership**: working together with iwi, hapū, whānau and Māori communities to develop strategies for Māori health gain and appropriate health and disability services.

b. **Participation at all levels:** involving Māori at all levels of the sector in decision-making, planning, development and delivery of health and disability services.

c. **Protection and improvement of Māori health status**: working to ensure Māori have at least the same level of health as non-Māori and safeguarding Māori cultural concepts, values and practices.

Providing high-quality ethnicity data will assist in ensuring Government is able to track health trends by ethnicity and effectively monitor its performance to improve health outcomes and achieve health equity. It will also provide Māori with quality information about their health status.

---

2   *New Zealand Health Strategy: Future direction*. **http://www.health.govt.nz/publication/new-zealand-health-strategy-2016**

3   **http://www.health.govt.nz/about-ministry/what-we-do/new-zealand-health-strategy-update**

2 Ethnicity data context

2.1 Definition of ethnicity

Stats NZ is responsible for the definition of ethnicity across all-of-government. The official definition of ethnicity, contained in the document *Statistical Standard for Ethnicity* (statistical standard) and used by the health and disability sector, is:

> "Ethnicity is the ethnic group or groups that people identify with or feel they belong to. Ethnicity is a measure of cultural affiliation, as opposed to race, ancestry, nationality or citizenship. Ethnicity is self-perceived and people can belong to more than one ethnic group.

> An ethnic group is made up of people who have some or all of the following characteristics:

> - a common proper name
> - one or more elements of common culture which need not be specified, but may include religion, customs, or language
> - unique community of interests, feelings and actions
> - a shared sense of common origins or ancestry, and
> - a common geographic origin."[4]

There are key elements of the definition provided by Stats NZ in the **statistical standard** and associated documents that need to be reflected in the health and disability systems and processes covered by these protocols. The three key elements are that:

- ethnicity must be self-identified
- people may identify with more than one ethnic group (multiple ethnicities)
- ethnicity may change over time (ethnic mobility).

The protocols operationalise the statistical standard and provide further information that is specifically relevant to the health and disability sector.

---

[4] Stats NZ's *Statistical Standard for Ethnicity 2005*. **http://www.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/ethnicity.aspx**. Accessed 25 November 2015.

2.2 Data quality

The protocols were initially introduced in 2004 and required a standardised process for ethnicity data collection, recording and output across the health and disability sector. More than a decade on from the first release of the protocols, issues with the quality of ethnicity data in different sources persist. Inaccurate or misclassified ethnicity data can impact on health statistics, planning, funding and monitoring.

It is important ethnicity data from the health and disability sector is collected in the same way as data in the Census (collected by Stats NZ). Health statistics are frequently based on the census' population figures. For example, rates of hospitalisation are calculated by comparing hospital and census-based data to determine proportions of hospitalisations in different populations. The ability to compare data is known as numerator and denominator consistency. This consistency allows the comparison of health data by ethnicity where it is collected in different health and disability settings. However, the statistical standard notes that a person may give a different response depending on the context. Where this is acknowledged, a decision on the denominator source needs to be made on a case-by-case basis.

The objective of improving ethnicity data quality is to ensure that when such data is used, it has the same relevance and meaning throughout the health and disability sector. Achieving this goal requires the implementation of a standardised process that is used by all collectors, recorders (including data entry staff, administrators, clerks, health professional interviewers and researchers who use the classification structure to record ethnicity responses) and users of ethnicity data. Each step of the process must be undertaken in a standardised manner. The ethnicity data process includes three discrete steps: collection, recording and output.

Information systems, organisational processes and contract specifications (where appropriate) must ensure that the minimum requirements in the protocols are met at each relevant step. It is recognised some organisations will only collect and record ethnicity data, while other organisations will only output ethnicity data. Each step of the process may be audited for compliance with the protocols and data quality improvements made based on issues identified. Systematic audit tools have been developed for this purpose.[5] Audit activities include feedback, review, education and training.

2.3 Additional or related information

Iwi/hapū, ancestry, country of origin, nationality and religion are not required under these protocols. A collector may wish to request this information for specific purposes.

---

[5]   For example, Primary Care Ethnicity Data Audit Toolkit (EDAT) 2013,
      https://www.health.govt.nz/system/files/documents/publications/primary-care-ethnicity-data-audit-toolkit-jun13-v2.pdf

3 Protocols for collecting and confirming ethnicity data

This section details the processes relating to the collection and confirmation of ethnicity data from a respondent. In most cases, 'asking' means giving the respondent a form (paper or electronic) to fill out although in some instances it may mean reading out the question, such as over the telephone.

Data collectors in the health and disability sector tend to be administrators, clerks and health professionals. Respondents tend to be patients, clients and members of the health workforce when filling in human resources' forms. Ethnicity data are also collected in most surveys, along with other demographic information such as age and gender.

> **Collection** is to present the standard ethnicity question to a respondent.
>
> **Confirmation** is to present a respondent with their previous ethnicity response for confirmation or correction.

3.1　Protocol requirements for collection

1　The standard ethnicity question for the health and disability sector is the Statistics New Zealand 2018 Census ethnicity question.

2　Where a respondent may not be able to fill in a form or questionnaire themselves due to disability, incapacity, being deceased or being a newborn or child, the approach should be adjusted (see 3.3.4 Proxy response).

3　The respondent must identify their own ethnicity (called self-identification) regardless of collection method – for example, face-to-face contact, use of a form, electronic collection or telephone.

4　The collector must not guess ethnicity on behalf of the respondent or limit the number of ethnicities given.

5　Ethnicity may be collected at any time but must be collected at least every three years.

6　Ethnicity may be confirmed at any time when other personal details such as gender and contact details are also confirmed.

3.2　Standard ethnicity question

The standard ethnicity question for the health and disability sector mirrors the Statistics New Zealand 2018 Census ethnicity question. The ethnicity question has been rigorously tested by Stats NZ to establish the most effective wording, layout and font, and is re-analysed after every census. It must not be changed.

To maintain consistency of responses the following requirements must be met:

1　Whenever possible, use the actual graphic as shown in Figure 1 when presenting the ethnicity question. The font size, format and dimensions (83 mm x 78 mm) are to remain the same as in Figure 1 where practical. In a few circumstances it is appropriate to increase the size of the graphic, such as in presenting it on a laminated card to be given to respondents in an interview.

2　Where the actual graphic is not used, the minimum requirements are:

　　a.　The opening words of the question must be the same, ie, "Which ethnic group do you belong to? Mark the space or spaces that apply to you."

　　b.　The listed ethnicities must all be present and in the order shown in Figure 1. No additional categories may be added. It is preferable that the categories are listed vertically.

　　c.　Any collection mechanism must allow multiple ethnic groups to be selected and must allow multiple ethnic groups to be entered in the 'other' section.

Figure 1: Standard ethnicity question

Which ethnic group do you belong to?
Mark the space or spaces which apply to you.
- New Zealand European
- Māori
- Samoan
- Cook Island Maori
- Tongan
- Niuean
- Chinese
- Indian
- other such as *DUTCH, JAPANESE, TOKELAUAN.* Please state:

Source: Statistics New Zealand, 2018 Census

3.3 Process for collecting ethnicity data

When collecting ethnicity data, self-identification must be the process used to identify a respondent's ethnic group(s). The standard ethnicity question allows the respondent to state as many ethnicities as they feel they identify with.

It is unacceptable for the collector to guess any respondent's ethnicity or to complete the question on behalf of the respondent based on what they perceive to be the respondent's physical appearance, name or nationality. It is also unacceptable for a collector to correct what ethnicity a respondent identifies with if they disagree or to ask the respondent to identify a single ethnicity they most identify with (eg, principal ethnicity).

The generic process outlined below describes the basic steps involved in collecting ethnicity in four different situations: self-completion (paper or electronic form/ questionnaire); verbal response, assisted response and proxy response.

3.3.1   Self-completion

1   Paper form/questionnaire

    a.   Give or send the respondent the form or questionnaire that contains the ethnicity question to complete.

    b.   Advise the respondent (in person or by letter) that:

- additional information about ethnicity is available if required[6]
- where appropriate they can have access to an interpreter.

    c.   Collect the form or questionnaire.

    d.   Check that the ethnicity question has been completed on the form or questionnaire.

    e.   If the question has not been filled in, check the respondent has not accidentally omitted it. If the respondent doesn't wish to state their ethnicity or ethnicities, they should be asked to confirm whether or not they have left the question intentionally blank.

2   Electronic collection (eg, internet, portal or tablet)

    a.   Advise the respondent onscreen or within a help tab that:

- links to information about ethnicity on the Stats NZ and Ministry of Health websites are available if required
- where appropriate they can have access to an interpreter or links to where information about ethnicity is held in other languages.

    b.   The standard ethnicity question must be used. A button or similar selection option for each response is to be provided. The ability to select multiple responses must be provided.

    c.   The question format must include the text of the standard ethnicity question:

> Which ethnic group do you belong to? Mark the space or spaces which apply to you.
> - New Zealand European
> - Māori
> - Samoan
> - Cook Islands Māori
> - Tongan
> - Niuean
> - Chinese
> - Indian
> - Other (Please state, eg, Dutch, Japanese, Tokelauan).

---

[6]   See the following websites for further information: Ministry of Health Ethnicity Data Protocols for the Health and Disability sector: **http://www.health.govt.nz/publication/ethnicity-data-protocols-health-and-disability-sector**; and Stats NZ **http://www.stats.govt.nz**

d. The 'Other' response should have a visible area for entry of other ethnic groups so respondents are aware they can enter information here. If 'Other' is selected, the respondent must be allowed to enter one or more ethnic groups. Lists of frequently identified ethnic groups may be used but the respondent must not be limited to only the groups in the list.

e. If the respondent does not enter any ethnicity, the information system should prompt them to enter it and offer the following alternative responses:

  • 'I don't know my ethnicity'

  • 'I do not want to state my ethnicity'.

  The information system should not allow this question to be left blank.

3.3.2 Verbal response

If you are required to collect ethnicity data by relaying the information to the respondent verbally and completing the form yourself (eg, by telephone), identify a standard place in the call where the ethnicity question is to be asked. This place is most likely to be at the beginning or end of the call when other demographic information such as name, address, gender and age is collected.

At the start of the conversation, you should explain why you are phoning/collecting this data.

When asking the ethnicity question, the interviewer should state: "I am going to read out a list of ethnic groups. Can you tell me which ethnic group or groups you belong to:

• New Zealand European?

• Māori?

• Samoan?

• Cook Islands Māori?

• Tongan?

• Niuean?

• Chinese?

• Indian?

• Another ethnic group such as Dutch, Japanese or Tokelauan? Please say what it is."

The interviewer should read out each of the categories and wait for a yes/no answer to each. When an answer is given, the interviewer continues asking the rest of the list until it is completed.

Asking the question in this way allows for more than one ethnicity to be selected. It also allows reporting of all other ethnic groups chosen by the person in the 'Another ethnic group' category. It facilitates self-identification and allows the person to pick one or a number of categories that they identify with. This method reduces interviewer bias.[7]

---

[7] Extract from Stats NZ's *Statistical Standard for Ethnicity 2005* "Verbal response".

3.3.3   Assisted response

Where the respondent has a disability that will hinder their ability to complete the ethnicity question, an appropriate aid should be provided.

3.3.4   Proxy response

Where the respondent is unable to complete the ethnicity question them self, it is desirable to collect ethnicity data using a proxy response. The method to follow in three different circumstances is described below.

a)   Incapacity

If the respondent is incapable of completing the ethnicity question, where possible the nominee or next of kin should answer the ethnicity question on behalf of the respondent. If there is no one accompanying the respondent, undertake one of these:

- Locate the next of kin and ask them to provide a proxy response.
- Wait until the respondent is able to complete the ethnicity question.

b)   Deceased

Where the respondent is deceased, the standard question should be presented to the next of kin to provide a proxy response about the respondent's ethnicity.

c)   Newborns and children

Where the respondents are newborns or children, the parent/guardian(s) must always be given the opportunity to complete the standard ethnicity question. Systems must not, for example, default ethnicity to that of the mother (eg, by transfer from the maternal record). There should be sufficient space on the form for completion of the standard ethnicity question.

As noted in the statistical standard, there is no legal or recommended age at which a child can respond on their own behalf. When children are capable of understanding the concept of ethnicity, they should be given the opportunity to complete the question themselves. The appropriate age for such understanding is a matter of judgement.

3.4   Process for confirming ethnicity data

Confirming ethnicity data means printing (or reading) out all recorded variables, stating that this is the ethnicity information recorded currently for a respondent, and asking the respondent to confirm their ethnicity data.

If respondents alter or add detail to their response during confirmation, judgement should be used as to whether they should be asked to complete the standard ethnicity question or whether to update the system directly from their response. It may be possible to design forms used to confirm demographic details with a printout of the details and include a standard ethnicity question with instructions to complete this if the current details are incorrect.

The classification descriptions do not always match a person's response, for example a response of Okinawan is classified to level 4 code 44211 Japanese or level 2 code 44 Other Asian. Where systems record only the classified ethnicity codes and their descriptions, a note should be added when presenting the recorded ethnicities for confirmation that says:

> 'Your ethnicity may not be recorded as you provided it because we group some ethnicity responses together using the Stats NZ classifications system.'

If ethnicity has already been noted to be incorrect, refer to section 3.5 Process for correcting ethnicity data.

3.5 Process for correcting ethnicity data

The requirements of the *Health Information Privacy Code 1994*[8] apply to ethnicity data.

- Respondents have the right to know what ethnicity data is recorded about them. Respondents must be able to provide the correct data if it is incorrectly recorded.

- Data collectors must make sure they have accurate data and to correct that data if it is determined not to be accurate.

Where ethnicity data already held about a respondent is noted to be incorrect by:

- the respondent – it is best practice for the ethnicity to be collected from the respondent as outlined above and recorded correctly

- the user of the data – it is best practice to collect the data from the respondent as outlined above (the user may correct the ethnicity data if the ethnicity response is known for example, where there is a data entry error).

3.6 Frequency of collecting and confirming ethnicity data

Self-identification means people can change their ethnicity over time and report it differently in different contexts. In addition, there are known issues with ethnicity data quality that mean a person's previous response may not have been correctly collected

---

or recorded. To ensure recorded ethnicity data is of the highest quality, respondents must be regularly given the opportunity to supply and confirm their ethnicity response.

The following points outline guidance on the frequency of collection and confirmation:

1   Ethnicity data must be collected during the first interaction with the health sector or agency, or where there is no available ethnicity data for a respondent.

2   Ethnicity data may be collected at any time and must be collected at least every three years.

3   Ethnicity data must be collected if the ethnicity already held for a respondent is currently coded as any of the following: Not stated (99), Not elsewhere classified (nec) or Not further defined (nfd).

4   Ethnicity data may be confirmed at any time. Frequency of confirmation is up to the collector's discretion. A health service may choose to collect all demographic information, including ethnicity, more frequently. It may be appropriate to confirm ethnicity data at every interaction with a health service, however if a respondent is presenting frequently to a health service this may not be appropriate.

4 Protocols for classifying, recording and storing ethnicity data

This section details how ethnicity data are classified and recorded once they have been provided by a respondent. The recording process uses the classification structure to identify the appropriate codes and then stores the identified codes. Where possible, information systems should allow collectors to enter ethnicity directly as provided by respondents and automatically identify the code to record it without manual classification.

The recording process in the health and disability sector uses the Stats NZ's Ethnicity classification structure. Stats NZ reviews the classification categories periodically, for example after the census. Mapping is provided by Stats NZ when labels or codes change between classification versions.

See Stats NZ's website for the current version of the **Statistical Standard, Ethnicity**. Protocol requirements for recording.

1   The Stats NZ Ethnicity New Zealand Standard Classification 2005 structure must be used to code ethnicity data.

2   Ethnicity data must be recorded at level 4 (the most detailed level of the classification).

3   Information systems must be capable of recording up to six responses.

4   Where a respondent reports more than six ethnicities, the Stats NZ standard process for reducing multiple ethnic responses must be followed.

5   Where a respondent reports multiple ethnicities, a 'principal ethnicity' must not be recorded.

6   'New Zealander' and like responses should be coded to the 'New Zealander' code at level 4.

4.2   Classification structure

A classification structure is a way to group a set of related categories in a meaningful, systematic and standard format. Some important principles of classification are to use:

- mutually exclusive categories – that is, every response will fit into only one category in the classification

- a complete list of possible responses

- a framework to show how to classify responses.

A classification has a structured system and contains rules for aggregating data. Where they relate to an evolving concept like ethnicity, classifications are periodically updated so they reflect the contemporary situation as well as allowing comparisons over time. Stats NZ is responsible for oversight of the categories described in the classification

The Stats NZ's Ethnicity New Zealand Standard Classification is a hierarchical structure with four levels. The minimum requirement in the protocols is recording at the greatest level of detail at level 4. Level 4 contains more than 231 codes. This includes detailed ethnic group codes, the 'Other' code and residual category codes (see below).

The Stats NZ code system starts with a single digit at level 1. Further digits are added with each move to a more detailed level, thereby increasing granularity of information. Each more detailed level can be mapped up or aggregated to a higher level.

Figure 2: Ethnicity classification level relationships (Asian example)

Note: Current number of codes at date of publication.
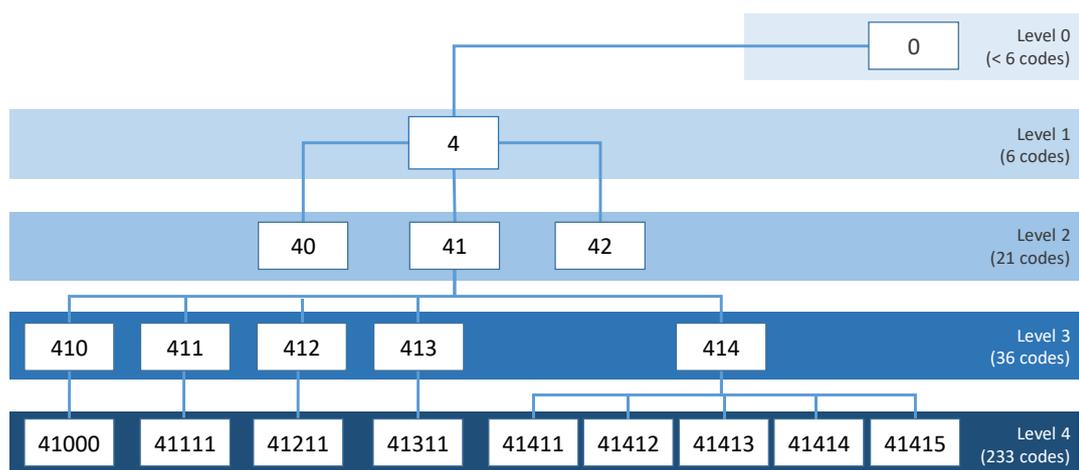
Example:
Level 1: (least detailed level) Code '4' is Asian
Level 2: Code '41' is Southeast Asian
Level 3: Code '414' is Other South East Asian
Level 4: (most detailed level), code '41411' is Burmese, code '41412' is Indonesian, '41413' is Lao, '41414 is Malay, and '41415' is Thai

4.3　How to record ethnicity

4.3.1　Determining the right code

Match the response with the ethnicity description and note the associated code. All the ethnicities printed on the standard question have the same description on their equivalent level 4 code. Where a respondent identifies with 'Other (such as Dutch, Japanese, Tokelauan)', the response should be matched with the description of the classification structure at level 4 and the associated level 4 code recorded.

4.3.2　Multiple responses

Respondents are able to identify as many ethnicities as they wish during collection of ethnicity data. Information systems must be able to classify and record up to six ethnicities as a minimum.

Where the respondent identifies more than six ethnicities, the method determined by Stats NZ (see Appendix A) should be used to reduce the number to six.

Recording up to six ethnicities means that level 1 ethnic groups will not be lost when reducing responses with more than six ethnicities.

Where a respondent reports multiple ethnicities, it is not appropriate to prioritise multiple ethnicities during recording or to ask respondents to choose a principal or primary ethnicity.

Prioritisation is only appropriate on output when using the ethnicity data for analysis and reporting (see '5.5.2 Prioritised output').

In situations where systems are not able to record up to six ethnicities and reduction of multiple responses to three recorded ethnicities is required, it is recommended that the prioritisation for output method is applied (as described in the Prioritised output section). Health information systems should not use the Stats NZ random method of reduction when reducing a response to less than six ethnicities because ethnic groups that have specific health needs may be lost in the process.

- Align responses coded at level 4 with their corresponding level 2 codes (first two digits).
- Prioritise using the level 2 prioritisation table.[9]
- Once the three prioritised ethnicities are determined, these should be recorded at the highest level of detail (level 4).

---

[9]　**http://www.health.govt.nz/nz-health-statistics/data-references/code-tables/common-code-tables/ethnicity-code-tables.**

4.3.3 Process for responses not matching standard classification descriptions

Level 4 classification is very detailed and most responses will be able to be classified. However, if a response is not in the classification structure, a decision must be made as to which is the most appropriate category for the response to be coded to.

Coding tools provided by Stats NZ (codefile) and Ministry of Health should be used to assist in finding appropriate codes. These should be used wherever possible to ensure consistent coding by all collectors.

- Stats NZ Classification Code Finder – **http://www.stats.govt.nz/tools_and_services/classificationcodefinder.aspx**
- Ministry of Health – **http://refraction.nz/eths**

Most sections of the classification contain 'Not elsewhere classified (nec)' categories that are used to code responses that do not match the standard description but can be identified as belonging to that part of the classification at level 4. For example, a response of 'Bosnian' should be coded to '12999 – European nec'.

Where responses are vague, it may be possible to classify them to a 'Not further defined (nfd)' category. For example, a response of 'European' should be coded as '12000 – Other European nfd'.

The Stats NZ standard contains advice on coding:

- multiple-worded responses like 'Fijian Indian' and 'Malaysian Chinese' that are one ethnic group
- hyphenated or linked responses like 'Tongan-Māori' or 'French/Austrian' that need to be classified as two responses
- where the response is an iwi, country or religion.

4.3.4 'New Zealander' responses

In the health and disability sector when a respondent has identified their ethnicity as 'New Zealander', 'Kiwi' or a variation of this response, this has historically been recorded as 'New Zealand European' (code 11) as outlined in the protocols and the supplementary notes 2009.

With the move to level 4 coding, this will no longer occur as 'New Zealander' has a separate level 4 code (code 61118). However, 'New Zealander' should continue to be aggregated to 'New Zealand European' in the process of output (see section 5.4) to maintain time series in the sector. It may also be output to the 'Other' category where this will enable comparability or maintain time series with a dataset.

4.3.5   Fijian Indian

There have been data quality issues noted with the collection, classification and recording of 'Fijian Indian'. The Ethnicity New Zealand Standard Classification codes 'Fijian Indian' as level 4 code 43112 (which aggregates at level 1 output to 'Asian'). Some respondents and some providers have chosen to alter collection forms or allow respondents to select 'Fijian' and 'Indian' separately. This creates two codes – 'Fijian 36111 (level 1 Pacific Peoples)' and 'Indian 43100 (level 1 Asian)' – with prioritised output this aggregates to 'level 1 Pacific Peoples'. This has implications for funding formulae and health status monitoring for both Pacific and Asian populations. Respondents identifying as 'Fijian Indian' must be coded 43112.

4.3.6   Residual codes

The residual codes, a series of codes beginning with 9, are used to record a meaningful description of some responses or outcomes of the collection process.

Table 1: Residual codes

Code	Description
94444	Don't know
95555	Refused to answer
96666	Repeated value
97777	Response unidentifiable
98888	Response outside scope
99999	Not stated

A copy of the Stats NZ standard definitions on use of residual codes can be found in the appendix (see 'Residual categories'). These coding guidelines are given to ensure consistency between collections. There are a number of ethnic groups that are multiple-worded responses but are one ethnic group. Some common examples are given here.

- Fijian Indian
- Turkish Cypriot
- Cook Islands Maori
- French Canadian
- Malaysian Chinese
- American European
- South African European.

There are responses that may be hyphenated or linked in some way, or written without linkage, that need to be classified as two responses. For example:

- Polish-Hungarian
- Tongan-Māori
- French/Austrian
- Australian/Lebanese
- Chinese New Zealander
- SerboCroatian.

There have been data quality issues noted with the use of the 'not stated' code (99 level 2; 99999 level 4). Historically this code has been used where data collectors have not been able to easily determine an appropriate ethnicity category or there have been issues with data transfer or extraction between systems. This code has also been used to populate blank data fields where ethnicity has not been collected but complete data is required. Systems where there is a high level of code 99 use (for example greater than 1 percent) should be examined for data quality issues and respondent's ethnicity data collected. Coding at level 4 ethnicity should virtually eliminate the need for the 'not stated' coding.

4.4 Coding non-ethnicity responses

4.4.1 Iwi

An iwi response to an ethnicity question is coded to 'Māori'. 'Iwi' responses may be collected by some providers but recording, classification and storage of iwi data isn't covered here. For further information on collecting iwi data, refer to the *Iwi Statistical Standard*.[10]

4.4.2 Country

A country response is coded to an appropriate ethnic group term – for example, 'Korea' is coded to 'Korean'.

---

[10]   See the following website for further information: **http://www.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/iwi.aspx**

4.4.3　Religion

Religious responses to the ethnicity question indicating an ethnic group are coded to the specific category in the classification. For example, 'Jewish' and 'Sikh' have separate categories at the most detailed level. Religious responses which are not an ethnic group – for example, 'Muslim', are not coded to an ethnic group but to 'response outside scope'.

5 Protocols for output of ethnicity data

Standard output provides comparable data about groups of interest for monitoring, planning, and the development and evaluation of policy for health care provision and the health status of New Zealanders. This section covers the ways in which ethnicity data can be output for uses such as analysis, funding formulae, health outcome evaluations or service access profiles. In output, there is no change to any of the recorded responses, although some aggregation is likely to occur in the reporting.

Data users tend to be researchers or analysts in primary health organisations, community providers, Māori or Pacific providers, district health boards (DHBs), non-government organisations, the Ministry of Health, the Accident Compensation Corporation and other government agencies. Output of ethnicity data may be flexible in terms of the level of detail required and the comparator categories presented. Standard categories are provided by the Stats NZ's standard classification structure as outlined in section 4. The protocols require classification and recording of ethnicity data at level 4. However, users may wish to output the data at a higher level of aggregation as indicated in the classification (level 1, level 2 or level 3). Other outputs may be used in health. Any output requires clear description of the method and categories used, and users of the data need to be aware of the limitations of any method used.

Sometimes ethnicity data is transferred from one system to another. For example, DHBs provide ethnicity data to national systems like the National Health Index (NHI). In most cases, this process is a simple transfer of recorded ethnicity. However, if more than six ethnicities at level 4 are being stored, the rules for classifying and recording must be followed.

5.1 Protocol requirements for output

1   The same output method and aggregate categories must be used for both numerator and denominator datasets.

2   'New Zealander' type responses should continue to be aggregated to 'New Zealand European' in the process of output to maintain time series in the sector. They may also be included in the output to the 'Other' category where this will enable comparability or maintain time series with a dataset.

3   Where there are multiple ethnicities for a respondent, one of the following three methods of output must be used:

    a.   total response (overlapping)

      b.   prioritised (the most common form of output in the health and disability sector)

      c.   sole/combination.

4   The method used must be described or noted along with any analysis. Standard descriptions are available in the statistical standard.

5   Up to six ethnicities at level 4 must be reported to Ministry of Health National Systems.

5.2 Stats NZ output aggregation groups

Users may choose to output ethnicity data at level 4. However, for many purposes in the health and disability sector, ethnicity data may be aggregated for output. Decisions about aggregation should take into account the balance between providing detailed ethnicity information and practical considerations such as available denominators and potential ability to identify with small numbers. Level 4 data can be aggregated within the Stats NZ's classification structure to the appropriate level for the user's purpose.

Codes can be aggregated by removing the last digits – from a five-digit code at level 4 (eg, 41111 Filipino), a three-digit code at level 3 (eg, 411 Filipino), a two-digit code at level 2 (eg, 41 Southeast Asian) and a one-digit code at level 1 (4 Asian) (refer to Figure 2). It is important the same aggregation is used for both numerators and denominators, and the categories described in any outputs.

5.3 Non-standard output groupings

In the health and disability sector, a 'super-aggregate' level 0 grouping is sometimes used. This is not an aggregation group under the Stats NZ's classification. There has been variability in the application of this level 0 category. For specific purposes particular super-aggregation categories may be preferred. For example Māori, Pacific Peoples, Asian, European/Other; or Māori, Pacific Peoples, European/Other. Where appropriate and practical, output is recommended at level 1 rather than level 0 (Māori, Pacific Peoples, Asian, Middle Eastern, Latin American and African (MELAA), European and Other) unless the subgroupings are too small for analysis or identify respondents.

The user may wish to use alternate non-standard Stats NZ aggregate grouping in order to make comparisons between populations. The appropriate comparison population will depend on the purpose for which the user is making the comparison. Examples of common comparison groupings in health include:

- Māori/Non-Māori (this is a Treaty of Waitangi-based comparison)

- Māori/Pacific Peoples/Non-Māori, Non-Pacific (this is a comparison used where inequalities for Pacific populations obscure inequalities for Māori, particularly in areas with high Pacific populations or in specific conditions)
- Māori/Total Population
- Māori/New Zealand European/Other
- Māori/Pacific Peoples/Chinese/Indian/Other (for some specific health conditions where inequalities for Asian sub-populations are high enough to warrant specific comparison, for example cardiovascular disease and gestational diabetes).

It is important the same aggregation is used for both numerators and denominators, and the categories are described in any outputs.

5.4 Output for 'New Zealander'

With the move to level 4 recording, 'New Zealander' responses will be recorded as code 61118. For the purposes of output however, it is recommended users continue to aggregate to 'New Zealand European' in the process of aggregation output (code 1 at level 1, code 11 at level 2) to maintain time series comparability.

If a collection currently outputs New Zealander to another category (eg, 'Other'), the user may wish to continue this practice to maintain time series.

Description of the output choice for New Zealander should be included in the data notes in any reporting.

5.5 Considerations for output of multiple ethnicity

There are a number of ways that multiple ethnicity responses can be output for analysis purposes. The three standard forms of output are described below.

These protocols require that one of these three forms is used for output, depending on which form of output is the most appropriate for the purposes of the data being reported. They also require that the form of output used in any particular table, graph or written analysis is made clear to readers.

All output methods have limitations that should be considered in analysis and reporting. For example, output methods generally include some level of aggregation and grouping which means ethnicity data does not strictly align with how respondents' self-identified.

5.5.1   Total response (overlapping) output

In total response output, each respondent is counted in each of the ethnic groups they reported. For example, where ethnicity is output at level 1, the respondent is counted once in each of the broad level 1 categories they identified with. Because individuals who indicate more than one ethnic group are counted more than once, the sum of the ethnic group populations will exceed the total population of New Zealand.

This form of output can be a useful option because it has the potential to represent people who do not identify with any given ethnic group, depending on the level of detail reported. In many cases in the health sector, total response output is reported at level 1.

Conversely, the approach has limitations in some situations in the health and disability sector. For example, it can create complexities in the distribution of funding based on population numbers or in monitoring changes in the ethnic composition of a population. In addition, it can create issues in the interpretation of data reported by ethnic groupings, where comparisons between groups include overlapping data. Further guidance on total response can be found in the statistical standard.

5.5.2   Prioritised output

One of the main criteria stipulated in the definition of ethnicity is that a person can belong to more than one ethnic group. The ethnicity question caters for multiple responses. However, the question does not ask people to indicate the ethnic group with which they identify the most strongly.

In prioritised output, each respondent is allocated to a single ethnic group using the prioritisation tables below. There are prioritisation orders for both level 1 and level 2 of the classification. The aim of prioritisation is to ensure that where some need exists to assign people to a single ethnic group, ethnic groups of policy importance or of small size, are not swamped by the New Zealand European ethnic group. Prioritisation is a reduction process for output and analysis purposes and does not assume this is the ethnic group that a respondent identifies most strongly with.

For example, if a data provider has indicated four ethnicities and these have been aggregated to level 2 as 40 – Asian, 21 – Māori, 51 – Middle Eastern and 11 – New Zealand European, the prioritised responses would be:

1   21 – Māori

2   40 – Asian

3   51 – Middle Eastern

4   11 – New Zealand European.

The following tables are available on **http://www.health.govt.nz/nz-health-statistics/data-references/code-tables/common-code-tables/ethnicity-code-tables**.

Table 2: Prioritisation for level 1 codes

Priority order	Ethnic group code (Level 1)	Ethnic group code description
1	2	Māori
2	3	Pacific Peoples
3	4	Asian
4	5	Middle Eastern/Latin American/African (MELAA)
5	6	Other Ethnicity
6	1	European
9	9	Residual Categories

Table 3: Prioritisation for level 2 codes

Priority order	Ethnic group code (Level 2)	Ethnic group code description
1	21	Māori
2	35	Tokelauan
3	36	Fijian
4	34	Niuean
5	33	Tongan
6	32	Cook Island Māori
7	31	Samoan
8	37	Other Pacific Peoples
9	30	Pacific Peoples not further defined
10	41	Southeast Asian
11	43	Indian
12	42	Chinese
13	44	Other Asian
14	40	Asian not further defined
15	52	Latin American
16	53	African
17	51	Middle Eastern
18	61	Other Ethnicity
20	12	Other European
21	10	European not further defined
22	11	New Zealand European
94	94	Don't know
95	95	Refused to answer
97	97	Response unidentifiable
99	99	Not stated

This output type is the one most frequently used in Ministry of Health statistics and is also widely used in the health and disability sector for funding calculations, monitoring changes in the ethnic composition of service utilisation and so on. It produces data that is easy to work with, as each individual appears only once. This means the sum of the ethnic group populations will add up to the total New Zealand population. Denominator data must also be prioritised to ensure numerator denominator consistency.

Limitations with prioritised output include that it places people in specific ethnic groups (high priority because of policy importance), which simplifies yet biases the resulting statistics as it over-represents some groups at the expense of others in ethnic group counts because of the order of prioritisation. It is also an externally applied single ethnicity which is inconsistent with the concept of self-identification including multiple ethnicities and should therefore not be used in processes of data collection or recording.

5.5.3   Sole/combination output

The sole/combination form of output (also referred to as single/combination) is rarely used in the health and disability sector. For completeness, this output has sole ethnic categories for respondents who report only one ethnic group and combination categories for respondents who state more than one ethnic group. Examples of combination categories are Samoan/Tongan, New Zealand European/Māori and Māori/Pacific Peoples.

The standard Stats NZ single/combination minimum output has nine groups: European, Māori, Pacific Peoples, Asian, Other, Māori/ European, Māori/Pacific Peoples, 'Two groups Not Elsewhere Identified' or the category titled 'Three groups'.

Limitations with sole/combination output include that it is relatively uncommon as an output method. In addition, some combination categories (eg, Two groups Not Elsewhere Identified) mean some ethnicities will not be identifiable from the data. Further guidance on sole/combination output can be found in the statistical standard. Sole/combination reporting at levels other than level 4 simplifies the results by placing some groups in aggregate categories.

5.6   Clear definition of output method

The aggregation level, comparison populations and method used for output of ethnicity analysis needs to be defined clearly for the user or reader in the data notes provided with any reporting. If different methods of analysing ethnicity at the output

stage are used, what method was used and how to interpret the results should be made explicit. Caveats and explanations should also be provided.

Below are some suggested examples to follow in two different contexts.

- Examples of how to include clear output information in titles:
    – *Male Life Expectancy (Prioritised Māori)*
    – *Hospitalisation Rates by Ethnic Group (Total Response) and Gender*
- Examples of how to include clear output information in the source or as a note:

    For total response (overlapping) analysis:

    'The ethnic data in this table allow for up to three responses per person. Where a person reported more than one ethnic group, that individual has been counted in each applicable group. Totals therefore do not add up to 100 percent.'

6   History of changes

This document reflects updated and revised ethnicity data protocols for the health and disability sector. It intends to more closely align health and disability sector processes and systems with the whole-of-government ethnicity data standard, as well as provide additional guidance to support high-quality ethnicity data. Table 4 outlines the changes to the protocols since their release in 2004, and the current alignment between the protocols and the statistical standard. Some further detail on the specific areas where the protocols do not fully align with the statistical standard is provided in the text below the table.

Table 4: Protocol revisions and alignment with Statistical Standard for Ethnicity 2005

Area	Stats NZ *Statistical Standard for Ethnicity* 2005 review 2009	Ministry of Health *Ethnicity Data Protocols for the Health and Disability Sector*		
		2004	2009 revision	2016 revision
Applies to	Whole-of-government administrative collections	Health and disability sector	Health and disability sector	Health and disability sector
Collection	Self-identified Standard census question	Self-identified Standard census question	Self-identified Standard census question	Self-identified Standard census question
Process of collection	Self-administered forms Internet Interviewer administered Telephone interviewing Proxy response	Self-completion (form/ questionnaire) Telephone Proxy response	Self-completion (form/ questionnaire) Telephone Proxy response	Self-completion (form/ questionnaire) Electronic (including online/ internet/portal) Verbal assisted response Proxy response
Electronic/ online collection of 'other' ethnicities	Allows an exemption process for collection where there is no facility to record free text 'other' categories (an identified list of level 2– 4 ethnicity categories)	No guidance	No guidance	Response to standard question 'other' to be recorded directly and either manually entered or electronically converted into the correct code
Frequency of collection	Census No guidance for other collections	No guidance	No guidance	3 yearly
Process for collection	No guidance	No guidance	No guidance	Guidance for collection, confirmation and correction

Area	Stats NZ *Statistical Standard for Ethnicity* 2005 review 2009	Ministry of Health *Ethnicity Data Protocols for the Health and Disability Sector*		
		2004	2009 revision	2016 revision
Classification structure	Four levels ETHNIC05 classification changes Recommend recording at level 4, minimum level 2	Five levels ETHNIC classification changes Recording minimum level 2	Five levels ETHNIC05 classification changes Recording minimum level 2	Five levels ETHNIC05 classification changes Recording minimum level 4
Recording multiple ethnicities	Recommend capture of up to six ethnicities, minimum up to three ethnicities	Minimum up to three ethnicities	Minimum up to three ethnicities	Minimum up to six ethnicities
Recording and output of 'New Zealander'	New category code 61 (61118) Advised to output to maintain time series	To code 11 (New Zealand European)	To code 11 (New Zealand European)	Level 4 code 61118 Output to maintain time series
Output groupings	Aggregate levels 1–4	Aggregate levels 1–4 plus additional level 0	Aggregate levels 1–4 plus additional level 0	Aggregate levels 1–4 plus additional level 0 Or other non-standard output groupings as determined by the user
Output method	Total response Sole combination	Total response Prioritised Sole combination	Total response Prioritised Sole combination	Total response Prioritised Sole combination
Method for reducing ethnicities if more than six	Random reduction manual method	Prioritise to level 2 and reduce if possible, prioritise and aggregate to level 1 if necessary to avoid losing a prioritised ethnic group	Prioritise to level 2 and reduce if possible, prioritise and aggregate to level 1 if necessary to avoid losing a prioritised ethnic group	Random reduction method if more than six ethnicities
Method of reducing ethnicities if system can only store three (transition period)				Use level 2 priority list, store three prioritised responses at most detailed level (level 4)

The protocols align with the statistical standard except in the following respects (in bold in the table above):

a.   The *Statistical Standard* recommends recording of ethnicity data at level 4 (the most detailed level possible) and up to six ethnicities, but does not require it. The health protocols require ethnicity data to be coded at level 4 and up to six ethnicities.

b.   While the *Statistical Standard* does not recommend the prioritisation method for output of ethnicity data, the protocols include prioritisation as an output method and it is commonly used. Prioritisation is maintained in the revised protocols.

c.   The *Statistical Standard* only has code levels 1-4. The revised protocols allow for the use of the health and disability sector of the super-aggregate code level 0 where this is appropriate, however it is recommended that routine data is reported at level 1. Alternate non-standard output groupings may also be used as determined by the user.

Note: Previously when the minimum multiple ethnicity requirement was recording up to three ethnicities, the method for reducing more than three ethnicities was prioritisation rather than the random method outlined in the statistical standard. This was so policy-prioritised ethnicities were not removed. The minimum requirements in the current protocols are for up to six ethnicities to be recorded, therefore it is unlikely that reduction will be routinely required and reducing to six will not lose any level 1 groupings. Where this is required, the protocols now align with the statistical standard and recommend random manual reduction.

Appendix A

Reducing multiple ethnic responses, manual methodology

1   Method for recording six ethnicity responses

This methodology paper is a technical paper for the development of software systems that support the inputting of large numbers of responses in surveys and administrative data sets. It outlines the treatment of responses where the number of ethnic groups given by an individual exceeds the number being output. Scenarios are described for reducing the number of multiple responses to six per individual, and to three responses, and examples are given of each.

If there are more than six responses per individual, then a random method for reducing the number of responses selects the six ethnicities to be retained. This manual method mirrors the software application method outlined in Appendix 2 in the link above.

To make your selection random, use a random number chart and methodically assign a number by either choosing a column or row to follow. After assigning the random numbers choose the lowest number as the response to be removed. This is explained fully in the method below.

Every level 1 ethnic group category that is represented by an individual's ethnicities must be represented in the final selection of responses. All level 1 ethnicity categories will be retained when the number of responses is reduced to six, as there are six categories at level 1.

Responses which would be coded to a residual category are removed first. For example, a response of vegetarian would be coded to the residual category 98888 response outside scope and would be the first response removed. If there are still more than six ethnicities then identify the level 1 categories they belong to.

Retain the responses from the level 1 categories that have just one ethnicity response belonging to them. All responses with the same first digit belong in the same level 1 category.

The level 1 categories that are represented by more than one ethnicity response are selected to reduce the number of responses through a random method. At least one ethnicity response representing each level 1 category is retained. Randomly select a level 1 category with more than one response. Do this by assigning each level 1 category a random number from a chart and select the lowest random number. This is the category from which an ethnicity will be selected to be removed. Next, randomly remove one response from the level 1 category selected. Do this by assigning each ethnicity in this category a random number and then remove the ethnicity with the lowest random number.

If this reduces the number of ethnicities to six then the random selection procedure can stop as the maximum number of responses has been retained. Otherwise, repeat this process until six responses remain.

The final selection must meet the requirement of no more than six ethnicities for processing and also retain information at level 1 of the classification for ethnicity.

2 Example of reducing to six responses

An individual's responses are French, Niuean, Cambodian, Vietnamese, English, Algerian and New Zealander.

There are no residual categories to remove.
- Classifying each ethnicity to their respective level one category has the following result:
- French and English are classified within the level 1 European category.
- Niuean is classified within the level 1 Pacific Peoples category.
- Cambodian and Vietnamese are classified within the level 1 Asian category.
- Algerian is classified within the level 1 Middle Eastern/Latin American/African (MELAA) category.
- New Zealander is classified within the level 1 Other Ethnicity category.

For three of the level 1 categories there is only one response given and these responses must be retained. They are Niuean, Algerian and New Zealander. That leaves two level 1 categories with more than one response from which to select the remaining responses to be retained.

Assign each level 1 category a random number. In this example, European is assigned 393 and Asian is assigned 214. The Asian category has the lowest random number and is the category from which an ethnicity will be selected to be removed.

Assign each ethnicity in the Asian category a random number: Cambodian is assigned 149 and Vietnamese is assigned 613. Cambodian has the lowest random number and so this response is removed.

The six ethnicities to be retained from the given responses are:
- French and English within the level 1 European category.

- Niuean within the level 1 Pacific category.
- Vietnamese within the level 1 Asian category.
- Algerian within the level 1 MELAA category.
- New Zealander within the level 1 Other Ethnicity category.

This selection meets the requirement of six ethnicities for processing and retains all the level 1 ethnicity information given by the individual.

Coding multiple worded responses

The following is an extract from the statistical standard.

These coding guidelines are given to ensure consistency between collections. There are a number of ethnic groups that are multiple-worded responses but are one ethnic group. Some common examples are given here.

- Fijian Indian
- Turkish Cypriot
- Cook Islands Maori
- French Canadian
- Malaysian Chinese
- American European
- South African European.

There are responses that may be hyphenated or linked in some way, or written without linkage, that need to be classified as two responses. For example:

- Polish-Hungarian
- Tongan-Māori
- French/Austrian
- Australian/Lebanese
- Chinese New Zealander
- SerboCroatian.

Residual categories

The following is an extract from the statistical standard.

Table 5: Definitions for Residual codes

Code	Description	Definition
94444	Don't know	The use of this category is necessary when the respondent is unsure of their ethnic group in an interviewer-administered survey (ie, asked verbally) or writes this in as a response.
95555	Refused to answer	This category is only used when it is known that the respondent has intentionally chosen not to respond to the question or wishes their objection to the question to be recorded. Its use is most applicable in face-to-face or telephone interviews, but may be used in self-completed questionnaires if the respondent has clearly indicated they refuse or object to answering the question.
96666	Repeated value	It is used when a respondent has given two responses that have the same code. This may be two written responses, or one tick box response and one written response. For example, someone may tick the New Zealand European tick box and write in New Zealand European.
97777	Response unidentifiable	This category is used when there is a response given, but is illegible, or it is unclear what the meaning or intent of the response is. This most commonly occurs when the response being classified contains insufficient detail, is ambiguous or vague.
98888	Response outside scope	This category is used for responses that are positively identified, that is, the meaning and the intent are clear but fall outside the scope of the classification/topic as defined in the standard. For example, a response of 'vegetarian' falls outside the scope of the ethnicity classification.
99999	Not stated	This category is only used where a respondent has not given any response to the question asked in a self-administered questionnaire, that is, it is solely for non-response.

Ethnicity classification levels

The Stats NZ's Ethnicity New Zealand Standard Classification is a hierarchical structure with four levels, as illustrated in Figure 2 (see section 4.2). For the current details of the classification, see **Ethnicity New Zealand Standard Classification 2005**.

For mapping details for those systems using an earlier version of the Ethnicity New Zealand Standard Classification 2005 is provided by Stats NZ. See Stat NZ's **Ethnicity New Zealand Standard Classification 2005 V2.0.0 to V2.1.0 V1.0.0**.

Each level is described in more detail below, along with an example of the codes. In the following tables 'not further defined' is documented by 'nfd' and 'not elsewhere classified' is documented by 'nec'.

> **Note:** The following information has been extracted from Stats NZ's Ethnicity Standard Classification 2005 – v2.1.0 (valid from 18 June 2020). For the most recent version of this information, refer to **Stats NZ's Classifications**.

Level 1

Level 1 is the top level of the Stats NZ numbering hierarchy. As indicated in Figure 2 in section 4.2, all codes at the more detailed levels derive from five level 1 codes.

Level 1 – alphabetical order		Level 1 – code order	
Code	**Descriptor**	**Code**	**Descriptor**
4	Asian	1	European
1	European	2	Māori
2	Māori	3	Pacific Peoples
5	Middle Eastern/Latin American/African	4	Asian
6	Other Ethnicity	5	Middle Eastern/Latin American/African
3	Pacific Peoples	6	Other Ethnicity
9	Residual Categories	9	Residual Categories

Level 2

Level 2 is a two-digit grouping. It includes six '9' series codes.

Level 2 – alphabetical order		Level 2 – code order	
Code	**Descriptor**	**Code**	**Descriptor**
53	African	10	European nfd
40	Asian nfd	11	New Zealand European
42	Chinese	12	Other European
32	Cook Islands Maori	21	Māori
94	Don't Know	30	Pacific Peoples nfd
10	European nfd	31	Samoan
36	Fijian	32	Cook Islands Maori
43	Indian	33	Tongan
52	Latin American	34	Niuean
21	Māori	35	Tokelauan
51	Middle Eastern	36	Fijian
11	New Zealand European	37	Other Pacific Peoples
34	Niuean	40	Asian nfd
99	Not Stated	41	Southeast Asian
44	Other Asian	42	Chinese
61	Other Ethnicity	43	Indian
12	Other European	44	Other Asian
37	Other Pacific Peoples	51	Middle Eastern
30	Pacific Peoples nfd	52	Latin American
95	Refused to Answer	53	African
96	Repeated Value	61	Other Ethnicity
98	Response Outside Scope	94	Don't Know
97	Response Unidentifiable	95	Refused to Answer
31	Samoan	96	Repeated Value
41	Southeast Asian	97	Response Unidentifiable
35	Tokelauan	98	Response Outside Scope
33	Tongan	99	Not Stated

Level 3

Level 3 is a three-digit grouping and includes six '9' series codes.

Level 3 – alphabetical order		Level 3 – code order	
Code	**Descriptor**	**Code**	**Descriptor**
531	African	100	European nfd
400	Asian nfd	111	New Zealand European
128	Australian	121	British and Irish
121	British and Irish	122	Dutch
412	Cambodian	123	Greek
421	Chinese	124	Polish
321	Cook Islands Maori	125	South Slav
944	Don't Know	126	Italian
122	Dutch	127	German
100	European nfd	128	Australian
361	Fijian	129	Other European
411	Filipino	211	Māori
127	German	300	Pacific Peoples nfd
123	Greek	311	Samoan
431	Indian	321	Cook Islands Maori
126	Italian	331	Tongan
442	Japanese	341	Niuean
443	Korean	351	Tokelauan
521	Latin American	361	Fijian
211	Māori	371	Other Pacific Peoples
511	Middle Eastern	400	Asian nfd
111	New Zealand European	410	Southeast Asian nfd
341	Niuean	411	Filipino
999	Not Stated	412	Cambodian
444	Other Asian	413	Vietnamese
611	Other Ethnicity	414	Other Southeast Asian
129	Other European	421	Chinese
371	Other Pacific Peoples	431	Indian
414	Other Southeast Asian	441	Sri Lankan
300	Pacific Peoples nfd	442	Japanese
124	Polish	443	Korean
955	Refused to Answer	444	Other Asian
966	Repeated Value	511	Middle Eastern

Level 3 – alphabetical order			Level 3 – code order	
Code	**Descriptor**		**Code**	**Descriptor**
988	Response Outside Scope		521	Latin American
977	Response Unidentifiable		531	African
311	Samoan		611	Other Ethnicity
125	South Slav		944	Don't Know
410	Southeast Asian nfd		955	Refused to Answer
441	Sri Lankan		966	Repeated Value
351	Tokelauan		977	Response Unidentifiable
331	Tongan		988	Response Outside Scope
413	Vietnamese		999	Not Stated

Level 4

Level 4 is a five-digit grouping and includes six '9' series codes.

At the most detailed level of the classification structure (level 4), larger groups are disaggregated or differentiated according to:

- geographic locality or origin (country, regions within a country, or island within a particular island group)
- cultural differences (which include distinctions such as language and religious belief)
- size.

Individual ethnic groups are classified into progressively broader groups according to geographic location or origin, cultural similarities, and size (in New Zealand). Many of the terms used for ethnic groups reflect geographic location or origin.

The size criterion helps to determine whether an ethnic group should be classified under an Other category at a particular level. Generally, the larger ethnic groups are disaggregated, while smaller ethnic groups fall into Other categories. Ethnic groups with very small numbers fall into Not elsewhere classified categories at level 4.

The category of Māori stands alone at all levels of the classification. This is in recognition of Māori as the tangata whenua (original inhabitants) of New Zealand and New Zealand's unique position as the only country where there is a commitment to the status, preservation and continuity of Māori cultural traditions (including language).

At levels 2, 3 and 4, distinctions between Pacific Island groups are made. This approach arose out of Recommendation 8 of the Ethnic Review Committee's report on ethnic statistics (Department of Statistics 1988). It reflects historical relationships with Pacific Island groups who represented a significant proportion of the migration flow into New Zealand. It also recognises the predicted increase in the proportion of the New Zealand population who are people of Pacific Island descent.

In the health and disability sector, Pacific Peoples, like Māori, have poorer on average health status. Both ethnic populations are the focus for one of the Government's key goals to reduce inequalities.

Level 4 – alphabetical order		Level 4 – code order	
Code	Descriptor	Code	Descriptor
44411	Afghan	10000	European nfd
53116	African American	11111	New Zealand European
53199	African nec	12100	British nfd
53100	African nfd	12111	Celtic nfd
12949	Afrikaner	12112	Channel Islander
12911	Albanian	12113	Cornish
51111	Algerian	12114	English
12943	American	12116	Irish
43117	Anglo Indian	12117	Manx
51112	Arab	12119	Scottish
52111	Argentinian	12121	Welsh
12912	Armenian	12199	British nec
44499	Asian nec	12211	Dutch
40000	Asian nfd	12311	Greek
51113	Assyrian	12411	Polish
12811	Australian	12500	South Slav nfd
12913	Austrian	12511	Croatian
44412	Bangladeshi	12512	Dalmatian
12914	Belgian	12513	Macedonian
12916	Belorussian	12514	Serbian
43111	Bengali	12515	Slovenian
44417	Bhutanese	12516	Bosnian
52112	Bolivian	12599	South Slav nec
12516	Bosnian	12611	Italian
52113	Brazilian	12711	German
12199	British nec	12811	Australian
12100	British nfd	12911	Albanian
12915	Bulgarian	12912	Armenian
41411	Burmese	12913	Austrian
53123	Burundian	12914	Belgian
41211	Cambodian	12915	Bulgarian
42112	Cambodian Chinese	12916	Belorussian
12945	Canadian	12918	Cypriot nfd

Level 4 – alphabetical order			Level 4 – code order	
Code	**Descriptor**		**Code**	**Descriptor**
53118	Caribbean		12919	Czech
12111	Celtic nfd		12920	Danish
12112	Channel Islander		12921	Estonian
52114	Chilean		12922	Finnish
41417	Chin		12923	Flemish
42199	Chinese nec		12924	French
42100	Chinese nfd		12926	Hungarian
52115	Colombian		12927	Icelandic
53124	Congolese		12928	Latvian
32100	Cook Islands Maori		12929	Lithuanian
12113	Cornish		12930	Maltese
12511	Croatian		12931	Norwegian
12918	Cypriot nfd		12932	Portuguese
12919	Czech		12933	Romanian
12512	Dalmatian		12934	Romani
12920	Danish		12935	Russian
94444	Don't Know		12937	Slavic
12211	Dutch		12938	Slovak
52118	Ecuadorian		12939	Spanish
51114	Egyptian		12940	Swedish
12114	English		12941	Swiss
53120	Eritrean		12942	Ukrainian
12921	Estonian		12943	American
53121	Ethiopian		12945	Canadian
44416	Eurasian		12947	New Caledonian
12999	European nec		12948	South African European
10000	European nfd		12949	Afrikaner
36111	Fijian		12950	Zimbabwean European
43112	Fijian Indian		12999	European nec
41111	Filipino		21111	Māori
12922	Finnish		30000	Pacific Peoples nfd
12923	Flemish		31111	Samoan
12924	French		32100	Cook Islands Maori
12711	German		33111	Tongan
53122	Ghanaian		34111	Niuean
12311	Greek		35111	Tokelauan

Level 4 – alphabetical order		Level 4 – code order	
Code	Descriptor	Code	Descriptor
37122	Hawaiian	36111	Fijian
42111	Hong Kong Chinese	37112	Indigenous Australian
12926	Hungarian	37122	Hawaiian
12927	Icelandic	37124	Kiribati
43199	Indian nec	37130	Nauruan
43100	Indian nfd	37135	Papua New Guinean
43114	Indian Tamil	37137	Pitcairn Islander
61113	Indigenous American	37138	Rotuman
37112	Indigenous Australian	37140	Tahitian
41412	Indonesian	37141	Solomon Islander
51115	Iranian/Persian	37144	Tuvaluan
51116	Iraqi	37145	Ni Vanuatu
12116	Irish	37199	Pacific Peoples nec
51117	Israeli/Jewish	40000	Asian nfd
12611	Italian	41000	Southeast Asian nfd
53113	Jamaican	41111	Filipino
44211	Japanese	41211	Cambodian
51118	Jordanian	41311	Vietnamese
41416	Karen	41411	Burmese
53114	Kenyan	41412	Indonesian
37124	Kiribati	41413	Lao
44311	Korean	41414	Malay
51119	Kurd	41415	Thai
41413	Lao	41416	Karen
52199	Latin American nec	41417	Chin
52100	Latin American nfd	41499	Southeast Asian nec
12928	Latvian	42100	Chinese nfd
51120	Lebanese	42111	Hong Kong Chinese
12929	Lithuanian	42112	Cambodian Chinese
12513	Macedonian	42113	Malaysian Chinese
41414	Malay	42114	Singaporean Chinese
42113	Malaysian Chinese	42115	Vietnamese Chinese
43118	Malaysian Indian	42116	Taiwanese
44418	Maldivian	42199	Chinese nec
12930	Maltese	43100	Indian nfd
12117	Manx	43111	Bengali

Level 4 – alphabetical order			Level 4 – code order	
Code	Descriptor		Code	Descriptor
21111	Māori		43112	Fijian Indian
61115	Mauritian		43114	Indian Tamil
52123	Mexican		43115	Punjabi
51199	Middle Eastern nec		43116	Sikh
51100	Middle Eastern nfd		43117	Anglo Indian
44419	Mongolian		43118	Malaysian Indian
51122	Moroccan		43119	South African Indian
37130	Nauruan		43199	Indian nec
44413	Nepali		44100	Sri Lankan nfd
12947	New Caledonian		44111	Sinhalese
11111	New Zealand European		44112	Sri Lankan Tamil
61118	New Zealander		44199	Sri Lankan nec
37145	Ni Vanuatu		44211	Japanese
53115	Nigerian		44311	Korean
34111	Niuean		44411	Afghan
12931	Norwegian		44412	Bangladeshi
99999	Not Stated		44413	Nepali
61199	Other Ethnicity nec		44414	Pakistani
61117	Other South African		44415	Tibetan
53127	Other Zimbabwean		44416	Eurasian
37199	Pacific Peoples nec		44417	Bhutanese
30000	Pacific Peoples nfd		44418	Maldivian
44414	Pakistani		44419	Mongolian
51124	Palestinian		44499	Asian nec
37135	Papua New Guinean		51100	Middle Eastern nfd
52127	Peruvian		51111	Algerian
37137	Pitcairn Islander		51112	Arab
12411	Polish		51113	Assyrian
12932	Portuguese		51114	Egyptian
52128	Puerto Rican		51115	Iranian/Persian
43115	Punjabi		51116	Iraqi
95555	Refused to Answer		51117	Israeli/Jewish
96666	Repeated Value		51118	Jordanian
98888	Response Outside Scope		51119	Kurd
97777	Response Unidentifiable		51120	Lebanese
12934	Romani		51122	Moroccan

Level 4 – alphabetical order			Level 4 – code order	
Code	Descriptor		Code	Descriptor
12933	Romanian		51124	Palestinian
37138	Rotuman		51125	Syrian
12935	Russian		51127	Turkish
31111	Samoan		51199	Middle Eastern nec
12119	Scottish		52100	Latin American nfd
12514	Serbian		52111	Argentinian
61116	Seychellois		52112	Bolivian
43116	Sikh		52113	Brazilian
42114	Singaporean Chinese		52114	Chilean
44111	Sinhalese		52115	Colombian
12937	Slavic		52118	Ecuadorian
12938	Slovak		52123	Mexican
12515	Slovenian		52127	Peruvian
37141	Solomon Islander		52128	Puerto Rican
53119	Somali		52129	Uruguayan
12948	South African European		52130	Venezuelan
43119	South African Indian		52199	Latin American nec
12599	South Slav nec		53100	African nfd
12500	South Slav nfd		53113	Jamaican
41499	Southeast Asian nec		53114	Kenyan
41000	Southeast Asian nfd		53115	Nigerian
12939	Spanish		53116	African American
44199	Sri Lankan nec		53118	Caribbean
44100	Sri Lankan nfd		53119	Somali
44112	Sri Lankan Tamil		53120	Eritrean
53125	Sudanese		53121	Ethiopian
12940	Swedish		53122	Ghanaian
12941	Swiss		53123	Burundian
51125	Syrian		53124	Congolese
37140	Tahitian		53125	Sudanese
42116	Taiwanese		53126	Zambian
41415	Thai		53127	Other Zimbabwean
44415	Tibetan		53199	African nec
35111	Tokelauan		61113	Indigenous American
33111	Tongan		61115	Mauritian
51127	Turkish		61116	Seychellois

| Level 4 – alphabetical order | | | Level 4 – code order | |
Code	Descriptor		Code	Descriptor
37144	Tuvaluan		61117	Other South African
12942	Ukrainian		61118	New Zealander
52129	Uruguayan		61199	Other Ethnicity nec
52130	Venezuelan		94444	Don't Know
41311	Vietnamese		95555	Refused to Answer
42115	Vietnamese Chinese		96666	Repeated Value
12121	Welsh		97777	Response Unidentifiable
53126	Zambian		98888	Response Outside Scope
12950	Zimbabwean European		99999	Not Stated

Appendix B

Revision history

Version updated	Updates
April 2021	The following updates have been included:

The following updates have been included:

- references to census 2013 ethnicity question changed to census 2018
- new links included for all Stats NZ's standards and classification website pages
- inclusion of links to Stats NZ's website for mapping details between Ethnicity classification versions
- changes to reflect latest version of the Ethnicity New Zealand Standard Classification 2005 V2.1.0

Code	Old label	New label	Version updated
12934	Gypsy	Romani	V2.1.0 (18/06/2020)
44411	Afghani	Afghan	V2.1.0 (18/06/2020)
44413	Nepalese	Nepali	V2.1.0 (18/06/2020)