Stats NZ
Tatauranga Aotearoa

Quality of ethnicity data in the experimental administrative population census (APC):

High level summary

Census Transformation

Te Kāwanatanga o Aotearoa
New Zealand Government

Crown copyright ©
This work is licensed under the Creative Commons Attribution 4.0 International licence. Except for any photographic images, or any content with a specific copyright statement, you are free to copy, distribute, and adapt the work, as long as you attribute the work to Statistics NZ and abide by the other licence terms. To ask permission to reuse a photographic image, send a request to info@stats.govt.nz. Please note you may not use any departmental or governmental emblem, logo, or coat of arms in any way that infringes any provision of the Flags, Emblems, and Names Protection Act 1981, or would infringe any provision if the use occurred within New Zealand. Use the wording 'Statistics New Zealand' in your attribution, not the Statistics NZ logo.

Disclaimer:

Access to the data used in this study was provided by Stats NZ under conditions designed to give effect to the security and confidentiality provisions of the Data and Statistics Act 2022. These results are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI), which is carefully managed by Stats NZ.

Integrated data has more information about the IDI.

Liability statement
While all care and diligence has been used in processing, analysing, and extracting data and information in this publication, Statistics New Zealand gives no warranty it is error free and will not be liable for any loss or damage suffered by the use directly, or indirectly, of the information in this publication.

Citation
Stats NZ (2025). *Quality of ethnicity data in the experimental administrative population census (APC): High level summary.* Retrieved from www.stats.govt.nz.

ISBN: 978-1-991307-82-8 (online)

Published in October 2025 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 5

Summary 5

Key findings 5

Background 6

Research questions 7

Data sources 7

Results for quality of ethnicity in the APC 8

Discussion and future work 14

References 16

Appendix: Proportions of level 3 and level 4 ethnicites 17

Tables and figures

List of tables

Table 1 Source ranking and accumulated ethnicity coverage rates for the APC, 2023 7
Table 2 Level 1 ethnicity counts by ethnicity in the APC, 2023 Census, and 2023 estimated resident population (ERP) 9
Table 3 Quality summary for ethnicity at levels 2 to 4 10
Table 4 Counts and percentages of level 2 ethnicity in the APC in 2023 and 2023 Census 11

List of figures

Figure 1 Coverage of level 1 ethnicities in the APC by single age (years) and sex, 2023 9
Figure 2 Counts of Pacific level 2 ethnicities by five-year age groups, APC in 2023 and 2023 Census 12
Figure 3 Distribution of the 15 largest level 3 ethnic groups, APC in 2023 and 2023 Census 13
Figure 4 Distribution of the 20 largest level 4 ethnic groups, APC in 2023 and 2023 Census. 14

Purpose

This paper assesses how well admin data in the Experimental administrative population census (APC) is currently able to provide ethnicity data from administrative (admin) sources compared with ethnicity data collected in the 2023 Census.

Summary

Ethnicity is a core demographic variable.

Coverage of ethnicity across all levels in the APC is very high when compared with the 2023 Census.

Our evaluation compares the APC with the direct census responses (86 percent of all census values). It excludes responses from 2023 Census that were derived from historical censuses, admin data, and statistical imputation.

Individuals can have more than one ethnicity. Even though most of the population has at least one associated ethnicity in the APC, not all ethnicity combinations are captured. From this paper, it is evident that at the more detailed levels of ethnicity, there are undercounts in the APC, particularly for 'New Zealand European' at levels 2 to 4.

Ethnicity-specific data issues that affect multiple data providers are also apparent in the APC. In particular, the overcounting of 'not further defined' and 'other' group categories, along with the 'Fijian' and 'British and Irish' ethnicities is evident.

Key findings

- Coverage of ethnicity in the APC across all ethnicity levels is greater than 99 percent for level 1 and 2 ethnicities and 98 percent and 95 percent for level 3 and 4 ethnicities respectively.
- At the national level, proportions of level 1 and 2 ethnicities in the 2023 APC are similar to those in the 2023 Census.
- At levels 3 and 4, we see an undercount of many ethnic groups in admin sources compared with the 2023 Census, due to how ethnicity is being collected and classified.
- Due to high coverage, overall quality measures across all ethnicity levels are largely due to consistency between individual APC and 2023 Census responses. Consistency of admin data with 2023 Census data shows promise across all ethnicity levels, with quality ratings greater than 0.95 for all level 1 groups and overall quality ratings of 0.77 for level 2, 0.75 for level 3, and 0.71 for level 4. Key drivers of this reduced quality in lower-level ethnicities overall is the undercounting of 'New Zealand European' and corresponding overcount of 'European not further defined' categories.
- This work will inform improvements to data sourcing and admin sources. Future work to improve ethnicity derivations include adding 2023 and 2018 Census responses to the APC

ethnicity outputs and investigating alternative methods to derive ethnicity from multiple admin sources.

Background

Ethnicity classification

Ethnicity is a hierarchical classification of four levels. Level 1 of the classification has six categories: European; Māori; Pacific Peoples; Asian; Middle Eastern, Latin American, and African (MELAA); and 'Other'. Level 1 is used solely for outputs, not collection. The 'Other' category is very small and is combined with European for the official estimated ethnic resident population.

Apart from Māori, level 1 categories are groupings of ethnicities, not ethnicities as such. Level 2 has 21 categories, level 3 has 36 categories, and level 4 has 180 categories, excluding residual categories. Ethnicities in New Zealand with larger populations appear in level 2 while those with smaller populations are only reported as separate ethnicities at level 3 or level 4. For example, for the level 1 Pacific Peoples group, Samoan and Tongan are two of the six Pacific ethnicities at level 2. Smaller Pacific ethnicities are coded as 'Other Pacific', which at level 4 includes ethnicities such as Hawaiian and Tahitian. Ethnicity New Zealand Standard Classification provides details.

Previous research

Individual admin data sources for ethnicity have documented quality issues associated with them. For example, when health datasets were linked with 2018 Census data, the health datasets undercounted Māori by 16 percent, with similarly large undercounts for the Pacific ethnic groups (Harris et al, 2022). A 2018 investigation found that overcounting 'not further defined' and grouped categories such as 'Other Pacific' was prevalent in certain agencies (Stats NZ, 2018). Other data quality issues found in the 2018 study were over-reporting of Fijian ethnicity across most agencies, as 'Fijian Indians' are being incorrectly coded to Fijian, and over-reporting of the African ethnic group in some agencies. Dixon et al (2024) noted that benefit data from the Ministry of Social Development (MSD) available in the IDI incorrectly codes Latin American and African ethnicities under the 'Other Ethnicity' group instead of under the MELAA group.

Dixon et al (2024) provides a detailed analysis of the quality of individual admin data sources for deriving information about ethnicity, at all four levels. The report looks at three measures for the data sources: individual-level consistency, precision (in the case of ethnicity comparisons with 2018 Census, the ratio of the number of consistent responses for an ethnicity to the total number of that ethnicity in the admin data), and recall (the ratio of the number of consistent responses for an ethnicity to the number of that ethnicity in the census data) and highlights differences between each administrative source and ethnicity.

Dixon et al (2024) noted that:

- Admin data sources tend to undercount individuals with multiple ethnicities, meaning there is a larger proportion of individuals with only a single ethnicity than in the 2018 Census. The proportion of multiple ethnicities are between 20 and 80 percent of the proportion in census. The exception to this is DIA birth registration data, which has approximately twice the number of individuals with multiple ethnicities.

- In general, admin data sources have improved since the initial work of Reid et al (2016), though issues remain for some sources and for specific ethnic groups.

In Dixon et al (2024), precision and recall scores highlight some differences between each admin source and each ethnicity. For example, at level 1, Māori, Pacific Peoples, and MELAA have, on average, lower precision and recall scores compared with European and Asian ethnic groups. But the 'Other ethnicity' group has considerably lower scores for both measures compared with all other level 1 ethnic groups. For the most part, the 'other', 'not further defined', and 'not elsewhere classified' ethnic groups have lower scores in admin data sources than the more specific ethnic groups.

Research questions

The main questions in this paper assessing quality include:

1. What is the coverage and overall quality ratings for ethnicity at levels 1 to 4 in the 2023 APC compared with 2023 Census?
2. How accurate are 2023 counts for APC, census and estimated resident population (ERP) ethnicities at levels 1 to 4?
3. What do different level 2 and 3 ethnic group counts look like by age?
4. What ethnicity levels are most concerning in terms of quality?

Data sources

This analysis was carried out using Stats NZ's Integrated Data Infrastructure (IDI). The IDI is a large research database that holds de-identified microdata about people and households. The IDI's data comes from government agencies, Stats NZ surveys, and non-government organisations. The IDI also contains 2013, 2018, and 2023 Census data. All data is linked at the individual level. Source agencies provide data periodically, and the IDI is updated three times a year. This update is known as the IDI refresh. This paper used the June 2025 refresh in the IDI.

Admin data in the IDI has also been used to produce the Experimental administrative population census (APC). The APC is an annual time series that derives census-type information from linked admin data. Stats NZ has released three iterations of the APC: the first and second iterations in 2021 and 2022, and most recently the third iteration in 2023 (Stats NZ, 2021, 2022, 2023a).

The APC derives ethnicity from multiple administrative sources, using a source-ranking method (Stats NZ, 2023b). Responses from the 2013 Census are also used, but by design are ranked lower in the APC ranking than they would be in an admin-data-first census. The ranking method applied is described in Reid et al (2016) and refined in Stats NZ (2018). The ranking of the admin sources was determined by how well these sources agree with census and has not changed from 2018 to 2023. Table 1 shows the resulting ranking and accumulated coverage rates for the APC.

Table 1

Source ranking and accumulated ethnicity coverage rates for the APC, 2023

Source	Minimum level of ethnicity collected	Accumulated ethnicity coverage (percent)		
		Level 1 and 2	Level 3	Level 4
DIA (birth registrations)	Level 4	52.5	52.5	52.5

MoE (tertiary)	Level 3	76.4	76.4	73.6
MoH	Level 2	99.0	96.6	92.6
MoE (primary & secondary)	Level 3	99.1	97.2	93.3
MSD	Level 2	99.5	97.9	94.3
2013 Census	Level 4	99.5	98.1	94.7

Note: DIA = Department of Internal Affairs; MoE = Ministry of Education; MoH = Ministry of Health; MSD = Ministry of Social Development.

As some ethnicities, such as Māori, do not break down into smaller categories at more detailed levels, agencies can also contribute to level 3 and 4 information, even if they only collect at level 2 or 3.

Source: Stats NZ

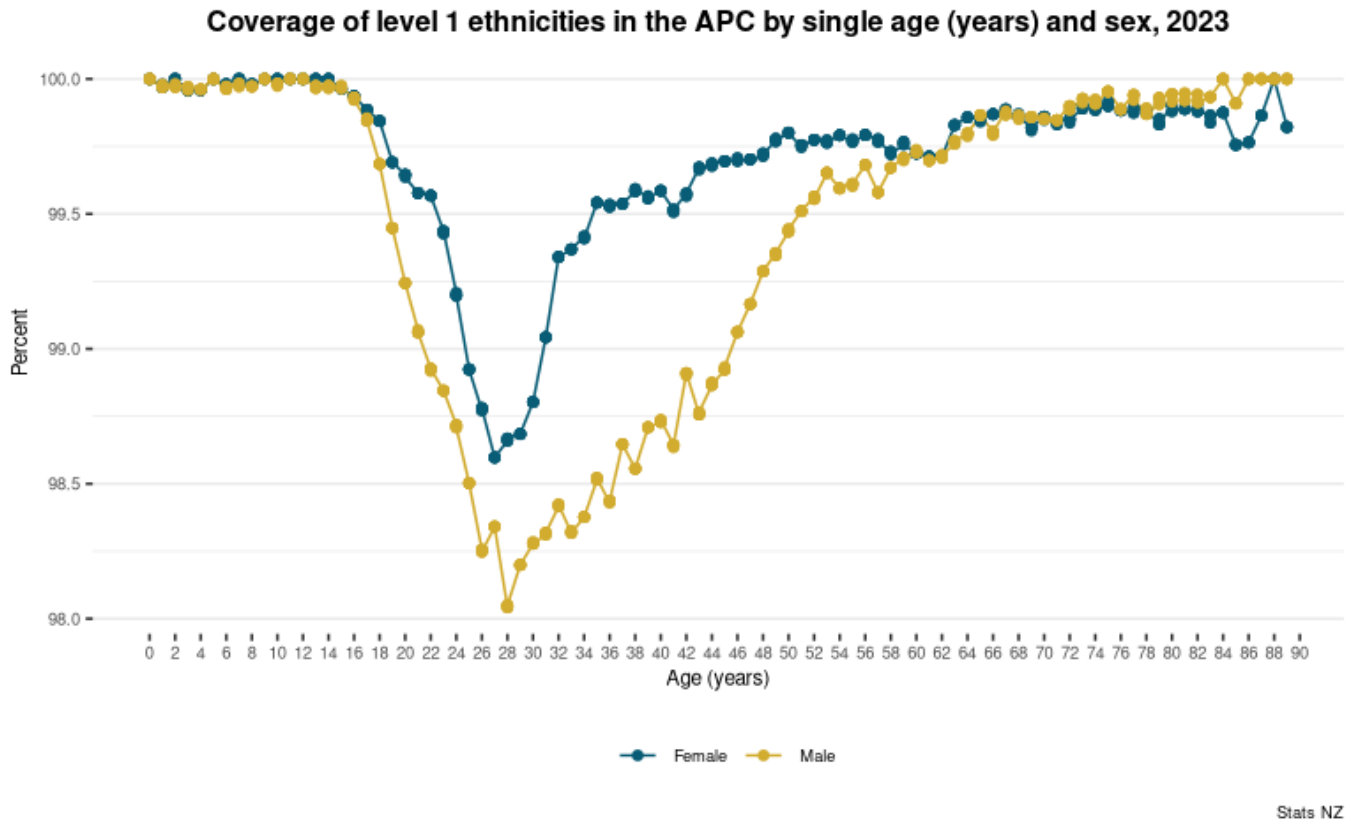
Results for quality of ethnicity in the APC

The APC mainly uses coverage and consistency measures to determine data quality. Coverage is the measure of the proportion of a population that has a response. Consistency is a measure of the proportion of responses that match some reference (often the reference is the census response).

Overall coverage and level 1 ethnicity counts

Coverage for ethnicity in the APC in 2023 is greater than 99 percent for levels 1 and 2, with coverage for level 3 at 98 percent and level 4 at 95 percent. Figure 1 shows coverage of level 1 ethnicities by single age and sex in 2023. Overall, females have slightly higher coverage than males. There is a distinct drop for people aged between 16 and 45 years, but coverage is above 98 percent for all age groups.

Figure 1



The APC also outputs quality ratings. Quality measures for level 1 ethnic groups have been derived using Dempster-Shafer Theory (DS), which measures the level of agreement between sources. Details of the DS approach and the derivation of the input quality of the data sources is described in appendix 4 (Stats NZ, 2022). Average DS scores are very similar to consistency scores but can be derived without direct comparisons to census. At level 1 we consider each ethnic group as a separate indicator variable (either a person belongs to or does not belong to that group). We derive a quality rating score for those belonging to that ethnic group. Since missingness for level 1 ethnicity is very low, the output ratings are mostly driven by the non-missing quality rating.

At level 1, the output quality ratings are 0.96 for European, 0.98 for Māori, 0.98 for Pacific Peoples, 0.99 for Asian, and 0.97 for the MELAA group.

Table 2 compares the counts and percentages of the population for level 1 ethnicities between the APC in 2023, the 2023 Census, and the ERP in 2023. There are some differences between the APC and census, especially for Māori and Other ethnicity, with the APC counting fewer individuals with Māori ethnicity and more individuals with Other ethnicity than the 2023 Census. The APC records more Pacific Peoples than both the census and the ERP. In the APC, the percentage of individuals with MELAA ethnicity is more aligned to the ERP percentage than the census is.

Table 2

Level 1 ethnicity counts by ethnicity in the APC, 2023 Census, and 2023 estimated resident population (ERP)

Ethnicity	APC in 2023		2023 Census		2023 ERP	
	Count	Percent (of total)	Count	Percent (of total)	Count	Percent (of total)
European[1]	3,399,195	66.1	3,353,190	67.1	3,493,800	67.2
Other ethnicity[1]	66,612	1.3	13,497	0.3		
Māori	862,500	16.8	886,515	17.8	911,200	17.5
Pacific Peoples	467,514	9.1	442,170	8.9	464,200	8.9
Asian	912,798	17.7	860,139	17.2	973,700	18.7
MELAA[2]	102,336	2.0	92,643	1.9	103,700	2.0
Total	**5,146,125**		**4,993,923**		**5,200,000**	

1. The estimated resident population (ERP) group 'Other ethnicity' and 'New Zealander' are counted in European ethnicity.

2. MELAA = Middle Eastern/Latin American/African.

Source: Stats NZ

Quality ratings for levels 2 to 4

Quality ratings for levels 2 to 4 are derived from the consistency between data for individuals with valid values in both 2023 Census responses and the APC in 2023. Consistency is defined as exact agreement between all ethnicities at a given level. For example, a response of Māori and Samoan in census and a derived ethnicity of only Samoan in the APC would count as disagreement. The output quality ratings are obtained by multiplying the coverage with the proportion of consistent values. Table 3 shows the output quality ratings for levels 2 to 4. About 0.05 of the reduced quality from level 1 to level 2–4 is driven by a large overcount of 'European not further defined' and a corresponding undercount of 'New Zealand European'.

Table 3

Quality summary for ethnicity at levels 2 to 4				
Ethnicity level	**Coverage (%)**	**Missing (%)**	**Non-missing quality rating**	**Output quality**
2	99.5	< 1	0.78	0.77
3	98.1	1.9	0.77	0.75
4	94.7	5.3	0.75	0.71

Source: Stats NZ

Level 2 counts

The aggregate counts for level 2 ethnicities from the APC match the 2023 Census counts well, except for some exceptions such as the APC overcounting 'not further defined' categories, and the APC recording more individuals than the census does with Fijian ethnicity (48,432 in the APC compared with 25,038 in the 2023 Census). Table 4 below has more detail.

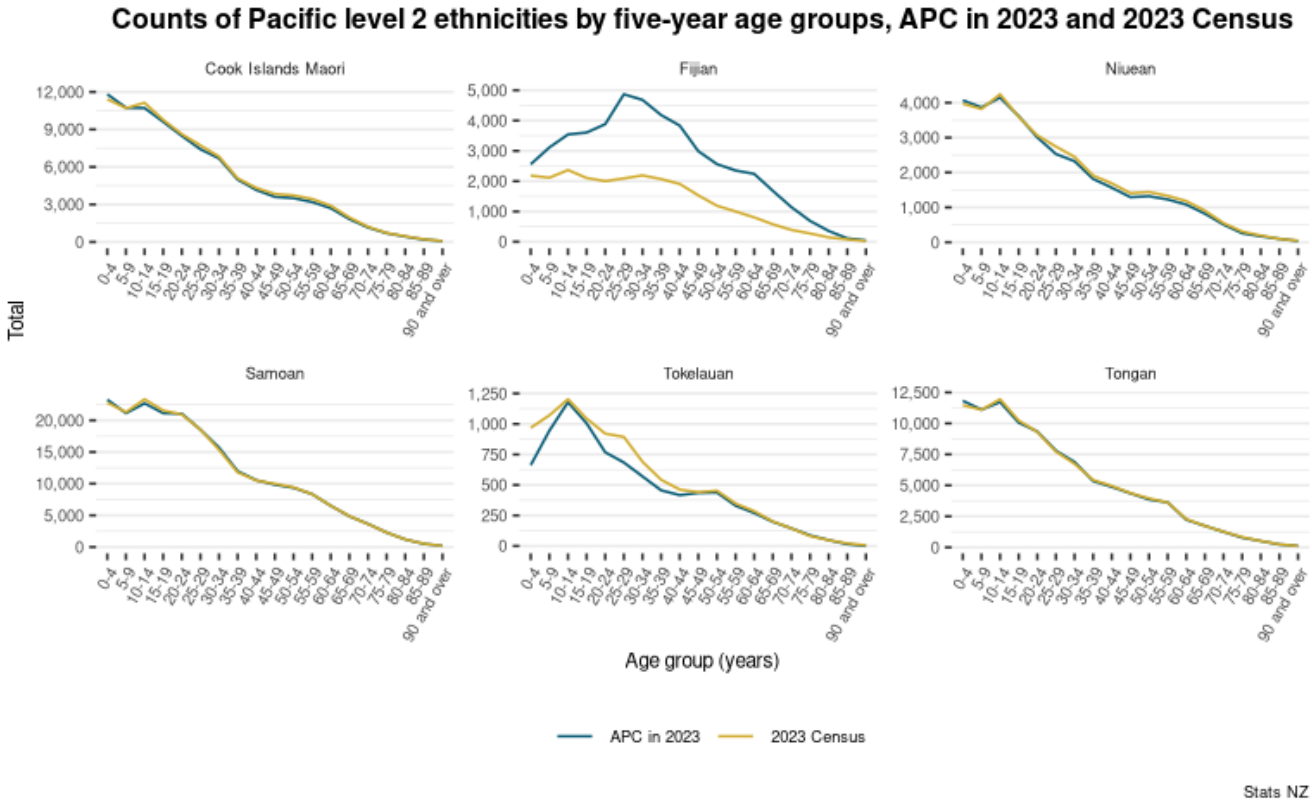
In terms of other major discrepancies, there are also more individuals recorded in the APC with 'Other European (level 2)', and Indian ethnicities. The 2023 Census records more individuals than the APC with Māori ethnicity.

Table 4

Counts and percentages of level 2 ethnicity in APC in 2023 and 2023 Census				
Level 2 ethnicity	**2023 Census**		**APC in 2023**	
	Count	**Percent (of total)**	**Count**	**Percent (of total)**
New Zealand European	3,099,858	62.1	2,810,448	54.6
Māori	887,493	17.8	862,530	16.8
Other European (level 2)	337,713	6.8	443,499	8.6
Indian	292,083	5.9	316,104	6.1
Chinese	279,045	5.6	276,171	5.4
European not further defined	21,837	0.4	265,842	5.2
Samoan	213,078	4.3	212,691	4.1
Southeast Asian	175,161	3.5	182,691	3.6
Other Asian (level 2)	112,248	2.3	135,603	2.6
Tongan	97,833	2.0	97,653	1.9
Cook Islands Maori	94,176	1.9	91,989	1.8
Other ethnicity (level 1)	56,136	1.1	66,591	1.3
Fijian	25,038	0.5	48,432	0.9
Latin American	38,160	0.8	40,098	0.8
Niuean	34,944	0.7	33,741	0.7
Middle Eastern	33,312	0.7	31,866	0.6
African	21,789	0.4	30,855	0.6
Other Pacific Peoples	20,157	0.4	23,565	0.5
Asian not further defined	12,645	0.3	15,798	0.3
Tokelauan	9,825	0.2	8,652	0.2
Pacific Peoples not further defined	2,274	0.1	1,116	< 1
Total population	**4,993,923**		**5,146,125**	
Source: Stats NZ				

Figure 2 shows a comparison of the APC in 2023 with 2023 Census for the level 2 Pacific ethnicities by five-year age groups. This clearly shows the overcounting of the Fijian ethnic group in the APC, while also showing that the other level 2 Pacific ethnicities are well counted in the APC in comparison with the 2023 Census.

Figure 2

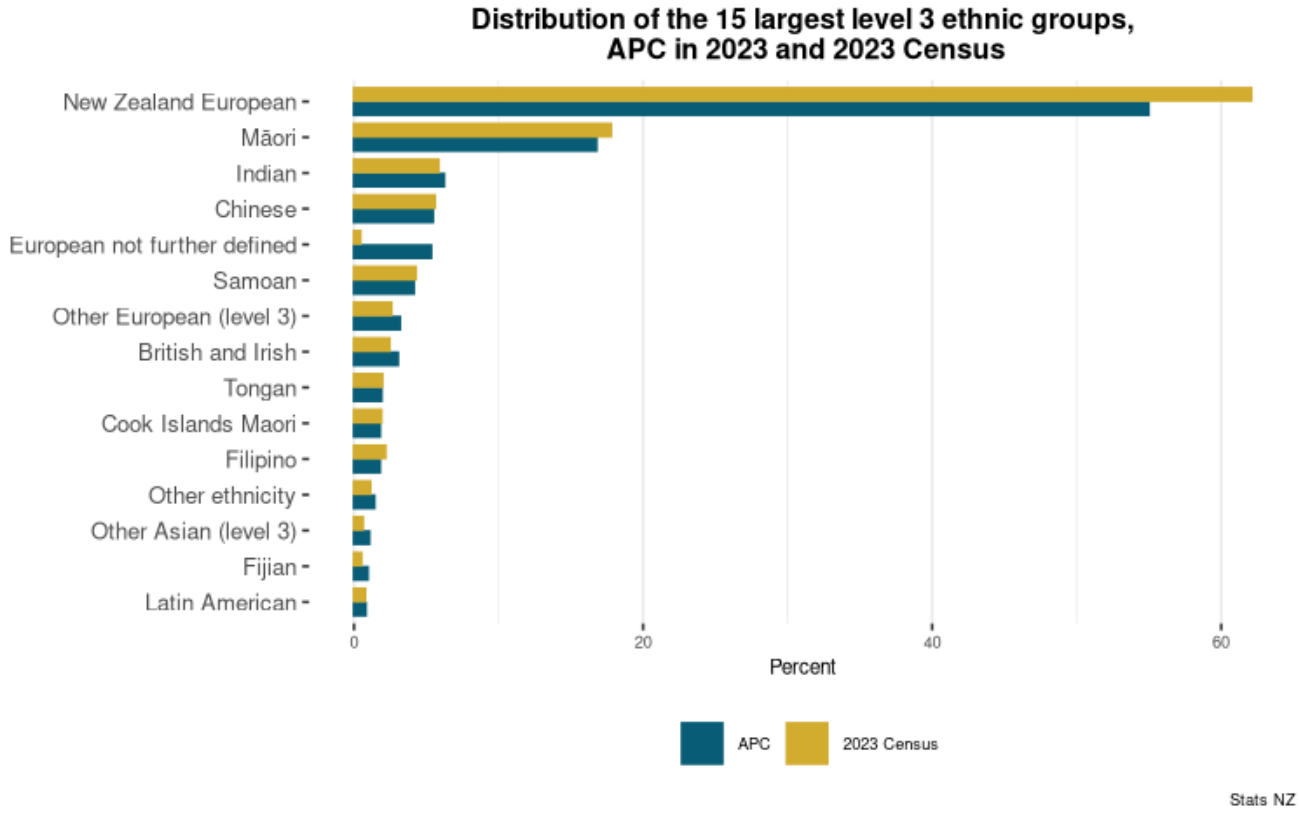


Level 3 counts

Figure 3 compares distributions between the APC in 2023 and the 2023 Census for the 15 largest level 3 ethnicities, with the corresponding numerical values given in table A1.

As the data shows, for level 3 ethnicities proportions between the APC and census are close except for the previously mentioned discrepancy for European not further defined/New Zealand European. However, the difference in the APC and census between the 'Other European' classification is not as large at level 3 as at level 2, because level 3 is separated into categories such as 'British and Irish'. See table A1.

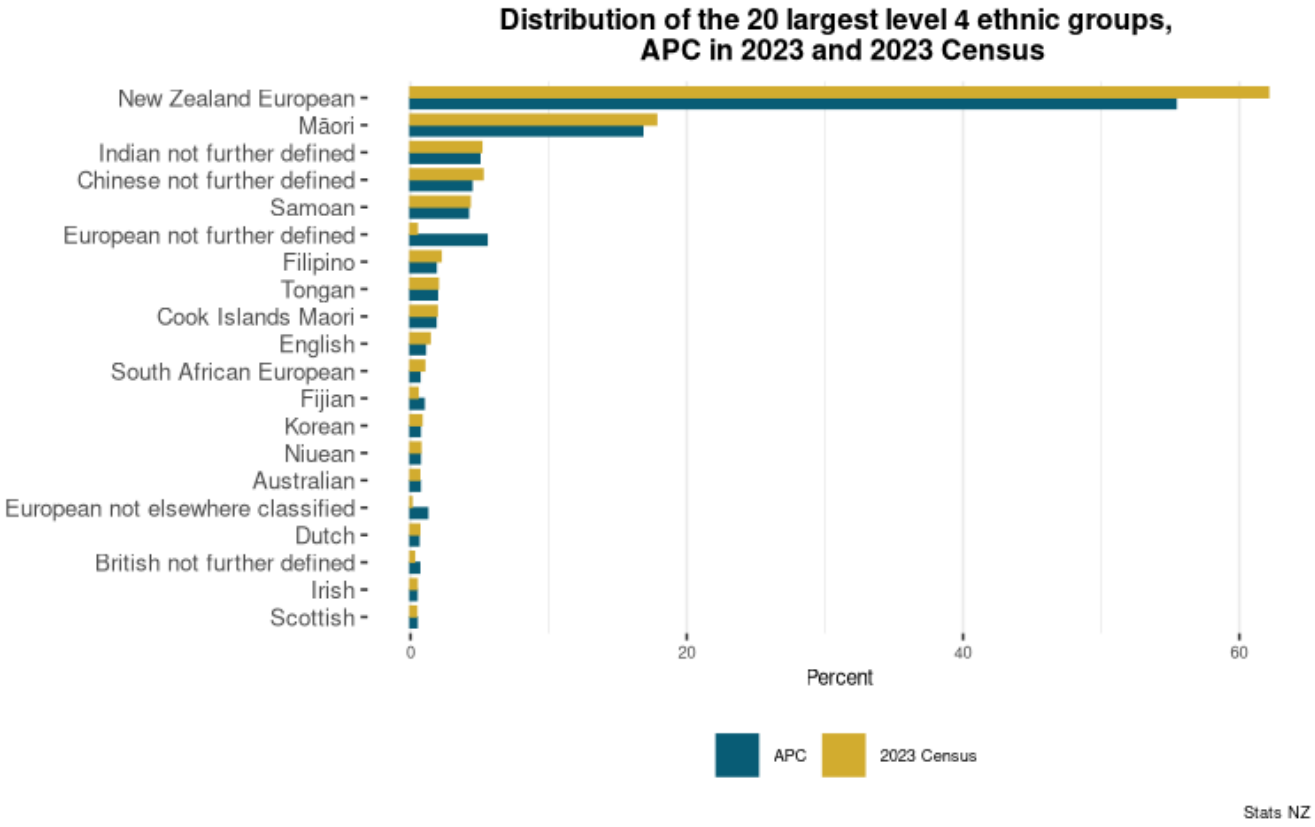
Figure 3



Level 4 counts

The undercounting of ethnic groups in the APC compared with the census is further demonstrated when looking at the 20 largest level 4 ethnic groups (see figure 4 and table A2).

Figure 4



Discussion and future work

Despite the quality of level 1 and 2 ethnicity counts in the APC being overall high, improvements in the quality of admin data will be necessary to support the production of detailed ethnicity counts in an admin-data-first census – particularly for level 3 and 4. There are three aspects to this:

- Coding and classifying to level 4.
- Updating admin collection and coding to align with changes that may come out of the Ethnicity Standard Review.
- Improvements to classification and coding practices. While there are a few areas of specific feedback Stats NZ can provide (for example, improving the coding of 'Fijian Indian' so they are not misclassified as 'Fijian'), improvements in the coding practice often align with data needs within each agency. For example, Te Aka Whai Ora developed an action plan for achieving high quality ethnicity data in the health and disability sector (McLeod 2023) with a particular focus on improving the quality of representation of Māori in key health datasets, to reduce systematic undercounting of Māori within the health datasets (Harris 2022).

The biggest challenge is that currently most agencies do not collect and code ethnicity data to level 4. This is required to provide data for many of the smaller ethnic groups. Improvements in the collection and coding of ethnicity is likely to add significant value to the ability of different agencies to monitor outcomes within their own data systems. But improving the collection and coding of ethnicity data is likely to take time and significant resource from the agencies

involved. Identifying where to prioritise quality improvements will need to be done in partnership with the agencies. Given their reach of the population, the health and education sector are likely to be key areas.

Improvements in the quality of ethnicity data collected in administrative settings is unlikely to have a large effect before 2028 because it will take some time and substantial improvements in the detail collected in admin data sources across the population. While data improvements are ongoing, historical census data can be used to alleviate quality issues, but this is not a long-term solution.

Methodological improvements can also play a crucial role to ensure different ethnicity values from different sources are used in a way that the combined APC values have better agreement with 2023 Census values. This might include refining the current ranking-by-source approach to one that allows different rankings for different ethnicities or using approaches such as latent class modelling. The latter was explored in Bycroft et al (2023). This work modelled the six level 1 ethnic groups as latent classes independently at the individual level and compared the results with the current source ranking method. The main benefit of the modelled latent class predictions over the source ranking method is attaining individual-level predicted values which are closer to those reported in the census. However, this analysis has only been run for level 1 ethnic groups, although there is potential for the model to be run at level 2 or level 3 of the ethnicity classification in future research.

References

Bycroft, C, Elleouet, J, & Tran, H (2023). Harmonising ethnicity from multiple administrative data sources using latent class modelling. Retrieved from www.stats.govt.nz.

Dixon, R, Batts, T, and Diener, H (2024). Reviewing the quality of ethnicity data from administrative sources: Comparison with the census. Retrieved from www.stats.govt.nz.

Harris, RB, Paine, SJ, Atkinson, J, Robson, B, King, PT, Randle, J, Mizdrak, A, & McLead, M (2022). 'We still don't count: The under-counting and under-representation of Māori in health and disability sector data'. *The New Zealand Medical Journal* 135 (1567), 54–78. doi: 10.26635/6965.5849. Available from journal.nzma.org.nz.

McLeod, M, Harris, R (2023) Action plan for achieving high quality ethnicity data in the H&D sector Retrieved from www.tewhatuora.govt.nz/

Reid, G, Bycroft, C, Gleisner, F (2016). Comparison of ethnicity information in administrative data and the census. Retrieved from www.stats.govt.nz.

Stats NZ (2018). Experimental ethnic population estimates from linked administrative data. Retrieved from www.stats.govt.nz.

Stats NZ, (2021). Experimental administrative population census: Data sources and methods. Retrieved from www.stats.govt.nz.

Stats NZ, (2022). Experimental administrative population census: Data sources, methods, and quality (second iteration). Retrieved from www.stats.govt.nz.

Stats NZ, (2023a). Experimental administrative population census (third iteration): Data sources, methods, and quality for household information. Retrieved from www.stats.govt.nz.

Stats NZ, (2023b). Experimental administrative population census (third iteration): Changes to data, methods, and quality. Retrieved from www.stats.govt.nz.

Appendix: Proportions of level 3 and level 4 ethnicites

Table A1

Proportions of the 15 largest level 3 ethnicity groups, APC in 2023 and 2023 Census		
Level 3 ethnicity	Percent	
	Census	APC
New Zealand European	62.1	55.0
Māori	17.8	16.8
Indian	5.8	6.2
Chinese	5.6	5.5
European not further defined	0.4	5.3
Samoan	4.3	4.1
Other European (level 3)	2.6	3.2
British and Irish	2.5	3.0
Tongan	2.0	1.9
Cook Islands Maori	1.9	1.8
Filipino	2.2	1.8
Other Ethnicity	1.1	1.4
Other Asian (level 3)	0.6	1.1
Fijian	0.5	0.9
Latin American	0.8	0.8
Source: Stats NZ		

Table A2

Proportions of the 20 largest level 4 ethnicity groups, for APC in 2023 and 2023 Census		
Level 4 ethnicity	Percent	
	Census	APC
New Zealand European	62.1	55.4
Māori	17.8	16.8
European not further defined	0.4	5.5
Indian not further defined	5.1	5.0
Chinese not further defined	5.2	4.4
Samoan	4.3	4.1
Tongan	2.0	1.9
Filipino	2.2	1.8
Cook Islands Maori	1.9	1.8
European not elsewhere classified	0.1	1.2
English	1.4	1.0
Fijian	0.5	0.9
Niuean	0.7	0.7
Korean	0.8	0.7
Australian	0.6	0.7
South African European	1.0	0.6
British not further defined	0.2	0.6
Dutch	0.6	0.6
Scottish	0.4	0.4
Irish	0.4	0.4
Source: Stats NZ		