Reporting health outcomes for distinct Pacific populations in New Zealand

Assessing the Integrated Data Infrastructure as a resource for detailed ethnic population outcomes

Nicole Satherley
Andrew Sporle

inzight analytics
inzight.co.nz

inzight.co.nz

Reporting health outcomes for distinct Pacific populations in New Zealand: Assessing the Integrated Data Infrastructure as a resource for detailed ethnic population outcomes

iNZight Analytics Limited

ISBN: 978-0-473-72711-6

Date: September 2024

How to cite this document:

Satherley, N., & Sporle, A. (2024). Reporting health outcomes for distinct Pacific populations in New Zealand: Assessing the Integrated Data Infrastructure as a resource for detailed ethnic population outcomes. Available from inzight.co.nz/projects/pacific-health-reporting

Disclaimers

Access to the data used in this report was provided by Stats NZ under conditions designed to give effect to the security and confidentiality provisions of the Data and Statistics Act 2022. The results presented in this report are the work of the authors, not StatsNZ or individual data suppliers.

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. For more information about the IDI please visit stats.govt.nz/integrated-data/

Data in this report have been reported in accordance with Stats NZ’s confidentiality rules for microdata use, and as such random rounding to the base 3 has been applied to all count data and counts of 5 or less have been suppressed (S).

Funding

Funded by Health Research Council (HRC) programme grant 17-610 led by Professor Parry Guilford, University of Otago.

Contents

List of Abbreviations

ACC	Accident Compensation Corporation
DIA	Department of Internal Affairs
DOL	Department of Labour
MBIE	Ministry of Business, Innovation, and Employment
MOE	Ministry of Education
MOH	Ministry of Health
MSD	Ministry of Social Development
NHI	National Health Index
NZCR	New Zealand Cancer Registry

Acknowledgements

The authors would like to thank Flax Analytics (Lis Ellison-Loschmann, Fiona McKenzie and Mona Jeffreys) and Lisa Underwood (COMPASS, The University of Auckland) for their comments and feedback on an earlier version of this report.

Funding was provided by Health Research Council (HRC) programme grant 17-610 led by Professor Parry Guilford, University of Otago.

Executive summary

Executive Summary

Pacific peoples are often treated as a single group for the purpose of reporting on health outcomes in New Zealand, but this ignores the diversity between specific Pacific ethnic populations.

This report summarises work conducted using Statistics New Zealand's (Stats NZ) Integrated Data Infrastructure (IDI) to better capture this diversity and enable more accurate reporting on cancer outcomes (all cancers and stomach cancer) among those who identify with "Level 2" Pacific ethnicities: Samoan, Cook Islands Māori, Tongan, Niuean, Tokelauan & Fijian.

This research linked information in the New Zealand Cancer Registry (NZCR) to administrative records within the IDI to provide enhanced coverage of Pacific ethnicities, and new information about cancer cases not recorded in the NZCR outside the IDI. This includes country of birth, years since arrival in New Zealand, and emigration from New Zealand following diagnosis. Within the IDI, relevant linked data from birth registrations, the mortality collection and border movements (arrivals and departures from New Zealand), 2013 and 2018 Census and Administrative Population Census data tables, were joined to the NZCR. We report the impact of adding this information on ethnicity data availability and counts for Pacific peoples. Additionally, different potential population denominators for reporting on population cancer rates are compared.

Supplementing NZCR ethnicity records with Census 2013, Census 2018, and Department of Internal Affairs (DIA) birth records (three high quality ethnicity data sources in the IDI) reduced missing ethnicity data in the NZCR by 58%. This increased Pacific ethnicity counts by 5% among the Pacific group in general to 13-15% for Niuean and Tokelauan ethnicity respectively. New Zealand vs. Overseas born status could be added without missing records based on the presence of a birth registration in DIA birth records. However, specific country of birth information (based on Census and border movement data) remained missing for 28% of the Pacific cohort (ranging from 24% among the Tongan cohort to 33% among the Tokelauan cohort).

For those not born in New Zealand, year of first arrival was missing for around 50% of Pacific peoples due to limited data availability for this information in the IDI. Availability of border movements data enabled us to examine rates of permanent emigration from New Zealand following a diagnosis (leaving the country within a specified period without any record of return). The data indicated 4.2% of Pacific peoples in the NZCR left the country within 5-years of their diagnosis without any record of return. This rate ranged from 4.3% (Samoan) to 2.1% (Niuean).

Finally, comparisons of the 2013 and 2018 Censuses and 2013 and 2018 Administrative Population Census (APC) as denominator sources for incidence rates indicated generally comparable rates between each source, with higher rates usually produced by the Census. However there is some variation in comparability across different subpopulations. The APC provides a higher count of the Pacific population and is available every year from 2006.

Overall, we find that the IDI is a useful tool for joining previously undocumented information to the NZCR for reporting on cancer incidence and mortality, such as country of birth and migration.

However, confidentiality requirements mean low case counts for some Level 2 Pacific cohorts must be suppressed or aggregated into larger groups (e.g. time periods of diagnoses). Random rounding of all counts to base 3 also has a larger effect on smaller counts, leading to less reliable data outputs for smaller, specific Pacific ethnicities. Data for these groups may be best documented outside the IDI, unless analysis involves data that are not available or difficult to link.

Introduction

Introduction

In New Zealand, Pacific peoples often experience worse cancer outcomes than non-Pacific/non-Māori peoples (Cleverley et al., 2023; Meredith et al., 2012). Although health outcomes among Pacific peoples in New Zealand are typically reported on for the broad Pacific ethnic category, Pacific peoples consist of several specific and demographically diverse ethnicities. This diversity is exemplified in Table 1, which documents demographic characteristics of Level 2 Pacific ethnicities (i.e., Level 2 codes for Samoan, Cook Islands Māori, Tongan, Niuean, Tokelauan, and Fijian, of the Standard Classification of Ethnicity; Statistics New Zealand [Stats NZ], 2005) at the 2018 New Zealand Census. These characteristics reveal diversity in age, migration histories, multi-ethnic identification, and population sizes among the Pacific population.

The unique circumstances of specific Pacific ethnicities may have important implications for the types of health outcomes they experience.

Table 1. Summary of demographic diversity among Level 2 Pacific ethnic groups, compared to the Level 1 Pacific group, at the 2018 Census.

Ethnicity	Population size	Median age	Born in NZ	Migrated 20+ years ago	Identified with 4 or more ethnic groups
Pacific peoples	381,642	23.4	66.4%	46.0%	3.2%
Samoan	182,721	22.8	66.7%	47.6%	4.1%
Cook Islands Māori	80,532	21.0	83.1%	67.6%	5.8%
Tongan	82,389	20.5	64.4%	44.3%	4.3%
Niuean	30,867	21.6	83.0%	76.2%	10.3%
Tokelauan	4,236	21.4	78.5%	69.0%	8.6%
Fijian	10,110	27.6	55.2%	21.9%	5.9%

Note. Each individual is counted in each Level 2 ethnic group they listed.
Census 2018 data obtained from stats.govt.nz/tools/2018-census-ethnic-group-summaries/

Although there is an expressed need for information on detailed Pacific ethnicities (Ministry of Pacific peoples, 2022), reporting of health outcomes seldom goes beyond the aggregate Level 1 Pacific grouping due in part to the analytic challenges of reporting on small samples.

In this technical report, we outline methods developed to explore the potential to supplement the New Zealand Cancer Registry (NZCR) with additional data from Stats NZ's Integrated Data Infrastructure (IDI). The kind of data available in the IDI covered in this report will not be added routinely to the NZCR, but the NZCR can be supplemented ad hoc with administrative data to help address more specific research questions. The key focus was to examine the extent to which cancer outcomes can be reported on among Level 2 Pacific ethnicities, and in relation to potentially relevant factors that are not currently measured in the NZCR. In other words, we looked to develop methods to better account for diversity in the Pacific population in cancer reporting.

We demonstrate this potential using the full NZCR (i.e., among those with any diagnosed cancer) as well as stomach cancer diagnoses among Pacific peoples in the NZCR. This work is part of a broader research programme on reducing the burden of stomach cancer in New Zealand. Stomach cancer was the 6th most prevalent cancer among Pacific peoples in New Zealand between 2017 and 2019, with 37 cases diagnosed per year (Cleverley et al., 2023). The smaller size of the stomach cancer cohort is also useful for demonstrating potential challenges in producing cancer incidence/outcome data for small populations experiencing rare events in the IDI.

We first focus on supplementing existing ethnicity records in the NZCR using additional administrative sources in the IDI. We then demonstrate the ability to add potentially relevant information to the NZCR from other sources, using country of birth and time in New Zealand as examples. We also explore available potential population denominator sources for reporting on cancer rates among Level 2 Pacific ethnicities. Although cancer is the focus of this report, the methods demonstrated here could be applied to a range of health outcomes.

Ethnicity coding in the New Zealand Cancer Registry

The NZCR contains details of people with cancer diagnosed in New Zealand, including diagnosis and cancer details and basic demographic details such as ethnicity (coded to Level 2), date of birth, and sex, but not country of birth. Records in the IDI are available from 1995, when data became more reliable with mandatory reporting introduced with the Cancer Registry Regulations Act 1994. The way ethnicity information is collected and assigned in the NZCR has changed over the years (see Shaw et al., 2009). Since 2009 however, ethnicity fields in the NZCR for new and retrospective cases (back to 1989) are periodically updated. Ethnicity information is sourced from the National Health Index (NHI), Mortality collection, and hospitalisation data (all Ministry of Health datasets) using an algorithm that selects up to three ethnicities per person if they are present on at least 20% of their records (see Health New Zealand, 2023). Prior to the implementation of this method, Shaw et al. (2009) analysed data in the NZCR data (up to 2004) linked to Censuses from 1981 – 2001 and identified an undercount of Māori, Pacific, and Asian ethnicities in the NZCR data, relative to Census data. This undercount was largest in earlier years and smallest between 2001 – 2004 (i.e., a 15% undercount of Māori and 10% of Pacific in 2001 – 2004). Missing ethnicities in the NZCR were predominantly (over 90%) non-Māori/Pacific/Asian in the Censuses.

As all data in the NZCR is now periodically updated, changes in collection and reporting standards for ethnicity are less likely to result in differences in over or under-counting of ethnicities at different points in time, provided additional sources of ethnicity are available in other health records.

However, the extent to which current methods of ethnicity reporting in the NZCR result in under- or over-reporting of ethnicities relative to official population counts (i.e., Censuses) or population estimates is yet to be assessed. It is also unknown to what extent reporting may differ for more detailed Level 2 (Pacific) ethnicities. Some ethnicity records remain entirely missing in the NZCR, suggesting no NHI, mortality, or hospitalisation records of ethnicity exist for some people – but these people may have ethnicity records present in other administrative sources in the IDI. Moreover, because the NZCR, like many administrative collections, only records up to three ethnicity records, those with more than three ethnicities only have the top three priority ranked ethnicities recorded. Thus, the NZCR is likely to exclude ethnicities for individuals who identify with more than 3 ethnicities (i.e., 3.2% of the Pacific population, but 8.6% and 10.3% of the Tokelauan and Niuean populations at the 2018 Census).

Using the IDI to supplement NZCR records

The IDI presents the opportunity to improve and expand reporting on cancer among Pacific peoples through ad hoc additions to, and updating of, NZCR records. While such updates will not be routinely conducted on the NZCR, they can be done to address specific research questions, with different data joined depending on the question within the IDI environment. The IDI is a research database that contains de-identified microdata from administrative government collections (e.g., from the Ministry of Health, Ministry of Education), Stats NZ collections (e.g., Censuses and surveys), and other data sources that have been linked across people and households (see Stats NZ, 2022a). This greatly increases the scope of information available about individuals and therefore the potential research questions that can be explored.

For the purposes of this report, the IDI enables the updating of ethnicity information (e.g., additions, and reductions of missing data) in the NZCR through linkage to ethnicity records from other collections in the IDI (such as the Census). Information that is completely absent in the NZCR, such as country of birth, migration data, and death records, can be added in by joining individual IDs ('snz uids') to ID records in collections that do contain this information. For example, Department of Labour (DOL; now the Ministry of Business Innovation, and Employment; MBIE) border movement records and Ministry of Health (MOH) death records.

A further benefit of using the IDI is that numerator and denominator data can be linked for the purpose of reporting population rates. This avoids numerator-denominator biases when using unlinked population denominators outside the IDI such as the Census (e.g., individuals diagnosed with cancer in 2013 may not necessarily have been in the 2013 Census). Data sources available outside the IDI (e.g., the Census, NZCR, and MOH mortality data) have previously been linked in the past to get around this issue (e.g., Blakely et al., 2009; Meredith et al., 2012) but this involves manual linking and there are limitations in the breadth of information that can be linked.

There are notable limitations of IDI-based data and analysis to keep in mind. Confidentiality rules around data outputting can be restrictive. All count data must be rounded (using random rounding to base 3) which has a larger impact on smaller values because the degree of potential rounding error makes up a higher percentage of the value. Counts smaller than 6 must also be entirely suppressed, meaning data often cannot be output when working with small subgroups – an issue particularly relevant for reporting on Level 2 ethnicity outcomes. Although additional data sources for specific variables are available, not everyone in the focal dataset (such as the NZCR) will have data on these variables in other data sources (e.g., country of birth can be sourced from numerous locations, but may still remain missing for some people). Timing of data availability also does not always align across data sources and may not be available for the required years of the study period in some sources.

Methods of coding ethnicity and data quality in the IDI

A key challenge to working with ethnicity data in the IDI is that there are often numerous sources of ethnicity information for the same individual. These are collected at varying times (but generally are not timestamped), at varying levels of completeness, and do not always record the same ethnicity (or ethnicities) for an individual (Reid et al., 2016). Encouragingly however, most major administrative collections, including MOH, code ethnicity to at least Level 2 detail (Reid et al., 2016).

Because of the availability of numerous sources of ethnicity, work has been conducted on how this data can be collated to provide a centralised and more complete source of ethnicity. Reid and colleagues (2016) examined Level 1 ethnicity records across core sources in the IDI (DIA, Accident Compensation Corporation [ACC], Ministry of Education [MOE] Tertiary and school, NHI, and Ministry of Social Development [MSD] Benefits) to compare an 'ever-recorded' and source-ranked approach to coding ethnicity across sources. The ever-recorded method assigns an individual an ethnicity if that ethnicity is recorded across any record in the IDI. By contrast, the source-ranked approach selects the highest rated quality source available for an individual and assigns ethnicity (usually up to three ethnicities) using only that source. They found an ever-recorded count of ethnicities in administrative sources produces an overcount of ethnicities, particularly for Pacific peoples (a ratio of 1.42) and Māori (1.20), relative to the 2013 Census count, compared to a slight undercount (0.97 – 0.99) using a source-ranked approach, and thus recommend a source-ranked over an ever-recorded approach.

Stats NZ (2018) further showed that the source-ranked approach performs comparably well (producing a slight undercount) as a method for collating Level 2 ethnicity information, relative to the 2013 Census.

A more recent data-driven approach to coding ethnicity in the IDI is Latent Class Analysis (LCA) demonstrated by Bycroft et al. (2023). Like the ever-recorded approach, LCA examines the ethnicity recorded for an individual in each administrative record. LCA can be conducted for each ethnicity (at Level 1 in their analysis), which determines a probabilistic 'latent' ethnicity score (either yes or no) for that given ethnicity based on the observed ethnicity (either yes or no) in each source included in the analysis. Rather than assigning preference for higher quality sources, LCA is more likely to identify an individual as a given ethnicity if they tend to have that ethnicity reported across multiple sources. LCA can make use of ethnicity records from all available sources to determine most likely ethnicity, in contrast to the source-ranked approach which discards all but the highest-ranked available source (regardless of whether those lower ranked sources contain accurate or useful information).

As this approach has been explored as an option for ethnicity coding more recently, it is not currently used by Stats NZ to determine ethnicity information for any data tables in the IDI. That said, the approach performs about as well as, if not better, than the source-ranked approach when comparing results to the 2018 Census.

Ethnicity coding method for the current research

Here we opted to use an ever-recorded approach but limited the number of sources used to the pre-existing NZCR ethnicity coding (based on MOH records as detailed earlier), 2013 and 2018 Census, and DIA birth records. The most recent Census ethnicity is assumed to be and treated as the highest quality source available in the IDI, hence its use as a benchmark for ethnicity comparisons (Bycroft et al., 2023; Reid et al., 2016). The DIA birth records are also considered a high-quality ethnicity source and produce a closer Level 1 ethnicity count match to the 2013 Census for Pacific peoples than any other data source in the IDI (e.g., Health, MSD; Reid et al., 2016). They also produce the closest count to the 2013 Census for Level 2 Pacific ethnicities, with a count ratio of 1.0 for Cook Islands Māori, Tongan, and Niuean ethnicities (Stats NZ, 2018). By contrast, Level 2 Pacific ethnicities are undercounted in MOH data (used in the NZCR) relative to the Census with ratios of 0.71 (Niuean) to 0.87 (Samoan), except for the Fijian category at a ratio of 2.14 (Stats NZ, 2018). This large deviation is highly likely due to coding discrepancies of Fijian Indian/ Indo-Fijian ethnicity (coded specifically only at Level 4 ethnicity coding) in health (and likely other administrative) data, which should be coded under Level 2 Indian ethnicity but has often been coded as both Fijian (Pacific) and Indian (Asian) separately (Ministry of Health, 2017; Stats NZ, 2018).).

Taking this ‘modified’ ever-recorded approach should therefore limit the extent of inflated ethnicity counts while also addressing missing and undercounted ethnicity in the current NZCR records using high quality data sources. Moreover, it is far less time consuming to construct relative to a full source-ranked approach as there is currently no pre-collated source-ranked table of Level 2 ethnicities available in the IDI. The exception is the Administrative Population Census (described in the following sections) which has up to three Level 2 ethnicities reported per individual obtained through source ranked IDI records. However, the APC data only contains individuals in the estimated administrative resident population from 2006 – 2022, and therefore likely misses a portion of the NZCR cohort, which includes diagnoses from 1995 onwards. It also does not include Census 2018 as an ethnicity source and only includes Census 2013 as a last resort, despite these being considered high quality sources. A full source-ranked approach comparable to Reid et al. (2016) or Stats NZ (2018a) would require IDI project access to all relevant data tables in the IDI which is not practical for most projects and time consuming to construct.

Finally, ethnicity coding methods have been compared at the whole population level. However, as with the data for this report, ethnicity may sometimes need to be sourced for a specific cohort. The NZCR cohort (and particularly stomach cancer cohort) are older in age and may have different ethnicity data coding quality for different IDI sources. In short, our method aimed to prioritise consistency in the coding of ethnicity for all individuals in the NZCR (any record across the given sources), high quality sources, and a relatively practical approach that does not require IDI project access to many data sources.

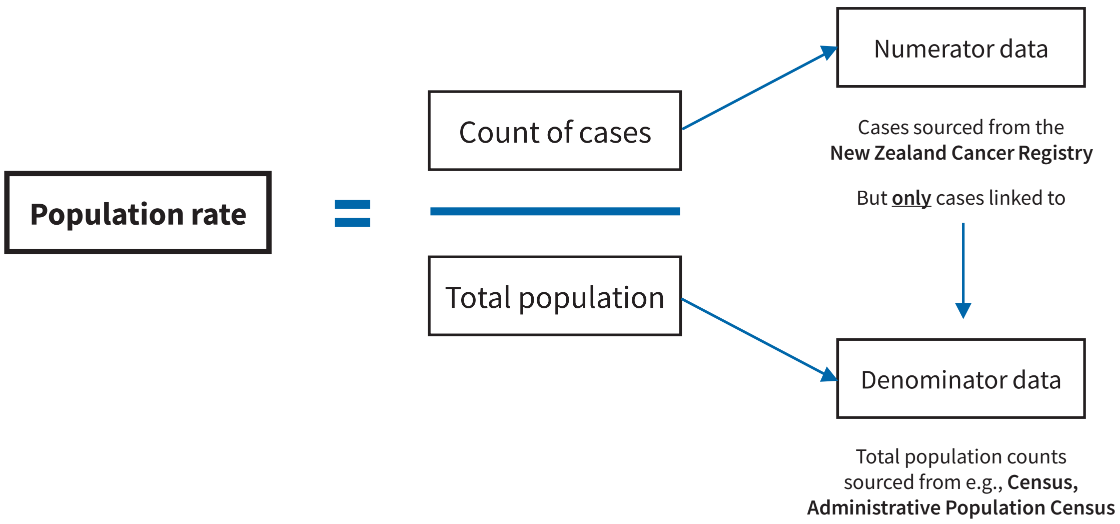
Data sources for Pacific population estimates

Denominator sources are important for reporting on cancer outcomes as they enable the calculation of population rates (such as prevalence and incidence; see Figure 1). Using data where individuals are linked between numerator and denominator data sources enables only case (numerator) data from those within the population of interest (denominator) to be analysed, which greatly reduces numerator-denominator bias. However, some bias may still occur due to data linkage error or missing records in numerator or denominator data. Issues such as these have previously been documented for existing commonly used denominators based on Census data, and Health Service User (HSU) populations (Sonder et al., 2024).

Census data sources

The Census is the key source of official population counts that can be used when reporting on cancer and other health outcomes and is available outside and within the IDI (the 2013 and 2018 Censuses specifically). Past research conducted outside the IDI has used census data linked to NZCR and

Figure 1.
Overview of how cancer rates in the population are generated from underlying data sources in this research.



mortality data, ensuring the numerator data (i.e., those with cancer or who die from cancer) are in the denominator (i.e., the usually resident population count at the given census; Blakely et al., 2009; Meredith et al., 2012). A key limitation with reliance on census data however is the 5-year intervals at which they are conducted. This means denominator information can be out of date by up to 5 years. Censuses also experience a degree of undercount, and net undercount in the 2013 and 2018 Censuses was particularly high for the Pacific ethnic group (4.8% and 4.9% respectively; Stats NZ, 2014; Stats NZ, 2020a).

An alternative potential denominator source produced from each Census is the Estimated Resident Population (ERP). The ERP includes residents temporarily overseas on Census night and adjusts for both population changes since Census night and net Census undercount (Stats NZ (2020b). It is therefore preferable to use as a denominator over the Census where possible (McLeod et al., 2023). However, ERPs are only produced for the broadest, Level 1 ethnic groups in New Zealand, and only during Census years for non-Māori ethnic groups (Stats NZ, 2020c), limiting their ability to supplement the Census as a denominator source for specific ethnic populations. The 2013-base ERP also continues to undercount Pacific peoples, even after adjusting for the estimated net undercount in the 2013 Census (Stats NZ, 2022c). Yearly population projections derived from the Census are also produced, however, projections by ethnicity are only produced in Census years using ERP base estimates, and projections are at 5-yearly intervals. Moreover, they are similarly limited to Level 1 ethnic groups, but also a few of the largest Level 2 groups (including the Samoan ethnic group; see e.g., Stats NZ, 2021). The issue of Census undercount also extends to using Best Available Populations (BAPs; population projections) as denominators (Sonder et al., 2024).

The IDI-ERP and APC data sources

The IDI-ERP (Estimated Resident Population) and Administrative Population Census (APC) are, in addition to the 2013 Census and 2018 Census, available to identify resident populations in the IDI. The IDI-ERP is based conceptually on the ERP published outside the IDI by Stats NZ. It is a yearly estimate of the resident population based on activity in administrative records, linked to specific individuals in the IDI (Stats NZ, 2022a). It includes people with administrative records in the two-years leading up to the ERP reference date (e.g., 30 June) but excludes those who died before the reference date, or who were not New Zealand residents on the reference date, determined by migration data (Stats NZ, 2022a).

A key advantage of using the IDI-ERP instead of Census data is that it can identify the resident population during intercensal years. This updated population count is available every year, avoiding reliance on out-of-date denominators. However, the IDI-ERP does not have pre-collated demographic characteristics (such as age, sex, and ethnicity) for individuals in the data table. Analysis of the IDI-ERP with regards to demographic characteristics therefore requires joining to other data sources in the IDI. For example, Brewer et al. (2020) examined cancer prevalence in the 2013 IDI-ERP, but sourced ethnicity information from the 2013 Census (however, around 25% of the 2013 IDI-ERP did not have records in the 2013 Census), and Stats NZ's core data tables (demographic information for people on the spine, or central linking table, that are pre-collated from IDI sources) for age and sex information. Stats NZ's core data tables can also be used to link ethnicity information to the IDI-ERP, but Level 2 ethnicity is not currently recorded in these tables. This means researchers would need to manually code Level 2 ethnicity from various administrative sources in the IDI to use the IDI-ERP as a denominator for Level 2 ethnicities.

The Administrative Population Census (APC) combines features of the IDI-ERP and the Census. Its goal is to identify a yearly resident population (available from 2006 onwards), and core demographic information about that population, using administrative records in the IDI (Stats NZ, 2023a). While it uses very similar methods as the IDI-ERP for identifying the resident population, it also contains pre-collated demographic information about the population from records in the IDI. An estimated resident population is easily attainable every year at June 30, however resident status is captured in a time-series which allows an estimated resident population to be attained on any desired date. We use the APC in this report, rather than the IDI-ERP, because of this additional demographic data and greater yearly coverage. Ethnicity in the APC data table is determined using the source ranked methods described by Reid et al. (2016). DIA birth records (for parents then children) are prioritised as the highest ranked and preferred source where available for an individual, followed by (in order) Ministry of Education (MOE) Tertiary, MOH, MOE schools, and Ministry of Social Development (MSD) collections, and Census 2013.

There are however limitations with the demographic data attached to the APC. Ethnicity in the APC is treated as fixed over time due largely to the lack of time stamping of ethnicity collection in the IDI (Stats NZ, 2022), with generally no indication of when the ethnicity record for a given individual was obtained. Because the APC is designed with a long-term vision of replicating Census-like information using only administrative records, 2018 Census data is not used to source demographics for the APC, and Census 2013 is only used as a last resort (Stats NZ, 2022) despite these being considered high quality data sources for ethnicity in the IDI.

Overview of methods

Overview of methods

In this report we provide an overview of the data joining process in the IDI to develop methods for, and assess the possibilities and limitations of, reporting on cancer outcomes (specifically stomach cancer) among Level 2 Pacific ethnicities (namely, Samoan, Cook Islands Māori, Tongan, Niuean, Tokelauan, and Fijian) in New Zealand.

We combined ethnicity information in the NZCR with information in the 2013 and 2018 Censuses and DIA birth records using an ever-recorded approach to allow for more complete reporting on cancer registrations by Pacific ethnicities (particularly by reducing the amount of missing data on ethnicity in the NZCR). We also document the degree of alignment between ethnicity recorded in the NZCR and ethnicity reported in the denominator source. Country of birth information was added using DIA birth records (for a New Zealand vs. overseas distinction) and then supplemented with specific country of birth using Census and border movements data. We then explore different denominator options in the IDI, namely, the 2013 and 2018 Censuses and the APC, for reporting on cancer outcomes at the population level. We document the percentage of cases in the NZCR that can be linked to each denominator source and counts of each Level 2 ethnicity by age and sex in the denominator populations, and among the cancer cohort numerators. Finally, we document differences and impacts of using different denominators on reporting cancer incidence among Level 2 Pacific ethnicities, including variability between denominators, and issues with low sample sizes (particularly in terms of ability to output data in line with confidentiality criteria for IDI data).

Overview of data joining process
Several datasets in the IDI October 2023 refresh were joined for this project as shown in Figure 2. Ethnicity information was sourced from Census 2013, Census 2018, and DIA birth records (either ethnicity recorded at birth or the ethnicity as a parent recorded on the birth certificate of their child). DIA records were also the primary source of country of birth information (supplemented with Census and border movements data for specific overseas country of birth). Border movements data was also used to code whether an individual had a record of leaving the country without return during

specified post-diagnosis follow-up periods, to investigate the extent to which some Pacific peoples may be lost to follow-up. The MOH mortality collection provides death dates for all registered deaths in New Zealand and was linked to the NZCR cohort. Joining the NZCR to the 2013 Census, 2018 Census, and APC (denominator data sources) also enabled reporting on cancer incidence among denominator populations. For the APC, yearly resident population estimates were taken at 30 June. Ethnicity information for reporting population rates within denominators was sourced solely from that denominator source to ensure numerator-denominator consistency. One issue with joining data in such a manner is that data in different collections are available for different years, as demonstrated in Table 2. The NZCR in the October 2023 refresh in the IDI includes all diagnoses from 1995 – 2022. Some collections that provide further contextual information to diagnoses have good years of coverage (e.g., DIA birth data, death dates, and hospitalisations). However, border movements data is only available from 1997 onwards, creating notable limitations for identifying how long

Figure 2. Overview of data joining process from data provision (solid lines) from different data providers to dataset joining within the IDI (dashed lines).

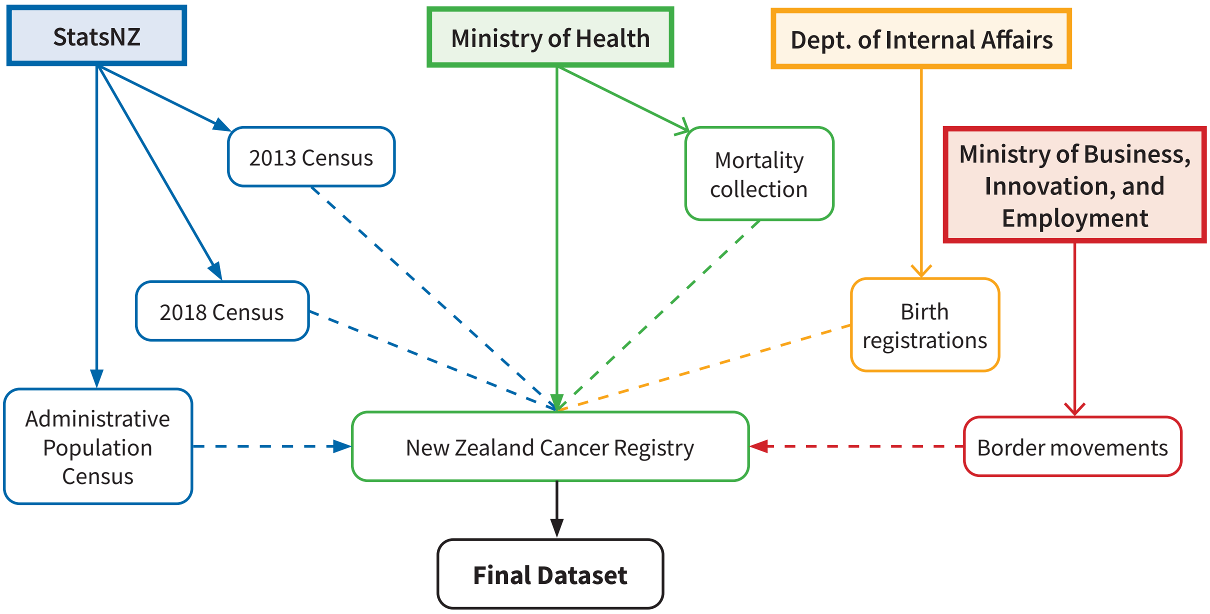


Table 2. Data availability by year for focal datasets in the October 2023 IDI refresh.

Data source	Years of coverage
New Zealand Cancer Registry	**1995 – 2022**
Mortality (Diagnoses)	Pre-1995 – 2018
Mortality (Death Dates)	Pre-1995 – 2023
DIA Births	Pre-1995 – 2023
Border Movements	1997 – 2023
Census 2013	2013
Census 2018	2018
Administrative Population Census	2006 – 2022

Note. Note. Data available in the October 2023 IDI refresh.

a person has been in the country. Diagnosis information for deaths was not available beyond 2018, meaning analysis of deaths past this time point is currently limited to all-cause mortality.

The NZCR contains entries for every diagnosis, and a given individual may have repeated diagnoses of the same or different cancers and therefore multiple entries. For this report, we include each individual once in any counts produced for the full NZCR or for those with stomach cancer. To do so, we limited all analysis to the full NZCR or stomach cancer cohort based on the earliest date of diagnosis (earliest diagnosed cancer of any type in the full cohort, or earliest stomach cancer diagnosis in the stomach cancer cohort). Ethnicity reporting uses total response reporting such that each individual is counted once for each Level 2 Pacific ethnicity reported (for example, the same individual could be included in both the Samoan and Niuean ethnicity counts). This means formal statistical comparisons between Level 2 ethnicity groups cannot be made due to non-independence.

Table 3. Counts and percentages of the NZCR (1995 – 2022) linkable to each IDI data source.

	All ethnicities				Pacific ethnicites			
Data source	**NZCR Linked**	**%**	**Stomach cancer cohort linked**	**%**	**NZCR Linked**	**%**	**Stomach cancer cohort linked**	**%**
Census 2013	320,787	60.0	4,374	39.8	10,713	50.0	336	37.0
Census 2018	280,536	52.5	3,153	29.1	10,191	47.5	255	28.1
APC (any year)	524,967	98.2	10,686	96.8	-	-	-	-
DIA births	381,582	71.4	6,906	62.6	-	-	-	-
DIA parent 1	81,825	15.3	1,020	9.2	-	-	-	-
DIA parent 2	96,696	18.1	2,070	18.8	-	-	-	-
DOL/MBIE Border movements	362,163	67.8	6,288	57	-	-	-	-
MOH Mortality	303,261	56.7	9,099	82.5	-	-	-	-

Data joining between IDI data tables

Table 3 displays the counts and percentage of the full NZCR (and stomach cancer cases specifically) that could be linked to each other source in the IDI. The overall proportion of individuals in the NZCR who are able to be linked to each data source is generally lower than 75%, reflecting limitations in attaining information for everyone from a given source. A higher proportion of the NZCR could be joined to Census 2013 compared to Census 2018, reflecting the time frame of data available in the NZCR.

That is, more people in the NZCR die, due to cancer or otherwise, as time goes on, and are less likely to be covered in data collected at later points in time. In terms of the mortality collection, 56.7% of cases in the NZCR could be linked to a death record. For stomach cancer cases specifically, 82.5% could be linked to a death record, reflecting the lower survival for stomach cancer compared to many other cancers.

Alignment of ethnicity between the NZCR, DIA birth records & population denominator data sources

Individuals in the NZCR may have different ethnicities recorded across different administrative records in the IDI. Table 4 displays a series of cross-tabulations of Level 1 and Level 2 Pacific ethnicities in the NZCR (yes/no) and each of 4 comparison sources – Census 2013, Census 2018, the APC, and DIA birth records, for those with non-missing records in each source. For example, of all the people recorded as Pacific in the NZCR and the 2013 Census, 82.4% were recorded as Pacific in each source, whereas 4.6% were only recorded as Pacific in the 2013 Census and 13.0% were only recorded as being Pacific in the NZCR. The Ratio columns provide the ratio of the count of each stated Pacific ethnicity recorded in the NZCR to the count recorded in the comparison source. The NZCR produced a slightly higher count of Pacific peoples than the 2013 and 2018 Census and APC (about 1.1 times higher than each) but a roughly similar count as DIA birth records. Looking to Level 2 ethnicities, a substantial amount of overcount may be caused by much higher counts of the Fijian and Other Pacific groups (i.e., other specific Pacific ethnicities than are not individually coded at Level 2) in the NZCR. The NZCR produced a higher count for each Level 2 ethnicity compared to the APC and 2013 Census (except for the Tokelauan group for the 2013 Census). Compared to the 2018 Census however, it produced a lower count for the Cook Islands Māori, Tongan, Niuean, and Tokelauan groups. Compared to the birth record data, it also produced a lower count of Tongans and Cook Islands Māori. In general however, there was less mismatch in ethnicity records between the NZCR and birth records.

Table 4. over page →

Table 4. Cross-tabulation of Pacific ethnicities in the NZCR, 2013 Census, 2018 Census, APC, and DIA birth records.

NZCR	Census 2013			Census 2018			APC			DIA birth records		
	No	Yes	**Ratio**	No	Yes	**Ratio**	No	Yes	**Ratio**	No	Yes	**Ratio**
Pacific												
No	305,487	513		267,579	807		494,832	636		5,664	48	
Yes	1,446	9,195	**1.10**	1,470	8,718	**1.07**	1,854	17,484	**1.07**	81	1,242	**1.03**
Samoan												
No	311,757	189		273,681	357		512,922	186		6,414	24	
Yes	309	4,386	**1.03**	366	4,170	**1.00**	669	7,917	**1.06**	24	573	**1.00**
Cook Islands Māori												
No	314,463	180		276,396	282		517,680	201		6,804	21	
Yes	207	1,797	**1.01**	231	1,665	**0.97**	516	3,297	**1.09**	15	195	**0.97**
Tongan												
No	314,682	99		276,627	186		518,145	90		6,651	12	
Yes	135	1,719	**1.02**	156	1,602	**0.98**	279	3,180	**1.06**	9	360	**0.99**
Niuean												
No	315,894	72		277,794	135		520,359	87		6,960	9	
Yes	96	582	**1.04**	87	555	**0.93**	189	1,056	**1.09**	9	57	**1.00**
Tokelauan												
No	316,407	33		278,346	48		521,265	27		7,002	6	
Yes	18	180	**0.93**	21	159	**0.97**	81	315	**1.16**	9	18	**1.13**
Fijian												
No	315,318	75		277,278	90		519,558	177		6,918	9	
Yes	828	417	**2.53**	810	393	**2.49**	729	1,227	**1.39**	57	54	**1.76**
Other Pacific												
No	315,735	87		277,677	102		519,957	225		6,960	15	
Yes	609	210	**2.76**	585	210	**2.55**	951	561	**1.92**	36	21	**1.58**

Note. Ratio is of total NZCR 'yes' records for the specified ethnicity relative to the total 'yes' records in comparison source. Red cells show undercount in NZCR relative to comparison source.

Updating ethnicity records

Table 5 shows the count of the Pacific (Level 1) and each Level 2 Pacific ethnicity in the NZCR (both the full registry and for stomach cancer cases specifically) and then with the addition of ethnicity information linked from the 2013 and 2018 Censuses and DIA records. By combining ethnicity records across all sources, the amount of missingness was reduced by around 58%, and the count of Level 2 ethnicities increased by over 100 in most cases in the full NZCR.

Notably, of the 2,187 individuals with missing ethnicity records in the NZCR that could be linked to ethnicity in the additional data sources, only 42 (1.9%) were of Pacific ethnicity in any of those sources. Of those who had ethnicity records in the NZCR but did not have a Pacific ethnic affiliation, 1,185 had a Pacific ethnicity recorded in another source. This means that the size of the Pacific cohort grew mostly due to the identification of Pacific ethnicities within administrative data sources for those not recorded as Pacific in the NZCR, rather than identifying Pacific ethnicities among those with entirely missing ethnicity records in the NZCR.

For stomach cancer cases specifically, missing ethnicity records were reduced by around 35% and counts of Level 2 Pacific ethnicities increased modestly, ranging from 2% (Samoan, Tongan) to 20% (Tokelauan; see Table 5).

Table 5. over page →

Table 5. Count of individuals of Pacific ethnicity using original NZCR records and updated, combined records.

	Full NZCR				Stomach cancer cohort		
Ethnicity	NZCR count	NZCR + CEN 2013 count	Final count (NZCR, CEN 2013, CEN 2018, DIA)	Total % increase/ decrease	NZCR count	Final count (NZCR, CEN 2013, CEN 2018, DIA)	Total % increase/ decrease
Pacific	21,444	21,987 (+543)	22,590 (+603)	5.3%	909	933 (+24)	2.6%
Samoan	9,480	9,675 (+195)	9,963 (+288)	5.1%	474	483 (+9)	1.9%
Cook Islands Māori	4,044	4,227 (+183)	4,422 (+195)	9.3%	123	132 (+9)	7.3%
Tongan	3,888	3,993 (+105)	4,143 (+150)	6.6%	168	171 (+3)	1.8%
Niuean	1,341	1,416 (+75)	1,518 (+102)	13.2%	87	90 (+3)	3.4%
Tokelauan	429	462 (+33)	495 (+33)	15.4%	15	18 (+3)	20.0%
Fijian	2,160	2,235 (+75)	2,304 (+69)	6.7%	48	54 (+6)	12.5%
Other Pacific	1,725	1,824 (+99)	1,899 (+75)	10.1%	66	72 (+6)	9.1%
Missing ethnicity	3,756	-	1,569 (-2,187)	-58.2%	69	45 (-24)	-34.8%

Adding country of birth information

Country of birth was linked to the NZCR data using DIA, Census, and border movement records. As DIA birth records are theoretically complete for all people born in New Zealand, this source was first used to code people born in New Zealand vs. overseas, a variable with no missing records. For specific birth country for those not born in New Zealand, Census 2013 then Census 2018 was used to identify specific country of birth. Consistent with Stats NZ methodology for the first and second-iteration APC (Stats NZ, 2022b), a birth country of New Zealand listed on the Census for those not in DIA birth records was accepted for those born prior to 1920, as birth records were less reliable before this date. Finally, border movement data was then used to obtain birth country (obtained either through passport information or visa applications; Gath & Das, 2019) for those whose country of birth remained missing.

Specific country of birth for Pacific peoples in the full cancer cohort is displayed in Table 6. This shows Cook Islands Māori and Niuean people in the NZCR were more likely to be born in New Zealand than the Cook Islands or Niue respectively. Tokelauans were roughly just as likely to have been born in New Zealand as Tokelau. This aligns with official Census population data on these groups showing a high proportion of New Zealand realm country ethnicities are born in New Zealand (see Table 1). These groups also had the highest proportion of missing country of birth information, which may reflect the migration histories of these groups. Specifically, those of a New Zealand realm country ethnicity are more likely than other Pacific ethnicities to have migrated to New Zealand more than 20 years prior to the 2018 Census. As border records are available from 1997 only, these individuals will not have country of birth information from this source, particularly if they both migrated prior to 1997 and have not left the country since.

Nonetheless, the data in Table 6 indicates general alignment between a person's ethnicity and their birth country – with either New Zealand or the 'ethnicity-matched' country being the most prevalent birth countries for

---

[1] Tokelau, Niue, and the Cook Islands are New Zealand realm countries. Here, we refer to 'realm ethnicities' as Level 2 Pacific ethnicities that 'match' those realm countries (i.e., Tokelauan, Niuean, Cook Islands Māori).

each Level 2 ethnicity. The presence of those of a Level 2 ethnicity born in a Pacific (or general) country not associated with their ethnicity is also to be expected as people may identify with more than one ethnicity. Overall, specific country of birth remained missing on a substantial number of records, ranging from 20.2% (among Fijian people) to 32.7% (Tokelauan).

For those who were not born in New Zealand, we linked year of first arrival sourced from Census 2013, Census 2018, and the APC (which sources arrival year from border movement data with missing records filled in by Census 2013) to the diagnosis year in the NZCR to determine time between arrival to diagnosis. Such data could be useful for identifying those arriving in New Zealand from realm countries specifically for treatment of symptoms for example. As shown in Table 7, first arrival years for some Pacific peoples were recorded as occurring after their first diagnosis in the NZCR (occurring for less than 2% of individuals with non-missing arrival data within each ethnic group). This may reflect linkage errors in the data (i.e., the wrong people have been linked across sources).

An additional significant limitation of this data is that year of first arrival in New Zealand is missing for around 50% of Pacific peoples not born in New Zealand. This is again due to the border movement records being available only from 1997. For people who first arrived before 1997, the only other sources of first date of arrival in the IDI are the Censuses. For those whose arrival date was available, diagnosis appeared to occur sooner after arrival among the Fijian cancer cohort compared to other ethnicities. Diagnosis within 0 – 4 years of first arrival appeared to be slightly more common among Cook Islands Māori and Tongans.

Tables 6. & 7. over page →

Table 6. Country of birth for each updated Level 2 ethnicity in the NZCR (any diagnosis, 1995 – 2022).

	Country of Birth										
Ethnicity	**New Zealand**	***NZ (stomach cancer)***	**Samoa**	**Cook Islands**	**Tonga**	**Niue**	**Tokelau**	**Fiji**	**Other Pacific**	**Other**	**Missing**
Pacific	**4,818** **(21.3%)**	108 (11.6%)	4,350 (19.3%)	1,293 (5.7%)	2,289 (10.1%)	384 -1.7%	123 -0.5%	1,560 (6.9%)	-	1,455 (6.4%)	6,312 (27.9%)
Samoan	2,172 (21.8%)	45 (9.3%)	**4,335** **(43.5%)**	-	-	-	-	-	156 (1.6%)	459 -4.6%	2,841 (28.5%)
Cook Islands Māori	**1,509** **(34.1%)**	33 (25.0%)	-	1,290 (29.2%)	-	-	-	-	45 (1.0%)	240 -5.4%	1,338 (30.3%)
Tongan	621 -15.0%	12 (7.0%)	-	-	**2,283** **(55.1%)**	-	-	-	111 (2.7%)	153 -3.7%	975 (23.5%)
Niuean	**489** **-32.2%**	12 (13.3%)	-	-	-	384 (25.3%)	-	-	81 (5.3%)	111 -7.3%	456 (30.0%)
Tokelauan	114 -23.0%	S	-	-	-	-	**123** **-24.8%**	-	54 (10.9%)	42 -8.4%	162 (32.7%)
Fijian	219 -9.5%	S	-	-	-	-	-	**1,503** **(65.2%)**	21 (0.9%)	96 -4.2%	465 (20.2%)
Other Pacific	423 -22.3%	6 (8.3%)	159 -8.4%	84 -4.4%	51 -2.7%	27 -1.4%	9 -0.5%	84 -4.4%	-	528 (27.8%)	531 (28.0%)

Note. Level 2 Pacific ethnicity birth countries are presented as either New Zealand, the matching Pacific country, and other Level 2 matched Pacific countries combined, to avoid data suppression issues with low counts.

Table 7. Time between first arrival in New Zealand and first cancer diagnosis (any cancer) among overseas born Pacific population.

Arrival to first diagnosis	Pacific	Samoan	Cook Islands Māori	Tongan	Niuean	Tokelauan	Fijian	Other Pacific
Less than 0 years	99 (1.1%)	39 (1.0%)	24 (1.8%)	18 (1.0%)	S (S%)	S (S%)	15 (1.0%)	6 (0.8%)
0 - 4 years	783 (8.7%)	240 (6.3%)	114 (8.4%)	159 (8.6%)	18 (4.0%)	12 (6.6%)	201 (14.0%)	72 (9.7%)
5 - 9 years	603 (6.7%)	165 (4.3%)	42 (3.1%)	153 (8.3%)	18 (4.0%)	9 (4.9%)	192 (13.4%)	63 (8.5%)
10 - 19 years	1,434 (15.9%)	441 (11.6%)	102 (7.5%)	354 (19.2%)	33 (7.3%)	27 (14.8%)	417 (29.1%)	150 (20.2%)
20+ years	6,126 (67.7%)	2,925 (76.8%)	1,071 (79.2%)	1,158 (62.9%)	384 (84.8%)	135 (73.8%)	609 (42.5%)	453 (60.9%)
Missing arrival year	8,682 (49.0%)	3,957 (50.9%)	1,548 (53.4%)	1,683 (47.7%)	567 (55.6%)	189 (50.8%)	651 (31.2%)	726 (49.4%)

Adding migration information

One difficulty in tracking health outcomes, particularly survival outcomes, is the potential migration of people out of the country during the follow-up period. Outcomes for individuals who left the country without return are unlikely to be captured in administrative data, and this may result in an over-estimation of survival over a specified period (Blakely et al., 2009). This may be particularly likely for Pacific peoples, with those from realm countries having freedom of movement between New Zealand and their birth nation, and access to publicly funded healthcare in New Zealand (Te Whatu Ora, 2024). Specifically, Pacific peoples may be more likely to return to a 'home' nation following diagnosis, particularly for cancers with poor survival such as stomach cancer. Alternatively, migration may be more likely for those from larger Pacific nations with greater resources (Blakely et al., 2009).

Here, we used border movement data in the IDI to identify those who had a record of departing the country without a record of return. Date of departure for this group was joined to the NZCR and then compared to diagnosis dates to determine the number of people in each ethnic group who left the country within 1, 2, 3 and 5-years of their first cancer diagnosis date. As shown in Table 8, there was some evidence of this occurring within the full NZCR cohort. Among all Pacific peoples diagnosed with any cancer, 4.2% of those who could be followed for 5-years from their diagnosis (that is, the diagnosis occurred at least 5 years before the end of mortality records in the October 2023 refresh) had a record of leaving the country within that period without a record of return (at any point in time). This percentage reduced to 3.7% when excluding those who left without a record of return, but still had a death record at any point in the IDI (thus verifying that they had survived the initial follow-up period).

The percentages of each Level 2 ethnic group who left without return were similar but with Niuean ethnicity the notable exception. Those of Niuean ethnicity were less likely to leave the country without return. Niueans and Cook Islands Māori are the most likely to be born in New Zealand among the Level 2 Pacific ethnicities, and overseas born Niuean people were most likely to have migrated to New Zealand more than 20 years ago (see Table 1). It is possible that Niuean people have more family in New Zealand than other Pacific ethnic groups, decreasing their likelihood of migrating back to Niue following a cancer diagnosis. Overall, the potential for return migration does not seem to be of particular relevance to those from the realm nations. Nonetheless, the data reveal differences between Level 2 ethnicities that would be masked by only analysing data among Pacific peoples in general. Joining this information to the NZCR in the IDI presents the opportunity to adjust survival analyses for loss to follow up. However, these data on permanent migration among different groups may also help inform adjustments to analyses conducted outside the IDI (e.g., Blakely et al., 2009).

Table 8. over page →

Table 8. Counts and percentages of those who left New Zealand without returning within different follow-up periods in the full NZCR.

	Pacific		Samoan		Cook Is. Māori		Tongan		Niuean		Tokelauan		Fijian		Other	
Follow-up period	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*
Total count of those in NZCR who can be followed up for the specified time for death records																
5-years	16,884	-	7,473	-	3,294	-	3,084	-	1,152	-	360	-	1,650	-	1,428	-
3-years	19,302	-	8,499	-	3,801	-	3,522	-	1,305	-	405	-	1,941	-	1,635	-
2-years	20,655	-	9,099	-	4,077	-	3,765	-	1,398	-	435	-	2,091	-	1,737	-
1-year	21,978	-	9,684	-	4,314	-	4,032	-	1,479	-	477	-	2,235	-	1,851	-
Left country within follow-up period with no record of return																
5-years	702	4.2%	324	4.3%	135	4.1%	123	4.0%	24	2.1%	15	4.2%	72	4.4%	54	3.8%
3-years	681	3.5%	300	3.5%	138	3.6%	123	3.5%	21	1.6%	15	3.7%	66	3.4%	57	3.5%
2-years	645	3.1%	282	3.1%	141	3.5%	117	3.1%	21	1.5%	12	2.8%	54	2.6%	60	3.5%
1-year	540	2.5%	243	2.5%	108	2.5%	102	2.5%	18	1.2%	12	2.5%	39	1.7%	48	2.6%
Left country within follow-up period with no record of return and no death record																
5-years	627	3.7%	288	3.9%	123	3.7%	108	3.5%	18	1.6%	12	3.3%	66	4.0%	48	3.4%
3-years	606	3.1%	261	3.1%	126	3.3%	108	3.1%	15	1.1%	12	3.0%	63	3.2%	48	2.9%
2-years	573	2.8%	249	2.7%	126	3.1%	99	2.6%	15	1.1%	9	2.1%	51	2.4%	51	2.9%
1-year	471	2.1%	207	2.1%	96	2.2%	87	2.2%	12	0.8%	9	1.9%	36	1.6%	42	2.3%

Note. Level 2 Pacific ethnicity birth countries are presented as either New Zealand, the matching Pacific country, and other Level 2 matched Pacific countries combined, to avoid data suppression issues with low counts.

Joining the updated dataset to population denominators

Linking cancer cases from the NZCR to specific population denominator sources is important for providing accurate information about cancer rates in the population (see Figure 2). The counts and proportions of people in the NZCR that can be linked to each denominator year are displayed in Table 9. The 2006 APC can be linked to the highest proportion of the full NZCR (76%) and stomach cancer cases (60.9%), with this proportion decreasing in later years. These linkage proportions include diagnoses that were made before or after the given denominator year, so the higher rates in earlier years reflect a) less people in the NZCR having died by any cause up to that year as well as b) there being more future years of potential diagnoses that are captured by the denominator in that year. Nevertheless, the percentages provide an indication of how much of the NZCR is available to work with, and what sample size, when using a given denominator.
Notably, in addition to the APC providing many years coverage of data, the 2013 APC also has a slightly higher (around 3%) coverage of the NZCR than the 2013 Census. This difference is consistent for the full registry and stomach cancer cases but is larger (roughly 6% - 9%) among Pacific peoples compared to all ethnicities. The 2018 APC and 2018 Census however had near identical coverage (although the 2018 APC had slightly higher coverage of those in the NZCR with Pacific ethnicity). This likely reflects documented net undercount in the 2013 Census which was worse for Pacific peoples (Stats NZ, 2014). Overall, linking the NZCR to the APC in the IDI provides both a greater available range of denominator years to utilise as well as slightly higher coverage compared to Census data.

Counts and proportions of denominator coverage for each Level 2 Pacific ethnicity in the full NZCR are provided in Table 10. Earlier years again have better coverage of NZCR cases, but there is some variability by ethnicity and between the APC and Census. In the 2006 APC, Niuean ethnicity has the highest coverage and Fijian and Other Pacific the lowest. The APC in 2013 tended to have higher coverage than the 2013 Census, but

---

[2] Note that percentages for Pacific peoples are based on Pacific ethnicity recorded in the Cancer Registry. The APC and Census may have additional Pacific cancer cases for people who were not recorded as Pacific in the NZCR.

the 2018 Census tended to have higher coverage than the 2018 APC. This again likely reflects the Pacific population net undercount in the 2013 Census. Census 2018 had better coverage of all Level 2 Pacific ethnicities, but clearly lower coverage of Fijian and other Pacific ethnicities than the 2018 APC. This likely reflects the inconsistent recording of Fijian ethnicities across administrative records compared to the Census, with administrative records likely to include Fijian Indian ethnicity in the Pacific Level 2 Fijian category, rather than or in addition to the Asian Level 2 Indian category (thus overcounting the ethnic group).

Coverage of stomach cancer cases across denominator years is displayed in Table 11. This Table demonstrates challenges of producing data on smaller populations within the IDI. Due to IDI confidentiality rules, much of the counts of Tokelauans in denominator years are below the threshold (6) to be outputted from the IDI. Counts for Tokelauan, Fijian, and other Pacific ethnicities are low across all denominator years, meaning further breakdowns of stomach cancer cases among these ethnicities according to factors like sex, age and country of birth, when specifically linked to a denominator, will be difficult or impossible to output from the IDI.

The 2018 Census and 2018 APC had more comparable coverage of Level 2 ethnicities as recorded in the NZCR for stomach cancer cases, with the 2018 APC having slightly higher coverage of the Samoan and Cook Islands Māori groups.

Tables 9., 10., & 11. over page →

Table 9. Counts and percentage of individuals in the NZCR (full and stomach cancer specific) that are linked to each denominator

	All ethnicities				Pacific ethnicity			
Denominator Year	Full NZCR linked	%	Stomach cancer cohort linked	%	Full NZCR linked	%	Stomach cancer cohort linked	%
APC 2006	406,461	76.0%	6,720	60.9%	13,335	62.2%	516	56.8%
APC 2007	398,088	74.5%	6,441	58.4%	13,245	61.8%	507	55.8%
APC 2008	389,205	72.8%	6,147	55.7%	13,131	61.2%	492	54.1%
APC 2009	380,304	71.1%	5,880	53.3%	12,945	60.4%	471	51.8%
APC 2010	370,950	69.4%	5,631	51.0%	12,732	59.4%	444	48.8%
APC 2011	360,651	67.5%	5,343	48.4%	12,432	58.0%	429	47.2%
APC 2012	350,031	65.5%	5,004	45.4%	12,135	56.6%	396	43.6%
CEN 2013	**320,790**	**60.0%**	**4,374**	**39.6%**	**9,735**	**45.4%**	**321**	**35.3%**
APC 2013	**339,432**	**63.5%**	**4,686**	**42.5%**	**11,763**	**54.9%**	**381**	**41.9%**
APC 2014	328,995	61.5%	4,416	40.0%	11,448	53.4%	354	38.9%
APC 2015	317,649	59.4%	4,080	37.0%	11,106	51.8%	333	36.6%
APC 2016	306,156	57.3%	3,786	34.3%	10,734	50.1%	306	33.7%
APC 2017	294,339	55.1%	3,471	31.5%	10,350	48.3%	282	31.0%
CEN 2018	**280,533**	**52.5%**	**3,153**	**28.6%**	**9,555**	**44.6%**	**234**	**25.7%**
APC 2018	**281,736**	**52.7%**	**3,156**	**28.6%**	**9,915**	**46.2%**	**258**	**28.4%**
APC 2019	268,800	50.3%	2,853	25.9%	9,477	44.2%	237	26.1%
APC 2020	256,017	47.9%	2,547	23.1%	9,057	42.2%	213	23.4%
APC 2021	242,091	45.3%	2,193	19.9%	8,538	39.8%	186	20.5%
APC 2022	226,743	42.4%	1,860	16.9%	7,947	37.1%	159	17.5%

Note. Percentages for Pacific ethnicity are the percent of Pacific ethnic group as recorded in the NZCR.

Table 10. Level 2 Pacific ethnicities in the full NZCR (any diagnosis) linked to each denominator.

Denominator Year	Samoan	%	Cook Islands Māori	%	Tongan	%	Niuean	%	Tokelauan	%	Fijian	%	Other Pacific	%
APC 2006	6,150	64.9%	2,577	63.7%	2,352	60.5%	903	67.3%	243	56.6%	930	43.1%	525	30.4%
APC 2007	6,078	64.1%	2,544	62.9%	2,352	60.5%	885	66.0%	237	55.2%	966	44.7%	525	30.4%
APC 2008	5,994	63.2%	2,511	62.1%	2,349	60.4%	876	65.3%	243	56.6%	987	45.7%	516	29.9%
APC 2009	5,865	61.9%	2,463	60.9%	2,325	59.8%	858	64.0%	237	55.2%	1,023	47.4%	516	29.9%
APC 2010	5,748	60.6%	2,403	59.4%	2,313	59.5%	840	62.6%	234	54.5%	1,026	47.5%	513	29.7%
APC 2011	5,598	59.1%	2,334	57.7%	2,283	58.7%	816	60.9%	222	51.7%	1,026	47.5%	501	29.0%
APC 2012	5,457	57.6%	2,277	56.3%	2,223	57.2%	798	59.5%	222	51.7%	1,020	47.2%	492	28.5%
CEN 2013	**4,584**	**48.4%**	**1,977**	**48.9%**	**1,827**	**47.0%**	**657**	**49.0%**	**216**	**50.3%**	**492**	**22.8%**	**306**	**17.7%**
APC 2013	**5,271**	**55.6%**	**2,217**	**54.8%**	**2,145**	**55.2%**	**762**	**56.8%**	**216**	**50.3%**	**1,026**	**47.5%**	**471**	**27.3%**
APC 2014	5,130	54.1%	2,151	53.2%	2,094	53.9%	729	54.4%	210	49.0%	1,002	46.4%	471	27.3%
APC 2015	4,977	52.5%	2,097	51.9%	2,010	51.7%	702	52.3%	195	45.5%	987	45.7%	465	27.0%
APC 2016	4,815	50.8%	2,022	50.0%	1,938	49.8%	666	49.7%	186	43.4%	981	45.4%	456	26.4%
APC 2017	4,662	49.2%	1,932	47.8%	1,854	47.7%	648	48.3%	177	41.3%	957	44.3%	447	25.9%
CEN 2018	**4,539**	**47.9%**	**1,953**	**48.3%**	**1,791**	**46.1%**	**690**	**51.5%**	**204**	**47.6%**	**486**	**22.5%**	**321**	**18.6%**
APC 2018	**4,446**	**46.9%**	**1,839**	**45.5%**	**1,782**	**45.8%**	**624**	**46.5%**	**171**	**39.9%**	**939**	**43.5%**	**441**	**25.6%**
APC 2019	4,254	44.9%	1,740	43.0%	1,707	43.9%	591	44.1%	168	39.2%	918	42.5%	414	24.0%
APC 2020	4,080	43.0%	1,644	40.7%	1,632	42.0%	570	42.5%	159	37.1%	882	40.8%	393	22.8%
APC 2021	3,840	40.5%	1,572	38.9%	1,530	39.4%	537	40.0%	147	34.3%	831	38.5%	378	21.9%
APC 2022	3,576	37.7%	1,452	35.9%	1,419	36.5%	492	36.7%	138	32.2%	786	36.4%	360	20.9%

Note. Percentages for Pacific ethnicity are the percent of Pacific ethnic group as recorded in the NZCR.

Table 11. Count of Level 2 Pacific ethnicities in each denominator linked to the NZCR (stomach cancer diagnoses).

Denominator Year	Samoan	%	Cook Islands Māori	%	Tongan	%	Niuean	%	Tokelauan	%	Fijian	%	Other Pacific	%
APC 2006	273	57.6%	72	58.5%	87	51.8%	51	58.6%	9	60.0%	18	37.5%	15	22.7%
APC 2007	270	57.0%	72	58.5%	87	51.8%	45	51.7%	6	40.0%	15	31.3%	15	22.7%
APC 2008	261	55.1%	72	58.5%	84	50.0%	45	51.7%	6	40.0%	18	37.5%	15	22.7%
APC 2009	249	52.5%	69	56.1%	78	46.4%	42	48.3%	S	S	18	37.5%	12	18.2%
APC 2010	234	49.4%	66	53.7%	78	46.4%	39	44.8%	6	40.0%	18	37.5%	15	22.7%
APC 2011	222	46.8%	63	51.2%	72	42.9%	36	41.4%	6	40.0%	18	37.5%	12	18.2%
APC 2012	210	44.3%	60	48.8%	66	39.3%	36	41.4%	6	40.0%	12	25.0%	12	18.2%
CEN 2013	**171**	**36.1%**	**54**	**43.9%**	**54**	**32.1%**	**27**	**31.0%**	**6**	**40.0%**	**6**	**12.5%**	**6**	**9.1%**
APC 2013	**198**	**41.8%**	**57**	**46.3%**	**63**	**37.5%**	**33**	**37.9%**	**6**	**40.0%**	**15**	**31.3%**	**15**	**22.7%**
APC 2014	186	39.2%	57	46.3%	57	33.9%	27	31.0%	S	S	15	31.3%	12	18.2%
APC 2015	177	37.3%	51	41.5%	54	32.1%	27	31.0%	S	S	12	25.0%	15	22.7%
APC 2016	165	34.8%	42	34.1%	51	30.4%	27	31.0%	S	S	12	25.0%	15	22.7%
APC 2017	153	32.3%	39	31.7%	42	25.0%	24	27.6%	S	S	12	25.0%	15	22.7%
CEN 2018	**126**	**26.6%**	**30**	**24.4%**	**39**	**23.2%**	**27**	**31.0%**	**S**	**S**	**6**	**12.5%**	**9**	**13.6%**
APC 2018	**141**	**29.7%**	**33**	**26.8%**	**39**	**23.2%**	**24**	**27.6%**	**S**	**S**	**9**	**18.8%**	**12**	**18.2%**
APC 2019	132	27.8%	27	22.0%	36	21.4%	24	27.6%	S	S	9	18.8%	12	18.2%
APC 2020	117	24.7%	24	19.5%	30	17.9%	21	24.1%	S	S	9	18.8%	9	13.6%
APC 2021	102	21.5%	18	14.6%	27	16.1%	21	24.1%	S	S	9	18.8%	6	9.1%
APC 2022	90	19.0%	18	14.6%	21	12.5%	15	17.2%	S	S	6	12.5%	9	13.6%

Note. Percentages for Pacific ethnicity are the percent of Pacific ethnic group as recorded in the NZCR.

Comparison of Census and APC denominator age and sex distributions and differences in Pacific population cancer rates

Full population denominator counts in each denominator year and source, and by age-group and sex, for Pacific populations (Level 1 and 2) are displayed in Tables 1 – 36 of the Appendix. Here, we focus on a comparison of the 2013 and 2018 Census with the 2013 and 2018 APC in terms of population counts produced and differences in crude cancer rates. Rates are based on diagnoses made during any year from 1995 – 2022, and are therefore based on both cases diagnosed prior to the denominator date and those diagnosed after the denominator date. Differences in rates between these denominators can be expected due to different target populations (therefore counting different people in the denominator) but also different coding of ethnicity (e.g., any given case may be recorded as Pacific ethnicity in one denominator population but not the other). The Census URP captures those resident in New Zealand on Census night in March, whereas here the APC captures all residents (whether temporarily overseas) on June 30 (but can be changed to specify any reference date).

Table 12 provides a comparison of the 2013 Census and 2013 APC for the Level 1 Pacific population. The total count of the Pacific population is considerably higher in the June 30 2013 APC compared to the Usually Resident Population (URP) 2013 Census count (i.e., around 77,000 higher). Some of this difference is explained by the different target populations of the denominators. However, as noted earlier the 2013 Census URP has a large net undercount of the target Pacific population. This difference in population counts appears to be highest for the 30 – 39 and 40 – 49 age groups, and percentage difference in within-age group rates are largest in the 30 – 39 (11.3%) age group, followed by 40 – 49 and 50 – 59 age groups. The difference in total population rate for all cancers is 120 people per 100,000 (with the 2013 Census producing the higher incidence rate). Stratified by sex, the 2013 Census produces a cancer incidence/prevalence rate of 135 per 100,000 more for men and 92 per 100,000 more for women, compared to the 2013 APC.

As shown in Table 13, rates produced by the 2018 Census and 2018 APC were much more aligned. The 2018 APC provides a higher population count (by around 30,000) than the 2018 Census but some of this difference can be explained by the exclusion of residents temporarily overseas in the 2018 Census URP, and natural population changes between the Census date in March and the APC estimate in June. For 2018, the difference in crude rates for all cancers between these denominators is 46 per 100,000 for the total Pacific population, and 66 for men and 18 for women per 100,000. The Census produced the higher rate overall (with the APC rate being approximately 1.8% lower) and a higher rate within the age groups examined (but a lower rate in the 80+ age group for Pacific women).

Tables 14 – 15 similarly show a greater difference in estimates between the denominators in 2013 than 2018 for stomach cancer rates. However, low counts mean some of the data required to produce crude rates within younger age groups cannot be outputted from the IDI/is suppressed. Looking to 2018 (Table 14), the Census provided a slightly higher rate overall which was consistent for men and women. Overall, the 2018 Census produced a lower rate in the 60-69 & 70-79 age groups.

Tables 12., 13., 14., & 15. over page ⟶

Table 12. Comparison of cancer rates (all cancers diagnosed 1995 – 2022 among Pacific peoples) using Census 2013 and APC 2013 denominators.

Age group	Census 2013 count	Linked cancer cases	Census 2013 Rate/100,000	APC 2013 count	Linked cancer cases	APC 2013 Rate/100,000	Census – APC rate difference
<30	182,085	786	432	227,919	960	421	10
30-39	35,640	840	2,357	46,656	1,041	2,231	126
40-49	33,285	1,725	5,183	42,645	2,109	4,945	237
50-59	22,974	2,361	10,277	29,127	2,862	9,826	451
60-69	13,563	2,298	16,943	16,644	2,796	16,799	144
70-79	6,276	1,332	21,224	7,497	1,593	21,248	-25
80+	2,124	348	16,384	2,493	405	16,245	139
Total	295,947	9,690	**3,274**	372,981	11,766	**3,155**	**120**
Men							
<30	91,272	330	362	115,713	423	366	-4
30-39	16,986	210	1,236	23,244	291	1,252	-16
40-49	15,846	516	3,256	21,141	645	3,051	205
50-59	11,061	981	8,869	14,310	1,191	8,323	546
60-69	6,561	1,212	18,473	8,058	1,458	18,094	379
70-79	2,769	705	25,460	3,375	855	25,333	127
80+	741	174	23,482	867	192	22,145	1,336
Total	145,236	4,128	**2,842**	186,708	5,055	**2,707**	**135**
Women							
<30	90,807	456	502	112,212	534	476	26
30-39	18,654	633	3,393	23,412	747	3,191	203
40-49	17,436	1,212	6,951	21,501	1,464	6,809	142
50-59	11,916	1,377	11,556	14,811	1,671	11,282	274
60-69	6,996	1,089	15,566	8,583	1,338	15,589	-23
70-79	3,507	627	17,879	4,122	741	17,977	-98
80+	1,383	171	12,364	1,626	213	13,100	-735
Total	150,699	5,565	**3,693**	186,267	6,708	**3,601**	**92**

Table 13. Comparison of cancer rates (all cancers diagnosed 1995 – 2022 among Pacific peoples) using Census 2018 and APC 2018 denominators

Age group	Census 2018 count	Linked cancer cases	Census 2018 Rate/100,000	APC 2018 count	Linked cancer cases	APC 2018 Rate/100,000	Census – APC rate difference
<30	231,909	738	318	246,657	705	286	32
30-39	46,320	645	1,392	50,889	681	1,338	54
40-49	40,122	1,359	3,387	43,650	1,410	3,230	157
50-59	32,298	2,157	6,678	35,778	2,310	6,456	222
60-69	18,852	2,394	12,699	21,054	2,571	12,211	487
70-79	8,892	1,545	17,375	9,789	1,692	17,285	90
80+	3,252	540	16,605	3,324	546	16,426	179
Total	381,645	9,378	**2,457**	411,141	9,915	**2,412**	**46**
Men							
<30	117,879	348	295	125,769	339	270	26
30-39	23,193	198	854	25,866	207	800	53
40-49	19,935	357	1,791	21,957	378	1,722	69
50-59	15,798	780	4,937	17,445	801	4,592	346
60-69	9,303	1,182	12,706	10,284	1,251	12,165	541
70-79	4,089	849	20,763	4,506	927	20,573	190
80+	1,191	243	20,403	1,200	240	20,000	403
Total	191,388	3,957	**2,068**	207,027	4,143	**2,001**	**66**
Women							
<30	114,027	390	342	120,888	366	303	39
30-39	23,127	447	1,933	25,020	471	1,882	50
40-49	20,187	999	4,949	21,696	1,035	4,770	178
50-59	16,497	1,380	8,365	18,333	1,512	8,247	118
60-69	9,552	1,212	12,688	10,767	1,320	12,260	429
70-79	4,803	699	14,553	5,283	765	14,480	73
80+	2,061	294	14,265	2,121	309	14,569	-304
Total	190,254	5,421	**2,849**	204,108	5,778	**2,831**	**18**

Table 14. Comparison of stomach cancer rates (diagnosed 1995 – 2022 among Pacific peoples) using Census 2013 and APC 2013 denominators.

Age group	Census 2013 count	Linked stomach cancer cases	Census 2013 Rate/100,000	APC 2013 count	Linked cancer cases	APC 2013 Rate/100,000	Census – APC rate difference
<30	182,085	9	5	227,919	12	5	0
30-39	35,640	24	67	46,656	30	64	3
40-49	33,285	51	153	42,645	54	127	27
50-59	22,974	84	366	29,127	96	330	36
60-69	13,563	87	641	16,644	102	613	29
70-79	6,276	57	908	7,497	75	1,000	-92
80+	2,124	9	424	2,493	9	361	63
Total	295,947	321	**108**	372,981	378	**101**	**7**
Men							
<30	91,272	6	7	115,713	6	5	1
30-39	16,986	12	71	23,244	21	90	-20
40-49	15,846	33	208	21,141	39	184	24
50-59	11,061	51	461	14,310	51	356	105
60-69	6,561	51	777	8,058	54	670	107
70-79	2,769	30	1083	3,375	42	1,244	-161
80+	741	S	S	867	S	S	S
Total	145,236	183	**126**	186,708	216	**116**	**10**
Women							
<30	90,807	S	S	112,212	S	S	S
30-39	18,654	12	64	23,412	9	38	26
40-49	17,436	18	103	21,501	18	84	20
50-59	11,916	33	277	14,811	42	284	-7
60-69	6,996	39	557	8,583	48	559	-2
70-79	3,507	27	770	4,122	33	801	-31
80+	1,383	6	434	1,626	S	S	S
Total	150,699	138	**92**	186,267	162	**87**	**5**

Table 15. Comparison of stomach cancer rates (diagnosed 1995 – 2022 among Pacific peoples) using Census 2018 and APC 2018 denominators.

Age group	Census 2018 count	Linked cancer cases	Census 2018 Rate/100,000	APC 2018 count	Linked cancer cases	APC 2018 Rate/100,000	Census – APC rate difference
<30	231,909	S	S	246,657	6	2	S
30-39	46,320	15	32	50,889	15	29	3
40-49	40,122	30	75	43,650	27	62	13
50-59	32,298	57	176	35,778	60	168	9
60-69	18,852	63	334	21,054	72	342	-8
70-79	8,892	51	574	9,789	60	613	-39
80+	3,252	15	461	3,324	15	451	10
Total	381,645	234	**61**	411,141	258	**63**	**-1**
Men							
<30	117,879	S	S	125,769	S	S	S
30-39	23,193	12	52	25,866	12	46	5
40-49	19,935	21	105	21,957	24	109	-4
50-59	15,798	42	266	17,445	39	224	42
60-69	9,303	39	419	10,284	42	408	11
70-79	4,089	24	587	4,506	33	732	-145
80+	1,191	6	504	1,200	9	750	-246
Total	191,388	150	**78**	207,027	159	**77**	**2**
Women							
<30	114,027	S	S	120,888	S	S	S
30-39	23,127	S	S	25,020	S	S	S
40-49	20,187	6	30	21,696	9	41	-12
50-59	16,497	18	109	18,333	21	115	-5
60-69	9,552	24	251	10,767	30	279	-27
70-79	4,803	24	500	5,283	27	511	-11
80+	2,061	9	437	2,121	6	283	154
Total	190,254	87	**46**	204,108	99	**49**	**-3**

Tables 16 – 23 present the same data for two of the larger Level 2 Pacific ethnic groups – Samoan and Cook Islands Māori. Overall similar comparisons between denominators are observed within these groups.

However, data suppression for stomach cancer rates is a much larger issue for both ethnic groups, with stratification by sex and age among Cook Islands Māori largely not possible. Aggregation of data (i.e., from narrower to broader age bands) can be used to avoid suppression issues to some extent. The 2018 Census produced a larger all-cancer rate for both groups overall than the 2018 APC, but the difference tended to be slightly larger among the Samoan and Cook Islands Māori population than the overall Pacific population. Specifically, the overall 2018 APC rate was 2.7% lower among Samoans and 5.4% lower among Cook Islands Māori.

In contrast, for stomach cancer rates specifically, the 2018 APC produced a slightly higher rate overall among Samoans and Cook Islands Māori (see Table 19 and Table 23). Despite producing a larger population count for these groups, more stomach cancer cases of each ethnicity could be linked to the 2018 APC compared to the 2018 Census.

Random rounding error

An additional issue when working with IDI data is the influence of the random rounding requirement for outputting all count data. Random rounding to base 3 means any given value outputted was generated from one of 5 possible underlying 'true' values within the IDI. The orange shaded cell of 9 in Table 18 (stomach cancer cases among Samoans aged 30 – 39) for example could have been rounded from any value from 7 – 11 (or remained unchanged if that underlying value was 9). Random rounding to base 3 is required regardless of the original count size (for counts greater than 5) which means that smaller counts are influenced to a much greater extent than larger counts.

The rate of 53 per 100,000 produced from the 2013 Census for this age group could therefore reflect a 'true' rate anywhere from 41 per 100,000 (if the true underlying count was a 7) to 65 per 100,000 (if the underlying count was an 11), ignoring the random rounding also present in the denominator value (16,890). This means rounding alone can contribute significantly to the difference in rates between the Census and APC, where the rates were based on smaller counts.

This is relevant to crude rates produced from the IDI, but not age-standardised rates, which can be calculated in the IDI direct from original values and outputted unrounded (as these values are generated from an underlying equation rather than raw counts). Thus, age-standardised rates can instead be produced where crude rates cannot. However, although age-standardised rates are helpful for comparisons between different populations (such as different Level 2 Pacific groups) or examining trends over time, they cannot speak to the actual level of prevalence and burden of cancer within populations (see Mathieu, 2023). counts mean some of the data required to produce crude rates within younger age groups cannot be outputted from the IDI/is suppressed. Looking to 2018 (Table 13), the Census provided a slightly higher rate overall which was consistent for men and women. Overall, the 2018 Census produced a lower rate in the 60-69 and 70-79 age groups.

Tables 16. to 23. over page →

Table 16. Comparison of cancer rates (all cancers diagnosed 1995 – 2022) for the Samoan ethnic group using Census 2013 and APC 2013 denominators.

Age group	Census 2013 count	Linked stomach cancer cases	Census 2013 Rate/100,000	APC 2013 count	Linked cancer cases	APC 2013 Rate/100,000	Census – APC rate difference
<30	89,784	369	411	107,508	444	413	-2
30-39	16,890	411	2,433	21,039	471	2,239	195
40-49	15,807	807	5,105	18,939	915	4,831	274
50-59	10,956	1,083	9,885	12,582	1,248	9,919	-34
60-69	6,909	1,146	16,587	7,698	1,311	17,030	-443
70-79	2,886	615	21,310	3,285	714	21,735	-425
80+	915	144	15,738	1,005	168	16,716	-979
Total	144,147	4,575	**3,174**	172,056	5,271	**3,064**	**110**
Men							
<30	44,937	156	347	54,435	192	353	-6
30-39	8,028	96	1,196	10,473	120	1,146	50
40-49	7,452	234	3,140	9,294	273	2,937	203
50-59	5,304	456	8,597	6,234	531	8,518	79
60-69	3,444	609	17,683	3,810	702	18,425	-742
70-79	1,284	336	26,168	1,506	393	26,096	73
80+	309	63	20,388	330	66	20,000	388
Total	70,758	1,947	**2,752**	86,082	2,277	**2,645**	**106**
Women							
<30	44,847	210	468	53,073	252	475	-7
30-39	8,862	318	3,588	10,563	351	3,323	265
40-49	8,352	573	6,861	9,645	645	6,687	173
50-59	5,649	627	11,099	6,351	714	11,242	-143
60-69	3,462	537	15,511	3,885	609	15,676	-164
70-79	1,602	276	17,228	1,785	321	17,983	-755
80+	609	81	13,300	675	99	14,667	-1,366
Total	73,383	2,622	**3,573**	85,977	2,991	**3,479**	**94**

Table 17. Comparison of cancer rates (all cancers diagnosed 1995 – 2022) for the Samoan ethnic group using Census 2018 and APC 2018 denominators.

Age group	Census 2018 count	Linked cancer cases	Census 2018 Rate/100,000	APC 2018 count	Linked cancer cases	APC 2018 Rate/100,000	Census – APC rate difference
<30	112,917	345	306	116,289	342	294	11
30-39	20,847	279	1,338	21,780	279	1,281	57
40-49	18,789	642	3,417	19,314	645	3,340	77
50-59	15,093	1,002	6,639	15,324	1,011	6,597	41
60-69	9,222	1,170	12,687	9,222	1,158	12,557	130
70-79	4,422	774	17,503	4,434	789	17,794	-291
80+	1,428	237	16,597	1,341	222	16,555	42
Total	182,718	4,449	**2,435**	187,704	4,446	**2,369**	**66**
Men							
<30	57,171	168	294	59,184	165	279	15
30-39	10,437	78	747	11,067	75	678	70
40-49	9,228	141	1,528	9,678	156	1,612	-84
50-59	7,386	372	5,037	7,491	360	4,806	231
60-69	4,587	585	12,753	4,554	576	12,648	105
70-79	2,112	450	21,307	2,100	444	21,143	164
80+	525	108	20,571	474	96	20,253	318
Total	91,446	1,902	**2,080**	94,548	1,872	**1,980**	**100**
Women							
<30	55,743	177	318	57,108	177	310	8
30-39	10,416	201	1,930	10,716	204	1,904	26
40-49	9,564	498	5,207	9,636	486	5,044	163
50-59	7,701	633	8,220	7,836	654	8,346	-126
60-69	4,638	585	12,613	4,671	582	12,460	153
70-79	2,313	324	14,008	2,334	348	14,910	-902
80+	903	129	14,286	864	126	14,583	-298
Total	91,278	2,547	**2,790**	93,165	2,577	**2,766**	**24**

Table 18. Comparison of stomach cancer rates (diagnosed 1995 – 2022) for the Samoan ethnic group using Census 2013 and APC 2013 denominators.

Age group	Census 2013 count	Linked stomach cancer cases	Census 2013 Rate/100,000	APC 2013 count	Linked cancer cases	APC 2013 Rate/100,000	Census – APC rate difference
<30	89,784	6	7	107,508	6	6	1
30-39	16,890	9	53	21,039	9	43	11
40-49	15,807	24	152	18,939	24	127	25
50-59	10,956	48	438	12,582	54	429	9
60-69	6,909	48	695	7,698	57	740	-46
70-79	2,886	36	1,247	3,285	45	1,370	-122
80+	915	S	S	1,005	S	S	S
30-59	43,653	78	179	52,560	87	166	13
60+	10,710	87	812	11,988	108	901	-89
Total	144,147	171	**119**	172,056	201	**117**	**2**
Men							
<30	44,937	S	S	54,435	S	S	S
30-39	8,028	S	S	10,473	6	57	S
40-49	7,452	15	201	9,294	15	161	40
50-59	5,304	24	452	6,234	27	433	19
60-69	3,444	30	871	3,810	30	787	84
70-79	1,284	15	1,168	1,506	24	1,594	-425
80+	309	S	S	330	S	S	S
30-59	20,784	45	217	26,001	48	185	32
60+	5,037	45	893	5,646	57	1,010	-116
Total	70,758	93	**131**	86,082	108	**125**	**6**
Women							
<30	44,847	S	S	53,073	S	S	S
30-39	8,862	S	S	10,563	S	S	S
40-49	8,352	6	72	9,645	6	62	10
50-59	5,649	24	425	6,351	27	425	0
60-69	3,462	21	607	3,885	27	695	-88
70-79	1,602	21	1311	1,785	24	1,345	-34
80+	609	S	S	675	S	S	S
30-59	22,863	36	157	26,559	39	147	11
60+	5,673	42	740	6,345	54	851	-111
Total	73,383	78	**106**	85,977	93	**108**	**-2**

Table 19. Comparison of stomach cancer rates (diagnosed 1995 – 2022) for the Samoan ethnic group using Census 2018 and APC 2018 denominators.

Age group	Census 2018 count	Linked cancer cases	Census 2018 Rate/100,000	APC 2018 count	Linked cancer cases	APC 2018 Rate/100,000	Census – APC rate difference
<30	112,917	S	S	116,289	S	S	S
30-39	20,847	9	43	21,780	9	41	2
40-49	18,789	15	80	19,314	12	62	18
50-59	15,093	30	199	15,324	33	215	-17
60-69	9,222	30	325	9,222	39	423	-98
70-79	4,422	33	746	4,434	39	880	-133
80+	1,428	9	630	1,341	6	447	183
Total	182,718	126	**69**	187,704	141	**75**	**-6**
Men							
<30	57,171	S	S	59,184	S	S	S
30-39	10,437	6	57	11,067	S	S	S
40-49	9,228	9	98	9,678	9	93	5
50-59	7,386	18	244	7,491	18	240	3
60-69	4,587	18	392	4,554	21	461	-69
70-79	2,112	18	852	2,100	21	1,000	-148
80+	525	S	S	474	S	S	S
30-59	27,051	30	111	28,236	30	106	5
60+	7,224	36	498	7,128	45	631	-133
Total	91,446	69	**75**	94,548	78	**82**	**-7**
Women							
<30	55,743	S	S	57,108	S	S	S
30-39	10,416	S	S	10,716	S	S	S
40-49	9,564	6	63	9,636	S	S	S
50-59	7,701	12	156	7,836	15	191	-36
60-69	4,638	15	323	4,671	18	385	-62
70-79	2,313	15	649	2,334	18	771	-123
80+	903	6	664	864	S	S	S
30-59	27,681	21	76	28,188	21	74	1
60+	7,854	36	458	7,869	42	534	-75
Total	91,278	57	**62**	93,165	63	**68**	**-5**

Table 20. Comparison of cancer rates (all cancers diagnosed 1995 – 2022) for Cook Islands Māori ethnic group using Census 2013 & APC 2013 denominators.

Age group	Census 2013 count	Linked stomach cancer cases	Census 2013 Rate/100,000	APC 2013 count	Linked NZCR	APC 2013 Rate/100,000	Census – APC rate difference
<30	40,464	186	460	48,969	219	447	12
30-39	6,525	165	2,529	8,007	195	2,435	93
40-49	6,180	363	5,874	7,221	426	5,899	-26
50-59	4,671	528	11,304	5,109	585	11,450	-147
60-69	2,352	405	17,219	2,505	444	17,725	-505
70-79	1,263	264	20,903	1,311	285	21,739	-837
80+	387	63	16,279	414	63	15,217	1,062
Total	61,842	1,974	**3,192**	73,536	2,217	**3,015**	**177**
Men							
<30	20,238	84	415	24,696	96	389	26
30-39	3,012	45	1,494	3,948	63	1,596	-102
40-49	2,814	102	3,625	3,480	132	3,793	-168
50-59	2,211	222	10,041	2,415	234	9,689	351
60-69	1,101	207	18,801	1,167	219	18,766	35
70-79	564	138	24,468	588	150	25,510	-1,042
80+	141	39	27,660	153	36	23,529	4,130
Total	30,081	837	**2,782**	36,447	930	**2,552**	**231**
Women							
<30	20,220	102	504	24,276	123	507	-2
30-39	3,510	120	3,419	4,056	135	3,328	90
40-49	3,366	261	7,754	3,738	294	7,865	-111
50-59	2,463	306	12,424	2,691	351	13,043	-620
60-69	1,251	198	15,827	1,335	225	16,854	-1,027
70-79	699	123	17,597	720	132	18,333	-737
80+	243	24	9,877	264	27	10,227	-351
Total	31,752	1,134	**3,571**	37,080	1287	**3,471**	**101**

Table 21. Comparison of cancer rates (all cancers diagnosed 1995 – 2022) for Cook Islands Māori ethnic group using Census 2018 & APC 2018 denominators.

Age group	Census 2018 count	Linked stomach cancer cases	Census 2018 Rate/100,000	APC 2018 count	Linked cancer cases	APC 2018 Rate/100,000	Census – APC rate difference
<30	52,341	183	350	53,382	159	298	52
30-39	8,805	156	1,772	9,114	153	1,679	93
40-49	7,293	282	3,867	7,215	270	3,742	125
50-59	6,450	474	7,349	6,189	456	7,368	-19
60-69	3,477	453	13,028	3,324	423	12,726	303
70-79	1,542	270	17,510	1,470	282	19,184	-1,674
80+	627	102	16,268	585	93	15,897	371
Total	80,535	1,920	**2,384**	81,279	1,836	**2,259**	**125**
Men							
<30	26,394	81	307	26,976	72	267	40
30-39	4,269	66	1,546	4,515	60	1,329	217
40-49	3,468	90	2,595	3,552	81	2,280	315
50-59	3,090	174	5,631	2,946	153	5,193	438
60-69	1,659	207	12,477	1,569	198	12,620	-142
70-79	693	144	20,779	651	153	23,502	-2,723
80+	252	48	19,048	228	45	19,737	-689
Total	39,825	810	**2,034**	40,437	762	**1,884**	**149**
Women							
<30	25,947	99	382	26,409	87	329	52
30-39	4,536	90	1,984	4,599	96	2,087	-103
40-49	3,825	195	5,098	3,666	189	5,155	-57
50-59	3,357	303	9,026	3,243	303	9,343	-317
60-69	1,815	243	13,388	1,758	222	12,628	760
70-79	855	129	15,088	819	129	15,751	-663
80+	372	54	14,516	360	48	13,333	1,183
Total	40,707	1,113	**2,734**	40,854	1,074	**2,629**	**105**

Table 22. Comparison of stomach cancer rates (diagnosed 1995 – 2022) for Cook Islands Māori ethnic group using Census 2013 and APC 2013 denominators.

Age group	Census 2013 count	Linked stomach cancer cases	Census 2013 Rate/100,000	APC 2013 count	Linked cancer cases	APC 2013 Rate/100,000	Census – APC rate difference
<30	40,464	S	S	48,969	S	S	S
30-39	6,525	6	92	8,007	S	S	S
40-49	6,180	12	194	7,221	12	166	28
50-59	4,671	15	321	5,109	15	294	28
60-69	2,352	15	638	2,505	15	599	39
70-79	1,263	6	475	1,311	6	458	17
80+	387	S	S	414	S	S	S
30-59	17,376	33	190	20,337	33	162	28
60+	4,002	21	525	4,230	24	567	-43
Total	61,842	54	**87**	73,536	57	**78**	**10**
Men							
<30	20,238	S	S	24,696	S	S	S
30-39	3,012	S	S	3,948	S	S	S
40-49	2,814	6	213	3,480	6	172	41
50-59	2,211	12	543	2,415	9	373	170
60-69	1,101	9	817	1,167	9	771	46
70-79	564	S	S	588	S	S	S
80+	141	S	S	153	S	S	S
30-59	8,037	21	261	9,843	18	183	78
60+	1,806	12	664	1,908	15	786	-122
Total	30,081	36	**120**	36,447	33	**91**	**29**
Women							
<30	20,220	S	S	24,276	S	S	S
30-39	3,510	S	S	4,056	S	S	S
40-49	3,366	S	S	3,738	9	241	S
50-59	2,463	S	S	2,691	6	223	S
60-69	1,251	6	480	1,335	6	449	30
70-79	699	S	S	720	S	S	S
80+	243	S	S	264	S	S	S
30-59	9,339	12	128	10,485	15	143	-15
60+	2,193	6	274	2,319	9	388	-115
Total	31,752	21	**66**	37,080	27	**73**	**-7**

Table 23. Comparison of crude stomach cancer rates (diagnosed 1995–2022) for Cook Islands Māori ethnic group using Census 2018 & APC 2018 denominators.

Age group	Census 2018 count	Linked cancer cases	Census 2018 Rate/100,000	APC 2018 count	Linked cancer cases	APC 2018 Rate/100,000	Census – APC rate difference
<30	52,341	S	S	53,382	S	S	S
30-39	8,805	S	S	9,114	S	S	S
40-49	7,293	S	S	7,215	S	S	S
50-59	6,450	9	140	6,189	6	97	43
60-69	3,477	9	259	3,324	12	361	-102
70-79	1,542	S	S	1,470	S	S	S
80+	627	S	S	585	S	S	S
Total	80,535	30	**37**	81,279	33	**41**	**-3**
Men							
<30	26,394	S	S	26,976	S	S	S
30-39	4,269	S	S	4,515	S	S	S
40-49	3,468	S	S	3,552	S	S	S
50-59	3,090	9	291	2,946	S	S	S
60-69	1,659	S	S	1,569	6	382	S
70-79	693	S	S	651	S	S	S
80+	252	S	S	228	S	S	S
30-59	10,827	12	111	11,013	9	82	29
60+	2,604	6	230	2,448	9	368	-137
Total	39,825	18	**45**	40,437	21	**52**	**-7**
Women							
<30	25,947	S	S	26,409	S	S	S
30-39	4,536	S	S	4,599	S	S	S
40-49	3,825	S	S	3,666	S	S	S
50-59	3,357	S	S	3,243	S	S	S
60-69	1,815	S	S	1,758	S	S	S
70-79	855	S	S	819	S	S	S
80+	372	S	S	360	S	S	S
30-59	11,718	S	S	11,508	S	S	S
60+	3,042	9	296	2,937	S	S	S
Total	40,707	12	**29**	40,854	9	**22**	**7**

Conclusion

Conclusion

This report identifies opportunities and limitations with assessing Pacific cancer outcomes among specific Pacific ethnic groups in the Stats NZ IDI. A key focus was on updating ethnicity records in the NZCR, but also comparing potential population denominators available in the IDI.

Supplementing ethnicity information in the NZCR with Census data and birth records reduced missing ethnicity in the NZCR by over half, and by about 35% among the stomach cancer cohort. Although we took an 'ever-recorded' approach across these sources, the overall Pacific cohort increased by just 5%, with Level 2 Pacific ethnicities increasing from 5% (Samoan) to 13-15% in size (Niuean and Tokelauan, respectively). Thus, ethnicity records were able to be supplemented with other high quality data sources in the IDI, minimising the chance of some ethnic affiliations being missed, but without excessively increasing the size of each ethnic cohort in the NZCR. Although the Niuean and Tokelauan groups increased in size the most, these groups were undercounted in the NZCR relative to the 2018 Census. They are also more likely than other Level 2 Pacific groups to identify with more than three ethnicities in the 2018 Census – that is, more than the maximum recorded per individual in the NZCR. Notably, most of the increase in size of the Pacific group was due to Pacific ethnic affiliations being found among those with non-missing, non-Pacific ethnicities recorded in the NZCR.

Comparisons of ethnicity coding between key sources did reveal a large over-count of the Fijian group in the NZCR relative to all other sources, likely reflecting a tendency of both Fijian and Fijian Indian ethnicities to be coded as Fijian in MoH records. Thus, using Fijian ethnicity coding from NZCR records will produce inflated counts

and likely does not reflect the intended Pacific ethnic group.
Although ethnicity from a given denominator source should be used when presenting population rates, these updates to ethnicity in the NZCR are important for general descriptive reporting of cancer cases. The data suggests IDI records can be effective for filling in missing ethnicity information and capturing affiliations potentially missed in NZCR records.

However, it is important to note that there is no agreed upon best method for combining ethnicity records for Level 2 Pacific ethnicities in the IDI. For this work we used an ever-recorded approach only using high-quality ethnicity sources however such methods may be changed or refined over time. For example, ethnicity records that remained missing may well be available in other administrative sources (e.g., education), and these sources may be more or less relevant to a particular cohort of interest. The IDI also provides the opportunity to link NZCR data to other factors not recorded in the registry, such as country of birth or time in the country prior to diagnosis. However, data availability limitations can significantly impact this. For example, around 50% of the overseas born Pacific peoples in the NZCR could not be linked to a year of arrival in New Zealand, primarily because border crossing data is only available from 1997.

For smaller ethnic groups, and rare health outcome data (e.g., stomach cancer), data suppression and error on small counts from random rounding requirements for IDI data alone create barriers to outputting data in much detail and will require large aggregations of data. For example, cancer outcomes need to be examined across broad diagnosis periods (e.g., whole period or 10-year diagnosis bands, as opposed to two-year diagnosis bands), limiting the extent to which changes in outcomes over time can be assessed. For specific Pacific ethnicities, this is problematic for smaller groups, particularly the Tokelauan ethnic group. A greater focus on outputting formal analysis rather than descriptive data can help circumnavigate this issue. For example, age-standardised rates are not subject to the same random rounding requirements as raw data because these are generated from equations rather than raw values (albeit cannot be used to replace crude rates entirely).

We focused on the stomach cancer cohort to demonstrate the possibilities and limitations of producing cancer incidence/outcome data for specific Pacific ethnicities. Stomach cancer was the 6th most prevalent cancer among Pacific peoples in New Zealand between 2017 – 2019 (Cleverley et al., 2023), but IDI analysis of more commonly diagnosed cancers, such as lung (approximately 2.8 times more prevalent than stomach) or breast cancer (approximately 4.5 times more prevalent; see Cleverley et al., 2023) may be conducted with greater ease and less impact of small and suppressed counts. Moreover, the methods of data joining outlined in this report can also be applied to various other health outcomes aside from cancer, such as general illnesses or disease.

Despite the limitations, the IDI can greatly expand the types of factors that can be examined in relation to cancer diagnoses. Notably, it also provides wider, yearly denominator coverage (in the form of the IDI-ERP and APC) than denominator sources outside the IDI (e.g., the Census).

References

Brewer, N., Atkinson, J., Guilford, P., Blakely, T., & Teng, A. (2020). An estimate of limited duration cancer prevalence in New Zealand using 'big' data. *The New Zealand Medical Journal (Online), 133*(1514), 49-6.

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

Cleverley, T., Meredith, I., Paotonu, D., & Gurney, J. (2023). Cancer incidence, mortality and survival for Pacific Peoples in Aotearoa New Zealand. *The New Zealand Medical Journal (Online), 136*(1586), 12-31.

Gath, M, Das, S (2019). *Potential for admin data to provide country of birth and years since arrival in New Zealand information.* Retrieved from www.stats.govt.nz.

Health New Zealand Te Whatu Ora. (2023). *New Zealand Cancer Registry (NZCR).* https://www.tewhatuora.govt.nz/for-the-health-sector/health-sector-guidance/diseases-and-conditions/cancer/new-zealand-cancer-registry-nzcr/

Mathieu, E. (2023). *How does age standardization make health metrics comparable?* OurWorldInData.org. Retrieved from https://ourworldindata.org/age-standardization

McLeod, M., Harris, R., Curtis, E. T., & Loring, B. (2023). *Considerations for Māori Data Analyses.* A report for Te Aka Whai Ora.

Meredith, I., Sarfati, D., Ikeda, T., & Blakely, T. (2012). Cancer in Pacific people in New Zealand. *Cancer Causes & Control, 23,* 1173-1184.

Ministry of Health (2017). *HISO 10001:2017 Ethnicity Data Protocols.* Wellington: Ministry of Health.

Ministry of Pacific peoples (2022). Improving Pacific data equity: Opportunities to enhance the future of Pacific wellbeing. *Long-term Insights Briefing 2022.* https://www.mpp.govt.nz/programmes/long-term-insights-briefing/

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

Shaw, C., Atkinson, J., & Blakely, T. (2009). (Mis) classification of ethnicity on the New Zealand Cancer Registry: 1981-2004. *The New Zealand Medical Journal (Online), 122*(1294), 10-22.

Sonder, G.J., Grey, C., Ryan, D., Cumming, J., Sporle, A. & Hill, P., (2024). Selective under-representation of Pacific peoples in population estimates for health indicator measurements in Aotearoa New Zealand misinforms policy making. *BMC Public Health. 24*(1), 564.

Stats NZ (2023a). *Experimental administrative population census (third iteration): Information by variable.* https://www.stats.govt.nz/research/experimental-administrative-population-census-third-iteration-information-by-variable/#ethnicity

Stats NZ (2023b). *Experimental administrative population census (third iteration): Changes to data, methods, and quality.* https://www.stats.govt.nz/research/experimental-administrative-population-census-third-iteration-changes-to-data-methods-and-quality/

Stats NZ (2022a). *Integrated Data Infrastructure.* https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/

Stats NZ (2022b). *Experimental administrative population census: Data sources, methods, and quality (second iteration).* https://www.stats.govt.nz/research/experimental-administrative-population-census-data-sources-methods-and-quality-second-iteration/#birthplace

Stats NZ (2022c). *Māori population under-estimation in 2013: Analysis and findings.* https://www.stats.govt.nz/reports/maori-population-under-estimation-in-2013-analysis-and-findings

Stats NZ (2021). *National ethnic population projections: 2018(base)–2043.* Retrieved from https://www.stats.govt.nz

Stats NZ (2020a). *Post-enumeration survey 2018: Methods and results.* Retrieved from https://www.stats.govt.nz

Stats NZ (2020b). *Estimated resident population 2018: Data sources and methods.* Retrieved from https://www.stats.govt.nz

Stats NZ (2020c). *Estimated resident population (2018-base): At 30 June 2018.* Retrieved from https://www.stats.govt.nz

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

Stats NZ (2014). *Post-enumeration survey: 2013.* Retrieved from https://www.stats.govt.nz

Te Whatu Ora. (n.d.). *Services for people from Pacific Island countries [Internet].* Health New Zealand Te Whatu Ora; Publication date unknown [updated 2024 Jan 8; cited 2024 Apr 23]. https://www.tewhatuora.govt.nz/for-health-providers/publicly-funded-health-and-disability-services/services-for-people-from-pacific-island-countries/