

Technical feasibility: Measuring population and dwellings using administrative data

2025 update





Crown copyright ©

[See Copyright and terms of use](#) for our copyright, attribution, and liability statements.

Citation

Stats NZ (2024). *Technical feasibility: Measuring population and dwellings using administrative data*. Retrieved from www.stats.govt.nz.

ISBN 978-1-991307-11-8 (online)

Disclaimer:

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

[Integrated data](#) has more information about the IDI.

Published in September 2024 by

Stats NZ Tatauranga Aotearoa
Wellington, New Zealand

Updated December 2025.

Contact

Stats NZ Information Centre: info@stats.govt.nz

Phone toll-free 0508 525 525

Phone international +64 4 931 4600

www.stats.govt.nz

Contents

Purpose and summary	5
Purpose	5
Scope.....	5
Summary	5
Introduction	11
Assessing the technical feasibility of an admin-first census model	17
How well can we count our core statistical units using admin data?	18
Identifying the New Zealand resident population in admin data.....	18
Assessing the technical feasibility of admin-based resident population counts.....	19
Identifying dwellings in administrative data.....	26
Assessing the technical feasibility of admin-based dwelling estimates	28
Creating household and family information from admin data.....	28
Assessing the feasibility of admin-based household and family units	30
Characteristics of core statistical units provided by admin data	31
How good are the characteristics that we can get from admin data?	31
Producing the official estimated resident population from admin data.....	34
How do we account for the change in census model in population estimation?	34
The next steps for testing the technical feasibility of admin-based population estimation.....	36
The role of surveys in an admin-first census model	36
Delivering for and with Māori, iwi, and hapū	39
Measuring key communities.....	41
Building trust in the new census model	44
Engaging with partners, customers, and communities	45
Public trust and confidence	45
Looking ahead	46
Opportunities.....	46
Sustaining admin-data quality	46
Progressing our methodology.....	47
Progress since initial publication	48
Next steps	49

References.....	50
Appendix.....	53

List of tables and figures

List of tables

1 Percentage of 2023 population with electoral Māori descent, with and without imputation	26
2 Quality targets for population estimation	53
3 Likelihood of a census topic being derived from administrative data sources, updated for 2024 ...	54

List of figures

1 Census counts by enumeration type and undercoverage as a percentage of the estimated resident population.....	14
2 Comparison between the resident population for APC and ERP, at 30 June 2023 and 2023 Census, by age	20
3 Consistency between predicted admin address and 2023 Census by single year of age, at different geographies.....	21
4 Relative difference of Census and the APC to ERP in 2023, by TALB and five-year age groups (excluding the Chatham Islands)	22
5 Distribution of 15 largest ethnicity level 3 groups, for the APC in 2023 compared with the 2023 Census	24
6 Quality of and potential to use admin data for deriving census attributes.....	33

Purpose and summary

Purpose

This report provides an update on the technical feasibility of using administrative (admin) data, supported by surveys, to measure New Zealand's population and dwellings. It outlines Stats NZ's readiness to shift towards an admin data-first approach for the census, and identifies key challenges and development areas as we progress through design and implementation.

In June 2025, following Cabinet endorsement, details were announced about upcoming changes to Aotearoa New Zealand's census. In future, census data will rely more on information already collected by government and other organisations (known as administrative or admin data).

Surveys will remain important for validating admin data and collecting additional information. Stats NZ will also work closely with iwi Māori and other priority communities to develop tailored approaches that help meet their data needs.

Scope

This paper focuses on how well core statistical units – people, dwellings, households and families – can be measured using admin data, supported by surveys, to provide census-type information.

Aotearoa New Zealand's next census is scheduled for 2030 (subject to legislative amendments). [Modernising the census](#) outlines the reasons for the change, the decision-making process, and planned next steps. [Proposed data collection approach and content for the census](#) provides further detail on Stats NZ's current thinking as at November 2025.

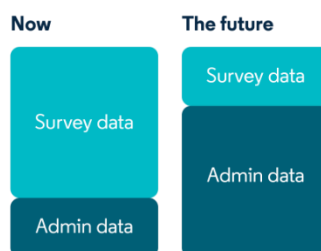
This paper updates the version in September 2024. Revisions reflect the 2025 decision on the next census model, summarised in [Modernising New Zealand's data system](#); methodological development since then, and updated comparisons with the 2023 Census and the 2023 estimated resident population (ERP). The methodological progress made since 2024 is summarised in the section [Progress since initial publication](#).

Summary

[Transforming the New Zealand Census of Population and Dwellings: Issues, options, and strategy](#) (2012) addressed concerns about the sustainability of the traditional census and the opportunities offered by the admin data and technology. At the time, an admin data-first census supported by surveys was aspirational.

Since then, sustainability risks have become evident. The 2018 Census fell short of historic response rates, and while the 2023 Census improved, it did not return to pre-2018 levels despite greater effort and investment. Producing reliable data in both censuses relied on using admin data to fill the gap, made possible by methodological research through the [Census transformation research programme](#).

The decision to change the model for the next census shifts from a survey-first to an admin data-first approach. Rather than administrative data supporting survey responses, surveys will be targeted to fill gaps in administrative data.



The transition will be gradual, with transitional outputs planned from 2026 and 2027; official key population measures expected in 2028; and a full set of annual census data from 2030. The shift is underpinned by extensive methodological research undertaken since 2012, experience from the 2018 and 2023 Censuses, and three releases of the experimental Administrative Population Census (APC).

New Zealand does not have a national identity number that links person-centred admin data to a central population file, unlike many countries with register-based censuses. Over the last decades, robust probabilistic linking methods have been developed to integrate multiple datasets. These methods, used to create the Integrated Data Infrastructure (IDI) and an admin-derived population file, have been critical to enabling an admin-first census, placing New Zealand ahead of other countries without a population register or national population identifiers that are seeking to use admin data in their census.

High coverage of the admin-derived population relies on combining data sources that provide comprehensive coverage and largely unbiased information – such as birth registrations, visa data, and tax records, which are less affected by the non-response biases typical of field-based censuses and surveys. Population statistics derived from linked admin data are more up to date, timelier, and better able to capture demographic change than those produced through a traditional survey-led census. This enables more responsive population information in a context of natural disasters, high migration flows, and an ageing population.

These changes increase demand for more frequent population statistics and for greater ability to integrate population data with other datasets that are only possible through an admin data driven approach. The IDI, and in the future, the Integrated Statistical Data System (ISDS), links admin datasets, Stats NZ surveys, and the historic Censuses at the individual level. Using multiple sources improves coverage and reduces reliance on the quality of any single dataset for key demographic variables.

Admin data allows for annual – or more frequent – production of information at more granular detail, and coverage down to small geographies. However, gaps remain for variables not captured in admin data, which require sample surveys. Collecting high-quality survey data is increasingly challenging and costly, so innovative survey approaches will be needed to sustainably produce detailed small-area data.

The success of an admin-first census depends on the trust of our Te Tiriti partners, customers, and the public. The combined census model and the release of the experimental APC products have increased awareness of how admin data supports the statistical system, but engagement insights consistently show that partners, customers, and communities are seeking more information about the quality of admin data. [Experimental administrative population census](#) has more information about the APC.

Several areas require further development, outlined under [Critical areas to progress](#):

- Limited coverage of iwi affiliation, and to a lesser extent, Māori descent.
- Lack of admin data identifying people in rainbow communities and providing information on disability.
- improvements required in the ethnicity data collection to support detailed population statistics for small ethnic communities.
- Development of the infrastructure to support a high-quality listing of dwellings.

Some of these gaps can be addressed through improvements to admin data and methodology, while others require a targeted survey programme. Delivering an admin data-first census supported by surveys is complex and involves managing change and risk. We are on track with key improvements, building on the existing foundations. Admin data continues to support timely population measures, while statistical methodologies and survey supplements are being developed to close remaining gaps and ensure robust, high-quality population statistics.

Methodological progress

We have demonstrated we can derive a high-quality list of people in the resident population using linked admin data. The admin-derived resident population closely tracks the age distribution of the official estimated resident population (ERP), with a relatively evenly distributed undercount of around 2 percent. This population provides the foundation for other census information: characteristics for each person can be derived from linked admin sources or surveys and individuals are grouped within dwellings to form households. The admin-derived resident population was used in both the 2018 and 2023 Censuses to include admin enumerations in the census file.

Although the admin-derived resident population is very close to the actual population, as measured by official population estimates, it is not accurate enough to meet user requirements for high-value uses, such as in health funding models and for local government electoral needs. As with the current census approach, methods to adjust for coverage error in the population listing will still be implemented to provide the accuracy required for official population statistics.

[The administrative population census \(APC\)](#) is an experimental product released by Stats NZ that demonstrates to customers our current ability to derive census-type information from linked admin data in the Integrated Data Infrastructure (IDI). As well as core demographic variables (age, sex/gender, ethnicity, location, and Māori descent), the APC includes variables on the topics of birthplace, income, work, and education, and in 2023 some household variables were included.

Our population estimation methods have become much more sophisticated in the last two decades (moving from a direct-weighting estimation approach to a Bayesian dual-system estimation model to adjust for census coverage errors). We have identified the key errors that our population estimation system needs to correct for under an admin-first census model supported by surveys, and are progressively developing a population estimation system that can correct for these errors. Core methodology has been internationally peer-reviewed and published (Graham et al, 2024), but more investigations are required to account for the error structures inherent when working with multiple admin data sources that differ from traditional census errors.

Population estimation models that account for subsets of the errors seen are currently in development, and these will eventually be merged and extended to correct for all known errors. Our current priority is to ensure that we have all the necessary components of the estimation system for a scenario where two admin lists of the target population are available.

Constructing admin-based households and families has progressed. Placing everybody in admin-derived households is methodologically challenging as errors compound. As addresses are used as a proxy for households, selecting the best address for a person for a given reference date is crucial. Developing a machine-learning model for this task has significantly improved accuracy compared to census data, outperforming the rules-based approach. It was used in the 2023 Census and to form APC households. There are still challenges, particularly with placing more mobile groups (for example, young people and recent migrants) in the correct households and identifying informal relationships and those formed overseas. Future work will include providing family information as well as determining the role that surveys will play in delivering household information.

Approximately half the census attribute variables for individuals can be derived using admin data. We have built up a comprehensive and detailed understanding of the quality of census characteristics that can be derived from admin data and identified the types of information collected by census that are unlikely to be obtained (Bycroft et al, 2021). High-quality variables have been released in the third iteration of the APC (Stats NZ, 2023b). For certain variables, information is of comparatively lower quality for some groups, such as recent migrants to New Zealand. Future work will focus on exploring additional data, including the sustained use of the electoral roll for Māori descent. Statistical imputation methods for any remaining missing values are being developed to improve the representativeness of the data.

A survey programme is required to deliver the full range of census information needs, and the initial design work for this is complete. Examples of information needs not currently well supported by admin data include gender, sexual identity, languages spoken, religious affiliation, activity limitations (disability), household tenure, and housing quality (such as dampness and mould). Surveys will be designed to provide information for small area geographies. While there will be loss of detail for those variables that rely on the new annual survey (Census Attribute Survey), compared with a high-coverage full-field enumeration census, the new survey will offer opportunities to improve timeliness, and to evolve content flexibly.

Critical areas to progress

Substantial admin data sources exist on properties, buildings, and addresses, and the development of the Places Index is underway. While we have not yet demonstrated the ability to provide a full census of dwellings and their attributes without field enumeration, the proof-of-concept phase is nearing completion, and the programme is now moving towards linking in historic census data for production testing. Significant work is required to complete sourcing the data, infrastructure, and methods required to produce national and subnational dwelling counts from admin data, and this will likely not include temporary dwellings. A shift away from an address-centric approach to a location-centric one allows a more detailed view, and the use of property, ratings, and buildings data. A coverage survey is planned to assess the quality of the derived dwelling list, and additional, complementing strategies will be needed to measure complex dwelling situations.

To maximise the value we get from our admin data-first census is to further improve how we form admin-based households. There is a vast array of regular address notification data in the admin data, and our methods for forming households from this data have been improving iteratively. The next step will be to expand the approach to exploit patterns of address changes that help us link individuals together to improve the chances of identifying those living together at a particular reference date. This, and the more extensive use of family data that can be found in admin datasets, will contribute to the next iteration of household and family data available in the admin population census.

Iwi affiliation is a core population identifier for Māori and an essential requirement for census that cannot yet be met through available admin data sources. The data collected in the census provides a rich resource for producing a wide range of population measures for iwi. These include core topics collected in the census, such as population size and location, language, and housing, as well as the ability to link population summaries to admin data. However, we also know that the traditional census model was not meeting all the data needs of iwi and Māori. We are actively working with our iwi partners on solutions to ensure iwi affiliation data meets their needs.

Further data sources will be required to improve the coverage of Māori descent. Without a full enumeration census, the combination of Māori descent information collected in electoral roll data and DIA birth registrations would provide the Māori descent indicator for around 90 percent of the population. The missingness mostly affects the overseas-born population. In the short term, historical census data will further reduce the amount of missing data. Statistical methods such as imputation and the population estimation process can account for missing data, although the level of adjustment would possibly be higher than some users would find acceptable. Wider collection of Māori descent in admin data would reduce the reliance on existing data sources and improve the coverage. Ongoing collection of Māori descent in survey collections will still be needed to validate data collected in admin sources, and to inform statistical imputation models.

The quality of ethnicity data (down to the most detailed level of the ethnicity classification, level 4) will need to be improved to support the sustainable production of population measures for ethnic communities. Improvements in the collection and coding of ethnicity are likely to add value to different agencies' abilities to monitor outcomes within their own data systems. These improvements will take time and resources from the agencies involved. Where to prioritise quality improvement will be identified in partnership with the agencies involved, to ensure the work aligns with their data strategy. Work with other agencies is already underway to improve data quality and how frequently it is collected. Ethnicity is a priority in these efforts.

The collection of representative and complete information for rainbow communities is limited in admin data settings, and may continue to be, because in many contexts it will not be necessary or appropriate to collect. With the census model moving away from a full enumeration approach, the loss of population-level identifiers for rainbow communities will reduce the quality of detailed population counts, the ability to report on attributes collected in the census, and the ability to link to admin data to report on measures such as health outcomes, income, and education. The new annual census survey (Census Attribute Survey or CAS) will provide support for measuring these outcomes, but will not provide the same level of detail as a full enumeration census, particularly for small populations such as trans and non-binary populations. The development of a census survey programme offers potential to deliver a wider set of social measures, such as wellbeing, that are not able to be provided by existing surveys due to sample size. It will be important to partner with rainbow communities to understand the impact and opportunities of a change in census model for supporting data needs for rainbow communities.

Data supplied by agencies is currently a mix of sex and gender. The mandated data standard for gender, sex, and variations of sex characteristics specifies the collection and output of gender data by default, as opposed to sex. Increasingly, agencies are collecting gender data, and Stats NZ is working with agencies to ensure that collection aligns with the data standard, and metadata is available describing the variable that is being collected.

[Data standard for gender, sex, and variations of sex characteristics](#) has more information.

The development of a survey programme will extend the range of information needs that can be met by admin data. The initial sample and collection design includes the development of the new annual (mixed-mode) census survey (CAS), sampling around 3 percent of the population annually. This will be supported by more targeted surveying of key communities, with the continued delivery of the New Zealand Household Disability Survey and Te Kupenga, and the development of a Pacific

Wellbeing Survey being examples. The development of a survey programme of this scale will introduce competing information needs, and a transparent process of balancing these will be important. Stats NZ has a lot of experience running social surveys; however, the development of a survey programme of this scale will require a careful approach to ensure the collection remains in budget and is able to deliver high-quality information for small communities. Investment in methods to reduce the decline in response rates, and to mitigate the impact of non-response, is needed for the CAS and our survey programme more widely.

The continued development of data and analytical infrastructure to support the use of integrated admin and survey data in a production setting is critical. The Integrated Data Infrastructure (IDI) has provided an invaluable test environment for research, but this environment is not suitable as an enterprise register-based statistical system, which an admin data-first census requires. The statistical location register (SLR) was developed and used as an operational frame for the 2018 and 2023 Censuses. As the SLR provides a list of addresses, and as its design does not include maintenance processes, it cannot function to provide dwelling statistics, because it does not have the functionality to adequately identify dwellings or allow determining changes to the dwelling stock over time. The development of the Integrated Statistical Data System (ISDS), a register-based statistical system that enables integrated data to be used in a production setting, is a critical component of an admin data-first census.



Introduction

In 2012, Stats NZ published the Census Transformation strategy (Stats NZ, 2012), which set out a plan for transforming the New Zealand Census of Population and Dwellings. The main drivers for change included the rising costs of field collection, declining response rates, the increased availability of admin data, and new technological advances. The strategy focused on modernising the traditional full-field enumeration census model in the short term, with a long-term focus on developing a new census model based on admin data.

[The Census transformation research programme](#) was then established to progress research into alternative ways of producing demographic, social, and economic statistics covering small area populations, including the feasibility of a census based largely on admin data, and supported by sample surveys.

In February 2014¹, Cabinet agreed that a modernised census would be conducted in 2018, with updated content, a 'digital first' data collection approach, and increased use of technology and admin data. While existing admin data was, at the time, unable to replace current census statistics, there was potential to use existing admin data within the census model. The main development for 2018 was construction of a register of addresses to support transitioning to a mail-out census (the Statistical Location Register (SLR)). Transforming the census to a model based primarily on admin data remained under investigation for longer-term change, through the census transformation programme.

In March 2015, Stats NZ presented to Cabinet the progress investigating future options for transforming the New Zealand Census beyond the 2018 Census (Stats NZ, 2015). The key design recommendations were as follows:

- A census model based on admin data *only* was identified as not plausible.
- The progression of a population register model (which would require a national unique identifier for every individual) was also not recommended, with the focus instead on making full use of the potential within existing integrated admin data sources.
- Stats NZ was directed to 'work towards a future census based primarily on government's administrative data, supported by redevelopment of its household surveys'.

A 2016 Cabinet paper (Stats NZ, 2016a) identified four key barriers to successful implementation of a census model based on admin data (with support from a sample survey) including:

- the quality of key admin data sources across government
- the need to further develop and test the innovative statistical methodologies required to produce admin-based population estimates
- legislative barriers within the Statistics Act 1975 prohibiting or restricting the provision of data to the Government Statistician for official statistics and research
- the use of census to trigger the electoral boundary reviews and the Māori Electoral Option (MEO) within the Electoral Act 1993.

Overall, the cabinet paper noted that there was significant opportunity to transform the New Zealand census with the promising future direction being a census based on admin data and supported by surveys. There was also the acknowledgement that an admin data and sample survey census was not expected to meet all information needs currently met by census; in particular, some

¹ 19 February 2014, the Cabinet Economic Growth and Infrastructure (EGI) Committee [EGI (14)9]

very detailed attribute information for small areas and small population groups would diminish. Stats NZ was directed² to continue active work towards an admin data census. Agencies responsible for key admin data sources in conjunction with Stats NZ, the Government Chief Information Office (GCIO), and Land Information NZ (LINZ) were also directed to improve the quality of core variables.

Since 2015, significant progress has been made to either reduce the impact or remove the key barriers identified, and we have made significant progress developing and testing statistical methodologies that derive census-type information from admin sources.

[Data and Statistics Act 2022](#) removed legislative barriers in the Statistics Act 1975.

[Register of government mandated data standards](#) for core variables provides a framework for improving the collection of data across government.

The Electoral Act 1993 specifies that electoral populations are to be calculated using census data. While the Data and Statistics Act 2022 gives the Government Statistician discretion as to the methods and data sources for producing official statistics, we need to ensure our electoral boundary setting responsibilities remain fulfilled.

[Modernising the census: the process for change](#) summarises the decision to adapt a new admin data-first model for the next census, which was publicly announced in June 2025. The new model is an evolution of the 2018 and 2023 combined census approach, from a model where administrative data supports survey responses to one where surveys are targeted at gaps in the information that can be derived from administrative data.

The change in census model will be phased in gradually, with transitional outputs planned from 2026, official key population measures planned for release in 2028, and a full set of census data produced annually from 2030.

This report outlines technical progress towards the new admin data-first census model. We summarise the work undertaken by the census transformation research programme to inform an assessment of the technical feasibility of a census based on admin data and supported by surveys.

In this introductory section, we provide a brief background on the census. In the context of moving towards an admin-first census, we also define what we mean by 'admin data', and describe the key features of an admin-first census model. We conclude this section by outlining the potential benefits and challenges of using admin data for official statistics.

The following sections focus on how to define technical feasibility, how well we can count the three core statistical units (people, dwellings, and households and families) that the census provides information about using admin data, and how well admin data can describe these units. The role of surveys within an admin-first census model is also discussed, along with the impact an admin-first census model will have on Stats NZ's ability to deliver for and with Māori, iwi, and hapū, and provide information for key communities. The remainder of the report discusses the opportunities and challenges involved in implementing an admin-first census model supported by sample surveys. It concludes with an overview of the key areas of development in terms of methodology, infrastructure, and data sourcing.

² [Cabinet Economic Growth and Infrastructure Committee minutes – Census transformation \(redacted\)](#)

The New Zealand Census of Population and Dwellings

The census provides an official count of how many people and dwellings there are in New Zealand. It gives a snapshot of the people in New Zealand and the places they live. The census is the only comprehensive dataset that covers everyone in the country and provides detailed breakdowns and information for small geographic areas and communities. The information collected in the census is used by government agencies, local authorities, businesses, iwi, community organisations, and the public to develop and implement new policies, research, planning, and decisions about services in areas like health, education, housing, and transport. Census data is also the fundamental data source for calculating electoral populations and setting electorate boundaries.

The New Zealand census has historically been conducted using a full-field enumeration model. This means everyone in the country on a 'census night' has been expected to complete a set of questions about themselves and their household and dwelling. However, over time the NZ census has evolved significantly to reflect changing needs and response patterns. Statistical methods have been essential in ensuring the data is representative. For example, statistical imputation was first employed in 1996 to complete missing electoral Māori descent information, and since 1996, a coverage survey (Post-enumeration Survey, or PES) has been conducted to evaluate the completeness of census coverage, and to report on census response rates.

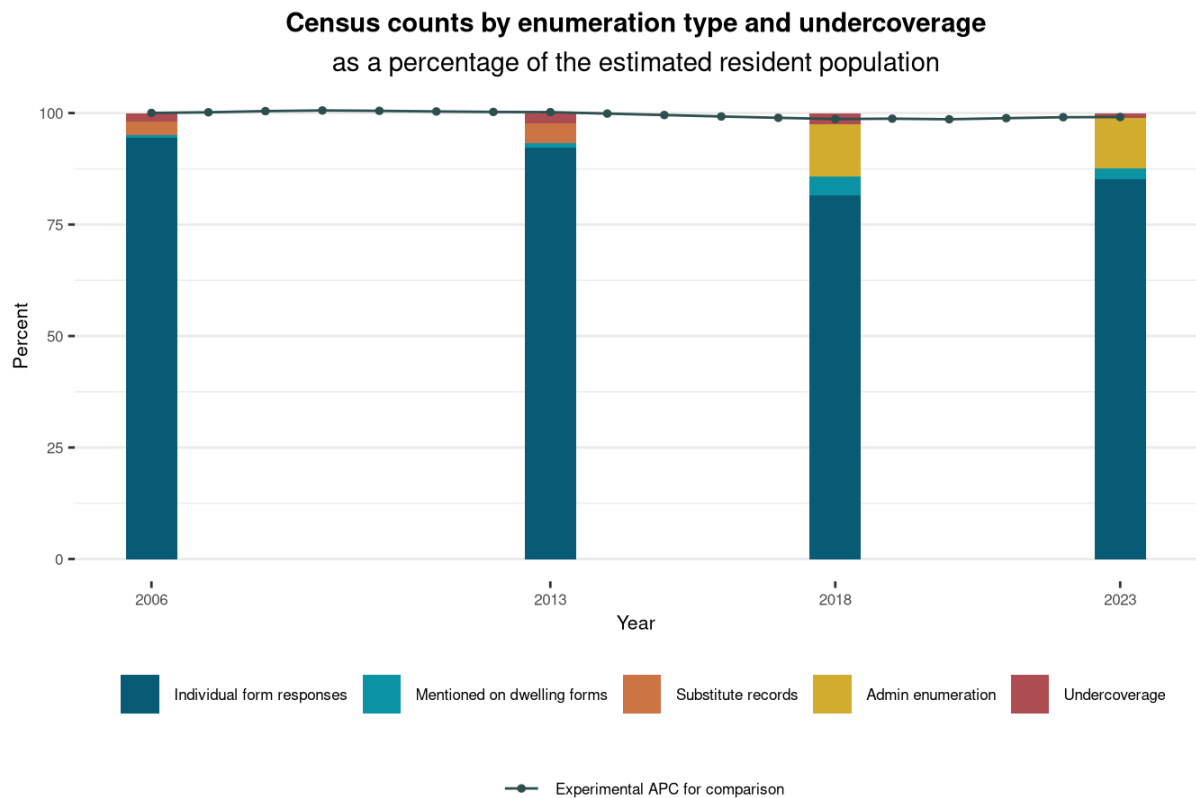
Over the last few census cycles, there has been a shift in the census model to incorporate increasingly more admin data. The 2018 Census achieved lower than expected response rates, which required the use of more admin data than was planned to produce data for the 2018 Census. The combined model was introduced, where the population is counted through a combination of field responses and admin data. High-quality admin records (referred to as admin enumerations) were added to the census file for people from whom we did not receive a response, and when we were confident in the quality of the admin record (Stats NZ, 2019a). Admin data was also used to fill in missing attribute information in the 2018 Census when the characteristics of people or dwellings were not provided on census forms (Stats NZ, 2019b). This innovative approach to mitigating collection issues was made possible by the methodological research that had been undertaken by the longer-term strand of the [Census transformation research programme](#).

In 2023, the census implemented a combined model by design. The use of admin data allowed us to produce census outputs that have higher coverage and are more representative across subpopulations (such as age groups, ethnic groups, or geographic areas) than if we relied only on census responses. The final 2023 Census individual dataset consists of 89 percent census responses and 11 percent admin enumerations (Stats NZ, 2024a).

The 2025 decision to modernise the census means that the next census will be produced using admin data first. That is, admin data will be used to enumerate individuals and dwellings and to derive attributes wherever coverage and quality are sufficient. Surveys will continue to play an important role, with a new annual survey asking questions of a small percentage of the population to collect information not available from admin data and to help confirm the accuracy of admin data. As part of the new approach, Stats NZ will work with priority communities to develop tailored solutions that help ensure their data needs are met.

While the changes are significant, the shift to an admin data-first census represents the next logical step in the evolution of New Zealand's census. Figure 1, (adapted from Olsen 2025), shows that response patterns and mitigation strategies have been shifting for a long time. These patterns will be even more significant for some subpopulations. Viewed this way, the new census model is a natural continuation of the use of admin data that ensures everyone is counted, and minimises undercoverage.

Figure 1 Census counts by enumeration type and undercoverage as a percentage of the estimated resident population



Stats NZ

Using administrative data for statistical purposes

Administrative data (or admin data) is data already collected by government agencies or private organisations while conducting their business or services. It is data that is not collected primarily for statistical purposes. Rather, it is collected for operations such as delivering a service (for example, health or education), or as a legal requirement to register events (for example, births, deaths, and marriages) or as a record of transactions or events (for example, tax payments and overseas travel journeys). The population (who is included in the data source) and data content (what information is collected) are defined by the organisation responsible, and they have primary control of the methods by which the admin data is collected and processed. As a result, admin data differs in nature, scope, and quality from data collected through the census or surveys, where control of who is asked for what information is in the hands of the statistical agency (Stats NZ, 2014).

Stats NZ has a long history of using admin data to produce population statistics. The ability to link data has changed the way we can use admin data and the value we can derive from it.

[The Integrated Data Infrastructure](#) (IDI) is a large research database by Stats NZ that holds de-identified microdata about people and households linked at the individual level. The data includes information on education, work, income, benefits and social services, migration, justice, and health gathered from a range of government agencies, Stats NZ surveys, and non-government organisations. Researchers use the IDI to conduct cross-sector research that provides insight into our society and economy. Source agencies provide data periodically, and the IDI is updated three times a year as part of a process called the 'IDI refresh'.

The basic structure of the IDI consists of a central 'spine' to which a series of data collections are linked (Black, 2016; Gibb et al, 2016). Broadly, the target population for the spine is all individuals who have ever been residents of New Zealand. The spine consists of the union of people in three data sources:

- all births registered in New Zealand since 1920
- all visas granted to migrants since 1997 (excluding visitor and transit visas)
- all individuals issued with an Inland Revenue (IR) tax number.

The range of integrated admin data available in the IDI replicates the variety of information collected in the census. This means the IDI can act as a test environment for examining the potential of linked admin data sources to produce census-type information. At present only data from central government agencies and local government have been investigated for use in the census context.

Benefits and challenges of using administrative data

Admin data can give us a broad picture of New Zealand, people, dwellings, households, communities, and families without having to ask every question in surveys like census. Admin data has high coverage of the resident population, and this level of coverage can be expected to be sustained over time. Some of the potential benefits of using admin data for official statistics include:

- reduced collection costs
- reduced respondent burden
- increased resilience against low response and vulnerability from external events (like pandemics, natural disasters, and severe weather events), which have impacted earlier censuses
- more frequent provision of up-to-date information about New Zealanders

- the potential to increase the frequency and timeliness of census information from what can be provided in a periodic census
- a new range of research possibilities and relevant statistical outputs; for example, longitudinal admin data design, allowing mobility analysis and the measurement of changes to an individual's attributes over time, which is currently not possible with the traditional census model.

At the same time, there are constraints and risks associated with using admin data to produce census-type information. The key challenges are related to the comprehensive nature of census information and the high-quality standards needed to meet users' needs. Difficulties may arise because admin data is not comprehensive by its nature (for example, certain data may only include a subset of the population) or may include people who are not part of the target population.

In addition to coverage issues, if information is not important or necessary for an administrative purpose, data may not be collected or may be of poor quality (Stats NZ, 2014). For example, as agencies move services online, residential addresses may become less important as a contact point and may not be captured well. This affects the ability to locate people where they live. Admin data only collects what is needed to deliver a service or to record an event. It does not capture everything we want to know from a census (for example, information needs for rainbow communities, iwi, and the disabled population). Further, admin data is susceptible to changes in definition or collection, which can adversely affect time series comparisons and the relevance of the data.

Admin data is not primarily collected for statistical purposes, so a careful assessment is needed to understand how it can be used for statistical purposes such as a census, and to measure the quality of derived statistical outputs. Quality issues that may affect the way in which admin data could be used for statistical purposes include:

- Coverage: admin sources generally include only part of the New Zealand resident population (undercoverage) and may also include people who are not members of the resident population (over-coverage).
- Different units: an admin reporting unit (for example, a rating property) may not align well to the statistical unit (for example, a dwelling).
- Misclassification: the variable may not align to the statistical concept we wish to measure, or there may be errors in the collection or processing of the data.
- Lack of high-quality variables suitable for linking to other data sources: missed matches or incorrect links can lead to coverage and misclassification errors.
- Timeliness: there may be long delays between the reference date and availability of data for statistical production.
- Accessibility: there may be difficulty obtaining access for legal, commercial, or other reasons.

Deriving census information from admin data sources requires understanding the strengths and limitations of admin sources, developing approaches to derive statistical information (which usually includes multiple sources), and overcoming any limitations. We also need to develop ways of describing the quality of admin-derived variables.

Work is underway to improve admin data quality and how frequently it is collected. Stats NZ has identified the key data sources required for delivering the census, and we are working with government agencies that can supply this data. For example, we will be working to improve admin data about ethnicity, younger and older people, and recent immigrants to New Zealand.

Assessing the technical feasibility of an admin-first census model

The basic requirement of an admin-first census supported by sample surveys is to derive high-quality lists of people, dwellings, and households in New Zealand from admin sources, without the need for full field enumeration. The decision pathway for determining the feasibility of an admin-first census model prioritises population estimation over dwellings and attribute information.

Attributes rely on good population estimates and high-quality values for core demographics (age, sex/gender³, location, ethnicity, and Māori descent). Thus, for attribute information to be technically feasible it is a prerequisite that we have determined population estimates. There is little merit in shifting to collecting attribute data from a sample survey unless we can produce population statistics without the full enumeration census. If population estimates are not feasible, there would be no need to proceed with assessing the feasibility of attribute or dwelling estimates, as the entire model would not be feasible.

In 2015, a set of quality standards were developed to assess the quality of population estimates produced from admin data (McNally & Bycroft, 2015). These quality standards were determined through consultation with core customers and provide a measure of the minimum accuracy acceptable to users. Separate standards were produced for both a survey-based and an admin-based census model. For the purposes of this report, we define the technical feasibility of producing population estimates as the ability to meet the admin-based census quality standards set by McNally and Bycroft (2015). For convenience they are included as table 2 in the appendix.

These standards were intended as a starting point only, but are useful to:

- describe the level of accuracy historically achieved in the population estimation process (over the 2001 to 2006, and 2006 to 2013 census cycles)
- provide a benchmark for assessing the strengths and limitations of admin data for producing population statistics.

These standards were based on which populations were included in the annual ERP publication at the time. When developing quality standards for future population statistics, a wider set will be included. For example, we would expect these to include quality targets for more detailed ethnic groups, Māori descent, gender, and sex at birth. Quality standards for dwelling and household estimates will be developed in consultation with key users to ensure they reflect information needs and customer requirements for accuracy. Furthermore, we currently assess feasibility based on comparison of admin estimates against a benchmark (for example, official estimates provided by the ERP). As we move towards an admin-first census model, quality standards that don't need to be compared with an alternative source of data will be required. For example, using credible intervals as an indication of the uncertainty surrounding the estimate.

The quality standards for population statistics are intended to be applied to the final population estimates, after implementing statistical models to adjust for any errors in the admin-derived resident population. These models are discussed in the section [Producing the official estimated resident population from admin data](#).

³ Increased adoption of our mandated data standard for gender will lead to improved gender data from admin sources over time. We use the term sex/gender in this document to refer to admin sources that contain a mixture of sex and gender information.

In this report we assess the quality of the admin data alone against these standards with the understanding that further improvements will be possible. To what extent the population estimation process can adjust for errors in the admin-derived resident population is yet to be determined. Unlike the recent censuses, the unadjusted counts are reasonably close to the quality targets, which provides a level of confidence that modelled estimates should meet these quality standards.

Along with the set of quality standards, the enduring census information requirements for and about Māori, which must be provided by future censuses regardless of the census model used, were identified (Gleisner et al, 2015). Variables that define Māori populations (Māori ethnicity, Māori descent, and iwi) and information about te reo Māori are all essential requirements for a census to provide. In addition, there is an established need for information across the broad topics of education, income, work, health, households and families, and housing.

How well can we count our core statistical units using admin data?

Identifying the New Zealand resident population in admin data

The first step in an admin-first census is to derive a list of people resident in New Zealand at a given date, including as many genuine usual residents as possible, while minimising the inclusion of people who are not usual residents, such as temporary international visitors. Methods for deriving a New Zealand resident population and assigning a location and ethnicity to the individuals within the population from admin data have been developed iteratively.

[Experimental population estimates from linked administrative data](#) released findings between 2016 and 2018 (Gibb et al, 2016, Stats NZ, 2016b; 2017a; 2018). The approach was used in both the 2018 and 2023 Census to include admin enumerations in the census file (Stats NZ 2019a, 2024a).

A ‘signs of life’ approach is taken to determine if an individual should be included as a usual resident at a given reference date. Individuals on the IDI spine are retained if they have records in selected admin sources over a two-year period prior to the reference date (for example, paid tax, used health services, or enrolled in education). Those who have died before the reference date are identified through death registration data and are excluded. Migration border-crossing data is used to exclude anyone who is not a usual resident on the reference date – for example, a resident who migrates to live overseas or a short-term visitor to New Zealand. Applying the ‘signs of life’ rules to the IDI spine provides an admin-derived population list, which is the admin equivalent of the people included in the census dataset. Unlike the five-yearly census, which provides a new list of people in the resident population only every five years for the census date, the admin-derived population can be derived for any given date.

Within the IDI, population tables of the admin-derived resident population for a given reference date are implemented as part of the IDI production process and made accessible to all researchers. The admin-derived resident population is now widely used as the population concept for research within the IDI. Since the release of the initial experimental series, further development work has gone into improving the methods used to construct the admin-derived resident population by refining the activity rules used to indicate an individual’s presence in New Zealand and introducing a step to remove duplicate records where possible. With these refinements, the admin-derived resident population list forms the basis of the administrative population census (APC), an experimental product released by Stats NZ to demonstrate to customers our current ability to derive census-type information from linked admin data sources.

The APC is an annual time series based on the admin-derived New Zealand resident population. Additional characteristics for each person can be derived from linked admin sources or surveys and individuals grouped within dwellings to form households. Stats NZ has released three iterations of the APC. The first iteration of the APC focused on demographic and identity variables: age, sex, geography, ethnicity, Māori descent, birthplace, and years since arrival in New Zealand (Stats NZ, 2021). Other topics such as work, income, and qualifications were added in the 2022 release (Stats NZ, 2022). The third iteration also included information on households (Stats NZ, 2023b). Currently, census data is a snapshot at one point in time, whereas the APC is held as a longitudinal record of people over time by design. With the APC, Stats NZ can assess the suitability of the admin-based resident population for use in the production of official population statistics.

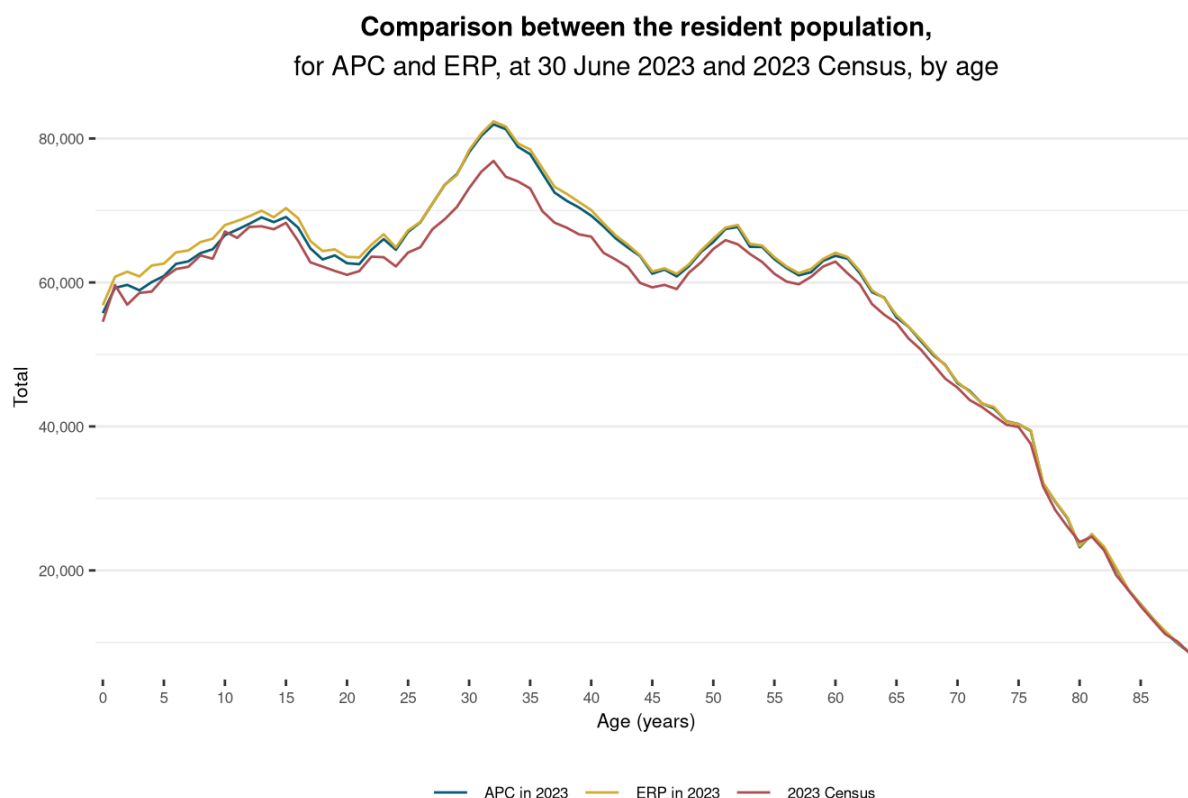
[Experimental administrative population census](#) has more information on the APC.

Assessing the technical feasibility of admin-based resident population counts

As outlined in the introduction, the leading requirement of a census is to count the population by core demographic variables, at a national level and down to small geographic areas. In this section we focus on how well the APC can capture the core demographics that feed into the production of the official estimated resident population (age, sex/gender, ethnicity, Māori descent, and location). We discuss additional attribute information in [Characteristics of core statistical units provided by admin data](#).

We compare the admin-derived resident population against official estimates provided by the ERP and use results to indicate the likely quality of our admin-derived resident population counts, while noting that the ERP also has uncertainty that could contribute to any difference (Bryant et al, 2016). The admin-derived resident population closely tracks the age distribution of the ERP (figure 2).

Figure 2 Comparison between the resident population for APC and ERP, at 30 June 2023 and 2023 Census, by age



Stats NZ

Even without any coverage adjustment, the APC counts in 2023 are, for some breakdowns, close to fulfilling the quality standard referred to above.

Compared with the 2023 ERP, there is a relative undercount of 0.9 percent. This does not quite meet the quality standard requirement of 0.5 percent. The standards for the Māori ethnic group are further away from being met. Analysis shows that the biggest differences in this group are for older people, which means that including more historical census data into the APC will improve the unadjusted counts. However, as outlined in the ethnicity data and Māori descent sections, admin data availability and quality will also need to improve in the longer term.

The standard for five-year age groups and sex is almost met with all of these groups within 5 percent, and 72 percent of these groups within 1.5 percent, which is lower than the 90 percent required. The standards for SA2s (statistical area 2s), and the ones for TALBs (territorial authority/Auckland local boards) by broad age groups are also very close to being met. The standards for TALB total populations are further from being met. Altogether this shows that the APC counts have a fairly uniform undercount compared to the ERP, which explains why the more granular the level the closer the standard is to being met.

Note, this is without any coverage adjustment. Overall, the admin-derived resident population is closer to the official ERP than the census usually resident population count (though this is partially due to the difference in target population between ERP and census⁴), and is substantially more

⁴ The official ERP is always higher than the census because the census target population excludes residents temporarily overseas on census night, and the census counts have not been adjusted for net census undercount.

representative than responses achieved from field enumeration in the 2018 and 2023 Censuses (prior to admin enumerations being added to the files). This provides confidence that we can count almost everyone in the population using admin data. The question to what degree we count them in the right place or as part of the right subpopulation is looked at in the next subsections.

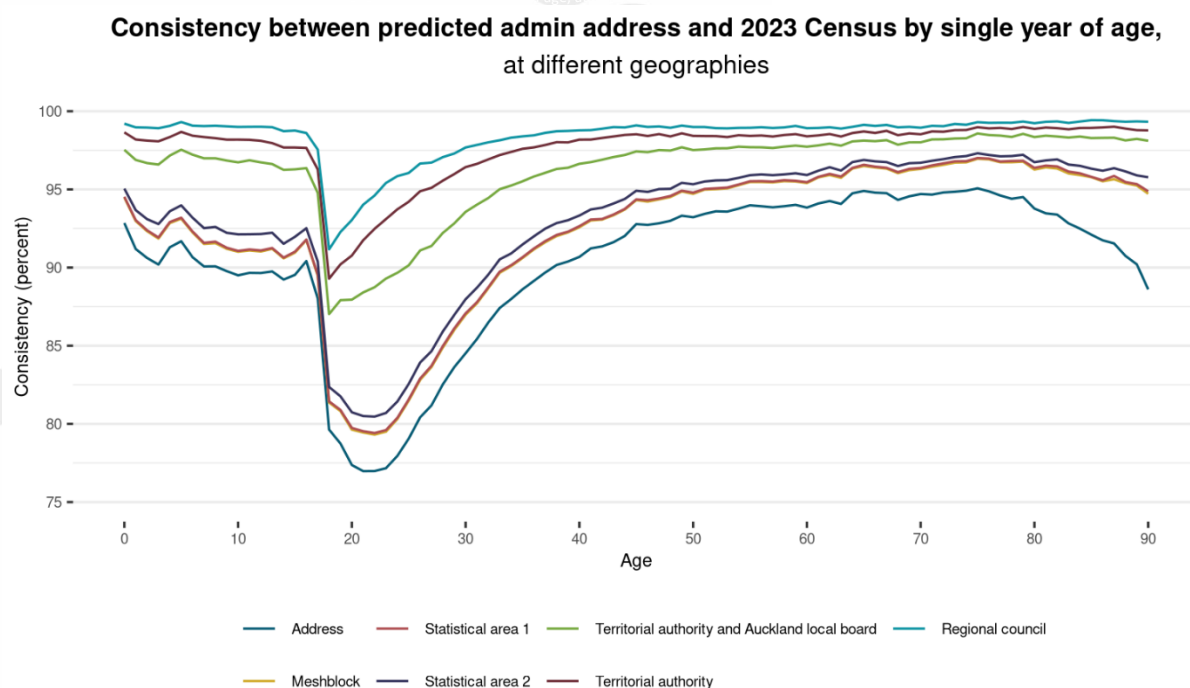
Producing subnational geographic counts

Information about where people live is key to understanding population structure. The accuracy of the admin-derived resident population counts, and hence population estimates, for subnational geographies is dependent not only on the accuracy of the admin-derived resident population, but also on the availability and accuracy of address information for the population.

The admin-derived resident population is a good approximation of the New Zealand-resident population for larger geographic areas, but accuracy naturally decreases at smaller geographies. Comparisons with the 2023 Census show that we can correctly identify the usual residence address for 77 percent of the census usually resident population, with consistency increasing for larger geographic areas (97 percent for regional councils). There are patterns of lower agreement for the mobile population, in particular young adults.

The APC currently selects usual residence address from multiple sources using the most recently notified address to provide individual information. A machine-learning approach to predicting the 'best' address has been implemented for the APC household information (Stats NZ, 2023b), and used in the 2023 Census (Stats NZ, 2024d). This achieves 89 percent agreement with 2023 Census usual residence address. This machine-learning approach will be developed further and implemented in the individual-level APC and the central IDI tables. The rates of address consistency vary by age-group (figure 3).

Figure 3 Consistency between predicted admin address and 2023 Census by single year of age, at different geographies

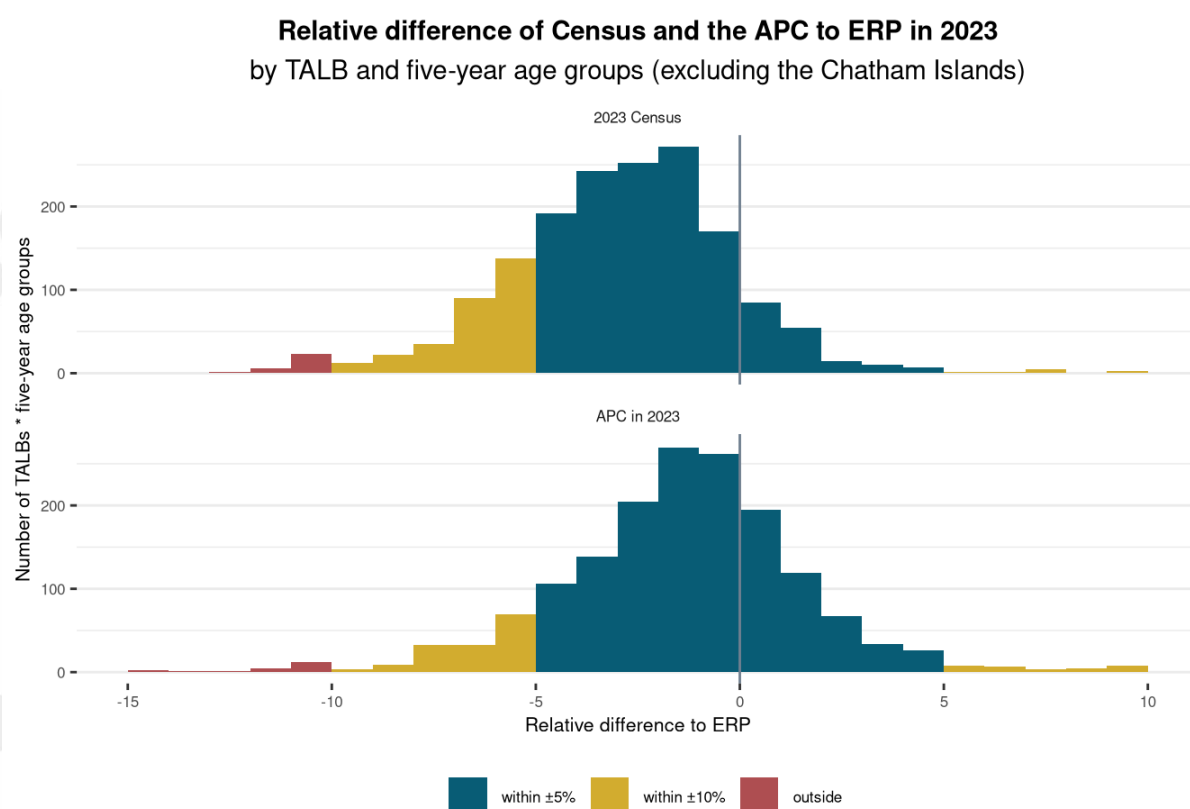


Stats NZ

Accurate addresses are important for placing people into the right household. Although not all errors (such as those resulting from misspellings or variations on house numbering) will place people in the wrong geographic area. Furthermore, aggregate counts are statistically more robust than individual-level analysis. Comparing the admin-derived resident population counts for territorial authority/Auckland local boards (TALBs) with the 2023 ERP, there is a relatively uniform undercount in the admin population across the TALBs. The majority of TALBs have admin-derived resident population counts between 1 to 3 percent lower than the ERP but are higher than the census counts. There are a small number of TALBs where the admin resident population count is between 5 and 10 percent lower than the ERP.

Figure 4 shows the percentage difference of TALB counts by five-year age groups for the admin-derived resident population (APC) and the 2023 Census, when compared with the ERP in 2023. The distribution of differences compared with the ERP is similar between the two sources, with the admin-derived resident population undercount typically smaller than the census undercount. Counts at the TALB level for 2023 do not meet the quality standard accuracy requirements of official population estimates, as specified in McNally and Bycroft (2015), but are close to fully meeting the quality standard for the SA2 geography, even without adjustments.

Figure 4 Relative difference of Census and the APC to ERP in 2023, by TALB and five-year age groups (excluding the Chatham Islands)



Stats NZ

Though the admin-derived resident population counts align more closely with the ERP, statistical methods of measuring and adjusting for systematic error in the underlying data will be required to ensure high-quality population estimates are produced and users have confidence in the data. These methods are discussed in the section [Producing the official estimated resident population from admin data](#). Coverage adjustment can account for variation in coverage across TALBs, which will be due to some misclassification of admin address information in addition to net undercount of the total population. There is also potential to improve the location information at source by making use

of additional data – for example, using education and employment data to provide more up-to-date information for young adults. This may reduce inconsistencies in location information for 20 to 30-year-olds (seen in figure 3Figure) who move from their family home to study or work but do not immediately update their address with government agencies and service providers.

Ethnicity data in admin sources

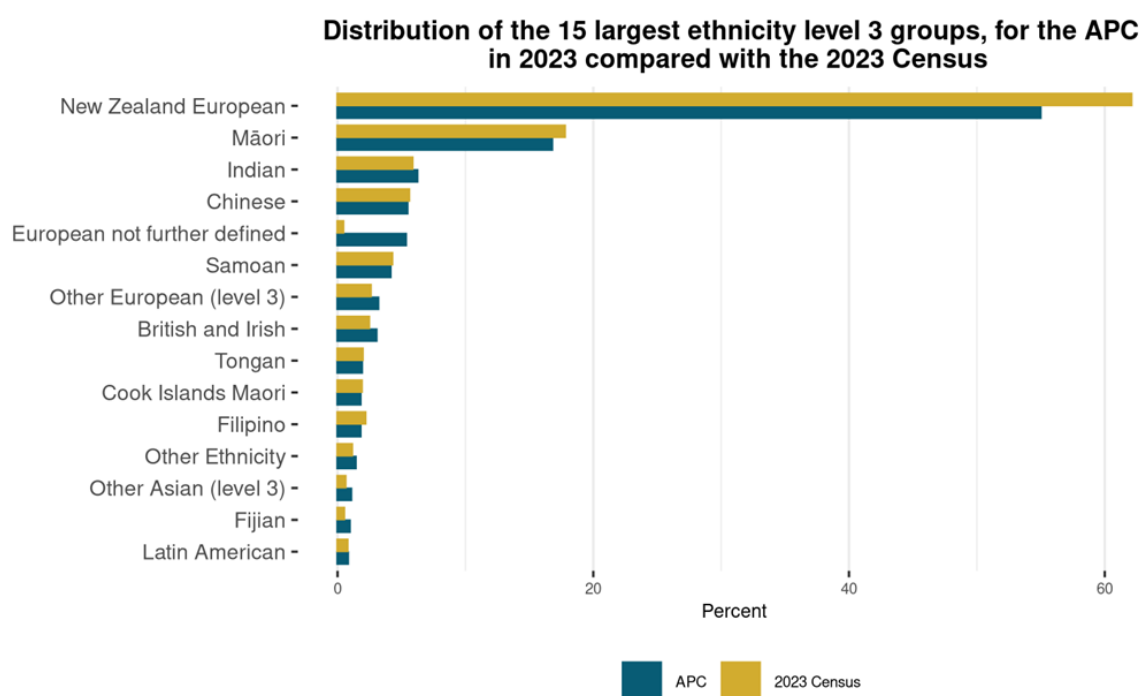
Ethnicity is a core demographic variable for describing the New Zealand population. Ethnic breakdowns of populations and communities are a fundamental and widely used aspect of census information. The ability to produce ethnicity data from admin sources is therefore a key consideration when determining if a census based on admin data is feasible.

[Ethnicity New Zealand Standard Classification 2005 V2.0.0](#) is a four-level hierarchy. Level 1 of the classification has six categories: 'European', 'Māori', 'Pacific Peoples', 'Asian', 'MELAA (Middle Eastern/Latin American/African)', and 'Other ethnicity' and is used solely for outputs, not collection. Ethnicities with larger populations in New Zealand appear at level 2 of the classification, while those with smaller populations are only reported as separate ethnicities at level 3 or 4. Agencies collect ethnicity information at varying levels, from the most detailed, level 4 (DIA, Stats NZ census and surveys), to level 3 (MoE) and level 2 (MoH, MSD). The nature of ethnicity also means that there will be legitimate differences in recorded ethnicity between collections.

Currently the APC derives ethnicity from multiple sources using the method described by Reid et al (2016) and refined by Stats NZ (2018). The admin data sources are ranked by quality, and the ethnicities provided in the highest-quality source available are selected for the individual. Ethnicity data from the 2013 Census is used for the less than 0.5 percent of individuals who have no ethnicity in admin sources. Additional modifications to the source-ranking method have been made for the APC to improve the derivation and usability of levels 3 and 4 (Stats NZ, 2022). If an individual was assigned a non-descript grouped category (for example 'Other Pacific Peoples') from an admin data source, and that individual has a 2013 ethnicity response that aligns with the response received in admin data but at a more detailed level, then the grouped category was replaced by the more specific ethnicity from 2013 Census data. For example, if the highest-ranking admin data source available for an individual was MoE tertiary data, which included the 'Other Pacific Peoples' ethnic group, and the 2013 Census recorded an ethnicity of 'Tuvaluan' for the individual, then the 'Other Pacific Peoples' response is substituted with the 'Tuvaluan' response.

By combining different sources, problems of specific admin sources can be somewhat mitigated. Overall, there is good agreement between APC and Census counts. Figure 5 compares distributions in the APC in 2023 with the 2023 Census for the 15 largest level 3 ethnicities. Proportions are close except for a large discrepancy between 'European not further defined' and 'New Zealand European'. Data issues that affect multiple data providers are apparent in the APC, in particular the overcounting of 'Not further defined' and 'Other' ethnic group categories, and the overcounting of the 'Fijian' ethnicity.

Figure 5 Distribution of 15 largest ethnicity level 3 groups, for the APC in 2023 compared with the 2023 Census



Stats NZ

While the use of 2013 Census data improves the coverage of level 4 ethnicities in the APC, and we will similarly be able to use 2018 and 2023 Census data in the short to medium term, this is not a long-term solution. There is a need for improved collection and coding practices of ethnicity across many government agencies to deliver high-quality data for the smaller level 4 ethnic groups.

[Reviewing the quality of ethnicity data from administrative sources: Comparison with the census](#), published in December 2024, and [Quality of ethnicity data in the experimental administrative population census \(APC\): High level summary](#), published in October 2025, provide more detail on ethnicity information from admin data. Together, both papers give a good overview of the current state of ethnicity information from admin data.

Small ethnic communities rely on the census for population counts and demographic summaries. These populations are unlikely to be well served by deriving this information from a sample survey due to population-size and sample-size restrictions. Improvements in the collection and coding of ethnicity collected in admin settings is likely to add significant value to the ability of different agencies to monitor outcomes within their own data systems, but improvements will take time and significant resources by the agencies involved. Identifying where to prioritise quality improvement will need to be done in partnership with the agencies involved to ensure the work aligns with their data strategy. Given the coverage of the population, the health and education sectors are likely to be key areas for focusing improvements.

Māori descent information

Māori descent information from the census provides important insights, informs decision-making activities, and enables the data aspirations of iwi and Māori. There is also a legislative requirement for census Māori descent counts because of their use in the electoral system.

The experimental APC currently derives Māori descent for approximately 86 percent of the resident population. The key sources used (and available) are DIA birth registrations, which provide information for the child and parents, along with 2013 Census responses. Currently the APC does not use 2018 or 2023 Census responses due to census being used for comparison in quality assessments. While the quality of Māori descent data is very good for people whom we have a value for, there are 14 percent missing a value.

Missing data is partly due to the lack of admin sources before 1995 when Māori descent was first collected in birth registrations. The use of 2013 Census data significantly improves coverage but there are still significant gaps, particularly for the overseas-born population who arrived since 2013. Coverage of the New Zealand-born population will gradually increase over time. Some migrants will be captured in the DIA births data if they have children born in New Zealand but, assuming an increasing portion of the resident population are born overseas, missing Māori descent information for migrants will continue to grow as we move further away from 2023, even if historical census responses are used.

Electoral roll data is now supplied to Stats NZ, including information on people aged 18 years and older who are eligible and enrolled to vote. The Electoral Roll also collects Māori descent information to support the eligibility for voting in Māori electorates. The recently released 2023 Census used electoral roll data to help fill in missing records for Māori descent (Stats NZ, 2024c). Our research shows that by including electoral roll data, the overall missingness for Māori descent in the experimental APC can be reduced by about 5 percentage points. It can also be expected to counteract future growth in missingness from migration, but it will not be able to completely eliminate it.

In the short to medium term, 2018 and 2023 Census data can be used to fill in missing Māori descent responses in the APC, but in the longer term, historical census data will become less relevant. There is potential to explore additional data sources to supply Māori descent information. For example, the Ministry of Business, Innovation and Employment (MBIE) visa data is a useful data source for migrants, but a change to the collection would be required as current visa applications do not ask about Māori descent. Ministry of Education (MoE) data from tertiary and school enrolments could also help to fill in the gaps for those aged 20 to 30 who have proportionally higher rates of missing Māori descent information, and for those under 18 years old who are not captured in electoral roll data. For example, 2023 Census used stated iwi affiliation in MoE data as a proxy for Māori descent when all other sources were missing that variable (Stats NZ, 2024c). Consulting and co-designing the methodology with Māori iwi partners is essential to ensure Stats NZ upholds its Te Tiriti obligations.

Even with additional sources, there will still be some people whose Māori descent data is missing. As with census currently, statistical imputation techniques will be used to fill the remaining missing values where appropriate (Stats NZ, 2024c). Ongoing collection of Māori descent in survey collections will still be needed to validate data collected in admin sources and to inform statistical imputation models for the proportion of the population without an admin-sourced Māori descent indicator. The patterns of missing data are likely to be different from those usually seen in census non-response. In future, those missing Māori descent are most likely to be overseas-born adults and older New Zealand-born people who do not enrol (or are ineligible) to vote, and overseas-born people who have not had a child in New Zealand.

Table 1 below compares the Māori descent proportions from the 2023 Census and the rebased ERP at 30 June 2023 with APC proportions. As the ERP only includes Māori descent and implicitly its complement, the table focuses on the *electoral* Māori descent variable. That leads to an even higher missingness in the APC as “don’t know” responses are treated as missing data. Based on this year’s research, incorporating data from the 2013 and 2018 Censuses and the Electoral Roll substantially reduces missingness. A simple statistical imputation taking into account very basic demographics results in proportions which are quite close to the ones from the ERP. This gives confidence that, similarly, for the next census, a combination of APC, 2023 Census, and Electoral Roll data, together with statistical imputation, will be able to provide reliable Māori descent counts. The exact method for combining these sources and for statistical imputation will need to be developed in collaboration with iwi partners.

Table 1 Percentage of 2023 population with electoral Māori descent, with and without imputation

Percentage of 2023 population with electoral Māori descent, with and without imputation						
Data source	Without statistical imputation			With statistical imputation		
	Māori descent	No Māori descent	Missing	Māori descent	No Māori descent	Missing
	Percent					
2023 Census	-	-	-	20.2	79.8	-
ERP at 30 June 2023	-	-	-	20.0	80.0	-
APC + Electoral Roll + 2013 and 2018 Census	18.8	72.0	9.2	19.7	80.3	-
APC	17.1	64.0	18.9	18.9	81.1	-
	-	-	-	20.2	79.8	-
Source: Stats NZ						

Identifying dwellings in administrative data

Currently, census is the only comprehensive source of information about housing stock and dwellings at lower levels of geography and other breakdowns.

[Dwelling and household estimates](#) are regularly derived based on the census count of dwellings, which gets updated using building consent information in the intercensal period to reflect changes in the housing stock. Estimated private dwellings are a measure of the private dwelling stock in New Zealand and include occupied and unoccupied private dwellings, as well as temporary dwellings. The method for deriving dwelling estimates considers the fact that only a certain percentage of consented dwellings get built and that there is a lag between issue of a building consent and completion of the dwelling. The estimates make no adjustment for private dwellings missed or counted more than once by the census. After each census the dwelling and household estimates for the preceding intercensal period are revised. National estimates are produced quarterly, whereas subnational dwelling counts (down to SA2 level) are only available for census years.

There has been a longstanding need for more frequent subnational information about the number of dwellings built over time and the size of the dwelling stock across New Zealand. An experimental series was published in 2017 that used admin-data sources such as building consents and city-

council data on inspections and code compliance to provide subnational estimates of unique residential private dwellings (Stats NZ, 2017b), but this series has not been further developed since then.

The key enabler of dwelling-count and attribute production in an admin-first census is having the necessary infrastructure of a national list of locations, and sound methods for identifying dwellings from this list at a given point in time. Stats NZ requires a maintainable list of uniquely identified dwellings.

Stats NZ created the Statistical Location Register (SLR) in 2017 to initially support the delivery of the modernised 2018 Census. While the SLR provides the basic infrastructure for a list of addresses and addressable objects, a Dwelling Register is required to support the delivery of dwelling counts. At present, the Dwelling Register based on the SLR exists as a static list of dwellings at the time of the 2018 Census and the 2023 Census. The dwellings are linked to an address string, and the dwelling-address links are held in the SLR. In this set-up, address is effectively used as a proxy for dwelling. There are shortcomings with this approach that mean it is not always clear where dwellings are located, how many dwellings should be at a location, or even if a dwelling exists.

While there are methods for updating addresses in the SLR, and regular updates occur via data supply from NZ Post and Land and Information New Zealand (LINZ), currently there is no system for updating and maintaining the dwelling population. Any updates that have occurred since the conception of the SLR have been based on fieldwork completed for the 2018 and 2023 Censuses. These updates only represent dwelling populations at fixed points in time. Under an admin-first model, all quality improvement processes need to be implemented both on a more regular basis and prior to producing counts of dwellings. This means our maintenance processes cannot rely solely on census-like field operations to mitigate error within the dwelling population.

At present, there is no mechanism for linking address information available in the SLR with person or dwelling-related data from within the IDI. This has limited our ability to investigate the potential for identifying dwellings from data held within the IDI.

Stats NZ is currently building the Integrated Statistical Data System (ISDS) which will have the ability to connect admin and survey data across three key registers: people, place, and business. A property-centric register known as the Places Index is being constructed, which will enhance features of the SLR by introducing the key entities of 'Property' and 'Building' and linking these entities with 'Address' and 'Dwelling' in a more precise and accurate manner. This work builds on a 2022 research collaboration between Stats NZ and LINZ (Toitū Te Whenua Land Information New Zealand, 2019), which was undertaken to help understand the importance of connected property data (a property spine linking addresses to properties, buildings, and dwellings), the quality of the data, the data gaps, and what indicators provide the right information relating to buildings/dwellings (Cutting et al, 2024). This cross-agency project was able to show that using the LINZ Property Data Management Framework (PDMF) (which describes the relationships between properties, addresses, buildings, and dwellings) and a property spine, dwellings could be identified using admin data.

[Property Data Management Framework](#) has more information on the PDMF.

Methods to accurately update and maintain a dwelling register over time are in development, based on available data. Sources of information include building consents, Stats NZ surveys, District Valuation Roll (DVR) data managed by LINZ, and residential address notifications from government agencies.

Assessing the technical feasibility of admin-based dwelling estimates

As the proof-of-concept for the Places Index within the ISDS nears completion, it has established the foundation for moving the register into a production-ready environment, beginning with the integration of 2023 Census data. Concurrently, we are developing an onboarding framework to integrate admin data sources in a structured and repeatable manner, starting with building consents.

However, it is clear from the outset that any admin-based dwelling count will be restricted mainly to permanent private and non-private dwellings. Temporary and mobile private occupied dwelling types classified as 'Other' are a small but important group and not generally well captured in admin data. However, there is evidence that some dwellings classified as 'Mobile' do in fact have formal addresses and could be included in admin dwellings. This has implications for the measurement of severe housing deprivation as discussed in the section [Measuring key communities](#).

While there are quality standards for population estimates produced from admin data (McNally & Bycroft, 2015), similar standards do not yet exist for dwelling and household estimates. The development of quality standards to measure the accuracy of dwelling and household estimates produced from an admin-first census is at an advanced stage.

Additional coverage adjustments may also be required in an admin-based model, either as a single exercise to inform coverage of an initial build of a dwelling register, or as a periodic survey to ensure the maintenance and updating of the register is sufficient for quality assurances. This would give reassurance to the shift to admin-based dwelling counts and potentially provide coverage-adjusted dwelling estimates, which have not been published since 2006. Design work has been initiated on a coverage survey for dwellings, accompanying the model change.

Creating household and family information from admin data

Information about the counts and key characteristics of households and families in New Zealand is an important output of the census. Census information about households and families is vital for informing public policy, meeting Te Tiriti o Waitangi / Treaty of Waitangi obligations to iwi and Māori, population projections, and the derivation of analytical measures, such as household crowding and social deprivation. To deliver these key statistics, we need to construct household units and family units (or family nuclei).

In the traditional census model these units start with the dwelling and are formed using address, relationship, living arrangement and resident information collected on census forms. Dwellings are identified first, and people are then counted within those dwellings so that the collection of data about people and dwellings are closely intertwined. In an admin-first census model this will no longer be the case, and data will be collected independently about people and about dwellings. This will result in increased complexity around placing people into the correct dwellings and households to ensure that we continue to produce a census file that is coherent across our core statistical units (people, dwellings, and households and families).

There are currently no admin-data sources that directly measure all households and family units; however, there are several admin sources that provide relationship information independent of where people live (DIA births, marriages, civil unions, and MBIE visa information), and information on where people live independent of their relationships. These types of information can be combined to construct households and families and allow for new opportunities that are not

available currently in census, such as communities of support outside of a household. There are gaps for informal relationships and those relationships formed overseas.

The third iteration of the experimental APC was the first to include admin-based census household information (Stats NZ, 2023b). Household size, household income variables, and relationship information is provided.

The target population of households in the APC is households usually occupying (or living) in private dwellings, the same target population as the official household estimates. This is different from the target population for census households, which includes only dwellings occupied on census night, and excludes dwellings where people are away on census night.

Admin households are constructed as one or more individuals in the admin-derived resident population who share the same unique address (excluding addresses for non-private dwellings). Address is used as a proxy for a dwelling, meaning that individuals are assumed to belong to the same household if they share the same address. The use of address as a proxy for a dwelling will combine households at addresses with multiple private dwellings that were previously counted separately and exclude households at addresses containing non-private dwellings. In the 2023 Census 1–2 percent of census dwelling addresses have more than one household at the same address, which affects about 3 percent of census households. People who live in mobile or temporary dwellings are less likely to have formal addresses, and so are less likely to be counted as admin households.

The ability to accurately place people into households depends on admin address assignment. The accuracy of admin usual residence address has improved over time. Several factors have contributed to improved performance since earlier work on forming households (Gath & Bycroft, 2018, Gibb & Das, 2015):

- improvements in Stats NZ addressing infrastructure (in 2019 the IDI shifted to a new address matching tool referencing the SLR)
- use of a machine-learning model to select the best admin address. See Stats NZ (2023b) and Stats NZ (2024d) for details.

Admin addresses are available for nearly all the admin-derived resident population, with an agreement rate of 89 percent with the usual residence address reported in the 2023 Census using the new machine-learning model. As discussed earlier and seen in figure 3Figure , there remains a noticeable drop in accuracy for highly mobile groups (for example, young adults from age 18 years) and people who have recently entered the resident population (for example, births, migrants). This may be partly due to lags in people reporting a change in address to government agencies.

Admin families are constructed as a couple, with or without child(ren), or one parent and their child(ren), all of whom are in the admin-derived resident population and share the same admin household/address. Multiple admin families can be identified within a single household address. Admin households created from multiple dwellings might be identified as a separate family nuclei. It is also possible for one admin family nucleus to be formed at one admin address, which may contain multiple dwellings. Accurate family nuclei depend on identifying accurate admin relationships (partnerships and parent-child relationships) within admin households. [Assessing the quality of partnership status attributes from admin data](#) was published in 2025 and will be followed by three families and households research papers, which are due to be published in early 2026. These papers will cover families, relationships, and network approaches to forming households from administrative data.

Assessing the feasibility of admin-based household and family units

There is currently no quality standard for households and families like the standard used for the admin-derived resident population, and we are yet to identify what level of accuracy users require for household and family-level statistics. In the interim we have evaluated our ability to count households and families through comparisons with official statistics published by census and the official household estimates.

Despite the differences, the total number of APC households is very close to the census household count (APC is 0.4 percent higher than the 2018 Census), and 93 percent of the official household estimates (Stats NZ, 2023b). Currently, the APC household dataset is a simple count of households, and there is no adjustment made for net undercount of individuals or households.

Several factors influence differences in identifying households between census and APC, including scope differences outlined above, addresses with multiple dwellings, errors in the Statistical Location Register or address matching, and census misidentification of occupancy status.

Of the *highly comparable* households, 72 percent of admin-derived households have the exact same household membership as reported in the 2018 Census. This an improvement on the 48 percent membership agreement reported in earlier studies of admin and households using the 2013 Census (Gath & Bycroft, 2018; Gibb & Das, 2015). Patterns of membership agreement are unevenly distributed, with lower agreement for larger households, households identified by census to contain a more complex family structure, households of unrelated people, and single-parent households. These types of households also tend to be more difficult for census field enumeration. Admin household size agrees with 74 percent of the highly comparable households in the 2018 Census. The most common difference in household size is for the admin household to have one extra person than census.

Admin family units were last published by Gath & Bycroft (2018), which showed that when admin families were found they had a high agreement with census. However, the coverage of admin relationships was low, which was a major hindrance to creating admin families. With the increased relationship data available, this analysis has been repeated, and significant gains have been made. The count of admin family units has increased to 90 percent of the 2018 Census families. Family nucleus membership was correct for approximately 82 percent of people in admin families. However, information is still lacking for people in single-parent families, and for people aged 80+ and those in their mid-20s.

In summary, since the initial publications (Gath & Bycroft, 2018; Gibb & Das, 2015) we have improved the ability to construct admin-based households and families, although there are still issues, particularly for placing more mobile groups in the correct households and forming family nuclei when relationships are missing. Further work will include extending the experimental household outputs to include family information, improving admin relationship selection, developing network-based approaches to household formation, and investigating new sources of address information. While we expect that the latter two pieces of work will result in further improvements to the quality of the admin households and families that we construct, we also need to develop methodology for deriving the unit-record-level quality of these units. This methodology, and its implementation, will be key in determining the role that surveys will play in supporting the delivery of household and family information in an admin-first census model (supported by surveys).

Characteristics of core statistical units provided by admin data

The preceding sections detail the considerable progress we have made in being able to identify our core statistical units for the population (individuals, dwellings, and households and families) from admin data as well as outlining some obstacles. Another key requirement in transitioning to an admin-first census model supported by surveys is assessing the ability of admin data sources to provide census-type information about the characteristics of these units.

The IDI contains many admin datasets, Stats NZ surveys, and the 2013, 2018, and 2023 Censuses, linked at the individual level. The system (as represented by the linked admin datasets within the IDI) as a whole is stronger than might be apparent from the point of view of a single agency. As well as the more obvious benefit of increasing coverage, the use of multiple sources for a variable has other benefits. Multiple sources provide an in-built redundancy so that the key demographic variables are not dependent on any single data source. The use of multiple sources also provides protection against quality limitations. For example, lower-quality sources may only be needed when higher-quality sources are not available.

We have built up a comprehensive and detailed understanding of the quality of census characteristics that can be derived from admin data and identified the types of information collected by census that are unlikely to be obtained from admin data (Bycroft et al, 2021).

As part of these investigations, we have transformed or combined data into variables that align with our statistical concepts and developed robust methods (and models) for combining information about a person, dwelling, household, or family from multiple admin sources into a single value. Quality frameworks have guided the assessment of the use of admin data for delivery of census information needs (Stats NZ, 2016c). We have consolidated the previous research and brought the admin data together in the APC.

How good are the characteristics that we can get from admin data?

Assessment of admin-data sources for delivering census-like characteristic information has focused on describing the key features of the admin sources and a quantitative assessment of the accuracy dimension of quality. Investigations have assessed both the representation (does admin data include the right people and dwellings; how well does admin data cover the target population) and errors of measurement (how well the admin variable measures the census concept). We have assessed representativeness by investigating the extent of the population that has coverage in admin data and have assessed measurement accuracy by comparing census responses with admin data on an individual level.

On the APC website we provide:

- [Experimental administrative population census: Data sources, methods, and quality \(second iteration\)](#)
- [Experimental administrative population census \(third iteration\): Information by variable.](#)

These give detail on the data sources, methods, and quality measures, and show a comparison with the 2018 Census distributions for each variable.

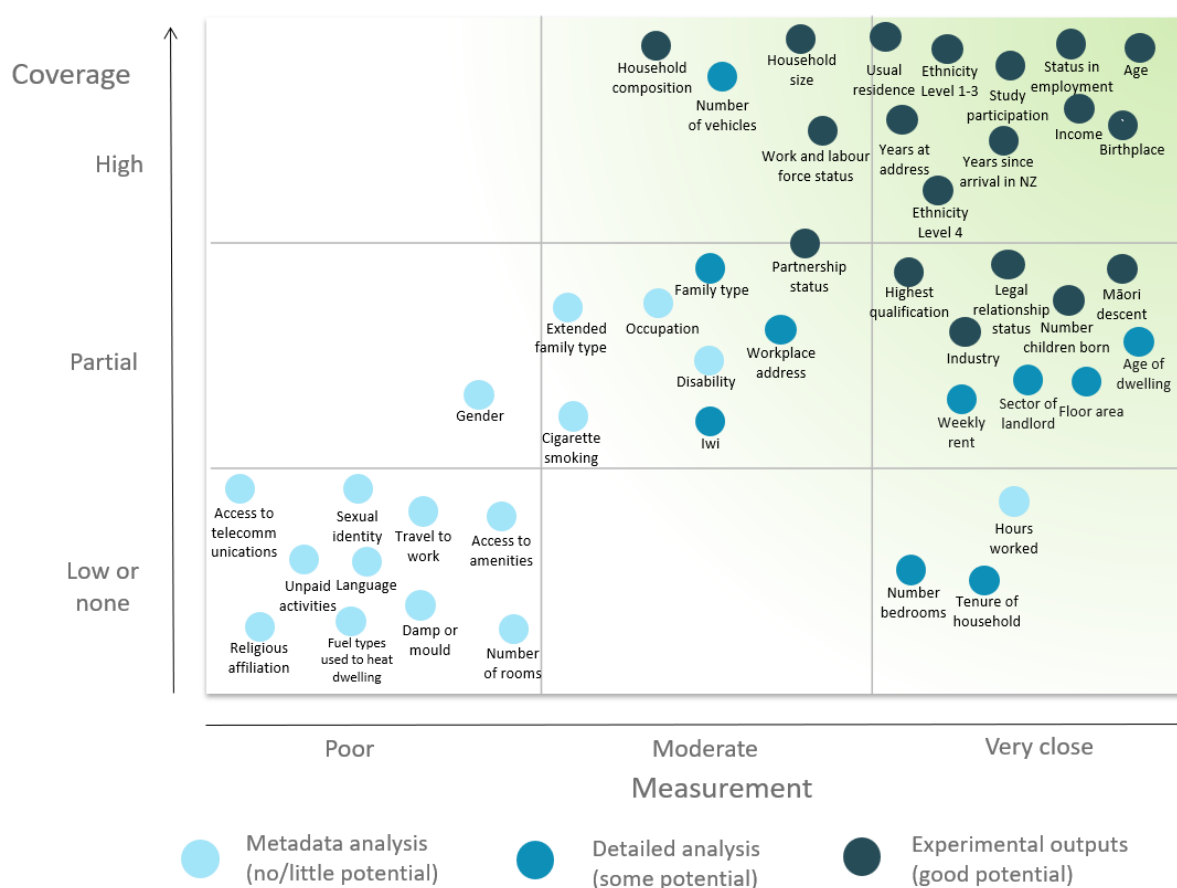
[Information for Māori about the experimental administrative population census](#) brings together information that is most directly relevant to iwi-Māori to support their use of the data and involvement in the development of the APC.

For all the variables included in the APC, admin concepts align closely with the concepts defined in statistical standards. In some cases, the admin variables are formal records of events from authoritative sources (for example, birth of a child, attainment of a qualification, payment of tax, and receipt of benefits), and the admin values are more accurate and precise than can be expected from a survey questionnaire. However, as discussed earlier, there are two important demographics (ethnicity and usual residence address) that are more likely to be inconsistently collected by agencies. Where there are multiple admin sources providing the same information, we have methods for choosing the best values and for resolving conflict, but these approaches could be developed further.

Around half of the census attributes variables for individuals can already be delivered through admin data. Figure 6 summarises to what extent admin data covers all census attributes. The placement of the bubbles is indicative, with vertical placement indicating the proportion of the population for which data is available and horizontal placement indicating the quality of the data that is available. Shading of the bubbles indicates methodological progress and availability of data. Dark blue bubbles are attributes for which work is already at a stage where they are included in experimental outputs in the APC, which allows for detailed analysis that can inform further, marginal improvements. There are some attributes for which there is documented potential, and work is ongoing to evaluate whether high-quality data can be produced. If this is deemed impossible, these concepts would need to be covered through surveys (see section below). In some cases (the light blue bubbles), no admin data is available, or the available data is not sufficient quality to produce accurate statistics. For example, we know we do not currently have sufficient admin data about New Zealanders' access to basic amenities, religious affiliation, languages spoken, housing quality, gender, sexual identity, and activity limitations (disability). For these attributes, the next census will rely on surveying. A tabular view of this information and how it compares to an original assessment (O'Byrne et al., 2014) can be found in table 3 in the appendix.

Stats NZ and census customers are familiar with errors typically encountered in a traditional census. These are mainly due to well-known non-response patterns, which decrease coverage across all variables, and variable-specific respondent or processing errors, which affect measurement. Error structures in variables derived from admin sources are different from those in the census and depend on the purpose of the agency and nature of the event or service. Some admin-derived variables achieved high coverage and were found to be highly accurate. Several variables showed good potential for providing census-type information, but there were caveats either in coverage or measurement error (Bycroft et al, 2021). We have since improved our ability to deal with some of the issues that had been identified. However, some gaps remain.

Figure 6 Quality of and potential to use admin data for deriving census attributes



One reason for lack of coverage in admin sources is the absence of digitised information from earlier periods. Use of 2013 Census as a proxy admin source has been effective in increasing the coverage for historical periods. There are still some gaps, particularly for older people who were not found in the 2013 Census. While the 2018 and 2023 Census can similarly be used to increase historical coverage further, as time goes on, historical data become less relevant as people leave the population through death or migration.

Migrants are a population group particularly affected by lower coverage at present because New Zealand admin data is limited in its ability to capture events that have occurred overseas (such as gaining a qualification, having a baby, getting married). As mentioned earlier, Māori descent information for migrants is also limited because New Zealand birth registrations are the main source.

Future work includes investigating and incorporating new data sources for known gaps (priorities are electoral roll data for Māori descent, visa applications for qualifications and family relationships, and new data sources for address information, with a focus on young adults). Statistical imputation methods for remaining missing values will be developed to improve the representativeness of the admin data.

Producing the official estimated resident population from admin data

Estimates of the New Zealand-resident population are the most critical output based on the census. Although the admin-derived resident population is close to the actual population as measured by official population statistics, statistical models are necessary to improve the representativeness of the admin data and to adjust for coverage errors. As with the current census approach, methods to adjust for systematic errors in the population listing will still be required for accuracy of official population statistics.

Population estimates are multi-source statistics, usually based on two or more partial listings of the target population. Currently, the official estimated resident population (ERP) in census years is derived using a version of dual systems estimation (DSE), which combines the census and the Post-enumeration Survey (PES), conducted soon after, but independently, of census. The PES data is linked to census data and used to estimate and correct for the undercoverage and overcoverage errors in the census using Bayesian methods (Elleouet et al, 2022; Stats NZ, 2023a). The resulting census-night ERP is further adjusted by adding residents temporarily overseas on census night and changes due to births, deaths, and migration between census night and the mid-year reference date (30 June) to produce the official 30 June base ERP. This base ERP is incrementally adjusted using demographic accounting methods to produce updated ERPs for the intercensal period until the next rebasing occurs following the subsequent census.

The potential linkage errors from linking the census and PES data are minimised by clerical reviews of links and non-links (which is possible since PES is a relatively small survey), and the remaining linkage error is assumed to be negligible and not corrected for.

How do we account for the change in census model in population estimation?

Since the current population estimation system at Stats NZ uses the census as a major input, a change to the census model has implications for the population estimation system. Developments in both administrative data, such as the development of the APC described in the preceding sections, and population-estimation methodology mean that it is likely the impact of a change in census model on population estimation can be accommodated, though with significant changes to the current system. The components of population change incorporated in the current population estimation system (particularly births, deaths, and migration) are not directly affected by changes to the census model. However, there will need to be inherent internal consistency between the population estimation system and these components of population change, as there is currently. Dynamic population models such as those being developed by the Office of National Statistics in the UK are one solution to this challenge.

[Dynamic population model for England and Wales: July 2022](#) has more information.

Research on methodology for the use of admin data for population estimation has been underway at Stats NZ for some years. Core methodology has been internationally peer reviewed and published (Graham et al, 2024). The critical inputs for point-in-time population estimation are two or more attempted listings (or captures) of the target population, which are then used via dual or multiple systems estimation (DSE/MSE) models to estimate the number of people missed by the lists (under-coverage) and their distribution over the attributes of interest, including sex, age, ethnicity, Māori descent, and area.

The current proposal being investigated is to use two (or more) administrative lists as inputs to a dual or multiple-systems estimation procedure to estimate a base ERP. Compared with the current census-based ERP, this would permit more frequent production of base ERPs, since population estimation would not be tied to a five-yearly census cycle. At the time of writing, it seems the most likely lists to use in a dual-systems set-up are a trimmed version of the APC and a list extracted from the national health index (NHI). Given the decades-long development of the latter and its 'cradle to the grave' population coverage, it is a natural candidate for consideration as a second list to use in conjunction with a trimmed version of the APC in a dual-systems estimation set-up. We are currently working on methods for trimming the NHI data of non-residents using the border-crossing data and developing health activity rules to extract an 'active-resident' group from the health data. This will have some conceptual similarity to the Ministry of Health's Health Service User Population (HSU) but with definitions tailored to our point-in-time population use case. The APC will be trimmed to further minimise the chance that non-residents are represented in the trimmed lists and to enhance the independence of the trimmed APC and health lists. We are also continuing to consider the potential of other admin sources, including the electoral roll, which may be an important source of information for the Māori descent population.

While we currently do not anticipate needing a specialist coverage survey (analogous to the PES) as part of the future population estimation system, survey data will continue to be an important input to an admin-based population estimation system. Since admin data can be prone to variation in how important variables such as ethnicity and location are recorded, an admin-based population estimation system needs to be supported by high-quality survey data to provide 'gold standard'⁵ respondent-reported data to validate admin data against and to inform the development of misclassification adjustment models and methods. We have been working on incorporating misclassification adjustment into a dual-systems estimation framework in a simplified setting, and now need to extend this work to accommodate a situation with error-prone data on both input lists. If only two lists are used as inputs to the population estimation, another important use of survey data is to check on the important conditional independence assumption usually invoked in dual-systems applications. If the assumption (inclusion on each of the lists is conditionally independent of inclusion on the other list, given covariates included in the model) is found not to hold, then a survey-based estimate of the degree of dependence can be incorporated into the dual-systems estimation, as described in (Graham et al, 2024).

Linkage error is an important consideration in using two or more admin lists for population estimation. Missed matches lead to inflated population estimates, while false links lead to under-estimates. False links can be minimised by following a conservative linking strategy whereby links are declared only if strict matching criteria are met. This strategy may increase the number of missed matches. In the current population estimation system, a conservative automated linking strategy is followed to link PES to census, followed by clerical review of non-linked records. The clerical review process adds a significant number of links to the total. This strategy is feasible because the relatively small size of PES means that only a modest number of records are passed to clerical review. Following a similar strategy in the case of two large administrative lists would lead to a much larger number of records going to resource intensive clerical review. Consequently, we are currently considering linkage error correction strategies. One option is to clerically check a sample of records not linked by a conservative automated linkage procedure and to use this clerically checked sample to estimate a missed-match rate that can be used to adjust the dual or multiple systems estimation procedure. We are also investigating a method based on Seber et al (2000), which uses multiple

⁵ It is assumed that responses received from a survey (either full or sample) carry higher-quality information, as demographic variables (like location and ethnicity) are believed to be close to their official standards and more up to date. Therefore, survey responses can be held as the 'truth', or at least the gold standard within the estimation system, and only admin-based listings require misclassification correction.

linkages to estimate the number of missed links, and an extension of this approach that makes use of a clerically reviewed sample to enhance the estimation of missed-match rate.

The next steps for testing the technical feasibility of admin-based population estimation

Applying our methods to our candidate administrative lists will inform next steps. We need to measure the extent to which errors in the admin-derived resident population can be adjusted to produce admin-based population estimates of sufficiently high quality.

Work on the extensions to the core population estimation model to account for misclassification and linkage error needs to continue. Progress on results from the latter were shared at the 65th ISI World Statistics Congress in October 2025 ([Estimating missed links between administrative data lists using dual systems estimation](#)). Error corrections for misclassification and linkage issues need to be merged and extended to account for overcoverage and missing attributes. Further, the current plan is to consider separately estimating the subnational population estimates by ethnicity and the Māori descent population.

Research to date has focused on developing models that can produce population estimates by single year of age and sex/gender, down to TALB. To meet the information requirements of official population estimates, the models will need to be extended to include additional covariates (for example, ethnicity, Māori descent, and New Zealand-born indicators). The adjustment will ideally also be extended to lower-level geographies, for example SA2. Smaller geographies, such as meshblocks, will still need to be adjusted proportionally, as is done currently within TALBs.

The role of surveys in an admin-first census model

Research completed to date, and recommendations provided to Cabinet, have highlighted that a future census model primarily based on admin data will need to be supported by surveys. In an admin-first census model, attribute information for a particular variable could be obtained either entirely from admin data, entirely from a large-scale attribute survey, or from a combination of admin data and survey data. Surveys are needed:

- to provide information that is not available in admin data
- to support information that is primarily sourced from admin data – for quality assurance or to fill gaps in admin data
- to support delivery of unbiased population estimates.

Some census attributes have limited admin data potential and in the absence of new data sources will continue to rely on survey collection. Dwelling-level census attributes are over-represented in this category compared with person-level attributes. Person-level census topics such as gender, sexual identity, languages spoken, religious affiliation, activity limitations, as well as housing characteristics such as fuel type used to heat dwelling, damp and mould, access to basic amenities, mortgage payments, and number of rooms will need to be collected through a survey for some time at least.

The role of surveys in supporting attributes that are derived from admin sources must be considered. If the admin data available is not good enough on its own for a particular attribute, or does not satisfy customer needs, then we would require the use of a survey to provide supplementary information. This includes those variables that, despite high coverage in admin data, are potentially missing information for specific groups (for example, for new migrants with overseas-

based information such as highest qualification, number of children born, and relationship information). When data is not missing at random, information from an external source such as a survey can be used to model the missing data. Survey data can also act to assess admin-data quality on a continual and sustainable basis.

To balance respondent burden alongside supporting a broad range of information needs, the Census Attributes Survey will include:

- primary content asked every year to allow comparisons over time and support granular analysis
- rotating content from 2029 onwards – covering new or more detailed topics, either across the full sample or targeted sub-groups.

Consultation with customers around what content is required will focus on the outcomes that users require and establishing how best to meet those information needs. Rather than simply continuing with what has historically been provided in census data, deeper questions can be asked about what information the census really needs to provide. For example, admin data provides very good information about those employed and not employed, but it does not cover categories such as 'Working without pay in a family farm or business' or distinguish between 'Unemployed' and 'Not in the labour force'. Depending on actual information needs, the census concept could be simplified with more detailed needs covered only as a sample by Stats NZ's Household Labour Force Survey.

Further, survey data will play a supporting role within the population estimation system. For instance, survey data could be used for checking the assumptions of our estimation system and for improving the quality of admin data (for example, correcting ethnicity and location for misclassification and measurement errors where appropriate). At this point we do not anticipate running a person-based coverage survey similar to the Post-enumeration Survey to support population estimates, due to the Census Attribute Survey providing information on misclassification. We are, however, planning the development of a Dwelling Coverage Survey to validate the quality of the dwelling estimates and the dwelling frame used to run the Census Attribute Survey.

An indicative high-level design for the new annual census survey (Census Attribute Survey or CAS) has been developed with the intention that it will provide survey estimates for individual, dwelling, and household variables that cannot be obtained from admin sources down to small geographic levels. The design work completed to date does not determine what will be collected and output in a future census, but the design does depend on specific user requirements on the level of accuracy for survey outputs.

Details of the sample design and the estimation strategy are currently being developed. The preliminary design is based on the approach used for the American Community Survey (ACS). Dwellings will be sampled from a dwelling register using an unclustered design to support flexible output geographies and small area statistics. A dwelling register is being developed as part of Stats NZ's development of the Integrated Statistical Data System. While this is established, the sample will be selected from the Statistical Location Register (SLR), which combines admin-sourced address information and enumeration from the 2023 Census to form an address register. The survey will be run continuously, in 12-month cycles from July to June.

Annual survey estimates will be supported for larger populations and areas, with data combined over multiple years to deliver information for smaller populations and areas. The planned achieved annual sample size of the Census Attribute Survey is 60,000 households. This will be reviewed following consultation, and as costing information improves through testing. The sample is designed to support basic analysis of populations between 30,000 and 50,000 with one year of data, and

populations of 5,000 to 9,000 with five years of data (300,000 households). The quality of survey estimates will vary across different measures depending on:

- the availability of related administrative data to improve estimation, and
- how similar people in the same household are to each other.

Statistical modelling techniques are likely to improve precision for key variables with high demand for small area or population data. Further work will be undertaken to determine how auxiliary data and small area estimation approaches can be used to reduce sample error and offset some of the loss in sample size compared to a full field enumeration census.

While Stats NZ has significant experience running social surveys, the attribute survey has several new features. These include the development of a multi-modal collection with a high level of self-response, selecting an unclustered sample from a dwelling frame, and developing appropriate estimation methods to support detailed small area estimates. Investment in survey methods is needed for both an attribute survey and our household survey programme more generally, in particular to reduce or mitigate increasing levels of non-response to surveys. There are limitations in how the attribute survey can support information needs for small populations, down to small areas, for census variables where direct collection is the only viable data source.

The move to an annual sample survey represents a significant shift for users of census data with outputs going from counts taken from a unit record file to estimates with sampling error. For small population groups and small-area estimates, where multiple years of survey collection are required, these estimates will also reflect responses received over a significant period. Additionally, as the census attribute survey data collection will be continuous, estimates based on a single year of data would also be thought of as pertaining to a period, rather than a 'census night' date.

Further work is required to finalise the estimation methodology to support small population and area statistics. It is likely that several approaches will be developed.

While the estimation process for multi-year estimates using pooled survey data is not complex, estimates will have reduced timeliness and can be conceptually challenging to interpret and explain. Using multi-year estimates to identify trends for small populations can also be challenging, due to the overlap in the pooled years between multi-year estimates. It is recommended when comparing multi-year estimates across time to not have periods that overlap; however, this does mean that data users will need to wait longer to identify trends. The advantages of pooling survey data is that a single dataset can support a broad range of uses and outputs, the outputs will be consistent with each other, and the pooled datasets produced are the same as standard survey datasets, which are more accessible to a wide range of users.

The alternative to pooling data is the use of statistical models to produce single-year estimates for each year. A model-based approach should be able to better utilise additional data sources to improve the quality of estimates. This approach would also remove the loss of timeliness from basing estimates directly on data collected over several years and would avoid the associated interpretation difficulties. However, a model-based estimation approach can be complex to design and run in a production setting. Unlike a survey-weighting approach, where a single set of weights is used to produce a wide range of measures, model-based estimation requires each outcome of interest to be treated separately. Further, modelled estimates also have the disadvantage that estimates produced over different margins may fail to be fully consistent. The systematic use of model-based estimation in a production setting would require significant methodological investment, but there are good foundations for this work. Model-based estimation was used in the

2018 and 2023 PES estimation processes and experimental domain estimates were produced for Te Kupenga and the NZ Disability Survey.

In summary, the shift to an admin data-first census model has implications for how estimates are published and the interpretability of such estimates. Estimates can be produced with increased frequency, though this comes at the cost of increased sampling error, for census variables where the sample survey is the main data source, and increased complexity in interpreting the results. Careful communication to users and the public of such a change is necessary.

Apart from meeting the current needs of census, one of the main opportunities of being able to measure the population more cost-effectively is that it creates options for meeting information needs that previously have not been able to be resourced. These may be concepts specific to different communities, they may make measurement of a census concept more useful (for example, asking a suite of questions on language usage), or they may be topics of great concern to public policy. These extensions of information available can be identified and traded off against the need to ensure core population measurement is sufficiently accurate.

Delivering for and with Māori, iwi, and hapū

Stats NZ recognises Te Tiriti o Waitangi as a founding document of Aotearoa New Zealand and is committed to honouring its Te Tiriti o Waitangi obligations. Delivering for and with Māori, iwi, and hapū is a strategic priority for Stats NZ. We are committed to providing accurate and relevant data for and with iwi-Māori, and to improving the data ecosystem with iwi and Māori. This goes beyond census and beyond the next census cycle, and for Stats NZ and the data system to better address the data aspirations for iwi and Māori, a comprehensive shift in elements spanning governance, ethics, methods, collection, and access will need to be considered. Greater use of partnership approaches is also likely required. For the scope of this paper, the technical feasibility of an admin-first census supported by surveys should be assessed on whether such a model can sustainably provide relevant data for iwi-Māori at least as well as the current model, and whether it ideally would improve upon that state.

Currently there is very little data that meets the variety of iwi needs, as stated in the iwi data needs paper (Te Kāhui Raraunga, 2021). This paper also stated that iwi data needs fall broadly under two roles – to achieve the aspirations of iwi and to influence the outcomes of Māori, iwi, and hapū. Iwi currently largely rely on census data to meet their data needs. However, this is problematic in light of the issues with the 2018 Census where we were unable to publish iwi data as official statistics due to insufficient data quality, with census data only being produced every five years. The response rate targets for Māori again were not met in 2023, and the iwi affiliation variable was given a ‘moderate’ quality rating overall.

[Data sources and methodology for iwi affiliation in the 2023 Census](#), released in September 2024, was co-developed and co-tested by the Data Iwi Leaders Group (Data ILG) and Stats NZ, working with Te Kāhui Raraunga Charitable Trust (the mahi arm of Data ILG). This ensured that iwi Māori data expertise was prioritised in the design of the methodology. The methodology uses alternative data sources to fill missing and residual values for the iwi affiliation variable. This builds on mahi undertaken by Data ILG and Stats NZ to release [Iwi affiliation \(estimated counts\): 2018](#) after the 2018 Census in collaboration with iwi technicians from the Data Iwi Leaders Group but not as official statistics.

Iwi affiliation is currently collected by the Ministry of Education (tertiary education data, early childhood education, and schools’ data) and the Ministry of Social Development (student loans and allowances data). While it is clear that iwi affiliation cannot be derived from these sources alone, the

work for the 2023 Census has built an understanding of the challenges and possible paths forward. The collaboration with Te Kāhui Raraunga can also serve as a model to build on in the future.

As identified earlier in this paper, the level of missing Māori descent information in the APC is currently too high. With the inclusion of electoral roll data and historical census data, we expect around 6 to 10 percent of the population to be missing Māori descent information for the next census. As with the current census model, statistical methods such as imputation will be required, but we can expect the level of undercoverage to be higher compared with the 2023 combined census model. Beyond next census, when historical census data will be less useful, additional collection of Māori descent information in admin data would be desirable to reduce reliance on DIA birth registrations and enrolments on the electoral roll. [Data sources and imputation for Māori descent in the 2023 Census \(updated\)](#), released in October 2024, was also done in collaboration with Te Kāhui Raraunga. This can also serve as a great base to continue work on Māori descent.

There are also known issues with differential reporting of ethnic group affiliation. Reporting of ethnic group affiliation is particularly sensitive and problematic, due to the reported experience of Māori having ethnicity fields completed for them in ways inconsistent with self-identification (Greaves et al, 2023). In addition, there is evidence that identification of ethnicity for some Māori may be context dependent (Greaves et al, 2022; Harris et al., 2022). There is work to do to understand the appropriate path forward for dealing with differential reporting of ethnicity, and the solutions are unlikely to be purely technical.

At present we expect that moving from a full-enumeration census model to an admin-first census supported by surveys would see both gains and losses in the availability of data for Māori, iwi, and hapū. The move to an admin-first census model can provide more frequent information about Māori populations at the same levels as the traditional five-yearly census. This timelier information would include some of the dimensions identified in Te Kāhui Raraunga (2021) as being important for iwi to assess their state of wellbeing (as in, basic demographics, location, income, education, and employment).

However, iwi would lose a nationwide update of iwi affiliation if the data system remains in its current form, as well as critical measures not available in other data sources. Examples of these measures include te reo Māori prevalence, disability measures, smoking rates, housing quality, and household tenure. While the annual census attribute survey will go some way to filling these information gaps, particularly for larger iwi, it will be unable to support detailed breakdowns for non-admin-based variables that census currently produces.

It is also important to acknowledge that the current mix of census-based population information and current social surveys does not meet iwi data needs. A model shift provides Stats NZ with an opportunity to invest in creating a data system designed with Māori, iwi, and hapū to better support their information needs. For example, an admin-first census model supported by surveys can provide more frequent information about Māori populations, which is a step towards meeting iwi data needs with more timely data beyond the five-yearly census.

Broader and meaningful engagement will be required with iwi and Māori around the appropriate re-use of admin data and the impact and opportunities of census options to better support the data aspirations of iwi and Māori.

[The Māori Data Governance Model](#) (Te Kāhui Raraunga, 2023) provides a good platform to guide how we might design and implement with a more equitable data system in mind.

Measuring key communities

The census provides population counts and measures for many diverse communities across New Zealand. Examples include small ethnic communities, rainbow communities (identified through combinations of sexual identity, sex at birth, gender, and variations in sex characteristics), iwi, and the disabled population.

We know that admin data does not always deliver to the information needs of these population groups. This is because admin data is generally collected from the perspective of service delivery without always considering the point of view or context of the individual who the data is about. While some admin-data sources provide very good information about these communities, some population groups are currently misrepresented, under-represented, or not present in admin-data sources.

A key component of the next census is the new annual Census Attribute Survey (CAS) that samples around 3 percent of the population each year. This is outlined in more detail in the section [The role of surveys in an admin-first census model](#) above. While the level of sampling is high, it will not provide the same level of detail as sampling the whole population. For population groups that are very small (say fewer than 9,000 members) the quality of data from even a large sample survey like the CAS will be poor in comparison to a full enumeration census. However, in many cases the traditional census model is also not meeting data needs.

While data collected in the census does not meet many of the information needs of these communities (Te Kāhui Rarunga (2021); Ministry for Pacific Peoples (2023)), the combination of a statistical population identifier that can be used with the broad set of census measures and linked to a rich set of admin-data sources such as health, education, and employment is extremely powerful. A move from a full-field enumeration model may mean that an admin-first census will no longer be able to reliably identify some of these groups in the data, and specific strategies will be required to ensure that a move from full enumeration does not create further data inequity.

A move from full-field enumeration also offers the potential to invest in options that may better support the information needs of these communities than the current model.

The solutions would need to be developed in partnership with community groups, but some possible strategies include:

- developing a data framework to identify and advocate for key information needs
- improving the quality of underlying admin data to develop a statistical population identifier (if possible and appropriate)
- prioritising content collected as part of the census survey programme – this could be either content collected in the Census Attribute Survey or the development of a separate survey such as Te Kupenga
- supporting the development of community data capability – an example of this is reducing barriers to enable iwi to design, collect, and make use of data collected for their individual needs.

While there will be common themes across different communities, the solutions will need to be developed in partnership, and there will be different challenges and approaches required.

A key challenge in moving from a full enumeration census will be balancing competing priorities, and an appropriate framework and governance will be required to ensure this is done well.

Below we briefly outline some issues and approaches for some key communities. Discussion on supporting iwi and Māori information needs is in the section [Delivering for and with Māori, iwi, and hapū](#).

Disability data

- Surveys will continue to be necessary to estimate the prevalence of disability in New Zealand and to understand disabled people's experiences, the level and type of support they need, and what facilitates or hinders their participation in important aspects of everyday life.
- The proposed Census Attribute Survey (CAS) will need to include the Washington Group Short Set (WGSS) questions and a question on self-identified disability or long-term condition (similar to the 2023 Census). The CAS can then produce outputs on the functional limitations covered by the six questions in the WGSS and a sample frame for a targeted disability survey, which would have a more comprehensive set of disability screening questions.⁶
- Moving from a full-field enumeration census will reduce the granularity of data that can be produced from the WGSS disability questions. Stakeholders highlighted the value of intersectional data about disability.
- The United Nations Convention on the Rights of Persons with Disabilities (UNCRPD) committee recommended that we 'develop a national disability data framework to ensure appropriate, nationally consistent measures for the collection and public reporting of disaggregated data on the full range of obligations contained in the UNCRPD'. Stats NZ has committed to lead this work in partnership with Whaikaha – Ministry for Disabled People. Disability data frameworks overseas have shown the value of combining surveys and admin data to monitor a wide range of outcomes. Developing this framework will help with prioritising information needs and can help to inform the design of our future social survey programme.

Ethnic communities

- Improvements in the quality of ethnicity data collected by key agencies to support level 4 ethnicity will be critical to an admin-first approach that does not include the use of a full enumeration survey. Without this information, Stats NZ will not be able to produce detailed population counts and summaries for these communities.
- While improvements in the collection and coding of ethnicity is likely to add significant value to the ability of different agencies to monitor outcomes within their own data system, this will take both time and significant resource by the agencies involved. Where to prioritise quality improvement will need to be identified in partnership with the agencies involved to ensure the work aligns with their data strategy. However, given the coverage of the population, the health and education sector are likely to be key areas.

Pacific data

- A move from a full-field enumeration model supported by admin data to an admin-based population census supported by surveying poses both significant risks and opportunities for and about Pacific ethnic groups.
- The biggest area of concern is that currently most agencies do not collect and code ethnicity data to level 4. This is required to provide data for many of the smaller Pacific ethnic groups.

⁶ The WGSS is an internationally standardised instrument. There is a need to explore how disability can be measured in alignment with Māori worldviews.

- In the short term, combining historical census data with admin data can reduce the impact from the lack of detailed ethnicity coding in admin data. However, while level 2 ethnicity counts are expected to be high quality, there is a risk that smaller Pacific ethnic groups included in level 4 of the classification would be under-represented in the long term, if the collection of ethnicity data by agencies does not improve.
- A move from a full-field enumeration model will also reduce the quality of the wider set of census attributes for smaller Pacific ethnic groups (for example, the proportion of people living in a damp home). Many of these populations have fewer than 2,000 people and, even with the scale of surveying intended, will not be well supported by a sample survey. Alternative options such as non-probability survey or traditional qualitative research methods may provide an effective method for delivering data for these groups without significant cost, but these options have yet to be explored.
- While a move from a full-field enumeration census model is likely to decrease the quality of data available for smaller Pacific ethnic groups, it also provides greater opportunity to deliver content that is designed for and with Pacific communities. There are a range of options for delivering new content, including the development of a Pacific Wellbeing Survey (similar to Te Kupenga), the addition of a Pacific wellbeing module to include in the census attribute survey, and the development of admin-based population measures relevant to Pacific data needs.

Rainbow data

- Rainbow communities in New Zealand are diverse, with transgender, non-binary, gay, lesbian, bisexual, takatāpui, and intersex populations among those within the rainbow umbrella. The experiences and information needs can vary between groups.
- Stats NZ has made significant progress over the last decade to improve representation for rainbow/LGBTIQ+ populations in official statistics. The 2023 Census was the first to ask everyone in Aotearoa New Zealand about their gender, sex at birth, sexual identity, and variations of sex characteristics.
- The collection of representative and complete information to represent rainbow communities is limited in admin-data settings, and may continue to be, because in many contexts it will not be necessary or appropriate to collect. Partnership with rainbow communities' experts and other agencies to identify valuable and suitable admin collections for collecting this data will be an opportunity to improve this. However, this would be a longer-term goal as it will take time and resources.
- A move away from full-field enumeration poses the highest risk for meeting rainbow/LGBTIQ+ population data needs into the future. The loss of population-level identifiers reduces the quality of population counts of rainbow communities, broken down by different characteristics and geography alongside the ability to report on attributes collected in the census and link to admin data to report on measures such as health outcomes, income, and education. The census attribute survey would provide support for measuring these outcomes, but it will not provide the same level of detail as a full-enumeration census, particularly for rainbow LGBTQI+ subpopulations.
- While a move from full-field enumeration decreases the ability to produce very detailed population counts for rainbow communities, the size of a census attribute survey provides an opportunity to collect and disaggregate a broader set of social measures, which are not able to be delivered by a five-yearly census, due to limitation on the questionnaire length, or existing social surveys, due to their smaller sample size. An example would be the production of wellbeing measures for different rainbow subpopulations.

- Delivery of content for rainbow communities needs to be represented in planning for the future census delivery. This is important across all the census models but will be particularly important if there is a move from full enumeration to ensure a change in census model does not disproportionately impact rainbow communities.
- There are a range of design choices for delivering for rainbow communities. Stats NZ is committed to working collaboratively with communities on these. Examples include prioritising information collected in a census attribute survey, the delivery of a rainbow module within the survey, or a follow-on survey collecting a wide range of content.

Migrants and refugees

- Recent immigrants will have higher levels of missing admin data for attributes that typically have high coverage in admin data (for example, ethnicity, field of study, highest qualification).
- Similar to the full enumeration model, CAS will provide nationally representative data on immigrants living in NZ.
- Data on refugees or specific migrant categories may be better supported by targeted sampling.

Homelessness and severe housing deprivation

- This is a complex and challenging area, and a lot of work beyond the technical will be required to do this well.
- Homelessness includes people living without shelter, living in temporary accommodation, sharing accommodation with another household, and living in housing that lacks basic amenities.
- The census is central to producing estimates of severe housing deprivation. Significant effort is put towards ensuring people living in a range of situations are included in the census. The census is usefully complemented by admin data for some variables (for example, income, study participation), but there is no admin data for other critical variables (for example, access to basic amenities).
- While service provision (for example, transitional housing) can provide useful specific information, homelessness and housing deprivation requires high-quality measurement, independent of service provision and access, to assess unmet need.
- It is likely that a range of strategies will be needed to measure different aspects of homelessness and housing deprivation. For example, 'sharing accommodation with another household' and 'dwelling lacking basic amenities' may be captured through household surveying, while measuring temporary accommodation would require a mix of service-provision data and sampling non-private dwellings. Measuring people living without shelter well will be challenging. These challenges exist for both full enumeration survey and ongoing sample survey models.

Building trust in the new census model

Beyond the technical there are emerging challenges when producing official statistics about people and dwellings within an 'admin-data first' framework. Stats NZ traditionally has a predominantly survey-based approach to producing official social statistics and therefore has a strong understanding of the challenges that that environment has. We will be working to build our capability in the understanding of challenges that are inherent with becoming an admin data-first organisation. From a technical point of view, we will also work to explain our methodology and the impacts of the change in census model on the data we produce, to support engagement with customers and the wider public.

Engaging with partners, customers, and communities

The benefits of the research to date are not realised if we cannot build a shared understanding of the change in census model with our Te Tiriti partners. Engagement conducted during the census transformation programme (2022 and earlier) found that overall, customers were satisfied with the current census model. While some issues exist, they are understood and can be accounted for, providing a foundation for progressing change collaboratively.

Engagement with customers and Te Tiriti partners, which has occurred since the release of the experimental APC products in 2021 and 2022, has focused on technical users (see Stats NZ, 2023c). Unsurprisingly there was a general interest in understanding the data quality delivered through admin data, with questions and concerns raised based on users' knowledge about the limitations and issues with single admin-data collections.

As we work to implement the new admin data-first model, we will continue to share information about our progress, and provide assurance to our customers.

Our engagement with customers will clearly acknowledge the trade-offs and opportunities in the quality dimensions in shifting to the new model. The information delivered by the new model will be different to that currently delivered under the full-field enumeration model. As stated earlier, there are strong opportunities to improve timeliness and the range of content delivered. However, we will also see loss of information for population characteristics that are not covered in admin-data sources. Continuing to engage with our partners, customers, and communities about these challenges will allow Stats NZ to have a greater understanding of New Zealand's data needs, and to prioritise improvements.

Public trust and confidence

As Stats NZ makes the shift to an admin data-first census, we will work to ensure we maintain the trust of the public and communities, particularly those with less trust in government and more likely to be impacted by the change in model. Transparency in how we will use people's admin data will be key, and we will ensure that the public, and in particular communities, can understand how data is being used, and the benefits of the shift.

Looking ahead

We have made significant progress in developing and testing the statistical methodologies described above. Implementation of the new admin data-first census model is advancing, and we are actively addressing any challenges that emerge along the way.

Opportunities

Quality is multi-dimensional, encompassing relevance, timeliness, accessibility, consistency, interpretability, and accuracy. The new admin-first census model offers opportunities to enhance quality across all these dimensions.

The APC enables longitudinal analysis, measurement of population flows, more frequent and timely data production, with greater flexibility well beyond a traditional census model.

Admin sources can provide more detailed and precise information than a census questionnaire. For example, admin-derived personal income is available in dollar amounts and by individual income sources. Some admin sources are also more accurate than census responses, as they record actual events, and avoid issues like recall error, misinterpretation of questions, or reluctant to answer. For example, respondents commonly report income net of tax and other payday deductions instead of gross income.

Admin data can capture information not collected in the census. For example, the census currently only records relationships between individuals at the same usual residence address whereas admin-data allow a more complete picture of household and family relationships, including those outside the household. Other examples include dwellings characteristics such as age and floor area, which relate to housing quality and are not collected in the census.

Unlike a fixed-length census questionnaire, there is no limit to the number of attributes that can be derived from admin data, enabling production of concepts tailored to users' needs such as variations of employment status over the previous year.

Admin data can also offer flexible in responding to new information needs. Once integrated into the wider statistical system, new variables can be derived from existing sources without designing and testing new questions or waiting for the next five-yearly census.

Sustaining admin-data quality

Data quality, and balancing trade-offs between different dimensions of quality, is always considered carefully within official statistics, regardless of the source of the data. However, the data quality challenges that present when using admin data differ to those that present when sourcing data through surveys. As an organisation we are increasingly becoming familiar with the sources of uncertainty and error in admin data. As the source of admin data is typically from another government department, and collected for administrative purposes rather than statistical, we have historically had little influence over how data is collected.

The Data and Statistics Act, 2022 gives the Government Statistician the mandate to improve the quality of key admin data across the government data system. This mandate, supported by the strategic priority to proactively improve the health of Aotearoa's data system will help to reduce the prevalence of some of the data-quality challenges seen within our investigations to date. However, it can be assumed that quality concerns will remain for some time and in some instances be persistent. For example, if information is not important or necessary for an administrative purpose, data may

stop being collected, may be collected for a slightly different concept, or how data is collected may change, which can all lead to poor statistical quality.

As an organisation we will continue to build and maintain our capability and understanding of the data-quality challenges of using admin data and ensure that we adapt our processes and procedures to support the production of high-quality admin-based population statistics.

While much admin data is already available within the IDI, and some property-level admin data is linked to the Statistical Location Register, further improvements parallel to technical ones are required to improve the quality of admin data for use in producing census-like statistics. This includes:

- extending permission for new data sources acquired for 2023 Census to more general research and census purposes – examples include electoral roll data and electricity data, among others
- obtaining new data sources to improve data quality and fill gaps for certain populations – examples include MBIE visa application data for information about migrants not currently available to Stats NZ and asking tertiary institutions for addresses for tertiary students
- initiating conversations with existing data suppliers to improve their data collection – for example, the district valuation roll (DVR) where there is currently inconsistency with councils collecting information on the number of bedrooms from building consents
- removing barriers that prevent integrating data held within the IDI with data within Stats NZ enterprise systems.

The continued development of data and analytical infrastructure to support the use of integrated admin and survey data in a production setting is critical. The Integrated Data Infrastructure (IDI) and Statistical Location Register (SLR) have provided an invaluable test environment for research, but this environment is not reflective of an integrated register-based statistical system, which an admin-first census supported by surveys requires.

Progressing our methodology

Delivering the next admin-first census represents a significant change, but we have made substantial progress in establishing the necessary technical and methodological foundations and are continually refining our approaches.

We have advanced the development of an admin-derived resident population and increasingly leveraged admin data to improve the quality of population statistics. Both the 2018 and 2023 censuses were already underpinned by the use of admin data, demonstrating its potential. Work to develop infrastructure and methods for producing dwellings and household estimates from admin data is ongoing. While underutilised data exist that could enhance admin-based dwellings counts and attributes, these sources are still being assessed and tested.

Admin data offer considerable promise for a range of census attributes. For some variables, however, challenges remain, including improvements required in ethnicity data collection, gaps in Māori descent coverage, a lack of iwi affiliation information, and data to represent rainbow and disabled communities. Many attributes can be produced directly from admin data, and others could be derived to a low geographical level via sample surveys, though some loss of detail for small populations is expected.

We recognise that admin data does not meet the diverse information needs of all our customers, including those of iwi and Māori, Pacific, rainbow, and disabled communities. A complementary

survey programme is needed to fill gaps and ensure equitable coverage. Broader engagement with customers and communities will also be essential to build trust and confidence in the use of admin data for official statistics.

Progress since initial publication

Since the initial paper in September 2024, Stats NZ has made significant methodological and infrastructure progress. We have accelerated data acquisition of new datasets, conducted rigorous evaluation of their usefulness for coverage and quality, and provided evidence to support strategic prioritisation of ingestion efforts.

Stats NZ has worked closely with a range of government departments and Crown entities to develop a cross-government plan, which will go to Cabinet in 2026. The plan includes a list of key census variables, admin data sources, and agency-specific commitments for the collection and supply of admin data to be delivered to agreed standards and timeframes. Initial efforts focus on, but are not limited to, improvements in location, detailed ethnicity, gender, and Māori descent.

Extensive technical analyses have supported [Future Census Independent Evaluation Panel](#) deliberations, the Government Statistician's decision on the future census model, the subsequent Cabinet process, and public transparency. Some of this information has already been shared publicly and we will continue to transparently make information available. For example, providing admin-based resident population counts: [Reviewing the quality of ethnicity data from administrative sources: Comparison with the census](#) was published in 2024; and [Quality of ethnicity data in the experimental administrative population census \(APC\): High level summary](#) and [Quality of geographic information in the experimental administrative population census \(APC\)](#) were both published in 2025. Collectively, we have strengthened our confidence in the proposed new census approach.

Dual System Estimation (DSE) using the APC and Census 2023 data aligns closely to the ERP estimates, providing confidence in the validity of the methodology. The next step is to implement DSE using two large administrative lists in preparation for production testing. In parallel, we have deepened our understanding of the impact of linkage error and misclassification within administrative data. While we are exploring the extent to which modelling can mitigate these issues, we also presented our methodological findings [Estimating missed links between administrative data lists using dual systems estimation](#) at the 65th ISI World Statistics Congress in 2025, and continue to investigate where system-level solutions may be needed.

Progress has also accelerated in household and family research. We are improving methods for deriving household and family constructs by incorporating additional datasets, exploring the use of graph and network-based modelling to better infer relationships, and refining address-assignment modelling at both the individual and household level.

The Places Index proof-of-concept phase – initiated in 2024 to link property, address, building, and related datasets into a listing of dwellings – is now nearing completion, covering around 70 percent of dwellings and producing broadly applicable methodologies. A dwelling strategy has been developed to guide the future direction of this area of work. The next step of production implementation is to link 2023 Census dwelling data with the existing linked admin data sources to enhance information, refine dwelling identification approach, and derive additional attributes.

Finally, engagement has been an increasing part of our work since the announcement in June 2025 about plans to modernise the census. In response to feedback from public consultation, academic stakeholders, and Māori, Pacific, and other community partners, we have endeavoured to increase transparency around our methodological development through publications, conference

presentations, responding to media requests, targeted discussions, and dedicated community engagement opportunities. These interactions have helped clarify misconceptions, address concerns, and refine our prioritisation of resources. We continue to expand our international collaboration to ensure alignment with international best practice while sharing New Zealand's experience to support global developments in admin-based population statistics.

Overall, we are on track with key improvements, building on strong foundations. While admin data continues to show promise for producing timely population measures, we are actively developing statistical methods, targeted surveys, and data acquisition strategies to address the remaining gaps.

Next steps

Our methodological roadmap will continue to evolve as we progress towards an admin-first census. Guided by this year's learning, our immediate focus is on strengthening core methods, infrastructure, and data acquisition to support the transition.

A key priority is supporting the transitional outputs planned for 2026 and 2027. The design and production of these outputs, along with user feedback, will inform any refinements needed ahead of 2028.

Significant work remains for households and dwellings. Establishing a reliable approach to identifying and enumerating dwellings is essential before we can expand into attribute development and new products. Where admin data cannot meet information needs, surveys – in particular CAS – will play a complementary role, albeit at a less granular level than a unit-record file.

We will continue the development of the infrastructure and data foundations needed for an admin-first model, including progression of the ISDS and the Places Index. This includes building a robust listing of dwellings, improving methods for classifying and maintaining dwelling records, and designing a dwelling-coverage survey to support quality assurance.

For household and families, we will work to refine methods for assigning membership, set quality targets, develop unit-record quality metrics, and determine how surveys can best supplement admin data.

We will evaluate new data sources to improve coverage for priority groups, such as ethnicity details, strengthen geographic accuracy, and work with Te Tiriti partners to address gaps such as iwi affiliation and Māori descent. Work will also continue on developing statistical solutions such as imputation, misclassification, linkage error correction, and population estimation methods suitable for admin-only production, including at small geographical levels.

Where admin data cannot meet information needs, particularly for small population and communities, we will work to develop tailored solutions that integrate survey and admin data in partnership.

Engagement will remain an increasing part of our mahi. We will work with census users, iwi and Māori partners, Pacific and other communities, and technical experts to ensure confidence in the new approach and to refine priorities as evidence accumulates.

As we deepen our understanding of the strengths and limitations of admin data, we will be better positioned to determine what system-level solutions are needed and where targeted collection is required to manage population measurement risks on an ongoing basis.

References

- Black (2016). [The IDI prototype spine's creation and coverage](#). (Statistics New Zealand Working Paper No 16–03) Retrieved from www.stats.govt.nz.
- Bryant, J, Dunstan, K, Graham, P, Matheson-Dunning, N, Shrosbee, E, & Speirs, R (2016). [Measuring uncertainty in the 2013-base estimated resident population](#). (Statistics New Zealand Working Paper No 16–04). Retrieved from www.stats.govt.nz.
- Bycroft, C, Miller, S, Gath, M, Matheson-Dunning, B, Simpson, K, & Das, S (2021). [The quality of administrative data for census variables: Strengths, limitations, and opportunities](#). Retrieved from www.stats.govt.nz.
- Cutting, C, Alspach, M, Cowell, S, Judd M, McBeth, S, & Page, M (2024). [A register-based statistical system in New Zealand: Progress and opportunities](#). *Statistical Journal of the IAOS*. 40:19–24.
- Elleouet, J S, Graham, P, Kondratev, N, Morgan, A K, & Green, R M (2022). 'Small domain estimation of census coverage – a case study in Bayesian analysis of complex survey data'. *Journal of Official Statistics*, 38(3), 767–792.
- Gath, M, & Bycroft, C (2018). [The potential for linked administrative data to provide household and family information](#). Retrieved from www.stats.govt.nz.
- Gibb, S J, & Das, S (2015). [Quality of geographic information in the Integrated Data Infrastructure](#). Retrieved from www.stats.govt.nz.
- Gibb, S, Bycroft, C, & Matheson-Dunning, N (2016). [Identifying the New Zealand resident population in the Integrated Data Infrastructure](#). Retrieved from www.stats.govt.nz.
- Gleisner, F, Downey, A, & McNally, J (2015). [Enduring census information requirements for and about Māori](#). Retrieved from www.stats.govt.nz.
- Graham, P, Varn, L Hendtlass, M, Green, R, & Richens, A (2024). [Bayesian dual systems population estimation for small domains](#). *Statistics Surveys*. 18: 1–81.
- Greaves, L M, Latimer, C L, Li, E, Hamley, L, Renfrew, L, Sporle, A, & Milne, B (2022). *Who are the Māori "in-between"? Indigenous diversity and inequity across descent, ethnicity, and iwi knowledge*. *Ethnic and Racial Studies* 46 (1), 166–189. Available from www.tandfonline.com.
- Greaves, L M, Lindsay, C, L, Latimer, C, Muriwai, E, Moore, C, Li, E, Sporle, A, Clark, T C, & Milne, B J (2023). 'Māori and the Integrated Data Infrastructure: an assessment of the data system and suggestions to realise Māori data aspirations' [te Māori me te Integrated Data Infrastructure: he aromatawai i te pūnaha raraunga me ngā marohitanga e poipoia ai ngā wawata raraunga Māori]. *Journal of the Royal Society of New Zealand*: 1–17. Available from www.tandfonline.com.
- Harris, R B, Paine, S J, Atkinson, J, Robson, B, King, P T, Randle, J, Mizdrak, A, & McLead, M (2022). 'We still don't count: The under-counting and under-representation of Māori in health and disability sector data'. *The New Zealand Medical Journal* 135 (1567), 54–78. Available from journal.nzma.org.nz.
- McNally, J, & Bycroft, C (2015). [Quality standards for population statistics: Accuracy requirements for future census models](#). Retrieved from www.stats.govt.nz.

Ministry for Pacific Peoples (2023). [Improving Pacific Data Equity: Opportunities to Enhance the Future of Pacific Wellbeing](https://www.mpp.govt.nz/programmes/long-term-insights-briefing/). Retrieved from <https://www.mpp.govt.nz/programmes/long-term-insights-briefing/>.

O'Byrne, E, Bycroft, C, & Gibb, S (2014). [An initial investigation into the potential for administrative data to provide census long-form information: Census Transformation programme](https://www.stats.govt.nz/). Retrieved from www.stats.govt.nz.

Olsen, B (2025). [The Census has already changed – we're going to need to live with it](https://www.infometrics.co.nz/). Retrieved from www.infometrics.co.nz.

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

Seber, G A, Huakau, J T, Simmons, D (2000). *Capture-recapture, epidemiology, and list mismatches: two lists*. Biometrics.56(4):1227–32. <https://doi.org/10.1111/j.0006-341X.2000.01227.x>.

Smith, P, van der Heijden, P, G, M, Cruyff, M, Pantalone, F, Diener, H, & Dunstan, K (2025). [Population size estimation using covariates having missing values and measurement error: estimating ethnic group sizes in New Zealand](https://www.austlii.edu.au/au/other/dfat/special/australian-new-zealand-journal-of-statistics/). Australian & New Zealand Journal of Statistics.

Stats NZ (2012). [Transforming the New Zealand Census of Population and Dwellings: Issues, options, and strategy](https://www.stats.govt.nz/). Retrieved from www.stats.govt.nz.

Stats NZ (2014). [An overview of progress on the potential use of administrative data for census information in New Zealand](https://www.stats.govt.nz/). Retrieved from www.stats.govt.nz.

Stats NZ (2015). [Census transformation – A promising future – Cabinet paper \(redacted\)](https://www.stats.govt.nz/). Retrieved from www.stats.govt.nz.

Stats NZ (2016a). [Census transformation progress update: realising the potential of government data](https://www.stats.govt.nz/). Retrieved from www.stats.govt.nz.

Stats NZ (2016b). [Experimental population estimates from linked administrative data: methods and results](https://www.stats.govt.nz/). Retrieved from www.stats.govt.nz.

Stats NZ (2016c). [Guide to reporting on administrative data quality](https://www.stats.govt.nz/). Retrieved from www.stats.govt.nz.

Stats NZ (2017a). [Experimental population estimates from linked administrative data: 2017 release](https://www.stats.govt.nz/). Retrieved from www.stats.govt.nz.

Stats NZ (2017b). [Experimental dwelling estimates](https://www.stats.govt.nz/). Retrieved from www.stats.govt.nz.

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

Stats NZ (2019a). [Overview of statistical methods for adding admin records to the 2018 Census dataset](https://www.stats.govt.nz/). Retrieved from www.stats.govt.nz.

Stats NZ (2019b). [Data sources, editing, and imputation in the 2018 Census](https://www.stats.govt.nz/). Retrieved from www.stats.govt.nz.

Stats NZ (2021). [Experimental administrative population census: Data sources and methods | Stats NZ](https://www.stats.govt.nz/experimental-administrative-population-census-data-sources-and-methods). Retrieved from www.stats.govt.nz.

Stats NZ (2022). [Experimental administrative population census: Data sources, methods, and quality \(second iteration\)](https://www.stats.govt.nz/experimental-administrative-population-census-data-sources-methods-and-quality-second-iteration). Retrieved from www.stats.govt.nz.

Stats NZ (2023a). [2023 Post-enumeration Survey: Standard design for coverage estimation](https://www.stats.govt.nz/2023-post-enumeration-survey-standard-design-for-coverage-estimation). Retrieved from www.stats.govt.nz.

Stats NZ (2023b). [Experimental administrative population census \(third iteration\): Data sources, methods, and quality for household information](https://www.stats.govt.nz/experimental-administrative-population-census-third-iteration-data-sources-methods-and-quality-for-household-information). Retrieved from www.stats.govt.nz.

Stats NZ (2023c). [Experimental administrative population census \(third iteration\): Changes to data, methods, and quality](https://www.stats.govt.nz/experimental-administrative-population-census-third-iteration-changes-to-data-methods-and-quality). Retrieved from www.stats.govt.nz.

Stats NZ (2024a). [Methodology for using admin data to count people in the 2023 Census](https://www.stats.govt.nz/methodology-for-using-admin-data-to-count-people-in-the-2023-census). Retrieved from www.stats.govt.nz.

Stats NZ (2024b). [Modernising our approach to the 2028 Census](https://www.stats.govt.nz/modernising-our-approach-to-the-2028-census). Retrieved from www.stats.govt.nz.

Stats NZ (2024c). [Data sources and imputation for Māori descent in the 2023 Census](https://www.stats.govt.nz/data-sources-and-imputation-for-maori-descent-in-the-2023-census). Retrieved from www.stats.govt.nz.

Stats NZ (2024d). [Predicting usual residence address from admin data in the 2023 Census](https://www.stats.govt.nz/predicting-usual-residence-address-from-admin-data-in-the-2023-census). Retrieved from www.stats.govt.nz.

Stats NZ (2024e). [Data sources and methodology for iwi affiliation in the 2023 Census](https://www.stats.govt.nz/data-sources-and-methodology-for-iwi-affiliation-in-the-2023-census). Retrieved from www.stats.govt.nz.

Stats NZ (2025a). [Reviewing the quality of ethnicity data from administrative sources: Comparison with the census](https://www.stats.govt.nz/reviewing-the-quality-of-ethnicity-data-from-administrative-sources-comparison-with-the-census). Retrieved from www.stats.govt.nz.

Stats NZ (2025b). [Assessing the quality of partnership status attributes from admin data](https://www.stats.govt.nz/assessing-the-quality-of-partnership-status-attributes-from-admin-data). Retrieved from www.stats.govt.nz.

Stats NZ (2025c). [Quality of ethnicity data in the experimental administrative population census \(APC\): High level summary](https://www.stats.govt.nz/quality-of-ethnicity-data-in-the-experimental-administrative-population-census-apc-high-level-summary). Retrieved from www.stats.govt.nz.

Stats NZ (2025d). [Quality of geographic information in the experimental administrative population census \(APC\)](https://www.stats.govt.nz/quality-of-geographic-information-in-the-experimental-administrative-population-census-apc). Retrieved from www.stats.govt.nz.

Te Kāhui Rarunga (2021). [Iwi Data Needs](https://www.kahuiraraunga.io/iwi-data-needs). Available from www.kahuiraraunga.io.

Te Kāhui Rarunga (2023). Māori data governance model. Available from www.kahuiraraunga.io/maoridatagovernance.

Toitū Te Whenua Land Information New Zealand (2019). [Property Data Management Framework: A model for connecting New Zealand's property data](https://www.linz.govt.nz/property-data-management-framework-a-model-for-connecting-new-zealand-s-property-data). Retrieved from www.linz.govt.nz.

Appendix

Table 2 Quality targets for population estimation

Quality targets for population estimation		
Geographic area	Population unit	Quality standard
National	Total population	The national total population estimate is within 0.5 percent relative error
	By 5-year age groups (0–4, 5–9, ... 85+) by sex	90 percent of all national population estimates by five-year age group and sex are within 1.5 percent relative error, and all are within 5 percent
National Māori ethnic group	Māori total population	The Māori national total population estimate is within 0.5 percent relative error
	Māori population by 5-year age groups (0–4, 5–9, ... 85+) by sex	90 percent of all national Māori population estimates by five-year age group and sex are within 2.5 percent relative error, and all are within 5 percent
TALB (population size greater than or equal to 100,000)	Total population	All total population estimates are within 2.5 percent relative error
	By broad age group (0–14, 15–24, 25–64, 65+)	All broad age group estimates are within 5.0 percent relative error
TALB (population size less than 100,000)	Total population	85 percent of total population estimates are within 2.5 percent relative error, and all are within 5.0 percent
	By broad age group (0–14, 15–24, 25–64, 65+)	All broad age group estimates are within 12.5 percent relative error
SA2 ¹ (population of 500 or more)	Total population	80 percent of total population estimates are within 5 percent relative error, and all are within 10 percent
1 At the time the quality standards were developed this standard referred to area units, which have since been replaced by SA2s (Statistical Area Level 2). Note Targets are as set in McNally & Bycroft (2015) Source: Stats NZ		

Table 3 below is an updated version of table 4, ‘Likelihood of a census topic being satisfied with administrative data sources’ (O’Byrne et al., 2014). This updated (2024) version was part of a report made available to IDI researchers in June 2025. Census concepts were assessed according to six quality measures: relevance, accuracy of coverage, accuracy of linkage, timeliness, accessibility, and consistency. These were then combined to determine the overall likelihood that available administrative data sources could satisfy a census topic. This was done by assigning one of three overall scores: ‘likely’, ‘possible’, or ‘unlikely’. The resulting assessments are:

- Likely: A census topic is ‘likely’ to be satisfied with administrative sources if one or more datasets combined are ‘good’ or ‘excellent’ for all six quality measures.
- Possible: It is ‘possible’ a census topic may be satisfied with administrative data if it relies on administrative data sources that would be suitable under certain conditions; for example, if certain changes happened (either changes to administrative collections, or changes to user needs), or modelling proved to be an acceptable approach to addressing quality issues. In some cases, these census topics may be ‘partly’ satisfied. A census topic with a ‘possible’ assessment usually has a mix of ‘excellent’, ‘good’, and ‘good to poor’ or ‘unknown’ quality measures.
- Unlikely: A census topic is ‘unlikely’ to be satisfied with administrative sources if the assessment of any quality measure for a census topic is poor, and there is little likelihood of improvement.

While attempts have been made to apply the quality measures in a consistent way across administrative data sources and topics, sometimes determinations have had a subjective element, particularly where an aspect of a data source was unknown.

These assessments are indicative only, using the information available at the time. They are not to be considered final verdicts on whether administrative sources could satisfy a particular census topic. Especially for attributes not currently in the APC, such as occupation and hours worked, more in-depth analyses would be needed to confirm any assessments.

Table 3 Likelihood of a census topic being derived from administrative data sources, updated for 2024

Likelihood of a census topic being derived from administrative data sources, updated for 2024				
Subject category	2023 Census priority	Census topic	Assessment 2014	Assessment 2024
Population structure				
	1	Population	Likely	Likely
	1	Age	Likely	Likely
	1	Sex (mixed concept)	Likely	Likely
	1	Gender	N/A	Possible ¹
	2	Legally registered relationship status	Unlikely	Likely

2	Partnership status in current relationship	Unlikely	Possible
3	Variations of sex characteristics	N/A	Unlikely
3	Number of children born	Unlikely	Likely
Location			
1	Usual residence address	Out of scope	Likely
2	Years at usual residence	Out of scope	Likely
2	Usual residence 1 year ago	Out of scope	Likely
2	Usual residence 5 years ago	Out of scope	Likely
Ethnicity, culture, and identity			
1	Ethnicity	Likely	Likely
1	Māori descent	Likely	Likely
1	Iwi affiliation	Possible	Unlikely ²
2	Birthplace	Likely	Likely
2	Languages spoken	Unlikely	Unlikely
3	Religious affiliation	Unlikely	Unlikely
3	Years since arrival in NZ	Likely	Likely
3	Sexual identity	N/A	Unlikely
Education and training			
2	Highest secondary school qualification	Possible	Likely
2	Highest qualification	Unlikely	Likely
2	Post-school qualification level of attainment	Unlikely	Likely

2	Post-school qualification field of study	Unlikely	Likely
2	Study participation	Likely	Likely
Income			
2	Sources of personal income	Likely	Likely
2	Total personal income	Likely	Likely
Work			
2	Status in employment	Possible	Possible ³
2	Work and labour force status	Unlikely	Unlikely/possible ⁴
2	Hours worked in employment per week	Unlikely	Unlikely
3	Industry	Likely	Likely
3	Sector of ownership	Likely	Likely
2	Occupation	Unlikely	Possible
3	Unpaid activities	Unlikely	Unlikely
Families and households			
2	Family type	Unlikely	Likely
2	Household composition	Unlikely	Likely
2	Extended family type	Unlikely	Possible
2	Child dependency status	Out of scope	Possible
Housing			
3	Access to telecommunication systems	Unlikely	Unlikely

1	Dwelling occupancy status	Out of scope	Possible
2	Dwelling type	Unlikely	Likely/ possible ⁵
3	Access to basic amenities	N/A	Unlikely
3	Dwelling dampness	N/A	Unlikely
3	Dwelling mould	N/A	Unlikely
3	Individual home ownership	Unlikely	Unlikely ⁶
3	Main types of heating and fuel types used to heat dwellings	Unlikely	Unlikely
3	Number of rooms	Unlikely	Unlikely
3	Number of bedrooms	Unlikely	Possible
2	Sector of landlord	Possible	Likely
2	Tenure of household	Unlikely	Likely ⁷
2	Weekly rent paid by household	Possible	Possible ⁸
Transport			
2	Main means of travel to work	Unlikely	Unlikely
2	Educational institution address	Out of scope	Likely
2	Main means of travel to education	Out of scope	Unlikely
3	Number of motor vehicles	Possible	Possible/ Likely ⁹
2	Workplace address	Possible	Likely ¹⁰
Health			
3	Activity limitations/disability	Unlikely	Unlikely
3	Cigarette smoking behaviour	Unlikely	Unlikely

N/A = not applicable because this variable was not a variable in the 2013 Census.

1. Dependent on all data sources identified collecting this information according to the standard. See above sections for more detail.

2. See above sections for more detail.

3. We can derive whether someone is a 'paid employee' or 'self-employed'. It might be possible to derive 'employer' from IRD data. It is unlikely we can derive 'self-employed and without employees' and 'unpaid family worker' from admin data.

4. It would be unlikely that 'employed full-time' and 'employed part-time' could be derived. 'Unemployed' might be able to be derived, but we would not be able to get the nuances of those 'not in the labour force' such as those seeking work, doing unpaid housework, etc.

5. Dwelling type is listed as likely/possible to be derived. This is based on Places Index design work.

6. It might be possible to derive individual home ownership from ISDS data that links individual and dwelling information.

7. We could probably derive whether a dwelling is 'owned or partly owned', but may have difficulty deriving whether a dwelling is 'not owned and not held in a family trust', or whether 'dwelling held in a family trust'.

8. The IDI has this information for some rentals, if they are in the DBH data, but not all.

9. It is possible to get motor vehicles by address and individuals, although this does not quite align with the census classification.

10. Can derive this if linking the IR EMS data to the business register. Geographic location might not be accurate for some people; for example, people who work from home.

Source: Stats NZ