Evaluation of the accuracy of Stats NZ population estimates and projections, 1996–2018

Citation
Stats NZ (2023). *Evaluation of the accuracy of Stats NZ population estimates and projections, 1996–2018*. Retrieved from www.stats.govt.nz.

ISBN 978-1-99-104949-0 (online)

Published in October 2023 by

Stats NZ Tatauranga Aotearoa
Wellington, New Zealand

Acknowledgement
The report was authored by Stephen Salzano, Cat Anderson, Helen He, and Kim Dunstan of the Population Insights unit – Ngā Māramatanga Taupori.

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

List of tables and figures

List of tables

List of figures

Purpose and overview

Purpose

This report is designed to help customers understand the accuracy of Stats NZ's population estimates and projections relative to observed populations, the reasons for inaccuracies, and discusses current developments that may improve future accuracy. It is an update of the more comprehensive 2016 report How accurate are population estimates and projections? An evaluation of Statistics New Zealand population estimates and projections, 1996–2013 (Stats NZ, 2016), with the update incorporating results obtained during the 2013-2018 intercensal period. This report evaluates accuracy at the national, regional council, territorial authority, and Auckland local board area level. The report focuses on the accuracy of estimates and projections of the total population produced between 2013 and 2018, with the accuracy of earlier estimates and projections as used in the 2016 report included to provide historical context.

Overview

How we produce estimates and projections

Stats NZ produces population estimates to indicate how our population is changing, typically after, and in between, a Census of Population and Dwellings. Our estimates are produced using the cohort component method which recognises that, fundamentally, only three factors can change the population: births (fertility), deaths (mortality), and migration:

$$P_t = P_{t-1} + B - D + NM$$

where P_t is the population at the end of a period, P_{t-1} is the population at the start of the period, B is the number of (live) births during the period, D is the number of deaths during the period, and NM is net migration (arrivals – departures) during the period. Inherent in the method is that each age-sex group ages by one year every year.

We also produce population projections to indicate how our population may change in the future, to inform planning and decision-making. Our projections are calculated using the same model as we use for estimates, except that assumptions are necessary about future births, deaths, and migration (or rates thereof).

Demographic projections are not predictions or forecasts; rather, they are an indication of what the population will look like in the future if the stated assumptions apply over the projection period, and given current policy settings. Typically, uncertainty is conveyed by estimated probabilities (stochastic projections) or alternative scenarios (deterministic projections).

Why we measure historical accuracy

Customers are generally interested in the most up-to-date estimates and projections. However, many also seek to understand how accurate past estimates and projections have been, partly to understand how accurate current estimates and projections may be. Assessing past accuracy can be useful in indicating the inherent uncertainty in estimates and projections, and how that uncertainty varies across time, geographic areas, and age groups.

Producers of estimates and projections focus on the ongoing production of quality, timely, and cost-effective information. However, assessing past accuracy can be valuable to understand the strengths and limitations of these methods, and of the estimates and projections themselves across different dimensions.

Quality dimensions

Most national statistical organisations provide guidelines and discussions of quality dimensions as they relate to statistics (eg, Australian Bureau of Statistics, 2009; Office of National Statistics, 2013; Statistics Canada, 2002; Stats NZ, 2007). These quality dimensions provide a useful framework for evaluating the usefulness of population estimates and projections.

- **Accuracy** – How do the estimated and projected trends compare with what is observed? Do the estimates and projections adequately illustrate changing demographic patterns?
- **Relevance** – Do the estimates and projections cover the necessary geographic areas, demographic characteristics (eg, age, sex, ethnicity), and time periods required by different customers? Are the estimates and projections produced to satisfy the expectations and aspirations of individuals or groups, or are they based on an objective assessment of demographic trends?
- **Timeliness** – Are the estimates and projections updated and available when they are needed?
- **Coherence** – Is the choice of methods, data, and assumptions consistent with accepted practices, and do they account for relevant factors? Are estimate and projection results plausible, given known constraints and limitations?
- **Accessibility** – Is the information readily available to everyone? Are there costs to access it?
- **Interpretability** – Is information about estimates and projections (eg, methods, assumptions, results) available, understandable, and even replicable? Do the estimates and projections provide measures of uncertainty?

Despite these diverse criteria, it is the accuracy of estimates and projections that is most often questioned and is the focus of this report.

Changes since the previous report

This report builds on and consolidates results published in How accurate are population estimates and projections? An evaluation of Statistics New Zealand population estimates and projections, 1996–2013 (Stats NZ, 2016), which itself was an update of *How accurate are projections? An evaluation of Statistics New Zealand population projections, 1991–2006* (Stats NZ, 2008). This updated report extends the analysis to include the latest 2013–2018 intercensal period.

For extensive information concerning the accuracy of population estimates and projections between 1991 and 2013, readers should refer to the 2016 report above. This includes evaluation of the accuracy of estimates and projections by age group, and at the small area level (area units).

Only a summary analysis is presented here on the accuracy of these population estimates and projections. Evaluations of other Stats NZ demographic estimates and projections – ethnic populations, families, households, and labour force – are not included. Their exclusion is because measuring these groups is further complicated by non-demographic factors, such as changes in ethnic definition, identification, and measurement.

Estimates and projections are always derived and published for the geographic boundaries existing at the time of publication. However, to help customers understand accuracy for current areas, this evaluation of accuracy focuses on the latest geographic boundaries. That is, the boundaries existing at 1 January 2023, updated from the boundaries used for the 2016 report, which were those existing at 1 January 2016 (Stats NZ, 2016). This has necessitated making some adjustments to the previous estimates/projections for boundary change between 2016 and 2023.

Report outline: a guide for readers

In helping to understand the accuracy of Stats NZ's estimates and projections, the following two main questions and their sub-questions are covered:

1. How accurate have past estimates and projections been?
    a. Is estimation and projection accuracy improving over time? Is past accuracy any indication of future accuracy?
    b. Do updated projections within an intercensal period improve on an initial set?
    c. What types of geographic areas are better/worse estimated and projected?
2. Which components of population change (births, deaths, net migration) contribute most to inaccuracy?
    a. What methodological changes might be considered as a result of this evaluation?

The report has three chapters.

- Chapter 2 (**Measures of accuracy**) briefly outlines the different measures of accuracy used in this report and possible accuracy standards for evaluation.
- Chapter 3 (**Our findings**) outlines the results of the analysis and provides answers to the first question and its sub-questions.
- Chapter 4 (**Discussion and future work**) focuses on the second question and what the next steps to improve accuracy may be.

Given the technical nature of the report, a **Glossary and abbreviations** section is provided.

See **Accuracy of estimated and projected populations 1996–2018** (interactive Excel file) under 'Available files'. It compares estimated (and projected) population change with observed population change nationally and for all individual regional council, territorial authority, and Auckland local board areas.

Summary of key findings

- Estimates and projections for larger geographic areas have smaller relative errors than estimates and projections for smaller geographic areas. Estimate and projection uncertainty therefore increases as geographic size decreases.
- For both estimates and projections, areas with the largest relative errors are generally those that have experienced the most rapid population change (either increase or decrease) and/or those that have experienced significant swings in net migration, with the pattern most pronounced for projections (over multiple periods).
- Most areas were under-estimated and under-projected due to the high net migration gains of the 2013–2018 period. This applied nationally and to most regional council areas and territorial authority areas.
- It is difficult to conclude whether estimate and projection accuracy is increasing for more recent estimates and projections. This is partly because accuracy is partly a function of the demographic variability of different periods, and partly because short-term accuracy may differ from long-term accuracy (in the case of projections).
- The range between the low, median/medium, and high projections was arguably too narrow, mainly reflecting the narrow range of the migration assumptions. This is true for both national and subnational projections.
- The intercensal projections update, produced two years after the first set of projections, has generally been more accurate than the first set.
- Historical projections have under-projected and over-projected the population at different times. Focusing on the 1991–2018 period, medium/median projections tended to under-project rather than over-project the population of geographic areas in New Zealand. However, there is no inevitability about the future direction of inaccuracy of the most recently published projections, as projection assumptions have been revised with each set of projections.
- Most regions, territorial authority areas, and Auckland local board areas have been both under-estimated and over-estimated at different times.
- Most regions, territorial authority areas, and Auckland local board areas have been both under-projected and over-projected at different times.
- Projections of deaths are more accurate than projections of births, while (net) migration is the most difficult component to either project or for assumption formulation.
- Projection accuracy generally decreases as the period from the base (starting point) increases. For example, relative errors tended to be higher after 10 years than after five years, and after 15 years than after 10 years.

Measures of accuracy

Several measures are used in this report to evaluate the accuracy of estimated and projected populations relative to observed populations. Accuracy is evaluated as at 30 June for each year for which estimated and projected populations are compared with observed populations.

Error

Error (E) is the numerical difference between the estimated/projected population and observed population in a given year:

$$E = EP_y - O_y$$

E = error

EP_y = estimated/projected population in year y

O_y = observed population in year y

An E of 500 indicates that a given estimate/projection was 500 higher (over-estimated/projected) compared with the observed population for that year. An E of -500 indicates a given estimate/projection was 500 lower (under-estimated/projected) compared with the observed population for that year.

The observed population is typically the revised or rebased population estimate in a census year, incorporating results from that census.

The error or difference between the original and revised or rebased census-year population estimates is commonly referred to as the ‘intercensal discrepancy’.

Corrected error

The 1996-base estimates and projections were adjusted (upwards) to account for the revision of the 1996-base population estimate in 2002. At the national level, the estimated resident population at 30 June 1996 was revised up by 18,000 (0.5 percent). To account for this revision, a corrected E is calculated (Keilman, 1997):

$$CE = EP_y - O_y - (EP_b - O_b)$$

CE = corrected error

EP_b = original population in base year b

O_b = revised population in base year b

If the aim of the estimates/projections is to measure population change over a given period, then the CE gives a better measure of the accuracy of the estimate/projection, without the conflating effect of a known inaccuracy in the base population estimate.

Relative error

Relative error (RE) is the percentage difference between the estimated/projected population (corrected if necessary, as above) and observed population in a given year relative to the observed population:

$$RE = \left(\frac{EP_y - O_y}{O_y}\right) \times 100$$

RE = relative error (percent)

A RE of 5 percent indicates that a given estimate/projection was 5 percent higher (over-estimated/projected) compared with the observed population for that year. A RE of -5 percent indicates that a given estimate/projection was 5 percent lower (under-estimated/projected) compared with the observed population for that year.

RE is a standard approach to assessing accuracy (eg, Bell and Skinner, 1992; Keilman, 1997, 2007; Wilson, 2007) although other measures can be used (eg, Swanson, Tayman, & Barr, 2000). RE facilitates comparisons between periods of different length, between areas of different population size, and between different population measures.

Mean relative error

Where multiple estimates/projections are being evaluated, such as those for regions, it can be useful to calculate the $mean\ RE$:

$$mean\ RE = \sum_{1}^{n} \left(\frac{EP_y - O_y}{O_y}\right) \times \frac{100}{n}$$

$mean\ RE$ = mean relative error (percent)

n = number of estimates/projections (eg, number of areas)

Absolute relative error

Absolute relative error (ARE) is the percentage difference, irrespective of sign (positive or negative), between the estimated/projected population and observed population in a given year relative to the observed population:

$$ARE = \left|\frac{EP_y - O_y}{O_y}\right| \times \frac{100}{n}$$

ARE = absolute relative error (percent)

ARE indicates the extent of inaccuracy, but not the direction (bias) of that inaccuracy.

Mean and median absolute relative error

Again, where multiple estimates/projections are being evaluated, such as those for regions, it can be useful to calculate the $mean\ ARE$:

$$mean\ ARE = \sum_{1}^{n} \left| \frac{EP_y - O_y}{O_y} \right| \times \frac{100}{n}$$

$mean\ ARE$ = mean absolute relative error (percent).

Alternatively, the $median\ ARE$ is the value where half the estimates/projections (eg, of regions) are above, and half are below, and is less subject to the influence of outliers with extreme ARE values than the $mean\ ARE$.

Possible accuracy standards

Estimates

Stats NZ has considered what 'fit for purpose' means for population estimates in the context of major changes to census (McNally & Bycroft, 2015).

The result of that investigation was some possible quantified quality standards for population estimates that represent the needs of key users. The standards build on our previous knowledge of key population statistics uses and reflect consultation with core customers of population statistics. The quality standards are expressed in terms of customers' minimum accuracy requirements for differing levels of geography and age group breakdowns. The standards provide a benchmark against which to assess the strengths and limitations of alternative approaches to producing population statistics.

These possible accuracy standards provide a useful framework for evaluating the accuracy of past population estimates, and reference is made to these standards in the evaluation sections (Chapter 3).

Projections

The demographic projections produced by Stats NZ are not predictions or forecasts. They indicate future population and change, if the stated assumptions were to apply over the projection period.

The value of population projections is less defined by whether they match reality, but more about whether they are plausible and useful to users at the time the projections were published. Projections aim to form a basis for developing reasonable expectations about the future; to help focus attention on potential events, risks, and opportunities; and for people and policy-makers to plan and make decisions accordingly.

Given the role and nature of projections discussed, we do not use the possible accuracy standards to evaluate projections the same way as we do for estimates.

Our findings

In this section, we present analysis for and discussion of the accuracy of both population estimates and projections between 2013 and 2018 in terms of the questions outlined in 'Report outline: a guide for readers', based on key findings from the analysis. We build on and consolidate results from the accuracy of estimates and projections from previous years (see An evaluation of Statistics New Zealand population estimates and projections, 1996–2013 (Stats NZ, 2016) for more in-depth discussion and analysis for periods up to 2013).

Estimates

This chapter reviews the accuracy of four sets of national and subnational population estimates published since 1996 against observed populations for the four most recent intercensal periods 1996–2001, 2001–2006, 2006–2013, and 2013–2018). Assessments are made for:

- New Zealand
- 16 regional council areas (regions)
- 67 territorial authority areas (TAs)
- and 21 Auckland local board areas (ALBAs)

Due to the delay of publication of 2018 Census results, we made an interim revision in the 2013-base population estimates to improve their accuracy and relevance after five years. This interim revision included two main parts:

- The national population estimates were revised back to June 2013 to incorporate the outcomes-based migration measure (see International migration uses new official measure).
- The subnational population estimates were initially revised for 2018 and 2019 only, for consistency with the revised 2013-base national population estimates, but also to reflect the geographic distribution evident from published 2018 Census counts.

As a result, we assess here the accuracy of the *revised* 2013-base national population estimates, but the *original* 2013-base subnational population estimates. The initially revised subnational population estimates have been partially adjusted using the 2018 Census counts, and this could bias the accuracy assessment if we used those.

New Zealand

National population estimates had a mean absolute relative error (ARE) of 0.8 percent over the four intercensal periods (see table 1.1).

This is within the 1.0 percent threshold suggested as a possible accuracy standard for these estimates (McNally & Bycroft, 2015).

However, for the most recent intercensal period (2013–2018), the 2013-base estimate in 2018 was an under-estimate of the observed population by 1.2 percent or 60,000, with these figures being the largest in absolute terms across the four intercensal periods.

Table 1. 1

Error and relative error of population estimates for New Zealand
1996–2018

Intercensal period	Error	Relative error (%)
1996–2001	-12,500	-0.3
2001–2006	-45,100	-1.1
2006–2013	29,000	0.7
2013–2018	-60,000	-1.2

Note: Errors above 0 indicate an overestimate, and errors below 0 indicate an underestimate. Shaded values are those not meeting the possible accuracy standards. The possible accuracy standard for total population is used here for each intercensal period.

Source: Stats NZ

While the 2013-base estimates in 2018 were lower than the observed population, the 2006-base estimates in 2013 were higher than the observed population. This pattern of accuracy suggests the 2013-base estimated resident population (ERP) was probably under-estimated, contributing to both the 2006–2013 over-estimate and the 2013–2018 under-estimate. This is supported by evidence given by other analyses including Māori population under-estimation in 2013: Analysis and findings (Stats NZ, 2022) and Experimental administrative population census: Data sources and methods (Stats NZ, 2021b, page 25).

The two intercensal periods (2001–2006 and 2013–2018) with the largest AREs had large net international migration gains:

- 161,000 in the five years ended June 2006
- 260,000 in the five years ended June 2018

This points to another likely source of inaccuracy. A larger pool of recent migrants increases the potential for inaccuracy if those migrants self-identify in the census as 'usually living' overseas and not in New Zealand. This may happen if they have recently arrived, are migrants on temporary visas and limited duration of stay, and/or made multiple trips into and out of New Zealand to the point that their usual residence is ambiguous.

Subnational

Patterns of accuracy

Over the four intercensal periods, the mean absolute relative error (ARE) was:

- 1.8 percent for regions
- 2.6 percent for TAs
- 3.6 percent for ALBAs.

Both table 1.2 and figure 1.1 show that:

- the mean and median AREs for regions were generally smaller than for TAs and ALBAs – with the percentiles and median presentation of ARE in figure 1.1 showing this clearly

- the pattern across intercensal periods for regions and TAs for the mean and median ARE was generally similar to the national results – highest for 2013-2018, followed by 2001-2006, then 2006-2013, with the lowest for 1996-2001
- ALBAs generally had higher mean and median AREs than for TAs, even though ALBAs are generally larger in population than most TAs
- ALBAs had the lowest accuracy across all of the various measures of accuracy for the intercensal period 2013-2018 – tables 1.2, 1.3, figure 1.1.

Table 1. 2

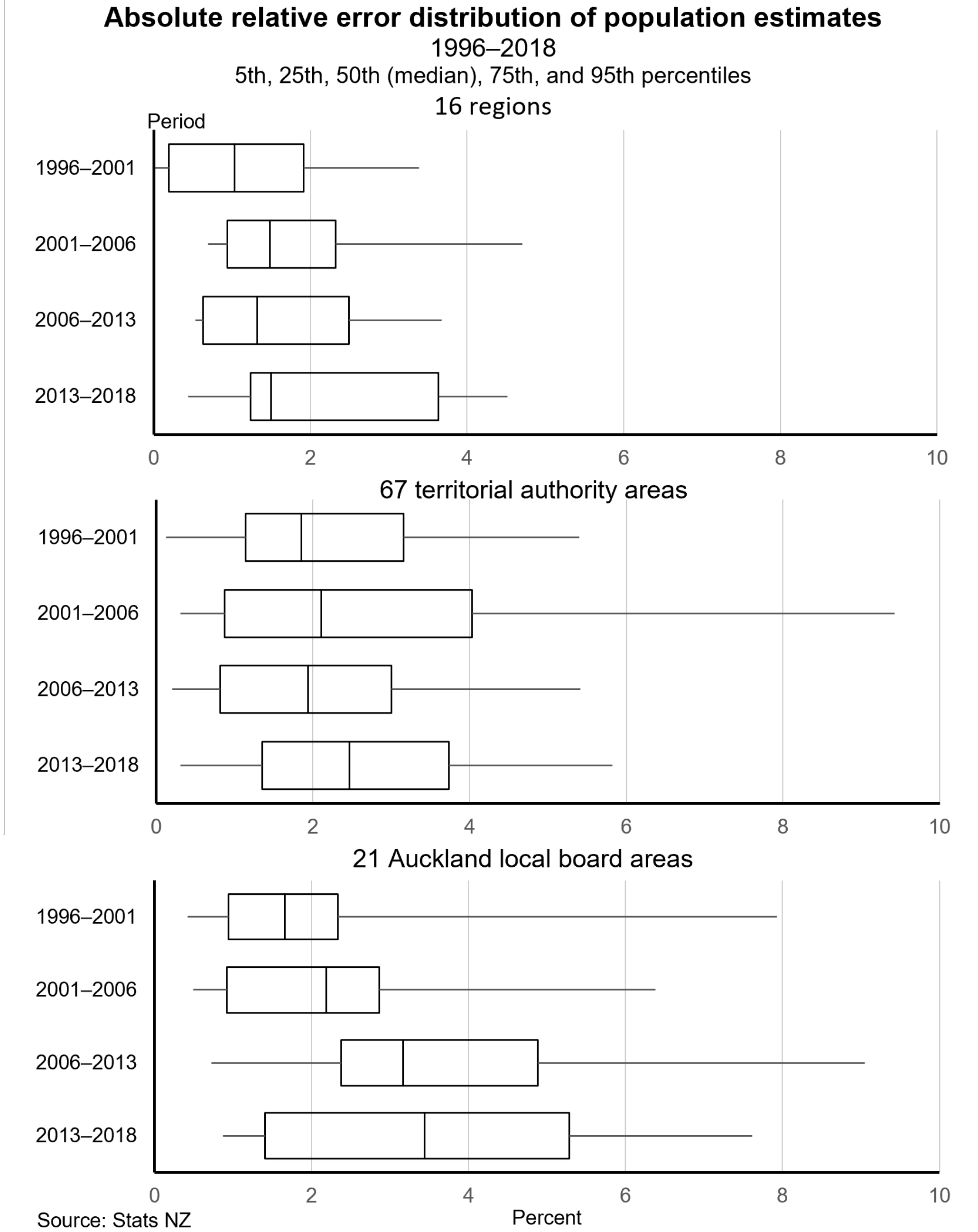
Mean absolute relative error of population estimates across 16 regions, 67 TAs, and 21 ALBAs

Intercensal Period	Area	Mean absolute relative error (%)		
		Total population	Population more than 100,000	Population up to 100,000
1996–2001	Regions	1.3	0.7	2.5
	TAs	2.3	0.9	2.5
	ALBAs	3.1	2.2	3.4
2001–2006	Regions	1.9	1.3	3.3
	TAs	2.9	0.6	3.2
	ALBAs	2.9	1.6	3.3
2006–2013	Regions	1.7	1.7	1.7
	TAs	2.2	2.2	2.3
	ALBAs	3.9	2.8	4.3
2013–2018	Regions	2.2	2.2	2.2
	TAs	2.8	2.0	2.9
	ALBAs	4.6	3.2	5.0

Note: Mean AREs are provided for areas with population more than and up to 100,000 based on 2018 populations.

Source: Stats NZ

Figure 1.1



Source: Stats NZ

Possible accuracy standards for subnational estimates (McNally & Bycroft, 2015) are split into two groups.

- For areas with population size of less than 100,000, 85 percent of total population estimates are within 5 percent relative error, and all are within 10 percent.
- For areas with population size of 100,000 or more, all total population estimates are within 5 percent relative error.

Based on these possible accuracy standards for the intercensal period 2013-2018 (table 1.3):

- all regions met the possible accuracy standards

- TAs and ALBAs did not meet the possible accuracy standards with one exception – TAs did meet the possible accuracy standards for the intercensal period 2006-13 only.

Table 1. 3

Percentage of absolute relative errors for population estimates with possible accuracy standards

16 regional councils, 67 territorial authority, and 21 Auckland local board areas

1996–2018

Period	Area	Population under 100,000		Population 100,000+
		Percentage of areas with absolute relative error (ARE) under:		
		5 percent	10 percent	5 percent
1996–2001	Regions	100	100	100
	TAs	92	98	100
	ALBAs	86	95	–
2001–2006	Regions	100	100	100
	TAs	80	98	100
	ALBAs	84	95	100
2006–2013	Regions	100	100	100
	TAs	92	100	100
	ALBAs	74	100	100
2013–2018	Regions	100	100	100
	TAs	83	100	100
	ALBAs	69	94	80

Notes:

1. There are 10 regions with population 100,000+ in 2001, 2006, and 2013, and 11 in 2018.
2. There are 5 TAs with population 100,000+ in 2001, and 7 in 2006, 2013, and 2018.
3. There are no ALBAs with population 100,000+ in 2001, 2 in 2006 and 2013, and 5 in 2018.
4. Shaded values indicate those that did not meet the possible accuracy standards.

Source: Stats NZ

The patterns of accuracy shown above for subnational areas can be explained in part by similar factors to those impacting the national results across time such as the large net international migration gains in the periods 2001–2006 and 2013–2018.

However, additional general factors come into play for subnational patterns of accuracy, including that areas with small populations are more sensitive to fluctuations in the components of population change (births, deaths, net migration) than areas with large populations.

There are also factors which are especially relevant to subnational patterns of accuracy related to the specific component of net international and internal migration; this factor impacts disproportionately on the Auckland region and as a consequence impacts on all of the other regions' patterns of accuracy since the Auckland region accounts for around one third of the total population of New Zealand:

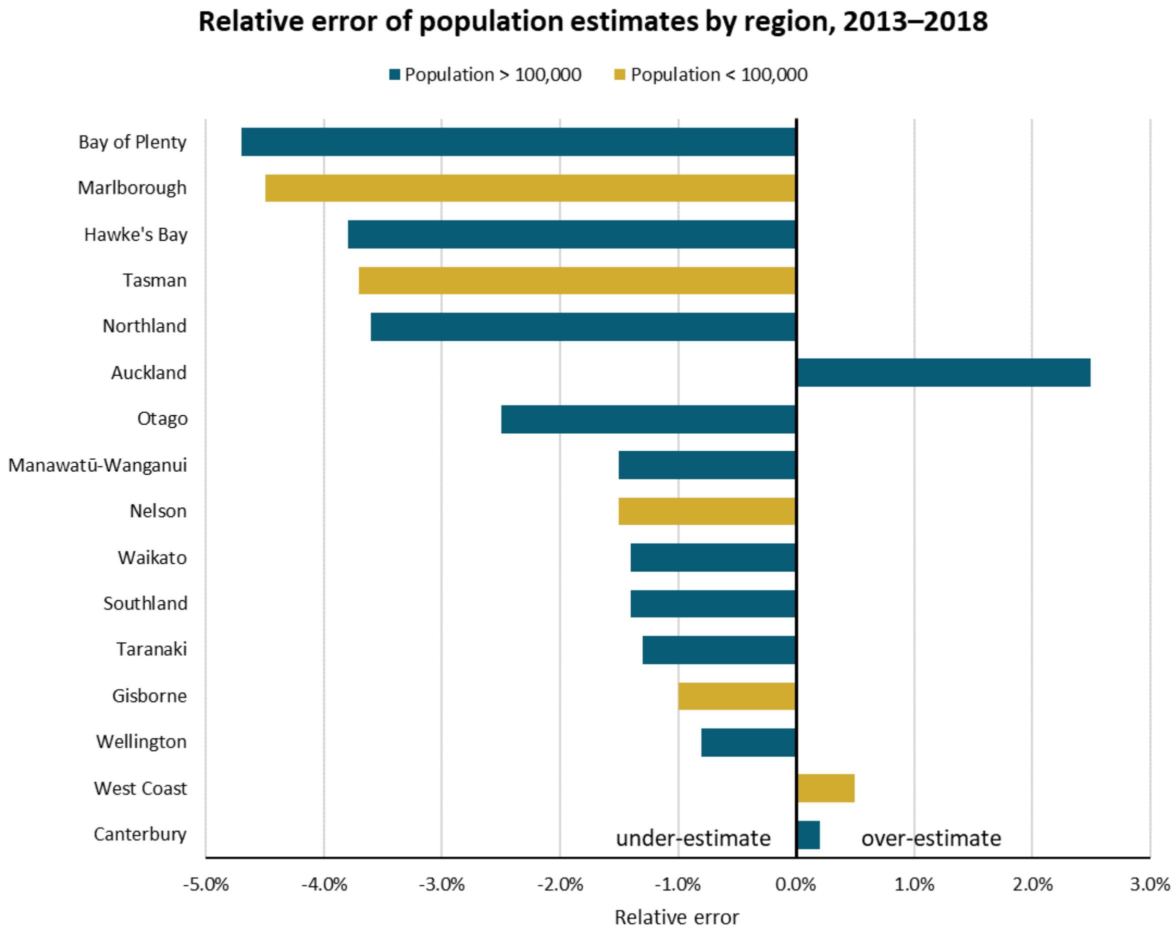
- measurement of international migration at a subnational level – while there is a strict definition of a migrant, their place of residence can be more ambiguous

- the large amount of international migration into and out of Auckland, with the challenge of distributing both migrant arrivals and migrant departures at the ALBAs level (further discussion on this is in the 'Regions' and 'Auckland local board areas' sections below)
- the challenges of measuring internal migration (movements of New Zealand residents between areas of New Zealand):
  - there is no direct data source that covers all people movements into and out of local areas on an annual basis
  - there are complications with the definition of place of residence – this overlaps with how international migration is measured, and it depends on how people interact with and identify themselves in administrative data.

Regions

For the 2013-base population estimates in 2018, Auckland had the largest over-estimation while most other regions were under-estimated. This is mainly explained by too much of the net international migration gain being attributed to Auckland between 2013 and 2018.

Figure 1.2



Note: Relative errors less than 0.0% correspond to an under-estimation of the population, while relative errors above 0.0% correspond to an over-estimation of the population.

Source: Stats NZ

During 2013−2018, Auckland was initially estimated to have received about 60 percent of the net international migration gain into New Zealand. This was an over-estimation caused by the following:

- The place of residence area was determined from responses on arrival/departure cards for those intending to stay in New Zealand/depart from New Zealand on a long-term basis: 43 percent of migrant arrivals indicated, and 38 percent of migrant departures indicated Auckland as their next/last place of residence. However, this may have been a temporary address, especially for arrivals who may have been uncertain or unfamiliar with where in New Zealand they would be living.
- Non-response on arrival/departure cards to the place of residence was higher for migrant arrivals (16 percent) than migrant departures (10 percent). Given no alternative data sources to indicate the place of residence for migrant arrivals and departures, place of residence for non-response was imputed based on the observed subnational distribution of place of residence for migrant arrivals and migrant departures. As a consequence of this imputation adjustment, the estimated net international migration gain for Auckland increased.

On the other hand, the observed population for Auckland in 2018 was influenced by where international migrants into New Zealand actually settled during the 2013-2018 period. Non-New Zealand migrants tend to be mobile and often travel within New Zealand before they settle. Their place of residence may therefore differ from their stated intended place of residence on arrival into New Zealand. This factor not only plays a role in explaining the over-estimation of the Auckland region population 2018, it is also relevant for explaining the over-estimation of all of the main cities with the largest net international migration gains (eg, Auckland, Wellington, Christchurch) during the period 2013-2018. This is shown in the section on territorial authority areas below.

There are also general limitations on the estimation of internal migration within New Zealand, which play a role in explaining the patterns of accuracy observed, not only for regions, but also for territorial authority areas and ALBAs as shown in the sections below:

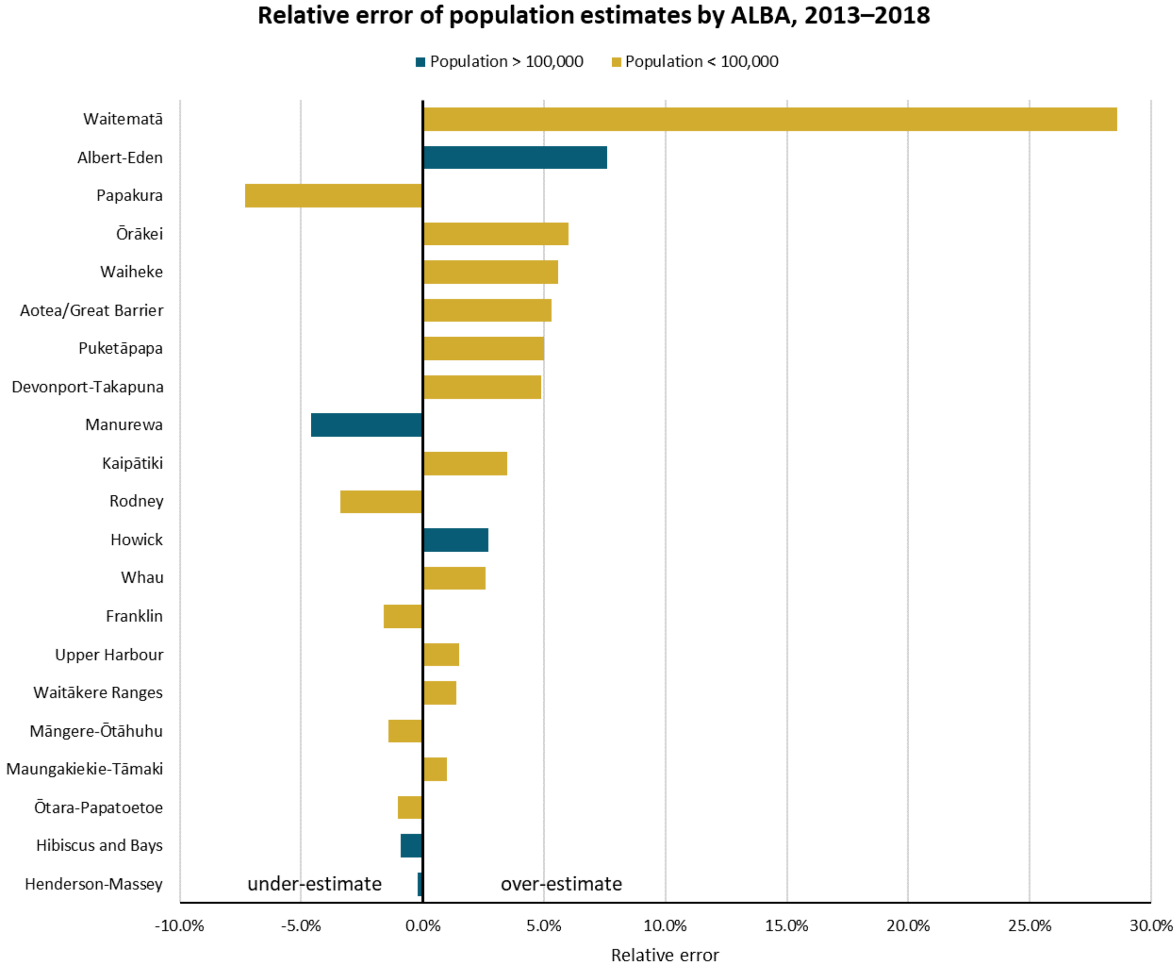
- There are limitations in the administrative data sources used to estimate internal migration within New Zealand, given that people migrating from one area of New Zealand to another are not required to register their movements with any agency.
- The five-yearly Census of Population and Dwellings has included a question on “usual residence five years ago” since 1971, and this has been the authoritative data source for measuring internal migration. However, the infrequency of the collection (every five years), and the ‘snapshot’ nature of a transition-based measure are limitations of this data source. This question was replaced in the 2018 and 2023 Census forms with a question on “usual residence one year ago” while usual residence five-years-ago information is still produced using a combination of previous census and administrative data. These changes are in response to the need for more accurate estimates of internal migration.

See also the later section ‘Discussion and future work’, which includes discussion on future developments for improving data on net migration, external migration, and internal migration.

Auckland local board areas

Although Auckland was over-estimated during the period 2013–2018, population estimates within Auckland had different results. Waitematā was significantly over-estimated, while other central ALBAs were over-estimated to a lesser degree. However, the southernmost and northernmost ALBAs were generally under-estimated.

Figure 1.3



Note: Relative errors less than 0.0% correspond to an under-estimation of the population, while relative errors above 0.0% correspond to an over-estimation of the population.

Source: Stats NZ

Inaccuracies in ALBA population estimates were mainly due to the additional challenge to allocate international migrant arrivals and departures to ALBAs. The intentions-based migration statistics (from arrival cards) only collect residence information for Auckland with no further breakdown.

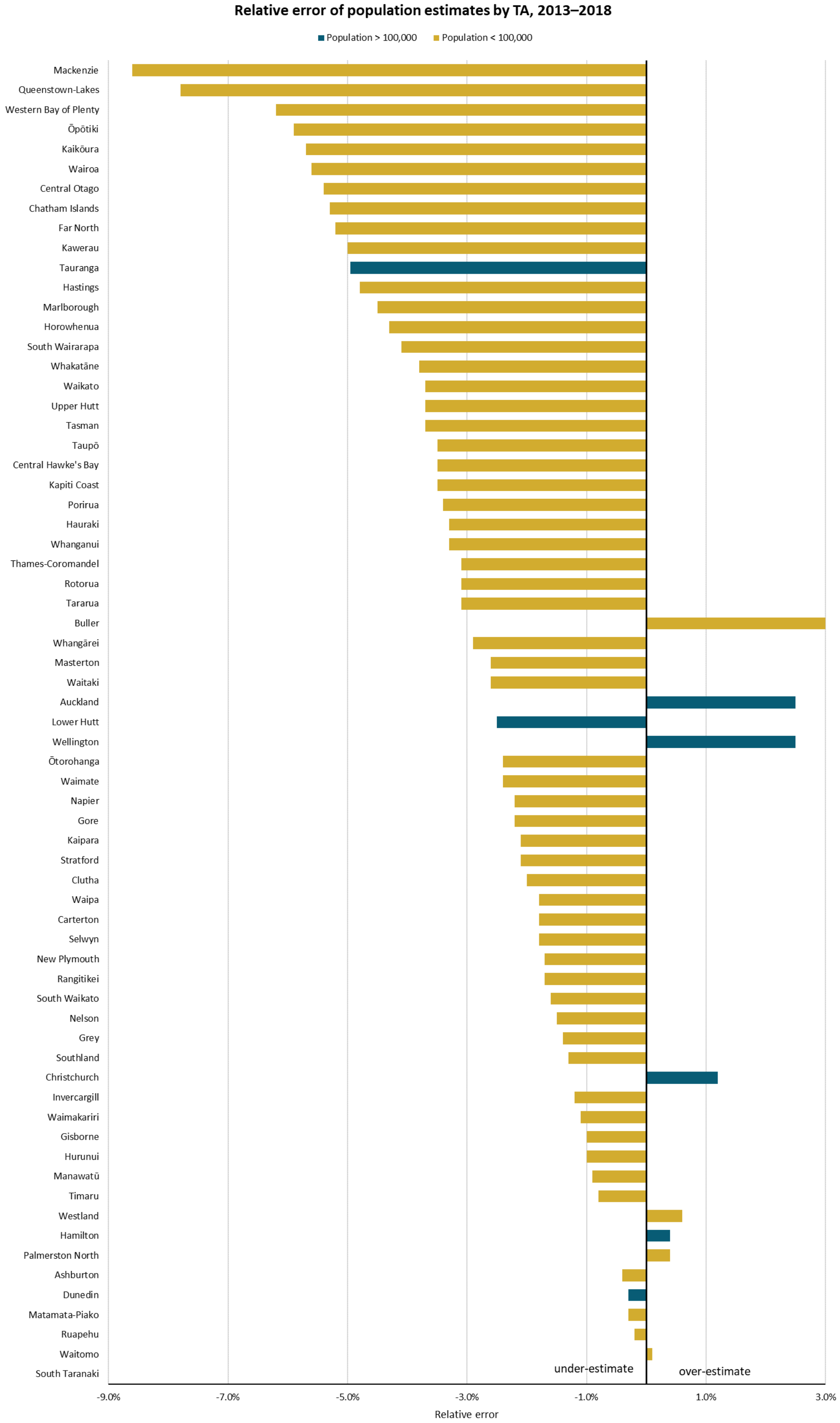
Possible contributions to the less accurate ALBA migration estimates for the 2013–2018 period, compared with earlier periods, are:

- Inconsistent methods used to derive ALBA international migration estimates from 2013−2018. The methods of estimating ALBA international migration have evolved over that period, from distributing largely based on 2013 census and historical information, to the use of observed migrant distributions from administrative data in the Integrated data infrastructure (IDI).
- Due to the timing of production and release of subnational population estimates, ALBA migration estimates were produced when the IDI only had limited address information for migrant arrivals. These migration estimates were heavily biased towards Waitematā.

Territorial authority areas

Like the pattern across regions and ALBAs, it was the main cities with the largest net international migration gains (eg, Auckland, Wellington, Christchurch) that were over-estimated during the 2013−2018 period. Other areas over-estimated were those that experienced population decrease (eg, Buller), while most of the other TAs were under-estimated.

Figure 1.4



Note: Relative errors less than 0.0% correspond to an under-estimation of the population, while relative errors above 0.0% correspond to an over-estimation of the population.

Source: Stats NZ

Projections

This chapter reviews the accuracy of national population projections published since 1991, and subnational projections published since 1996, against observed populations. It also includes a discussion of projections versus forecasts, deterministic and stochastic projections, New Zealand historic projections since the 1950s, as well as the components of population change in terms of births, deaths, and net-migration.

As with the accuracy of subnational population estimates, the accuracy of subnational projections is assessed for:

- 16 regional council areas (regions)
- 67 territorial authority areas (TAs)
- and 21 Auckland local board areas (ALBAs).

Stats NZ usually produces an initial set of national (New Zealand) and subnational projections in the year after each census year. These projections are typically updated two years later, three years after the census year. The precise timing depends on the availability of rebased population estimates, which in turns depends on the availability of the census results and Post-enumeration Survey results. When the 2011 Census was cancelled due to the Canterbury earthquakes, an additional set of national population projections was produced before the census in 2013. In this report we evaluate the accuracy of population projections for the initial sets, and for the updated sets of projections as follows:

- 11 sets of national projections since the 1991 Census
- nine sets of subnational projections since the 1996 Census – for regions, territorial authority areas, and ALBAs, noting that up until and including the 2006-base projection published 2010 these were derived retrospectively from area unit population projections.

Projections versus forecasts

Demographic projections produced by Stats NZ are not predictions or forecasts. The projections indicate future population and change, if the stated assumptions apply over the projection period. Stats NZ produces a range of alternative projections to illustrate different scenarios. In contrast, a forecast is one prediction of what the population will be at a given date.

Projections and forecasts share similar methods of production (eg, extrapolation, modelling, expert judgement), but the main differences relate to how they are communicated to customers and how they should be used.

Projections are based on current policy settings, and do not try to anticipate major policy changes. The difference between projections and predictions is provided in an analogy made by the Australian Productivity Commission (2005): someone sees a large boulder on a train track. The **projection** is that there will be a rail disaster and many deaths if the boulder is not moved or the train is not stopped. The **prediction** is that someone will move the boulder, averting the accident. The projection is much more useful for policy formulation and planning.

Organisations, which include central and local governments, often initiate strategies to avert the population trends implied by the projections. It is therefore illogical to criticise the projections if they do not match actuality, especially when the projections are used to inform those strategies. One role of projections is to enable future changes to be understood and managed, if not averted.

The value of population projections is therefore less defined by whether or not they match reality, but more about whether they are plausible and useful to customers at the time they are published. Projections form a basis for developing reasonable expectations about the future; help focus attention on potential events, risks, and opportunities; and allow people and policymakers to plan and make decisions accordingly.

The distinction between projections and forecasts is not necessarily shared by others (eg, Alho, 1997; Keilman, 1997) who argue the mid-range projection is inevitably used as a forecast. However, Stats NZ makes it clear when publishing projections that customers can (and should) make their own judgement as to which projections are most suitable for their purposes – this will not necessarily be the mid-range projection.

Deterministic and stochastic projections

Stats NZ always provides a range of alternative projections or scenarios. This is partly to indicate the uncertainty in projection outcomes, partly to indicate the significance of different projection assumptions for projection results, and partly to provide different projection scenarios for different customers.

Conventional deterministic projections convey uncertainty qualitatively – through alternative scenarios. A deterministic projection is a single projection from a given set of assumptions (eg, about fertility, mortality, migration).

For historical national population projections up to the 2009-base published in 2009 and all subnational population projections, Stats NZ typically produces 'low', 'medium', and 'high' projections. These descriptors relate to population growth rates and population size but not necessarily to other population characteristics (eg, births, deaths, median age, dependency ratios). This is one reason why Stats NZ has been developing stochastic projections, which better convey the uncertainty in all projected population characteristics.

Stochastic (probabilistic) population projections are produced to give estimates of uncertainty, although these estimates are themselves uncertain. Stochastic population projections are produced by combining 2,000 simulations of the assumptions for the 'medium' deterministic projections (births, deaths, and migration). These simulations can be summarised by percentiles, which indicate the probability that the actual result is lower than the percentile. We publish nine alternative percentiles (2.5th, 5th, 10th, 25th, 50th, 75th, 90th, 95th and 97.5th percentiles).

Official stochastic projections of the New Zealand population were published for the first time in 2012 with the 2011-base national population projections. Subnational projections continue to be produced as deterministic projections with 'low', 'medium', and 'high' projections.

For this report, for the evaluation of the accuracy of projections, we compare the following projected populations with observed populations:

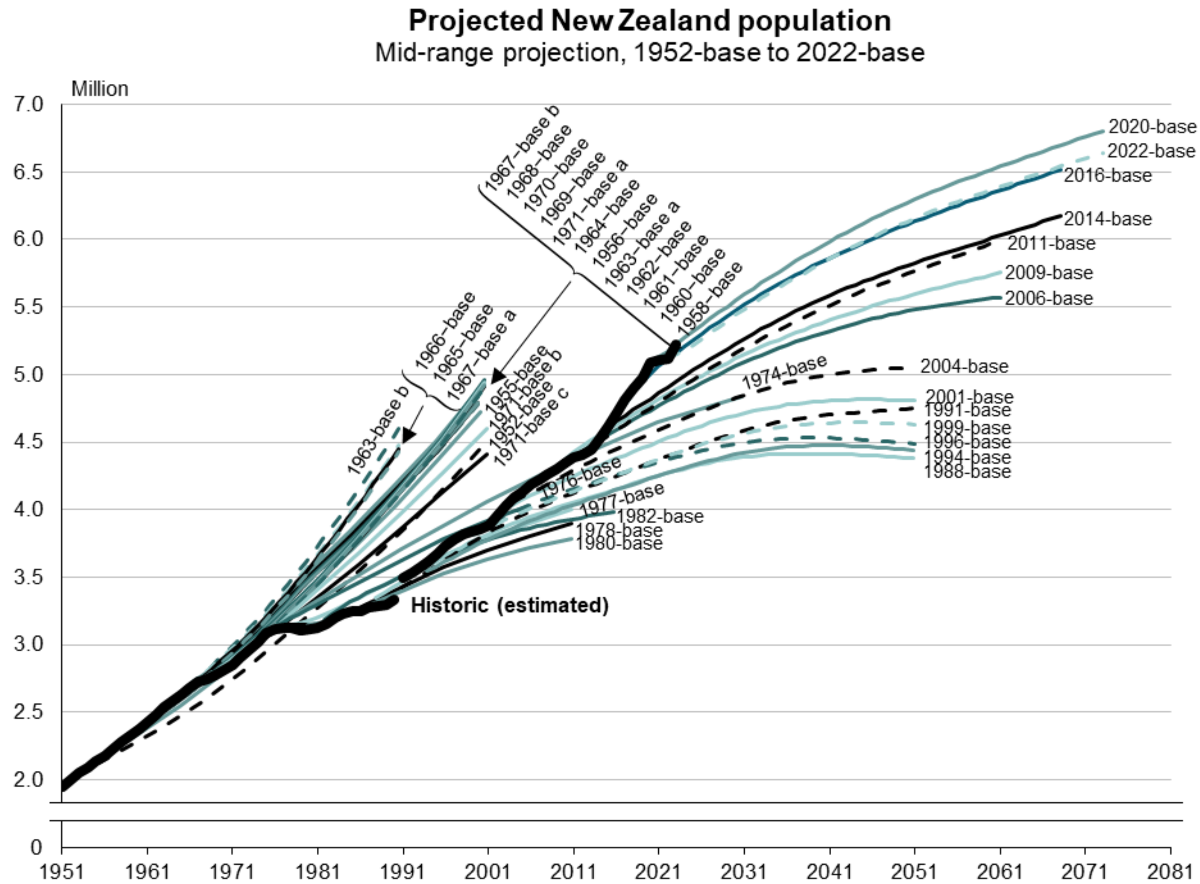
- medium projections for all subnational projections, and national projections up to the 2009-base published in 2009
- median (50th percentile) projection for the national stochastic projections for the 2011-base published in 2012, 2014-base published in 2014, and 2016-base published in 2016 projections.

New Zealand

Historic national projections from the 1950s

The mid-range projections of New Zealand's population have varied considerably over time (figure 1.5). Generally, the population tended to be over-projected during slower population growth periods, and under-projected during faster population growth periods. However, the explanation relates more to the underlying components (eg, lower fertility rates than assumed, and higher migration gains than assumed), than population growth per se.

Figure 1.5



Notes:

1. The discontinuity in the historical population estimates in 1991 denotes a shift from the de facto to resident population concept.
2. Different projections from a common base year are denoted by a, b, and c.

Source: Stats NZ

National projections from the 1990s

New Zealand experienced particularly fast population growth during the 2014−2019 period, with average annual population growth averaging 2 percent, driven by international migration gains. This contributed to the under-projection results evident in 2016/2018 (table 1.4).

Population growth even outstripped higher growth percentiles/scenarios during this period, except for the 2016-base projections published in 2016. See **Accuracy of estimated and projected populations 1996–2018** (interactive Excel file) under 'Available files'.

Between 1991 and 2018, Stats NZ published 11 sets of national population projections (table 1.4). For the mid-range projections, in terms of years between the most recent census at the time of publication of the national projections and the comparison year, they ranged from:

- 3.2 percent under-projection to 1.0 percent over-projection (based on 10 sets of projections) after 5 years
- 3.8 percent under-projection to 0.3 percent over-projection (based on 9 sets) after 10 years
- 2.9 to 7.1 percent under-projection (based on 6 sets) after 15 years
- 4.4 to 9.3 percent under-projection (based on 4 sets) after 20 years
- 8.2 to 10.3 percent under-projection (based on 2 sets) after 25 years.

Table 1. 4

Error and relative error of mid-range population projections
New Zealand
1991–2018

Projection[1]	Comparison year	Error (000)	Relative error (%)
1991-base published 1992	1996	-30	-0.8
	2001	13	0.3
	2006	-121	-2.9
	2011	-193	-4.4
	2016	-384	-8.2
1994-base published 1994	1996	-42	-1.1
	2001	-33	-0.9
	2006	-193	-4.6
	2011	-281	-6.4
	2016	-485	-10.3
1996-base published 1997	2001	38	1.0
	2006	-128	-3.1
	2011	-212	-4.8
	2016	-440	-9.3
1999-base published 2000	2001	-3	-0.1
	2006	-159	-3.8
	2011	-230	-5.2
	2016	-440	-9.3
2001-base published 2002	2006	-75	-1.8
	2011	-136	-3.1
	2016	-336	-7.1
2004-base published 2004	2006	-32	-0.8
	2011	-66	-1.5
	2016	-240	-5.1
2006-base published 2007	2011	9	0.2
	2016	-125	-2.7
2009-base published 2009	2011	42	1.0
	2016	-83	-1.8
2011-base published 2012	2016	-149	-3.2
2014-base published 2014	2018	-155	-3.2
2016-base published 2016	2018	-15	-0.3

1.The 1991-base published 1992 and 1994-base published 1994 are compared with observed populations at 31 March for each comparison year; all other base projections are compared with observed populations at 30 June for each comparison year.

Note: Errors above 0 indicate an over-projection, and errors below 0 indicate an under-projection.

Source: Stats NZ

Generally, errors compounded as the period from the base (starting point) increased. For example, both absolute and absolute relative errors tended to be higher after 10 years than after 5 years, and after 15 years than after 10 years. This was not always the case, because under-projection and over-projection errors sometimes offset each other.

Components of population change

Deaths

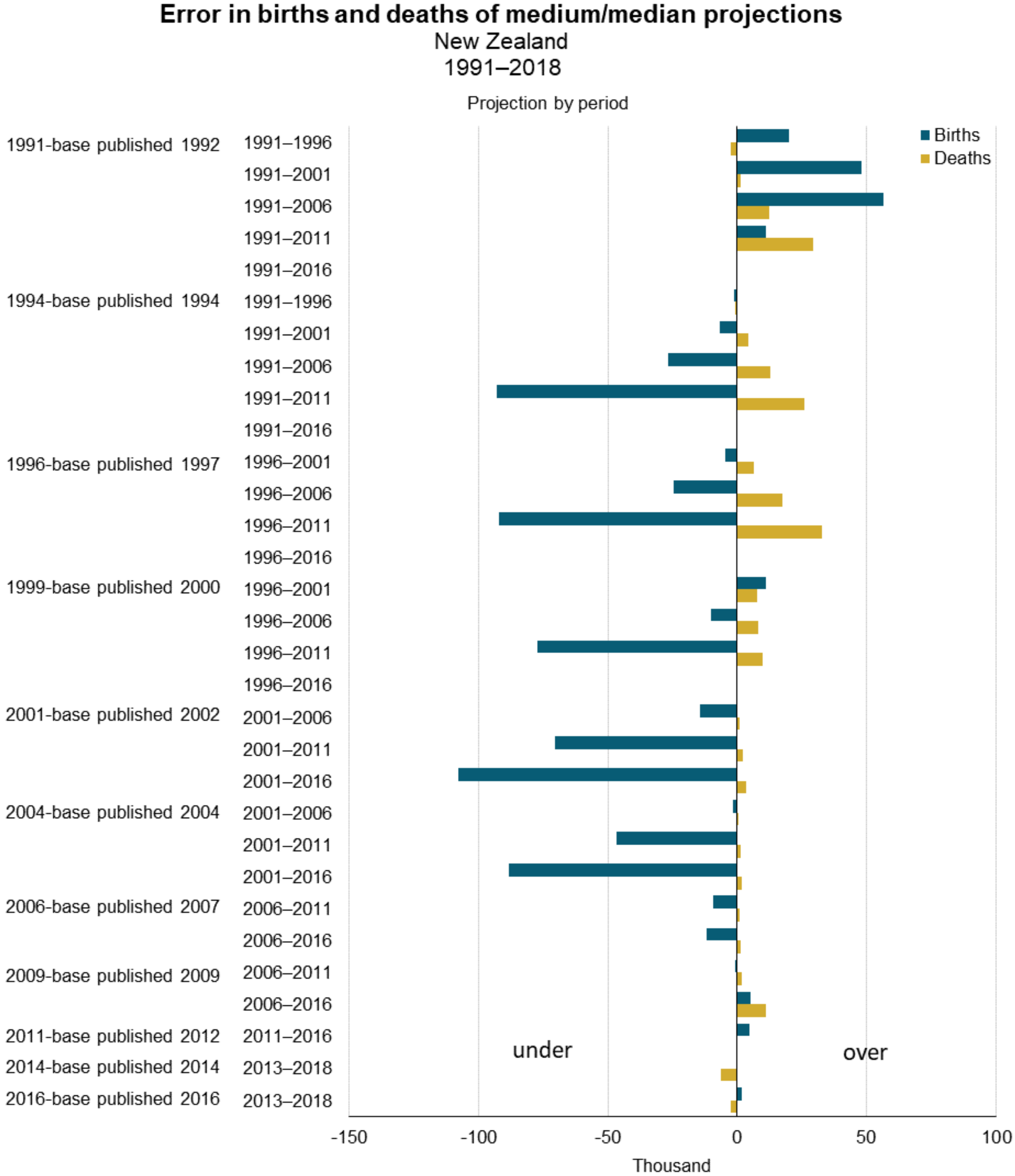
Death numbers from mortality/survivorship assumptions were projected most accurately from the three components in general. This reflects that death rates and death numbers, although subject to annual fluctuations, have fluctuated within a relatively narrow band. For example, annual death registrations in the 20 years ended June 2018 were 26,900–33,600 (mean 29,200, standard deviation 600). Over this period there were no major wars or epidemics to invalidate the mortality assumptions.

A shift from over-projecting to slight under-projecting of the death numbers was seen for the 2014- and 2016-base projections. Possible contributions to the change are:

- Death rates at all ages generally trended downwards. However, the gain in longevity has slowed down in recent years compared with historically. For example, period life expectancy at birth increased by about 1 month a year between 2013 and 2018, compared with about 3 months a year between 1996 and 2013 (Stats NZ, 2021a).
- Mortality assumptions using the coherent functional demographic method introduced in the 2011-base projections (Woods & Dunstan, 2014), do not fully reflect the slowdown in the longevity gain, due to the nature of the forecasting model based on historical trends.

Future improvements on model selections will help improve the accuracy of mortality/survivorship assumptions, and hence the number of deaths projected.

Figure 1.6



Notes:

1. Error: projected minus observed births and deaths.
2. All projections relate to June years except the 1991-base and 1994-base projections (March years).
3. All births and deaths relate to the resident population. Observed births and deaths are estimated occurrences in each period, based on birth and death registrations. Projected births and deaths are projected occurrences in each period.

Source: Stats NZ

Births

Up to 2011, national population projections tended to under-project birth numbers, except for the 1991-base. This reflects lower fertility assumptions (long-term medium total fertility rates of either 1.85 or 1.90 births per woman) than actually occurred (1.98 in 2006 and 2.01 in 2011).

Over the 2011–2018 period, birth numbers and fertility rates were projected more accurately for the more recent projections, with some fluctuation between small under-projecting and small over-projecting. This reflects that birth numbers have fluctuated within a relatively narrow band (with an average of 59,500 and standard deviation 800) compared with historically (with an average of 58,700 and standard deviation 2,900 for 20 years ending June 2011).

Migration

Assumptions for migration continue to be the most difficult to formulate. To a large extent, this reflects the volatility of New Zealand's external migration flows. In the 17 years ended June 2018, annual net outcomes-based migration (using the 12/16-month rule) ranged from -15,000 in 2012 to 64,600 in 2016 (mean 26,400; standard deviation 24,700). Such swings are difficult to anticipate, given so many factors affect both arrivals and departures, including immigration policy changes (which the projections do not try to anticipate).

Table 1.5 shows three different migration measures for the past, compared with the net migration assumed for each set of projections over the respective period. The outcomes-based migration estimates provide a more accurate measure of migration – it classifies a traveller as a migrant based on their actual movements. The PLT and residual migration are included for comparison.

Table 1. 5

Net migration estimates and migration assumptions
New Zealand

	1991–1996	1996–2001	2001–2006	2006–2011	2011–2016	2013–2018
Net migration (12/16)	...	…	161,200	42,400	137,300	260,000
Net migration (PLT)	81,300	-13,200	116,600	47,700	156,200	303,000
Net migration (residual)	88,000	1,200	160,800	25,300	181,900	318,100
1996-base published 1997	…	50,000	25,000	25,000	25,000	25,000
1999-base published 2000	…	-5,600	19,400	25,000	25,000	25,000
2001-base published 2002	…	…	100,00	25,000	25,000	25,000
2004-base published 2004	…	…	130,500	38,000	50,000	50,000
2006-base published 2007	…	…	…	46,000	50,000	50,000
2009-base published 2009	…	…	…	71,300	63,000	50,000
2011-base published 2012	…	…	…	…	19,900	55,000
2014-base published 2014	…	…	…	…	…	152,300
2016-base published 2016	…	…	…	…	…	297,600

Symbol:

… Not available

Notes:
1. The outcomes-based (12/16) migration measure is calculated from the actual travel histories of people travelling in and out of New Zealand.
2. Permanent and long-term (PLT) migration is an intentions-based migration measure based on travellers' stated intentions on arrival and departure cards.
3. The residual method calculates net migration by subtracting estimated natural increase (births minus deaths) from the estimated population change during each period.

Source: Stats NZ

Short-term migration assumptions since 2011 have been continuously revised up to reflect the record high net migration gains that New Zealand experienced.

The median long-term migration assumption also increased to 15,000 in the 2016-base projections, following the lifting of the mid-range assumption in the 2011-base projections (from 10,000 to 12,000).

These assumptions have also had a wider range to reflect larger uncertainty seen in the migration estimates. In the 2016-base projections, the assumptions imply a 50 percent chance that long-run net migration will be between -1,300 and 31,300 in any given year. They also imply that about 1 year in 4 will have more departures than arrivals, while 1 year in 14 will have net migration exceeding 50,000.

The most-recent (2016-base) mid-range assumptions have been improved to better reflect the current migration environment for immigration policy and more-temporary flows. The stochastic approach also better indicates the variability and uncertainty of New Zealand's migration balance.

Subnational

Patterns of projection accuracy observed for subnational areas are similar to those for New Zealand, but with a wider range of errors for lower geographies:

- Projections become less accurate as the time from the base increases.
- Projection updates during an intercensal period are more accurate than the initial set.
- Due to the high population growth experienced in 2014−2020, more areas were under-projected in 2016/2018, and more areas had observed populations higher than the high projection.

Regions

Of the nine sets of projections published between 1997−2017, five years out from the base population:

- the two sets of 2013-base projections had the largest mean ARE and the lowest percent of regions under a 5 percent ARE threshold (table 1.6)
- the 2013-base published 2015 ranked highest for the median ARE with the 2013-base published 2017 ranking third highest (figure 1.7).

This largely reflected net migration assumptions being lower than realised during the period 2013−2018, and partly an under-estimation of the 2013 base population (Stats NZ, 2022; 2021b).

Table 1. 6

Summary measures of relative error of medium projections
16 regions
1996–2018

Projection	Mean ARE	Percent of regions with ARE under:	
		5%	10%
5 years out from base population			
1996-base published 1997	2.2	94	100
1996-base published 2000	1.5	100	100
2001-base published 2002	2.3	94	100
2001-base published 2005	2.0	100	100
2006-base published 2007	1.7	100	100
2006-base published 2010	1.4	100	100
2006-base published 2012	1.2	100	100
2013-base published 2015	4.6	50	100
2013-base published 2017	2.9	75	100
10 years out from base population			
1996-base published 1997	3.0	88	100
1996-base published 2000	3.1	75	100
2001-base published 2002	4.6	44	94
2001-base published 2005	3.7	63	100
2006-base published 2007	4.7	56	100
2006-base published 2010	3.6	69	100
2006-base published 2012	3.9	63	100
15 years out from base population			
1996-base published 1997	4.5	50	100
1996-base published 2000	5.5	38	100
2001-base published 2002	9.3	25	44
2001-base published 2005	7.4	31	69
20 years out from base population			
1996-base published 1997	8.8	25	56
1996-base published 2000	9.9	13	50

Source: Stats NZ

Figure 1.7

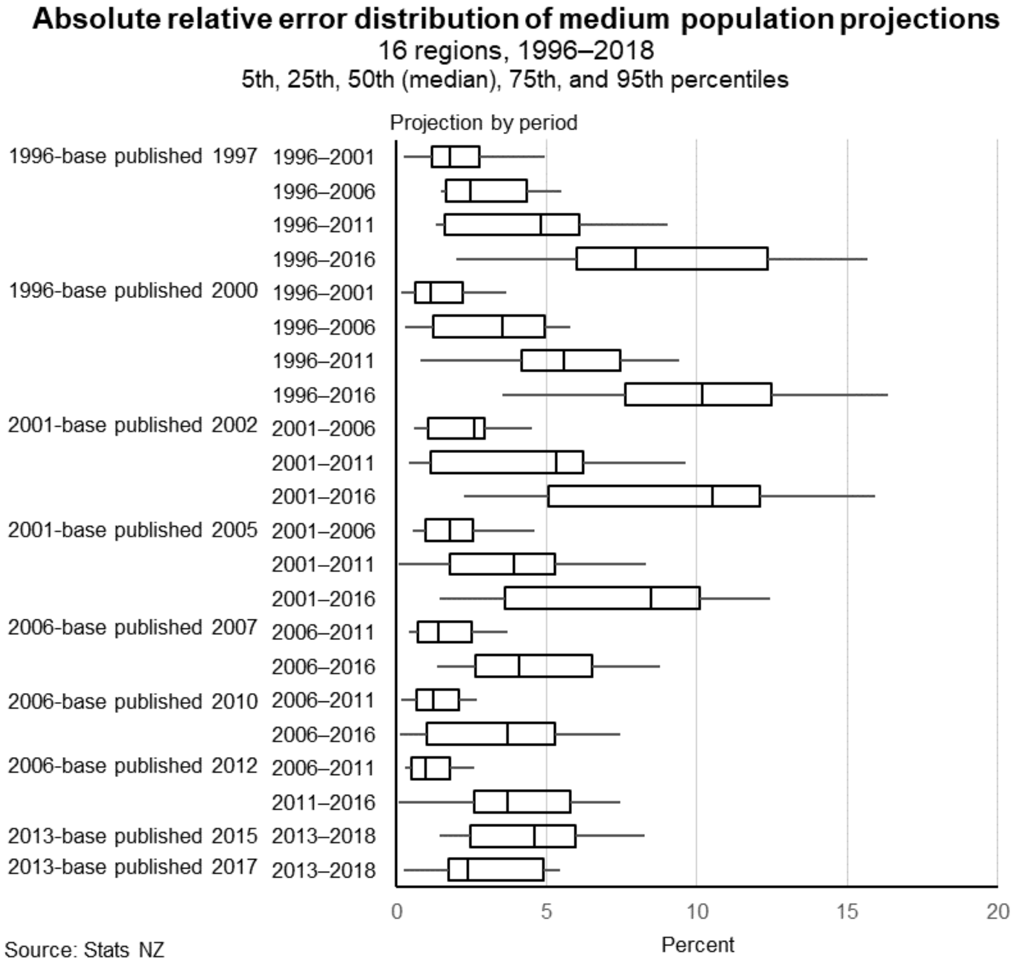
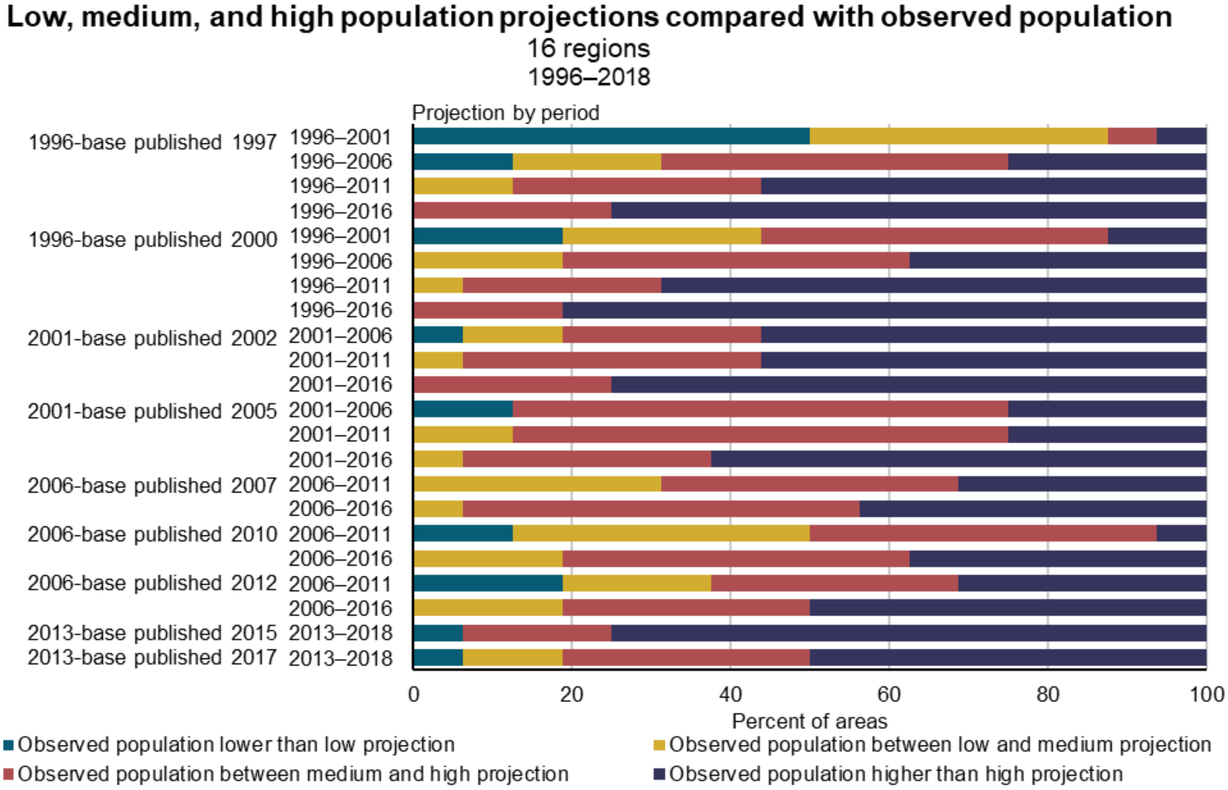


Figure 1.8 indicates the extent to which the collective 'low-medium-high-growth' projections encapsulate the observed population. This ranged from 81 percent of regions for 2006–11 (2006-base published 2010), to just 19 percent of regions for 1996–2016 (1996-base published 2000), and 2013–2018 (2013-base published 2015). This shows that the range between the low, medium, and high projections was arguably too narrow, largely reflecting too narrow a range of net migration assumptions. Most of the regions were under-projected in 2018.

Figure 1.8



Source: Stats NZ

Territorial authority areas

The results for TAs are similar to those for regions above except that the range of errors is wider.

Of the nine sets of projections published between 1997 and 2017, five years out from the base population, the two sets of 2013-base projections had the largest mean ARE and the lowest percent of TAs under a 5 percent ARE threshold (table 1.7), as well as the highest median AREs (figure 1.9).

In particular, five years out from the base population, the 2013-base published 2015 had a mean ARE over 5 percent, with less than half of the TAs under a 5 percent threshold.

The same explanation applies as for regions that this largely reflected net migration assumptions being lower than realised during 2013−2018, and partly an under-estimation of the 2013 base population (Stats NZ, 2022; 2021b).

Table 1. 7

Summary measures of relative error of medium projections
67 TAs
1996–2018

Projection	Mean ARE	Percent of TA areas with ARE under:	
		5%	10%
5 years out from base population			
1996-base published 1997	3.2	82	97
1996-base published 2000	2.5	87	99
2001-base published 2002	3.3	84	96
2001-base published 2005	2.7	88	97
2006-base published 2007	2.5	90	100
2006-base published 2010	1.8	97	100
2006-base published 2012	1.6	97	100
2013-base published 2015	5.9	43	88
2013-base published 2017	3.7	69	99
10 years out from base population			
1996-base published 1997	5.8	58	82
1996-base published 2000	5.3	58	88
2001-base published 2002	6.8	39	72
2001-base published 2005	5.6	54	81
2006-base published 2007	6.6	36	82
2006-base published 2010	5.6	51	93
2006-base published 2012	5.5	46	93
15 years out from base population			
1996-base published 1997	8.6	42	70
1996-base published 2000	8.2	31	70
2001-base published 2002	12.3	19	39
2001-base published 2005	10.3	27	54
20 years out from base population			
1996-base published 1997	13.1	28	45
1996-base published 2000	12.7	21	46

Source: Stats NZ

Figure 1.9

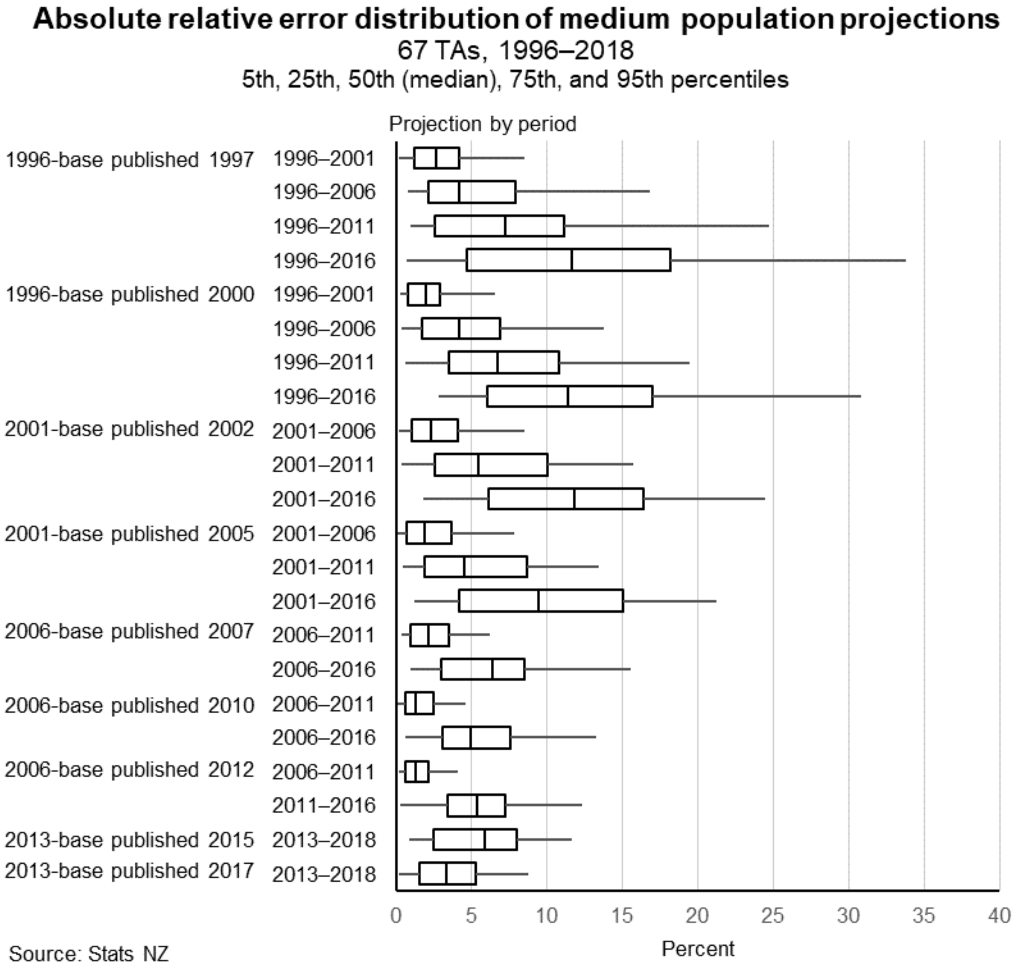
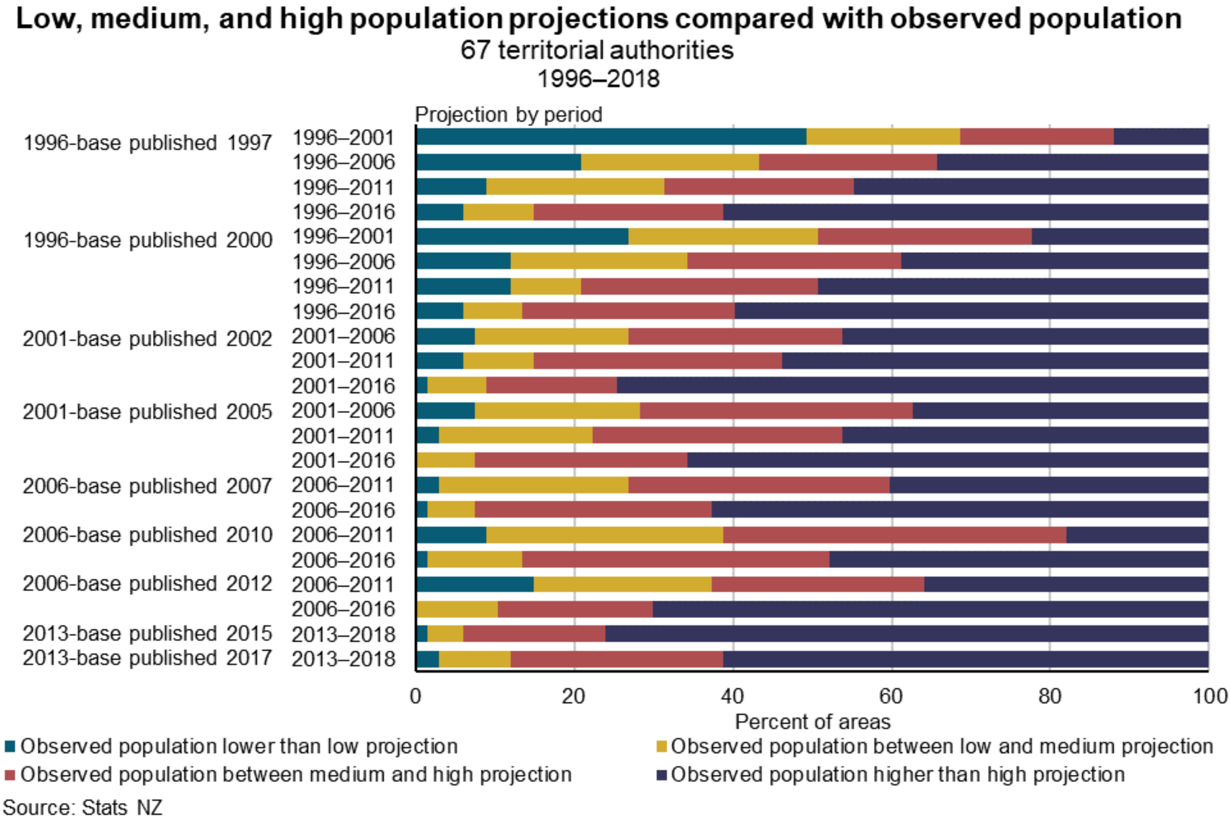


Figure 1.10 indicates the extent to which the collective 'low-medium-high-growth projections' encapsulate the observed population. This ranged from 73 percent of TAs for 2006–2011 (2006-base published 2010) to just 22 percent for 2013–2018 (2013-base published 2015). As with regions (figure 1.8) this shows that the range between the low, medium, and high projections was arguably too narrow, largely reflecting too narrow a range of net migration assumptions. Most of the TAs were under-projected in 2018.

Figure 1.10

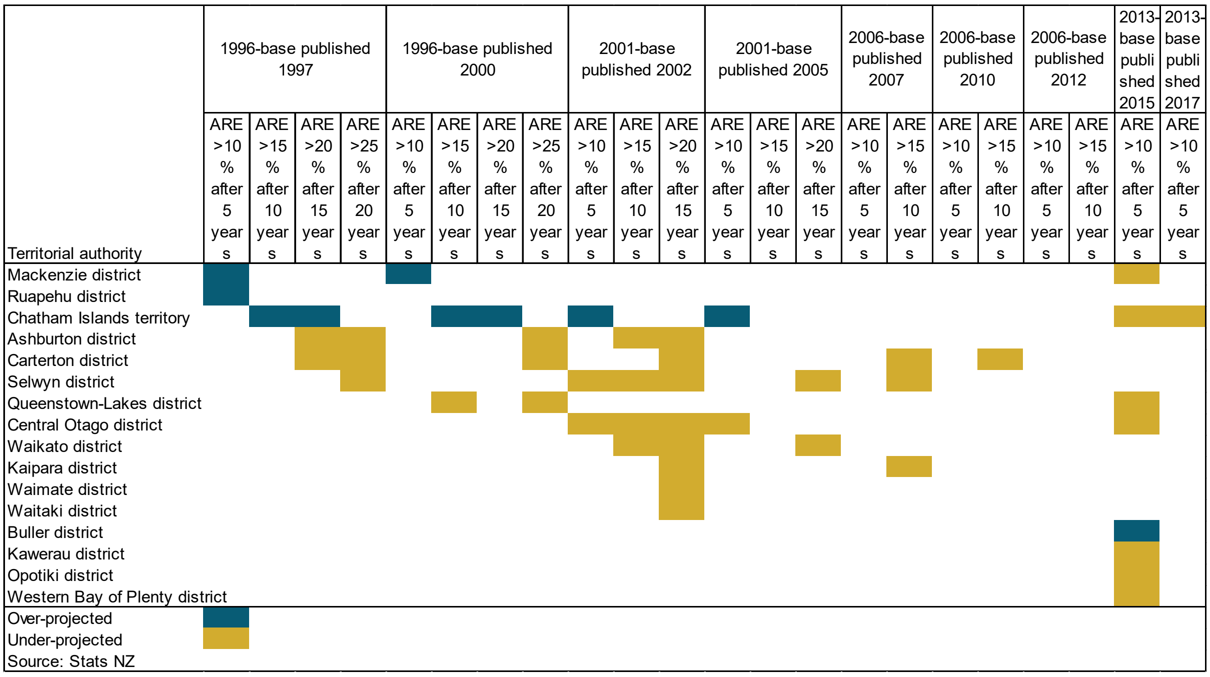


Outliers

It can be revealing to analyse TAs with particularly large relative errors (eg, AREs >10 percent after five years, >15 percent after 10 years, >20 percent after 15 years, >25 percent after 20 years). The choice of thresholds is arbitrary and is simply a way to highlight the largest AREs.

Figure 1.11

Territorial authority areas with largest absolute relative errors of medium population projections, 1996–2018



At the TA level, these large AREs generally did not persist. They were addressed in subsequent projections (ie, by incorporating new population estimates, especially those based on new census counts).

In general, fast-growing areas (eg, Waikato, Selwyn, and Queenstown-Lakes districts) are more likely to be under-projected, while areas experiencing population loss being over-projected (eg, Buller).

Auckland local board areas

In contrast to the results for regions and TAs, the two set of 2013-base projections for ALBAs had similar mean and median AREs and percent of ALBAs under a 5 percent ARE threshold as for the 1996-base, 2001-base, and 2006-base ALBA projections (table 1.8 and figure 1.12). This largely reflects an over-estimation in the 2013 base ALBA populations offsetting the lower-than-actual migration assumed for ALBAs.

Table 1. 8

Summary measures of relative error of medium projections
21 Auckland local board areas
1996–2018

Projection	**Mean ARE**	**Percent of ALBAs with ARE under:**	
		5%	10%
5 years out from base population			
1996-base published 1997	4.9	71	90
1996-base published 2000	3.6	81	95
2001-base published 2003	4.1	81	95
2001-base published 2005	3.7	81	95
2006-base published 2008	2.4	95	100
2006-base published 2010	2.9	86	100
2006-base published 2012	2.8	81	100
2013-base published 2015	3.3	76	95
2013-base published 2017	4.2	81	95
10 years out from base population			
1996-base published 1997	11.0	48	71
1996-base published 2000	9.2	48	76
2001-base published 2003	5.4	71	86
2001-base published 2005	4.2	81	90
2006-base published 2008	3.6	76	90
2006-base published 2010	3.6	71	90
2006-base published 2012	4.5	76	86
15 years out from base population			
1996-base published 1997	14.5	43	76
1996-base published 2000	11.5	52	71
2001-base published 2003	7.3	57	86
2001-base published 2005	5.9	71	86
20 years out from base population			
1996-base published 1997	18.0	38	62
1996-base published 2000	14.3	29	57

1. ALBA projections were published for the first time in Nov 2012 (2006-base). Projections for ALBAs prior to this were derived retrospectively by aggregating area unit projections. For ease of presentation, table 1.8 and figures 1.12, 1.13, and 1.14 refer consistently to the ‘published’ year across all of the ALBA projection series. However, the ‘published’ year corresponds to the year of the final month of completion of the area unit projections up until the 2006-base published 2010 series. Additionally, before 2006, area unit projections were not published as official projections.

Source: Stats NZ

Figure 1.12

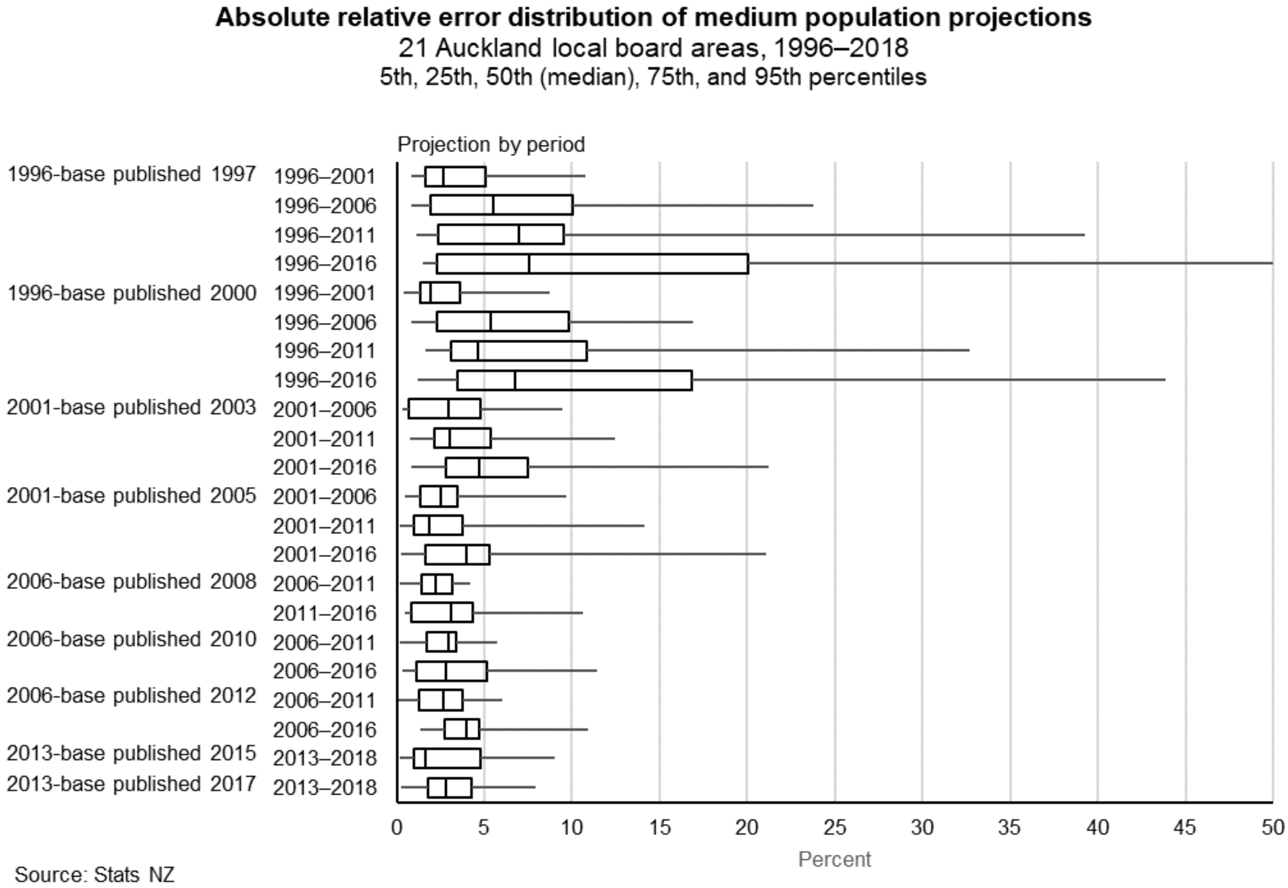
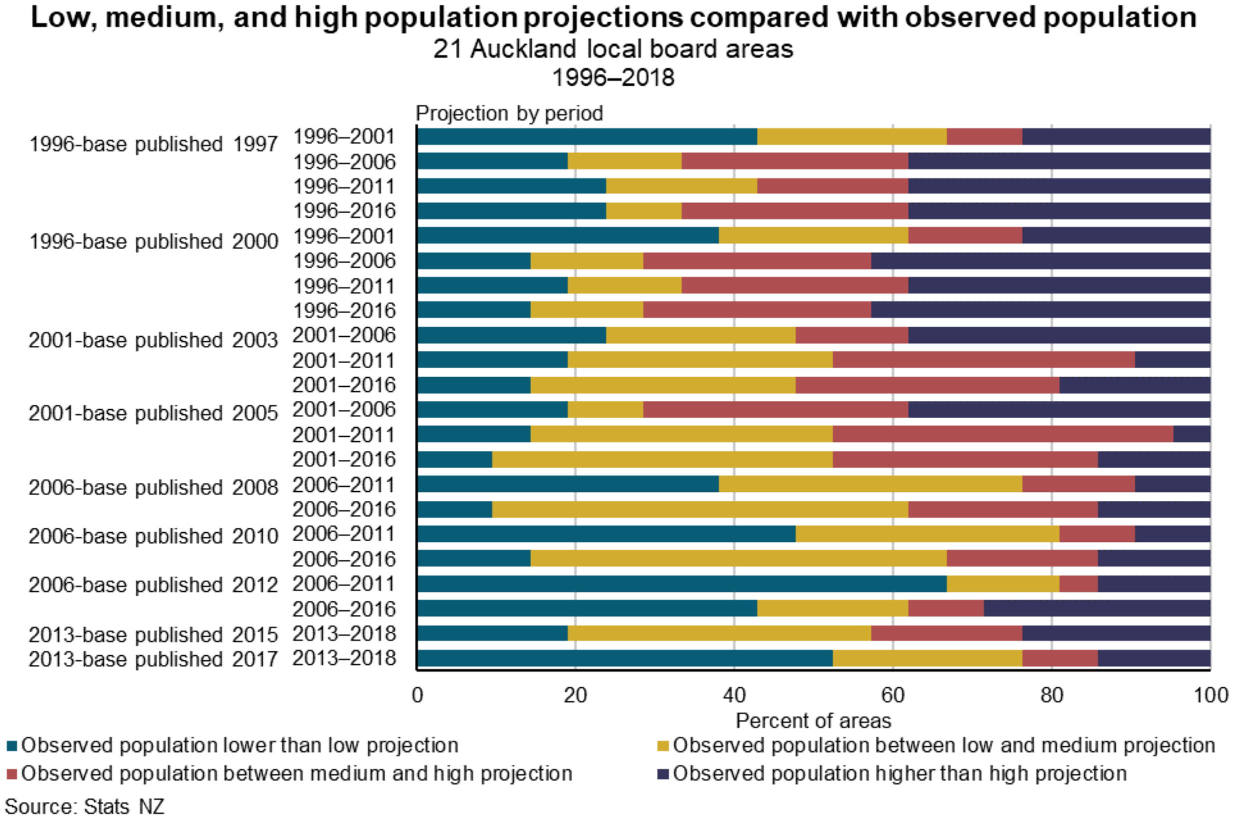


Figure 1.13 indicates the extent to which the collective 'low-medium-high-growth projections' encapsulate the observed population. This ranged from 57 percent of ALBAs for 2013–18 (2013-base published 2015) to just 19 percent of ALBAs for 2006–11 (2006-base published 2012).

Note that for ALBA projections, only those published after 2012 (the 2006-base published 2012, 2013-base published 2015, and 2013-base published 2017) can be used for the accuracy assessment against 'low-medium-high-growth' projections. The ALBA projections for earlier periods were derived from area unit projections. Historical 'medium' projections for ALBAs can be approximated by aggregating area unit projections. However, the low and high projections for area units were designed to be plausible scenarios for the geography, but are unlikely to remain plausible after aggregating to higher geographies (Stats NZ, 2016). We include all of the earlier ALBA projections series results in figure 1.13 for completeness with this caveat on interpretation of the 'low-medium-high-growth projections' results applying for ALBAs.

Figure 1.13

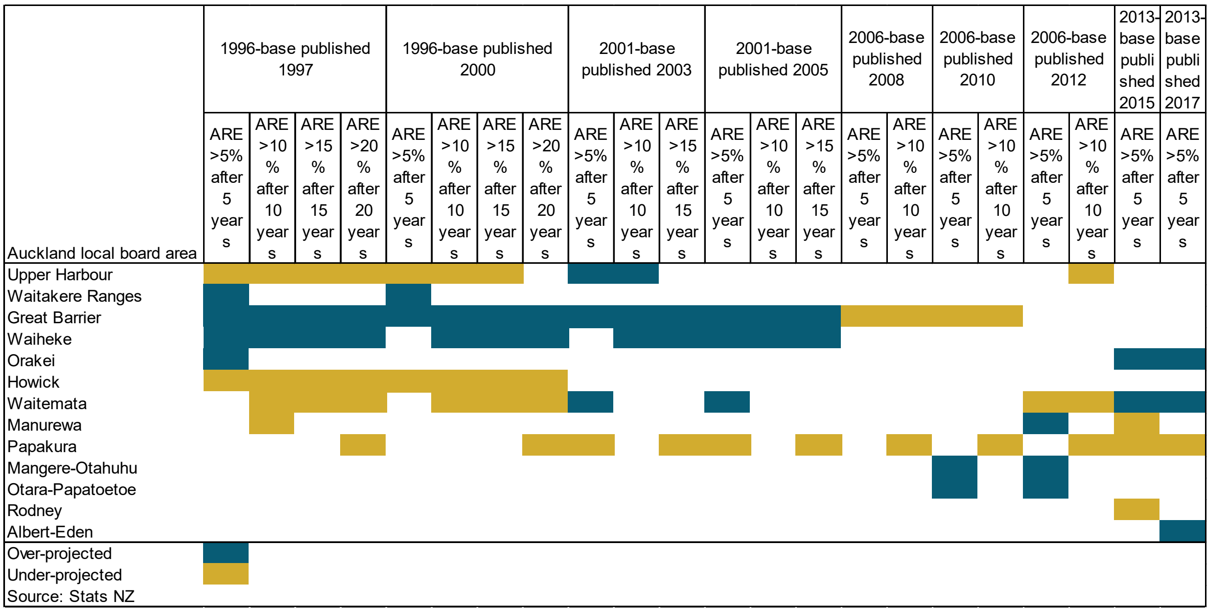


Outliers

It can be revealing to analyse ALBAs with particularly large relative errors (eg AREs >5 percent after five years, >10 percent after 10 years, >15 percent after 15 years, >20 percent after 20 years). The choice of thresholds is arbitrary and is simply a way to highlight the largest AREs.

Figure 1.14

Auckland local board areas with largest absolute relative errors of medium population projections, 1996–2018



These large AREs generally did not persist, except for the smaller Great Barrier and Waiheke ALBAs. Papakura was also consistently under-projected. Generally, the large AREs were addressed in subsequent projections (ie, by incorporating new population estimates, especially those based on new census counts).

There are some interesting reversals in the larger Upper Harbour and Waitemata ALBAs between over- and under-projection. Both areas had substantial net migration between 1996 and 2018 but projecting the precise level is difficult due to the high volume and volatility of the flows. In such cases, it would seem prudent to reflect this uncertainty in wider alternative ‘low’ and ‘high’ net migration levels (in deterministic projections).

Discussion and future work

Analysing the accuracy of our historical population estimates and projections has both retrospective and prospective value. It allows us to not only gain an understanding of the accuracy of historical data, but also to look for potential limitations in our methods and measures and find ways to ameliorate these limitations in the future. This is especially important when considering the accuracy of other derived measures which use population estimates as the base population (eg, as in our projections).

Given accuracy tends to deteriorate over time, away from the base, the accuracy results are a useful baseline for questions around how frequent a census should be. This is important as Stats NZ considers options for transforming the census (McNally & Bycroft, 2015).

Accuracy is not obviously deteriorating for more recent estimates and projections, but nor is it obviously improving. However, accuracy is partly affected by the prevailing demographic conditions. During periods of rapid population change, such as those resulting from high net migration gains, relative errors tend to be higher.

It is important to reiterate the nature of Stats NZ's projections. Demographic projections produced by Stats NZ are not predictions or forecasts. The projections indicate future population and change if the stated assumptions apply over the projection period. The value of population projections is therefore less defined by whether or not they match reality, but more about whether they are plausible and useful to customers at the time they are published. Projections form a basis for developing reasonable expectations about the future; help focus attention on potential events, risks, and opportunities; and allow people and policymakers to plan and make decisions accordingly.

Customers who use Stats NZ's estimates and projections should understand from this accuracy evaluation the importance of using the latest available estimates and projections. Estimates are revised in the light of new information, especially when new census data allows subnational population estimates to be recalibrated. Projections get updated following the availability of new population estimates and other demographic data.

No one should make decisions using estimates and projections that have been superseded by new information. In addition, while this assessment of past accuracy can be used to indicate the magnitude and pattern of future inaccuracies, there is no inevitability about the magnitude or direction (under or over) of those future inaccuracies.

Future developments

Net migration

The accuracy of net migration estimates and net migration assumptions have a major impact on the accuracy of population estimates and projections, respectively.

Since 2012, we have primarily used three data sources to estimate net migration for subnational areas:

- international 'permanent and long-term' migration statistics
- primary health organisation (PHO) enrolment data
- Inland Revenue (IR) tax data.

Several additional data sources were then used to validate the estimates, including:

- electoral enrolment data
- residential building consents
- information provided by local councils
- data on specific population sub-groups (ie, defence force personnel, prison population, and tertiary students).

Several developments also have the potential to improve our net migration estimates and provide more detailed breakdowns for external and internal migration.

External migration

In January 2019, Stats NZ adopted an outcomes-based measure of international migration, using an underlying 12/16-month rule. The 12/16-month rule allows border crossings to be classified as migrant or non-migrant crossings based on the actual time a traveller spends in New Zealand or overseas after crossing the border.

Migrant arrivals are people living overseas who spend 12 months or more within a 16-month period in New Zealand, regardless of their citizenship and legal entitlement to be in New Zealand. Similarly, migrant departures are people living in New Zealand who spend 12 months or more within a 16-month period outside of New Zealand, regardless of their citizenship. For more information about the new migration measure and the 12/16-month rule, see Migration data transformation.

We expect an outcomes-based migration measure to improve the accuracy of estimates and projections. The previous intentions-based migration series was a good indication of migration trends, but generally understated both migrant arrivals and migrant departures (see International migration estimates extended back to 2001). As a result, actual net migration (migrant arrivals minus migrant departures) was sometimes lower, and sometimes higher, from that suggested by measuring intentions.

Due to the delay in the release of 2018 Census data, the 2013-base population estimates were rolled forward to 2019 at the national level, but with initial revisions back to 30 June 2013 to incorporate the new outcomes-based migration measure. These estimates have been used in the analysis in this accuracy report.

The revision has also been applied to the 2018 and 2019 subnational population estimates. They are consistent with the revised national series which are based on the 2013 census usually resident population count, but reflect the 2018 Census subnational population distribution. For the purpose of analysis in this accuracy report, we used the initial 2013-base subnational population estimates which are independent of the 2018 Census results.

The outcomes-based subnational migration method has been adopted in the production of 2018-base subnational population estimates, and will be assessed after the 2023-base ERP is available.

See Outcomes versus intentions: Measuring migration based on travel histories and Estimated resident population 2018: Data sources and methods for more details.

Internal migration

In 2018 we demonstrated an approach for directly measuring movements of residents within New Zealand using linked administrative data in the IDI (see Internal migration estimates using linked administrative data: 2014–17).

This new method has been adopted in the production of 2018-base subnational population estimates, with continuous improvements to address coverage and timeliness limitations. This method will also be assessed after the 2023-base ERP is available.

Base population

Measurement of the base population as the starting point for all our estimates and projections is subject to various potential inaccuracies. However, we have introduced several developments across each step in the process of creating the ERP.

2023 Census

Census, as the cornerstone of population estimates and projections, is recognised as a critical data source (Bycroft, 2006). The 2023 Census strategies were designed to maximise census coverage (participation in the census) and response (valid answers to the census questionnaire). These strategies remain important even though census processes are being modified and transformed within a combined census model – merging data collected through the 2023 Census with alternative data sources. This approach was first introduced in the 2018 Census to mitigate the lower-than-expected level of responses.

Developing the combined model by design means we can consider the overall combination of both response and administrative records in our design, and that design will be driven by customer needs for output quality and understandable and statistically robust methods. Using a combined census model for the 2023 Census | Stats NZ has more detailed information about the 2023 Census model.

See 2023 Census for more details.

2023 Post-enumeration Survey

The Post-enumeration Survey (PES) is conducted by Stats NZ following each census to evaluate how complete the census coverage was. We then use data from the PES to provide more complete estimates of the population (ERP) (Stats NZ, 2021c).

The 2023 PES continues to make improvements to the way it measures census coverage. Significant changes from the 2018 PES include a move away from using sex-based questions to gender-based questions, more focus on te ao Māori and accessibility to these communities, and further improvements to design, methodology, and quality assurance.

See 2023 Post-enumeration Survey for more details.

Census transformation and administrative population census (APC)

The census transformation programme is investigating alternative ways of producing small-area population, social, and economic statistics.

In 2012, Stats NZ published Transforming the New Zealand Census of Population and Dwellings: Issues, options, and strategy (Stats NZ, 2012), which sets out a plan for transforming the census. The report proposed a short-term focus on modernising the current census model, with a long-term goal to develop a new census model based on government administrative data.

The APC is part of the ongoing census transformation programme looking at the potential for a future census based on administrative data supported by sample surveys.

The long-term objective of the APC is to merge census information with official population estimates. Under this approach, there would be only one target resident population measure.

Other advantages over a traditional census are that results can be compiled annually, and the underlying unit-record data is inherently longitudinal.

See Census transformation programme for more details.

Glossary and abbreviations

absolute error (AE) – the difference, irrespective of direction (positive or negative), between the estimated/projected population and observed population in a given year.

absolute relative error (ARE) – the percentage difference, irrespective of direction (positive or negative), between the estimated/projected population and observed population in a given year, relative to the observed population.

administrative data – data resulting from interactions of individuals and groups with government agencies. Data can be collected for, or as part of, service delivery, an application, membership, regulation, or registration. For example, administrative records are maintained to regulate the flow of goods and people across borders, to respond to the legal requirements of registering births and deaths, and to administer benefits such as pensions, or obligations such as taxation. Data is typically collected with a specific administrative purpose in mind, without regard to its statistical use.

ALBA – Auckland local board area. Given the size of Auckland region/city, which accounts for over one-third of New Zealand's population, it is necessary to provide some intermediate disaggregation between the region/city and area units. ALBAs provide this. They came into existence on 1 November 2010. There are 21 ALBAs based on latest boundaries.

assumption – a statement about a future course of behaviour (eg, fertility, mortality, migration) from which demographic projections (eg, of populations) are derived.

average annual population change – an annualised population growth rate calculated as a constant annual rate of population change over the stated period (ie, assuming geometric growth rates to allow for compounding growth).

base population – the starting population, usually distributed by age and sex, from which population estimates and projections are derived.

census usually resident population count – a count of all people who usually live in a given area and are present in New Zealand on a given census night. The census usually resident population count of New Zealand excludes visitors from overseas and excludes New Zealand residents who are temporarily overseas.

For subnational areas, the count excludes visitors from overseas and elsewhere in New Zealand (people who do not usually live in that area), but includes residents of that area who are temporarily elsewhere in New Zealand on census night (people who usually live in that area but are absent).

cohort component method – a method of estimating or projecting the population by updating the size of each age-sex group in the base population, for deaths and migration within each age-sex group, during the period between the base date and a given date. New birth cohorts result from births between the base date and the given date.

de facto population – a statistical basis for a population in terms of those present in a given area at a given time.

deterministic projection – a single projection from a given set of assumptions (eg about fertility, mortality, migration).

error – the difference between the estimated/projected population and observed population in a given year y. For population estimates, this is commonly referred to as the **intercensal discrepancy**.

estimate – an indication of the historical demographic characteristics (size and composition) of population, families, households, or labour force, typically using recorded data.

Stats NZ's population estimates are produced using data from the most recent Census of Population and Dwellings, updated for estimates of the components of demographic change (births, deaths, and net migration) since that census.

estimated population in year y – the estimated ERP at the reference date for measurement of the accuracy of estimated populations in year y, eg, 30 June 2018; also referred to more generally as the 'original estimates', or more specifically for a given inter-censal period such as 2013–2018 as the '2013-base population estimates in 2018' – these estimates are based on the census usually resident population count at census night Tuesday 5 March 2013, with updates for :

- net census undercount (as measured by the Post-enumeration Survey)
- residents temporarily overseas on census night
- births, deaths and net migration between census night and 30 June 2018
- reconciliation with demographic estimates at the youngest ages.

estimated de facto population – an estimate of all people present in a given area at a given date. The estimated de facto population of New Zealand includes all people present in New Zealand and counted by the census (census night population count). This estimate includes visitors from overseas who are counted on census night but excludes New Zealand residents who are temporarily overseas. This estimate includes no adjustment for net census undercount.

For subnational areas, the estimate includes visitors from overseas and elsewhere in New Zealand (people who do not usually live in that area) but excludes residents of that area who are temporarily elsewhere on census night (people who usually live in that area but are absent).

The estimated de facto population at a given date after census includes births, deaths, and net migration (arrivals less departures) of people during the period between census night and the given date.

estimated resident population (ERP) – an estimate of all people who usually live in a given area at a given date. The estimated resident population of New Zealand includes all residents present in New Zealand and counted by the census (census usually resident population count), residents who are temporarily overseas (who the census does not attempt to count), and an adjustment for residents missed or counted more than once by the census, or counted in error (net census undercount) as measured by the Post-enumeration Survey (PES). Visitors from overseas are excluded.

For subnational areas, the estimate excludes visitors from overseas and elsewhere in New Zealand (people who do not usually live in that area) but includes residents of that area who are temporarily elsewhere on census night (people who usually live in that area but are absent).

The estimated resident population at a given date after census includes births, deaths, and net migration (arrivals less departures) of residents during the period between census night and the given date.

forecast – a single prediction of what the population will be at a given date.

intercensal – the period between census dates. More generally, it can refer to the period between census years (eg, between population estimates at 30 June of one census year and 30 June of the next census year).

intercensal discrepancy – the difference between population estimates produced before the census and population estimates rebased after the census. It is a measure of the accuracy of population estimates and is the net combined effect of various factors including inaccuracies in:

- the census counts at the beginning and end of the period
- the adjustments to derive population estimates (from census counts) at the beginning and end of the period
- the components of population change (births, deaths, migration) during the period.

net census undercount (NCU) – the difference between undercount and overcount. It is usually expressed as a percentage of what should have been the complete count rather than as a percentage of what was counted.

net migration – the difference between arrivals and departures.

observed population change – change in population between two dates as measured by census counts or population estimates. Where available, the estimated resident population is used in preference to census counts because the former provides the best available measure of the resident population, as it includes allowances for net census undercount and residents temporarily overseas.

observed population in year y – the final revised or rebased ERP at the reference date for measurement of the accuracy of estimated and projected populations in year y, eg, 30 June 2018; also referred to more generally as the 'revised estimates'.

over-estimate – an estimated population that was higher than the observed population at the corresponding reference date.

over-projection – a projected population that was higher than the observed population at the corresponding reference date.

post-enumeration survey (PES) – the Post-enumeration Survey (PES) is conducted to measure the completeness of the census count. This allows population data produced after the census to be updated for the people who are missed in the census, counted more than once, or counted in error, as measured by the PES.

projected population in year y – the projected ERP at the reference date for measurement of the accuracy of estimated populations in year y eg, 30 June 2016, or 30 June 2018. In this report we evaluate the accuracy of population projections for 11 sets of national projections since the 1991 Census, and nine sets of sub-national projections since the 1996 Census for regions, territorial authority areas, and ALBAs. Projection series used in this report use descriptors which tell us the base projection year, and which year the projections were published in; for example in table 1.4 the term '2016-base published 2016' is included – this means that the mid-range national population projections referred to have the 2016 national ERP as their base and were published in 2016. In contrast, in table 1.6 the term '2013-base published 2017' is included – this means that the medium subnational population projections referred to have the 2013 subnational ERPs as their base and were published in 2017.

projection – an indication of the future demographic characteristics (size and composition) of population, families, households, or labour force based on an assessment of past trends and assumptions about the future course of demographic behaviour (eg, fertility, mortality, migration, living arrangement type, labour force participation).

region – regional council (region). There are 16 regions, as there were when constituted in 1989, although they have had several boundary changes since then.

relative error (RE) – the difference between the estimated/projected population and observed population in a given year, relative to the observed population.

resident population concept – a statistical basis for a population in terms of those who usually live in a given area at a given time.

residents temporarily overseas (RTO) – people who usually live in New Zealand but are overseas for a period of less than 12 months. Most RTOs on census night are away on holiday or business and return to New Zealand within a few weeks of departure.

stochastic (probabilistic) projection – a projection that varies randomly according to the probability distributions of the assumptions (eg about fertility, mortality, migration). Multiple stochastic projections (or simulations) are typically summarised by percentiles to indicate the distribution of values (eg projection results).

TA – territorial authority (cities and districts). Through amalgamation, the number of TAs has reduced from 74 in 1989 to 73 in 2006, then to 67 in 2010 until now.

under-estimate – an estimated population that was lower than the observed population at the corresponding reference date.

under-projection – a projected population that was lower than the observed population at the corresponding reference date.

References

Alho, J M (1997). Scenarios, uncertainty and conditional forecasts of the world population. *Journal of the Royal Statistical Society*, *160*(1), 71-85.

Australian Bureau of Statistics (2009). ABS data quality framework. Retrieved from www.abs.gov.au.

Bell, M, & Skinner, J (1992). Forecast accuracy of Australian subnational population projections. Journal of the Australian Population Association, 9(2), p207–235.

Australian Productivity Commission (2005). Economic implications of an ageing Australia. Retrieved from www.pc.gov.au.

Bycroft, C (2006). Challenges in estimating populations. New Zealand Population Review, 32(2), 21–47.

Keilman, N (1997). Ex-post errors in official population forecasts in industrialized countries. *Journal of Official Statistics*, *13*(3), 245-277.

Keilman, N (2007). UK national population projections in perspective: How successful compared to those in other European countries? Population Trends, 129, 20–30.

McNally, J, & Bycroft, C (2015). Quality standards for population statistics: Accuracy requirements for future census models. Retrieved from www.stats.govt.nz.

Office of National Statistics (2013). Guidelines for measuring statistical output quality. Retrieved from http://webarchive.nationalarchives.gov.uk.

Statistics Canada (2002). Statistics Canada's Quality Assurance Framework. Retrieved from www.statcan.ca.

Stats NZ (2007). Principles and protocols for producers of Tier 1 statistics. Retrieved from www.stats.govt.nz.

Stats NZ (2008). *How accurate are projections? An evaluation of Statistics New Zealand population projections, 1991–2006.*

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

Stats NZ (2016). How accurate are population estimates and projections? An evaluation of Statistics New Zealand population estimates and projections, 1996–2013. Retrieved from www.stats.govt.nz.

Stats NZ (2021a). National and subnational period life tables: 2017–2019. Retrieved from www.stats.govt.nz.

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

Stats NZ (2021c). Why do we do the Post-enumeration Survey (PES), and why is it hard? Retrieved from www.stats.govt.nz.

Stats NZ (2022). Māori population under-estimation in 2013: Analysis and findings. Retrieved from www.stats.govt.nz.

United Nations (1998). *Recommendations on statistics of international migration: Revision 1*. New York: Author.

Wilson, T (2007). The forecast accuracy of Australian Bureau of Statistics national population projections. Journal of Population Research, 24(1), p91–117.

Woods, C, & Dunstan, K (2014). Forecasting mortality in New Zealand. Statistics New Zealand Working Paper No 14–01.