

# **Session 10: Assessing the quality of population estimates and projections**

Cat Anderson & Kim Dunstan

# Revision Cycles

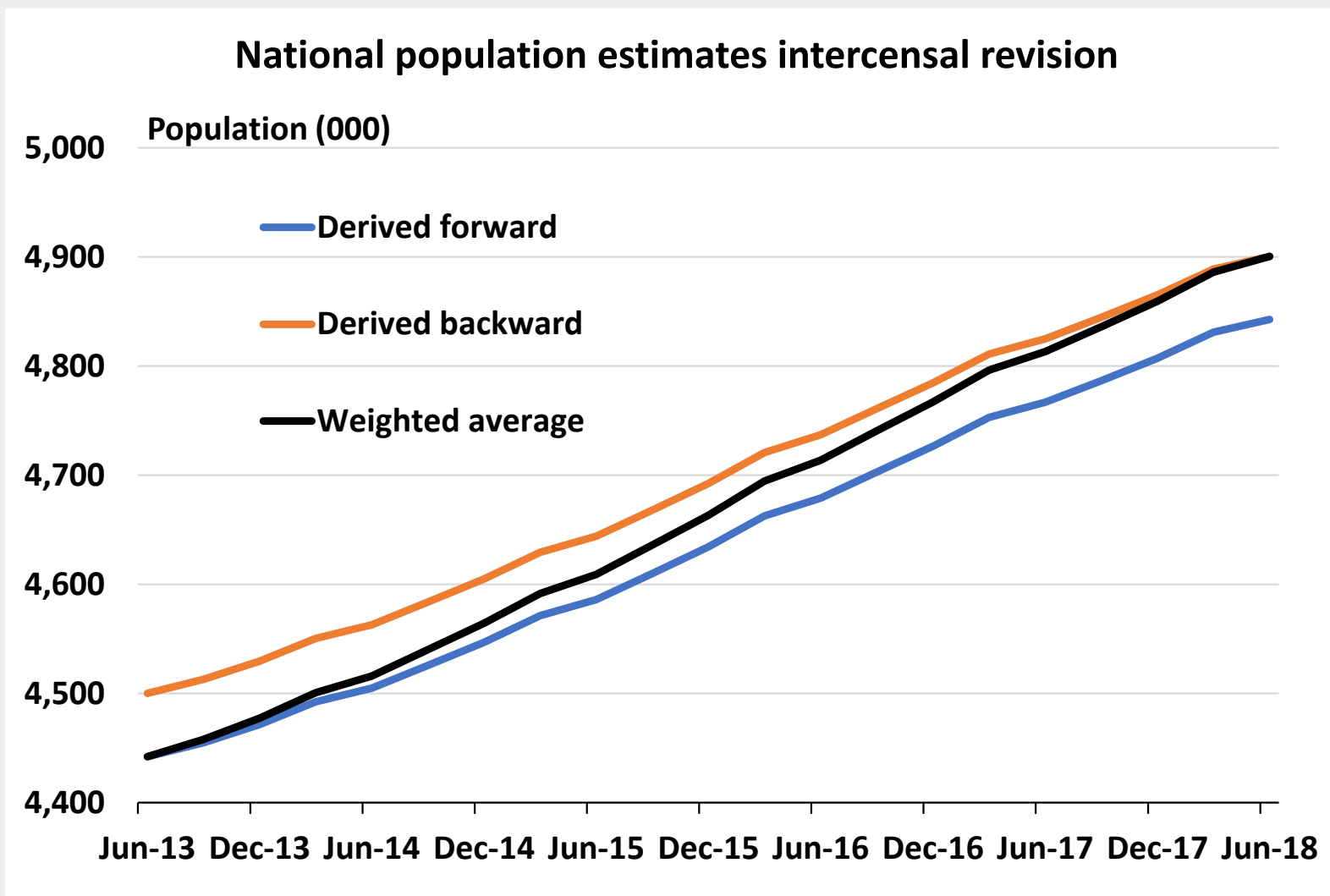
# Intercensal revisions

- Progressive updates to estimates as more data becomes available
  - Typically more accurate estimates of births, deaths and migration
- Revision date should be notified to customers in advance
- Advantage of regular release and revision cycle
  - For example, an annual release where the previous year's estimate is also updated
- Revisions have downstream impacts
  - Series using population denominators also need to be revised
- Data management: versioning files/programs to avoid confusion/error

# Post-censal revisions

- Update after new base population (eg, census count) becomes available
- Intercensal estimates need to be revised to give coherent time series
- Generally earlier intercensal periods will not be further revised
- Revision date should be notified to customers in advance
- Data management: versioning files/programs to avoid confusion/error

# Post-censal revision – weighted average of forward and backward estimates



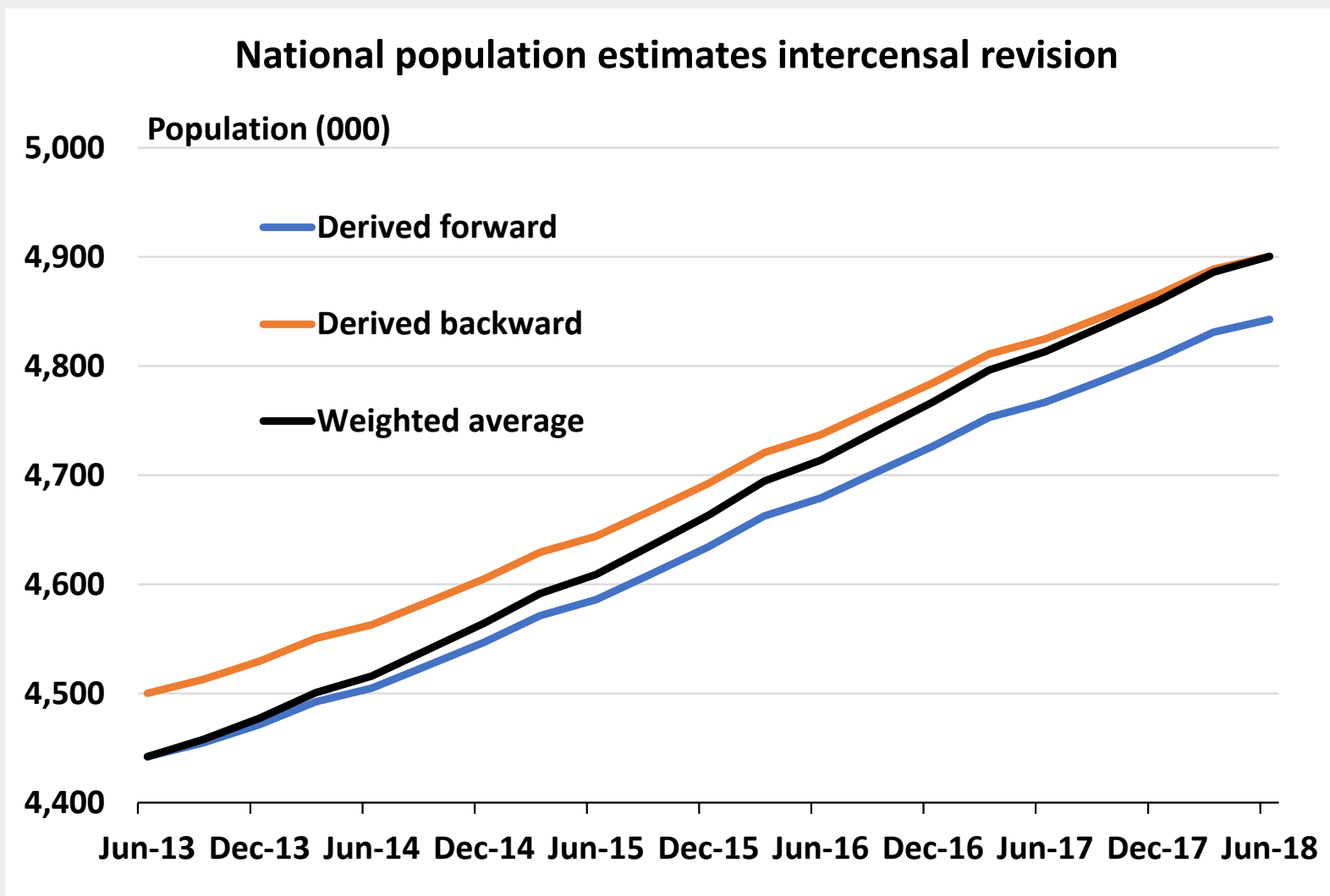
# Data quality dimensions

- Accuracy
  - Comparison with true values; errors and revisions
- Relevance
  - Demographic breakdowns
- Interpretability
  - Explanations of methods and assumptions
- Timeliness
  - Frequency and lags of updates
- Coherence
  - Use of best practice methods; consistency over time
- Accessibility
  - Cost or technology barriers

# Accuracy: Error

- Numerical difference between the estimated/projected population and observed population in a given year
- Also called '**intercensal discrepancy**' in the case of estimates
- $E = E_y - O_y$             or             $E = P_y - O_y$
- Where:
  - $E$  = Error
  - $E_y$  = estimated population in year y
  - $P_y$  = projected population in year y
  - $O_y$  = observed population in year y
- An error of 500 indicates the estimate/projection was 500 higher than the observed population for that year
- An error of -500 indicates the estimate/projection was 500 lower than the observed population for that year

# Post-censal revision – weighted average of forward and backward estimates





# Accuracy: Relative Error

- Percentage difference between the estimated/projected population and observed population in a given year
- Also called '**intercensal discrepancy**' in the case of estimates

- $$RE = \frac{E_y - O_y}{O_y} \times 100 \quad \text{or} \quad RE = \frac{P_y - O_y}{O_y} \times 100$$

- Where:

- RE = Relative error (percent)
- $E_y$  = estimated population in year y
- $P_y$  = projected population in year y
- $O_y$  = observed population in year y
- An error of 5% indicates the estimate/projection was 5 percent higher than the observed population for that year
- An error of -5% indicates the estimate/projection was 5 percent lower than the observed population for that year

# Examples - Estimates

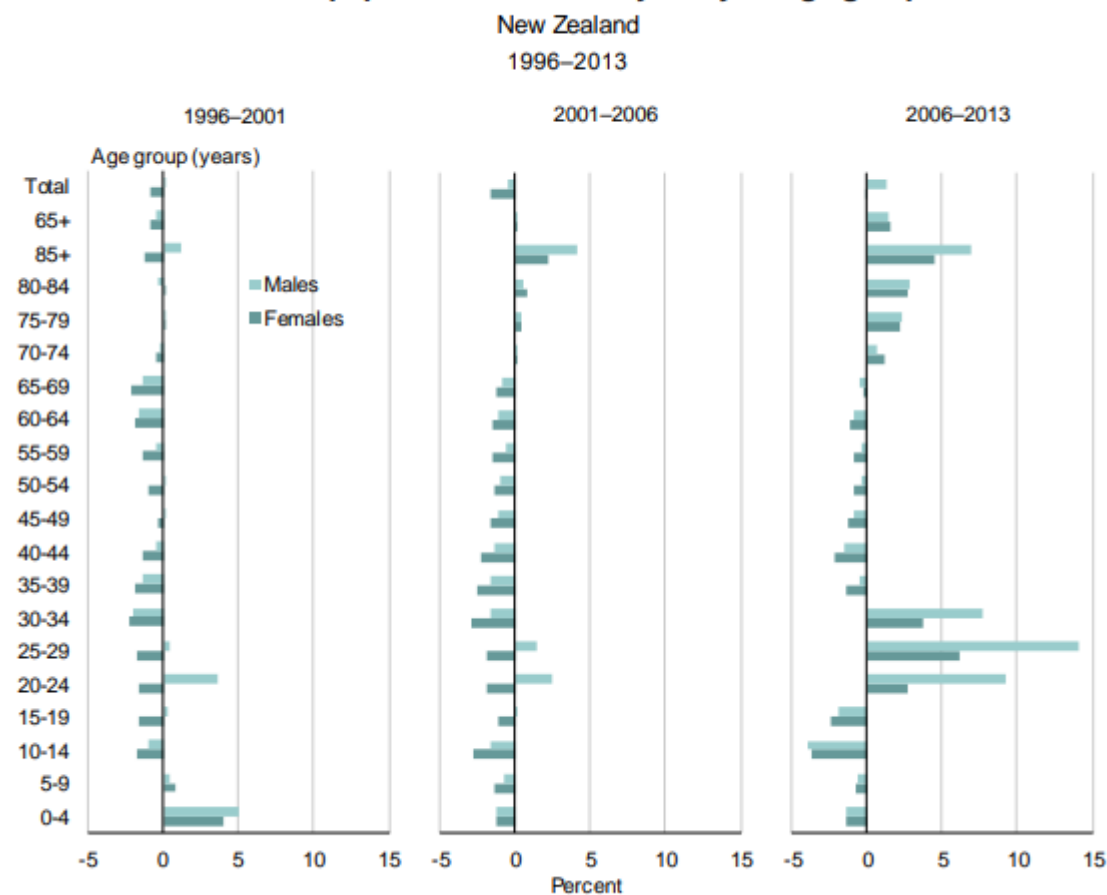
**Table 1.1**  
1.1 Error and relative error of population estimates for New Zealand, 1996–2018

Error and relative error of population estimates for New Zealand 1996–2018		
Intercensal period	Error	Relative Error (%)
1996–2001	-13,900	-0.4
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

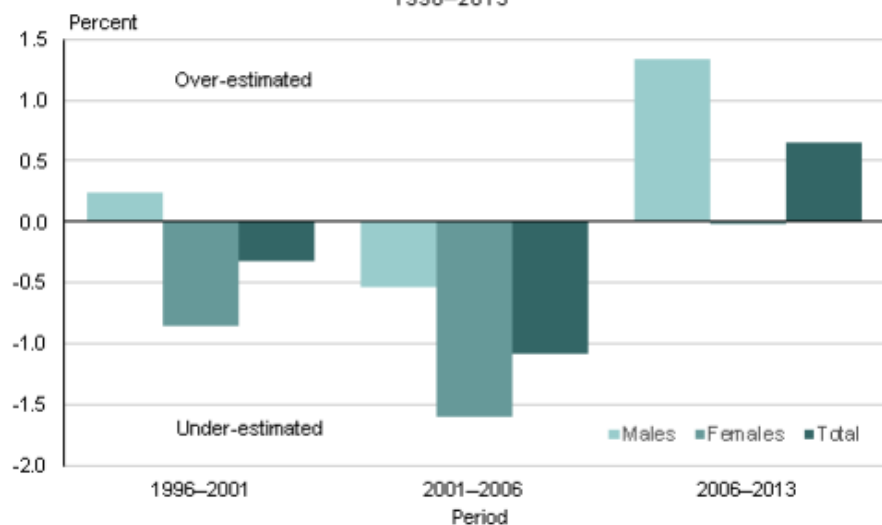
**Figure 5.2**  
Relative error of population estimates by five-year age group and sex



Source: Statistics New Zealand

**Relative error of population estimates by sex**

New Zealand  
1996–2013



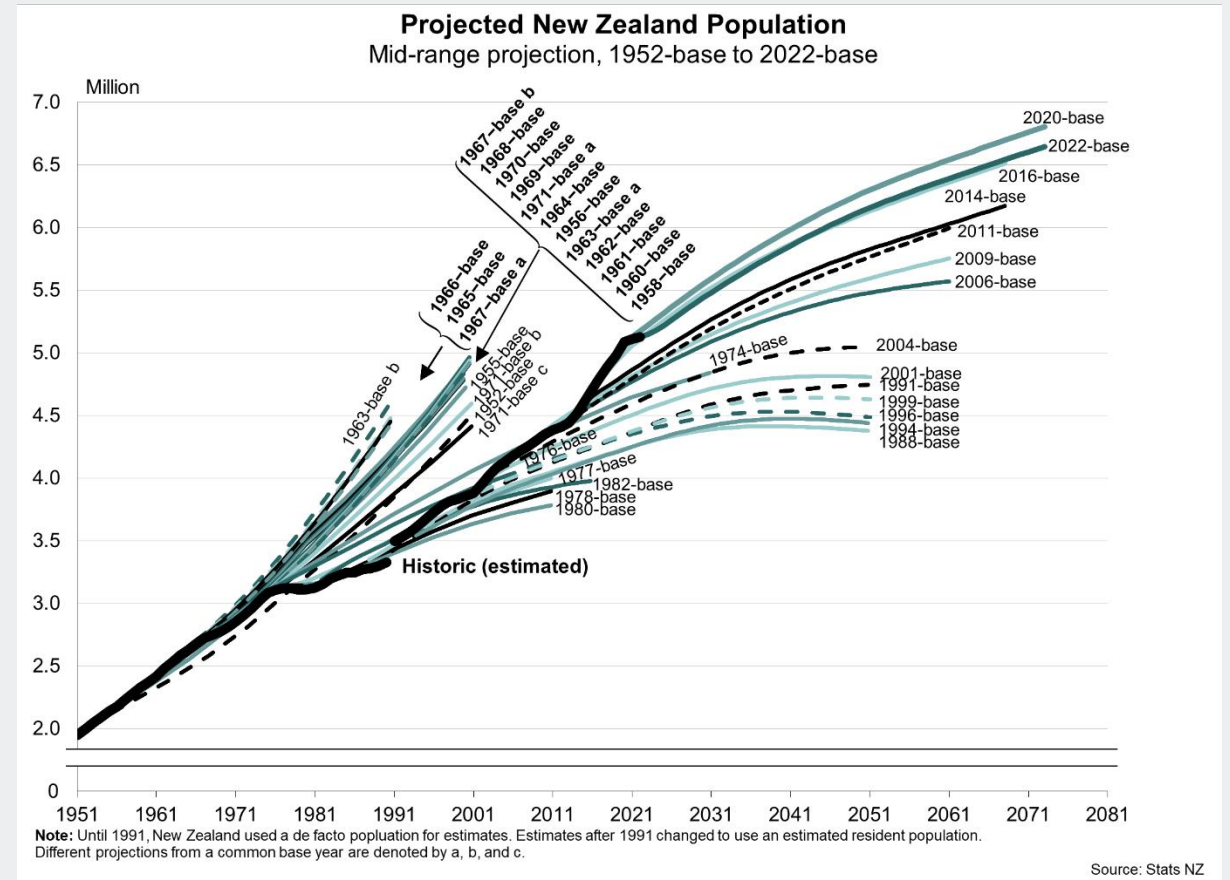
Source: Statistics New Zealand

# Examples - Projections

**Table 7.1**

Error and relative error of mid-range population projections New Zealand 1991–2013			
Projection	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
1994-base published 1994	1996	-42	-1.1
	2001	-33	-0.9
	2006	-193	-4.6
	2011	-281	-6.4
1996-base published 1997	2001	38	1.0
	2006	-128	-3.1
	2011	-212	-4.8
	1999-base published 2000	2001	-3
2006		-159	-3.8
2011		-230	-5.2
2001-base published 2002	2006	-75	-1.8
	2011	-136	-3.1
2004-base published 2004	2006	-32	-0.8
	2011	-66	-1.5
2006-base published 2007	2011	9	0.2
2009-base published 2009	2011	42	1.0

Source: Statistics New Zealand



# Sources of error

- For estimates, the intercensal discrepancy is the net combined effect of inaccuracies in:
  - census counts at the beginning and end of the period
  - any adjustments to derive population estimates (from census counts) at the beginning and end of the period
  - components of population change (births, deaths, migration) during the period.
- For projections, the error is the net combined effect of inaccuracies in:
  - census counts at the beginning and end of the period
  - any adjustments to derive population estimates (from census counts) at the beginning and end of the period
  - assumptions about the components of population change (births, deaths, migration) – or the underlying rates – during the period.

# Reference

- Stats NZ (2016). [How accurate are population estimates and projections?](#) An evaluation of Statistics New Zealand population estimates and projections, 1996–2013.
- Stats NZ (2023 - forthcoming). [How accurate are population estimates and projections?](#) An evaluation of Statistics New Zealand population estimates and projections, 1996–2018.
- Evaluating accuracy takes time, but can be useful:
  - Answering customer enquiries
  - Better understanding the strengths and weaknesses of estimates/projections and the underlying methods
  - Identifying improvements to data sources and methods