Background

Population estimates are fundamental to the conduct of governments, distribution of funding, and population-based policy and service delivery. The current approach to population estimation in Australia has performed well to date and has enabled the production of quality population estimates. However, the environment in which the Australian Bureau of Statistics (ABS) operates is changing and there are growing challenges to the current approach (similar to national statistical offices around the world). At the same time, the increasing prevalence, quality and understanding of integrated administrative data offers an opportunity to strengthen the sustainability of the production of population estimates.

As part of the current approach to population estimation, a new population base is created every five years through a combination of the Census and the Post-Enumeration Survey (PES) in a dual capture framework. Utilizing administrative component data, a quarterly series of estimated population moves off this base for the next five years, until the cycle repeats. Estimating population via a dual or multiple capture framework is generally regarded as the gold standard, and is the approach taken by most comparable National Statistical Organizations (NSOs). The method works by taking one capture of the population (in Australia's case the Census) and adjusting it by the results of a second, independent capture of the population (in Australia's case the PES). The same framework can be extended to multiple captures of the same population.

The current dual capture approach, which produces a modified (via weighting) dual system estimator (DSE), faces two main challenges: growing demand for small area population estimates, and the risk of low response and resulting response correlation bias.

PES is designed for the primary purpose of supporting population estimation at the state/territory level and Aboriginal and Torres Strait Islander population estimation at the national level. There is substantial demand for small area data and sub-national Aboriginal and Torres Strait Islander data, however, the PES sample size is inadequate to support these uses.

PES requires high response rates to minimize response correlation bias (where people or groups missed in Census also have a higher chance of being missed in PES). If response rates drop, the risk of correlation bias increases, as does the risk of lower quality population estimates. To continue to produce quality population estimates in this context, the ABS must develop methodologies and/or introduce new data sources to improve protection against low response rates and correlation bias.

Exploring a triple system estimator

The data integration landscape in Australia continues to evolve. The Person-Linked Administrative Data Asset (PLIDA) is increasingly well known and understood. PLIDA is a secure data asset which aims to cover all people who were resident in Australia at any point from 2006 onwards. It combines information on health, education, government payments, income and taxation, employment, and population demographics (including the Census) over time.

There is international precedent for using integrated administrative data for population estimation or as part of the Census and associated coverage survey. Integrated administrative data has been used for: enumeration support, quality assurance and data treatment; substitution for missing data; and adding or replacing data sources in a multiple capture system. ABS is exploring how

integrated administrative data could be used alongside Census and PES in a triple capture approach.

A triple system estimator (TSE) using Census, PES and PLIDA would:

- protect against low response and resulting correlation bias
- provide contingency options for low Census and/or PES response
- provide opportunities for data improvement by substantially increasing the amount and detail of data being used for population estimation
- be adaptable and enable use of new or different data sources in the future.

Initial findings

ABS has undertaken some initial analysis using 2021 Census, 2021 PES and PLIDA data. PLIDA data was scoped to the population of interest, that is persons resident in Australia at Census time, using the Administrative Snapshot of Population and Housing scoping method (Administrative data snapshot of population, methodology, 30 June 2021).

When stratifying the population by broad age groups (0-14, 15-29, 30-49. 50 and over) and state/territory, the TSE population estimate was within the 95% confidence interval of the published estimate for seven of the eight states/territories. When stratification included Indigenous status, the TSE was within the 95% confidence interval of the published estimate for four of the eight states/territories.

At a national level and with Indigenous status included in the stratification, the TSE produced a population estimate that was about 280,000 persons above the published estimate and outside its 95% confidence interval. The over-estimate for the Aboriginal and Torres Strait Islander population (about 150,000), which accounted for most of the national population over-estimate, was due to small PES sample sizes in Aboriginal and Torres Strait Islander strata.

Dropping indigenous status from stratification reduced the instability of the TSE estimate and resulted in a population estimate 124,000 higher than the official figure.

Next steps

To better understand how TSE could improve or otherwise impact population estimates, ABS plans include:

- Improving TSE methodology alignment with ABS requirements (and comparability of these tests) by incorporating PLIDA data into the current weighting and estimation system
- Linking Census-PES-PLIDA prior to the deletion of Census name and address following the 2026 Census and PES
- Methodological development and testing of the linkage strategy
- Undertaking privacy assessments and data governance requirements

Based on the results of these analyses, ABS will decide on the role of linked administrative data and TSE for population estimates.