University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Wilson, T;Grossman, I;Temple, J

Title:
Local area population projections dataset for Australia

Date:
2022-10-01

Citation:
Wilson, T., Grossman, I. & Temple, J. (2022). Local area population projections dataset for Australia. Data in Brief, 44, https://doi.org/10.1016/j.dib.2022.108559.

Persistent Link:
https://hdl.handle.net/11343/327075

License:
CC BY-NC-ND

Data in Brief 44 (2022) 108559



Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib



Data Article

Local area population projections dataset for Australia



Tom Wilson*, Irina Grossman, Jeromey Temple

Melbourne School of Population and Global Health, The University of Melbourne, 207 Bouverie St, Melbourne, VIC 3010, Australia

ARTICLE INFO

Article history:
Received 3 June 2022
Revised 1 August 2022
Accepted 24 August 2022
Available online 30 August 2022

Dataset link: Population projections data (Original data)

Keywords:
Population projections
SA3 areas
Australia
Population change
Population ageing

ABSTRACT

A new dataset of population projections for local areas of Australia is described. The areas comprise SA3 areas of the Australian Statistical Geography Standard, which mostly range in population between 30,000 and 130,000. The projections are launched from the 2020 Estimated Resident Populations published by the Australian Bureau of Statistics and extend out to 2035. They are available by sex and five year age group up to 80–84 with 85+ as the final age group and in five year projection intervals. The projections were prepared using the synthetic migration cohort-component model, a new model for population projections which requires much less input data than conventional projection models, and therefore involves much lower costs and production time. Despite this, a recent evaluation demonstrated respectable forecast accuracy, and greater accuracy than equivalent simple projection models producing populations by age and sex. The age-sex projections are constrained to independent age-sex national projections and local area projections of total populations. The dataset consists of local area projections for the whole of the country which is consistent in methods, input data, and projection outputs due to the use of one model. This is rare in Australia because local area projections are most commonly prepared by individual State/Territory Governments using different methods, data sources, projection assumptions (which can be influenced by State/Territory population policies), and time periods. These

* Corresponding author.
E-mail address: wilson.t1@unimelb.edu.au (T. Wilson).
Social media: @AppliedDemog (T. Wilson), @IriGrossman (I. Grossman), @JeromeyTemple (J. Temple)

https://doi.org/10.1016/j.dib.2022.108559
2352-3409/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

nationally consistent projection data should be useful for a wide range of local area planning, policy, and research purposes, such as childcare demand, school enrolments, power and water usage, aged care provision, store and business site selection, living arrangements and household projections, labour force projections, and transport modelling.

© 2022 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Specifications Table

Subject	Social science
Specific subject area	Local area population projections
Type of data	Table
How the data were acquired	Simulated
Data format	Excel and csv files of population projections
Description of data collection	Population projections were produced as the output of the synthetic migration cohort-component population projection program. This program is available at https://doi.org/10.6084/m9.figshare.19372784.v1
Data source location	Country: Australia Geography: SA3 areas of the Australian Statistical Geography Standard 2016. Launch year Estimated Resident Populations for 2020: available via ABS Data Explorer https://explore.data.abs.gov.au/ Model migration schedule: model fitted to SA3 area internal migration data available via ABS TableBuilder https://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder Local crude migration turnover rates: estimated using SA3 area census migration data available via ABS TableBuilder https://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder State Government total population projections for local areas within urban regions: New South Wales: Data supplied directly by the New South Wales State Government Victoria: https://www.planning.vic.gov.au/land-use-and-population-research/victoria-in-future Queensland: https://www.qgso.qld.gov.au/statistics/theme/population/population-projections/regions South Australia: https://data.sa.gov.au/data/dataset/population-projections-for-sa Western Australia: https://www.wa.gov.au/government/document-collections/western-australia-tomorrow-population-forecasts#sa2-australian-statistical-geography-standard Tasmania: https://www.treasury.tas.gov.au/economy/economic-data/2019-population-projections-for-tasmania-and-its-local-government-areas Northern Territory: https://treasury.nt.gov.au/dtf/economic-group/population-projections Australian Capital Territory: https://apps.treasury.act.gov.au/__data/assets/word_doc/0006/1305582/ACT-Population-Projections-Paper-FINAL.doc
Data accessibility	Local area total population projection model and data: https://doi.org/10.6084/m9.figshare.19744798.v1 All synthetic migration projection model input data available from https://doi.org/10.6084/m9.figshare.19744816.v1 Repository name: Figshare Data identification number: 10.6084/m9.figshare.19769560.v1 Direct URL to data: https://doi.org/10.6084/m9.figshare.19769560.v1 Instructions for accessing these data: click on URL for direct access to the data

Value of the Data

- The projections should prove useful by scientists, government, and business, for a wide range of planning, budgeting and research purposes.
- For scientists, the projections could provide inputs to specialist modelling in a wide range of projects, for example: projections of local labour force supply; changing living arrangements and household types; the number of older persons with chronic health conditions; the potential number of people affected by sea-level rise as the global climate changes; and transport modelling.
- For analysts in government and business, the data could be useful for assessing future: childcare needs; the number of school places required; dwelling construction; health and education staffing; transport demand; water, sewerage and power demand; aged care places; store, office or factory location decisions; and the identification of potential new markets for goods and services.
- Recent research [1] has demonstrated that the low-cost synthetic migration cohort-component model produces population projections with respectable accuracy, and greater accuracy than other simple models producing population projections by age and sex. They should be useful for studies which need projections up to 15 years ahead and with a 5 year age group breakdown.
- The dataset represents a nationally consistent and publicly available set of local area projections for the whole of Australia.

1. Data Description

The projections are available in the Excel file 'SA3 2020-based projections data.xlsx'. The dataset consists of population projections for SA3 areas of Australia for 30th June 2025, 2030 and 2035, along with 2020 Estimated Resident Populations (ERPs) published by the Australian Bureau of Statistics which form the jump-off (or launch) year populations of the projections. SA3 areas were designed by the Australian Bureau of Statistics to usually contain populations between 30,000 and 130,000 people [2], though some have populations which lie well outside this range. The SA3 areas in this dataset are those defined by the 2016 Australian Statistical Geography Standard [2]. All populations in the dataset are defined as the population usually resident in each local area (and thus *exclude* temporary visitors, and *include* residents temporarily away). The projections were produced by sex and five year age group from 0–4 to 80–84, with 85+ being the final age group.

A few SA3 areas have been merged in the projections dataset due to very small populations. Three SA3 areas in New South Wales were merged into 'NSW Remainder'. These were: Blue Mountains – South, Illawarra Catchment Reserve, and Lord Howe Island. The three SA3 areas in the Other Territories region, Christmas Island, Cocos (Keeling) Islands, and Jervis Bay, were also merged into simply 'Other Territories'.

The Excel file consists of three worksheets. Years are located in columns, while other variables occupy rows. The worksheet 'Totals' reports the total populations of each SA3 in 2020 (the jump-off year Estimated Resident Populations) and projections for 2025, 2030, and 2035. The 'AgeSex' worksheet contains the SA3 area populations by sex and five year age group for 2020, 2025, 2030, and 2035. The third worksheet, 'Components', lists the projected population accounts for each projection interval, 2020–25, 2025–30, and 2030–35. Included in this sheet are start-of-interval populations, births, deaths, net migration, and end-of-interval populations.

The data in each of the Excel worksheets is also provided in csv files for users who prefer this format. They are organised as flat files, with just one column of population data. The variables are summarised in Table 1.

Table 1
Variables in the local area projections dataset.

Variable Name	Summary	Type	Values
SA3 code	Unique ID assigned to each SA3 area by the ABS [1]. Additionally, the NSW Remainder has been assigned a code of 19999 and Other Territories 90109.	Integer (5 digits)	e.g., 10102 for Queanbeyan
SA3 area name	Names of the SA3 areas, the NSW Remainder area, and Other Territories	String	e.g., Queanbeyan
Year	The projected year. Projections are available in 5 year intervals out to 2035	Integer	2020, 2025, 2030, 2035
Population	The projected number of individuals for the indicated year	Numeric	e.g., the projected total population for Canterbury in 2035 is 172,229
Sex	Sex of the projected population	String	Males, Females
Age group	Projections are available for 18 five year age groups from 0–4 to 85+	String	e.g., 5–9
Period	This variable describes five year periods and is used in the Components worksheet/csv file. The number of births, deaths, and the net migration to occur within each period, for each SA3 region, are provided	String	2020–2025 2025–2030 2030–2035
Component	The Component variable is used in the Components worksheet/csv file. It indicates whether the value refers to births, net migration, mortality, start-of-period population, or end-of-period population	String	Births Deaths End-of-period population Net migration Start-of-period population
Value	The Value variable is used in the Components worksheet/csv file. The variable refers to the value of the projected component given by the components variable.	Numeric	e.g., -3328 for net migration in Mackay during the 2025–30 period.

2.1. Experimental Design, Materials and Methods

2.1. Projection Model

The population projections in the dataset were prepared using a projection program which implements the synthetic migration cohort-component model [1]. This new approach to projections incorporates the conceptually strong feature of modelling inwards and outwards migration (rather than net migration) but without the high data requirements of multiregional local area projection models [3]. It enables the projections preparation process to be much quicker and easier. Recent testing of the model by creating local area 'projections' for past periods and then evaluating these projections against actual Estimated Resident Populations demonstrated a respectable level of accuracy, and greater accuracy than simple benchmark models [1]. For example, three sets of test projections 10 years ahead produced Median Absolute Percentage Errors calculated across all age-sex population groups of 8.5%, 9.5%, and 7.1% compared to 9.3%, 11.9%, and 7.7%, respectively, with the Hamilton-Perry model [1]. The modelling process is summarised in Fig. 1.

The heart of the synthetic migration model consists of a bi-regional cohort-component model. The population of a cohort at time t+5 years is calculated as the size of the same cohort at time t minus the deaths that occur in the five year interval, minus outward migration,

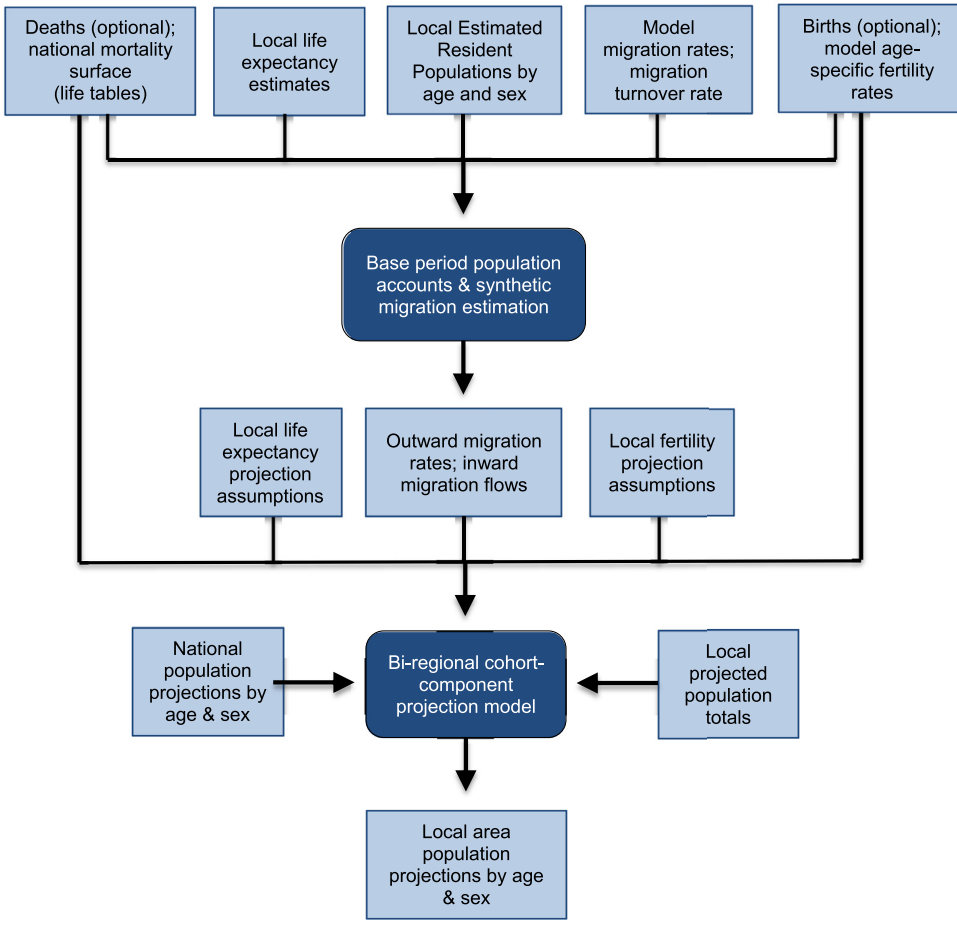

Fig. 1. An overview of the population projections process. Source: modified from Fig. 1 in Wilson & Grossman (2022) [4].

plus inward migration. For example, the male population aged 50–54 in 2025 is projected as the male population aged 45–49 in 2020, minus the deaths which occur to this cohort between 2020 and 2025, minus outward migration, plus inward migration.

To keep the model as simple as possible and data requirements to a minimum, migration is only modelled between each local area and the rest of the world. The model does not incorporate separate international and domestic migration streams. Inward migration flows refer to migration from the rest of the world *in* to the local area; outward migration flows describe migration *out* of the local area to the rest of the world. This two-way division of geographical space gives the model its bi-regional name. Outward migration is calculated by multiplying an outward migration rate by the local area population-at-risk. Inward migration is projected directly as numbers (rather than via rates) because the rest of the world population is not modelled.

Births and deaths are projected in a standard way for cohort-component models. Deaths are projected using age-specific death rates multiplied by populations. Births, which form the initial population of new cohorts, are projected by multiplying age-specific fertility rates by female populations-at-risk.

The local area age-sex projections are constrained by two sets of independent projections:

(1) Total local area populations.
(2) National projections by age and sex.

Previous research has shown how constraining projections to more aggregate independent projections can improve forecast accuracy [5–8].

2.2. Input Data and Projection Assumptions

The SA3 area population projections were launched from the 30th June 2020 Estimated Resident Populations by age and sex published by the ABS [9]. Fertility assumptions were formulated in terms of the Total Fertility Rate (TFR), the average number babies born per woman. Local area TFRs for the base period were estimated using the Hauer and Schmertmann xTFR indirect estimation method [10] and assumed to remain constant into the future. A fixed age profile of fertility, calculated from the average of recent national fertility rates, was used to generate age-specific fertility rates consistent with each area's TFR. Mortality assumptions were set in terms of life expectancy at birth. Life expectancy at birth by sex for each local area was assumed to the same as national projected life expectancy. Age-specific death rates were generated using a mortality surface of national life tables [11], from which a set of age-specific death rates consistent with each assumed life expectancy value was extracted.

The projection model uses synthetic migration flows as its most important data input. In many cases, real migration data is unavailable, too sparse, too expensive, or too time-consuming to adjust and smooth. Synthetic migration data is estimated over a five year base period prior to the projections. The estimation calculations are undertaken separately for each local area and consist of seven steps [1].

(1) Preliminary migration turnover (inward plus outward migration combined) by age and sex is estimated as the product of model migration rates and the base period population-at-risk. The model rates for the current projections were created by fitting a model migration schedule to all inter-SA3 area internal migration flows recorded by the 2016 Census. The same model schedule is used for all areas.
(2) Total migration turnover for the local area is calculated by multiplying a crude (all-age) migration turnover rate by the total base period population-at-risk. For these projections, local area-specific crude migration turnover rates were estimated from 2016 Census migration data.
(3) The preliminary migration turnover values by age and sex are scaled to sum to total migration turnover. The estimates are then split into two to produce preliminary values of separate inward and outward migration flows by age and sex. The inward and outward migration flows consist of the same values at this stage.
(4) Base period total net migration is calculated as an accounting residual, which is total population change over the five year base period after accounting for births and deaths.
(5) The preliminary inward and outward migration flows by age and sex are adjusted proportionally upwards or downwards to be consistent with the total net migration value. The two sets of migration flows are now different, but the age profiles are still identical.
(6) Base period residual net migration by age and sex is calculated.
(7) In the final step, inward and outward migration flows by age and sex are adjusted proportionally up or down to be consistent with the age-sex net migration values. This results in inward and outward migration age profiles which now differ in shape.

For use in the projection program, outward migration rates by age and sex are calculated by dividing the synthetic outward migration flows by the base period populations-at-risk. For inward migration, migration flows are applied directly. Both the outward migration rates and inward migration flows are effectively adjusted during the running of the projection program

due to the application of the two projection constraints of local area total population projections and national population projections by age and sex.

To obtain a cohort-component projection which sums over age and sex to the independent total population, initial projected inward and outward migration flows are adjusted so that the start-of-interval population plus births minus deaths plus inward migration minus outward migration equals the independent total population projection. Fertility and mortality rates do not change; only migration is adjusted. Then migration is further adjusted so that net migration by age and sex sums over all local areas to the independent national projection of net migration by age and sex. A customised iterative proportional fitting routine is employed to apply these constraints.

National projections were created using a standard cohort-component model. These projections assumed a long-run TFR of 1.65, and life expectancy increasing to 87.6 years for females and 84.0 years for males by 2035. Net overseas migration was assumed to increase to 225,000 per annum by 2025–26 and remain constant thereafter.

Local area projected population totals were prepared using an approach which takes the average of four extrapolative models. The four constituent models of this averaged model are:

(i) a constant share of population model in which local area populations are projected as the jump-off year proportion of the national population multiplied by the national projected population.
(ii) a linear/exponential model which projects local area population using linear extrapolation if base period growth is positive and exponential extrapolation if it is negative.
(iii) a share of growth model in which projected local population growth from the linear/exponential model is adjusted to match projected national population change and,
(iv) a modified exponential model [12] in which the exponential model is subject to floor and ceiling limits to avoid excessive growth or decline.

For local areas located within State and Territory Greater Capital City Statistical Areas (GCCSAs), the averaged model total population for the GCCSA was used, but it was distributed to local areas according the most recent State or Territory Government projections. This allowed State/Territory Government expectations about future residential dwelling growth to be indirectly incorporated into the projections. Outside the capital city regions, the averaged extrapolative local area population totals were used directly.

Ethics Statements

Ethics approval was not required. All data used in the preparation of the local area projection dataset was aggregate secondary data.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Population projections data (Original data) (ResearchData).

CRediT Author Statement

Tom Wilson: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing; **Irina Grossman:** Data curation, Validation, Resources, Writing – original draft; **Jeromey Temple:** Writing – review & editing.

Acknowledgments

Funding: This research was supported by the Australian Research Council Centre of Excellence in Population Ageing Research (project number CE1101029).

References

[1] T. Wilson, Preparing local area population forecasts using a bi-regional cohort-component model without the need for local migration data, Demogr. Res. (2022).
[2] Australian Bureau of StatisticsAustralian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas, July 2016, Australian Bureau of Statistics, 2016 https://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.001 (accessed 15 May 2022), doi:10.4054/DemRes.2022.46.32.
[3] A. Rogers, Multiregional Demography: Principles, Methods and Extensions, Wiley, 1995.
[4] T. Wilson, I. Grossman, Evaluation of a new simplified population projection model: a case study of local government area projections in Tasmania, Australia, J. Reg. Stud. (2022) forthcoming.
[5] J. Baker, D. Swanson, J. Tayman, The accuracy of Hamilton–Perry population projections for census tracts in the United States, Popul. Res. Policy Rev. 40 (2021) 1341–1354, doi:10.1007/s11113-020-09601-y.
[6] M. Reinhold, S.L. Thomsen, Subnational population projections by age: an evaluation of combined forecast techniques, Popul. Res. Policy Rev. 34 (2015) 593–613, doi:10.1007/s11113-015-9362-0.
[7] J. Tayman, D.A. Swanson, J. Baker, Using synthetic adjustments and controlling to improve county population forecasts from the hamilton–perry method, Popul. Res. Policy Rev. 40 (2021) 1355–1383, doi:10.1007/s11113-021-09646-7.
[8] T. Wilson, Evaluation of alternative cohort-component models for local area population forecasts, Popul. Res. Policy Rev. 35 (2016) 241–261, doi:10.1007/s11113-015-9380-y.
[9] Australian Bureau of StatisticsSA3 area Estimated Resident Populations, Australian Bureau of Statistics, 2021 https://explore.data.abs.gov.au/.
[10] M.E. Hauer, C.P. Schmertmann, Population pyramids yield accurate estimates of total fertility rates, Demography 57 (1) (2020) 221–241, doi:10.1007/s13524-019-00842-x.
[11] T. Wilson, Evaluation of simple methods for regional mortality forecasts, Genus 74 (2018) 14, doi:10.1186/s41118-018-0040-z.
[12] J. Baker, X. Ruan, A. Alcantara, T. Jones, K. Watkins, M. McDaniel, M. Frey, N. Crouse, R. Rajbhandari, J. Morehouse, J. Sanchez, M. Inglis, S. Baros, S. Penman, S. Morrison, T. Budge, W. Stallcup, Density-dependence in urban housing unit growth: an evaluation of the Pearl-Reed model for predicting housing unit stock at the census tract level, J. Econ. Soc. Meas. 33 (2-3) (2008) 155–163, doi:10.3233/JEM-2008-0301.