

Can reliable sub-state emigration estimates be obtained from an administrative population dataset? The case of the Person Level Integrated Data Asset (PLIDA) of Australia

Tom Wilson*

Gin Wu

Aude Bernard

Neil Argent

Anthony Kimpton

* Corresponding author. Contact details at the time the paper was written:

Address: Melbourne School of Population and Global Health, The University of Melbourne, 207 Bouverie St, Melbourne, Vic 3010, Australia; email: wilson.t1@unimelb.edu.au

ORCID: [0000-0001-8812-7556](https://orcid.org/0000-0001-8812-7556)

Current contact details:

Email: advanceddemographicmodelling@gmail.com

Funding

This work was supported by the Australian Research Council Discovery Projects scheme (Project no. DP200100760).

Conflict of Interest

The authors declare that they have no conflicts of interest

Abstract

To fully understand and forecast local and regional population change, it is necessary to obtain data for all the demographic components of change – births, deaths, internal migration, and overseas migration. In Australia, and many other countries, high-quality and geographically detailed data is available for most demographic components, but not emigration. The Australian Bureau of Statistics publishes good quality estimates of immigration and emigration at the national and State scales, but available sub-state emigration data is heavily modelled and suffers from quality issues. Good quality sub-state immigration data is available from the census. The aim of this study was to evaluate the quality of sub-state emigration data for Australia extracted from a large administrative population dataset, the Person Level Integrated Data Asset (PLIDA), which links several administrative datasets. We compared State and sub-state emigration flows from PLIDA to available emigration data, as well as sub-state immigration flows from PLIDA to census-derived immigration flows. We found PLIDA-based emigration flows to be considerably lower in volume than official State emigration estimates, likely due to undercoverage in PLIDA of

temporary migrants and residents of remote regions. We discovered sizeable geographical, temporal, age, and visa/citizenship variations in emigration coverage relative to available emigration statistics. Our view is that PLIDA does not yet allow sub-state emigration estimates of sufficient quality for most uses. But we are hopeful that improvements to the population coverage of PLIDA will allow this in the future.

Key words

Emigration; sub-state; data quality; PLIDA; Australia

1. Introduction

Population estimates and forecasts for local areas and regions are widely used across the public and private sectors. They provide, amongst other things, essential data inputs for planning services and infrastructure, denominators for health and demographic indicators, data to inform government policies and programmes, and information for the business sector on market size and site selection. Some countries (e.g. the Nordic countries; the Netherlands) are fortunate enough to possess population registers which offer spatially detailed demographic data (Poulain and Herm 2013). In Australia and many other countries without such registers, however, the census, usually conducted every 5 or 10 years, provides the basis for population estimates for the years in which the census is held. Population estimates for non-census years, and population forecasts for the future, are based on census-year estimates with allowance for the subsequent demographic components of change – i.e., births, deaths, internal migration, and international migration.

In many high-income countries, subnational data on births and deaths is generally high quality. Reasonable or acceptable quality data on internal migration is often available, and immigration to subnational areas from other countries is often measured by a question in the census and published for a variety of subnational geographies. Unfortunately, data on *emigration* (defined as international moves *to* other countries) from subnational areas is often unavailable, or estimated indirectly using bold assumptions and/or proxy data (e.g. distributing national emigration flows to subnational areas using the distribution of *immigration*). Attempting to estimate emigration as residual population change after all other components have been taken into account is prone to producing implausible or impossible numbers (e.g., ‘negative’ emigration) due to errors in estimating populations and other demographic components (Jensen 2013).

Emigration is also of interest in its own right. It forms a key component of local and regional population change that demographers and other researchers are interested in studying and understanding. It also has policy relevance in ensuring the quantity and quality of labour supply and maintaining populations in remote areas. In Australia and some other countries, regional migration policy aims to attract immigrants to less populated areas, but also retain them over the long run (e.g. Boese 2023; Dyrting et al. 2020; Hugo 2008; Raymer and Baffour 2018). Without good quality sub-state emigration data, it is difficult to determine what proportion and what type of migrants subsequently leave the country, and therefore assess the extent to which these regional migration policies are effective.

The focus of this paper is on emigration from subnational areas, using Australia as a case study. Australia is fortunate in that good quality emigration data is available at the national and State scales. Official statistics on emigration are derived from “various processing systems, passport documents, visa information, and incoming passenger cards” as travellers cross the Australian border (ABS, 2021). However, there is currently no data source which directly measures

emigration for sub-state areas. While some sub-state emigration data *are* available, they possess a number of limitations. The emigration estimates published by the Australian Bureau of Statistics (ABS, 2023a) are heavily modelled due to this lack of direct estimates of emigration at the sub-state scale. The ABS distributes State-scale emigration to local areas “based on Census data on the number of overseas arrivals in the previous year and the proportion of population born overseas, and SEIFA score” with SEIFA being the Socio-Economic Indexes for Areas, an index of relative socio-economic advantage and disadvantage (ABS, 2023a).

Previous academic research on the estimation of sub-state emigration in Australia has produced plausible numbers (Raymer et al. 2020; author reference), but they remain indirect estimates of emigration based on a number of assumptions and proxy measures. In the international literature, relatively little research attention had been directed towards developing robust estimation methods for sub-national emigration. Examples include emigration derived from the sample attrition of large-scale surveys which repeatedly sample people in the same dwelling units after accounting for mortality and internal migration (van Hook et al. 2006); and a regression approach which estimates emigration rates from numerous demographic and socio-economic variables at a large region scale, which is then applied to distribute large region emigration estimates to smaller sub-national areas (ONS 2009). While these methods are useful, a direct measure of sub-state emigration would be preferable, especially for a country like Australia which typically experiences large immigration and emigration flows (ABS 2023b).

In recent years, a new integrated population dataset for Australia has become available to authorised researchers under strict confidentiality provisions. The Person Level Integrated Data Asset (PLIDA), formerly the Multi-Agency Data Integration Project (MADIP), connects many large demographic, economic and social administrative, census and survey datasets at the individual level (ABS 2023c). It is managed by the ABS in partnership with several other federal and State government agencies. The linked data potentially offers new opportunities for understanding social and economic issues – and informing policy responses – which have not been possible previously. Importantly for the purposes of this study, it links individual-level international migration data and information on local area of usual residence (or former usual residence for those emigrating), enabling data on sub-state emigration to be extracted.

This paper addresses the question ‘Is emigration data for sub-state regions calculated from PLIDA of reasonable quality? Exactly what constitutes ‘reasonable’ quality will of course depend on individual users’ needs. We assess the estimates in terms of the extent to which they match the limited directly-measured emigration data as well as a more qualitative judgement of their plausibility in light of existing knowledge about migration trends and patterns. We primarily consider SA4 sub-state regions, large regions in the Australian Statistical Geography Standard (ABS, 2021). They are commonly used for reporting a wide range of social, economic, and demographic data, as well as typical regions for which population projections are prepared (e.g.,

QGSO 2023). There are 88 such regions across Australia (excluding the small islands of the ‘Other Territories’) and they mostly contain populations of between 100,000 and 500,000.

Following this introduction, we describe the PLIDA dataset, methods used to evaluate the PLIDA emigration estimates, and the ABS migration datasets used as comparison data sources. Results are presented in section 3, with section 4 including a discussion of the findings, implications for estimating sub-state emigration, and concluding remarks.

2. Data and methods

2.1. PLIDA

The Person Level Integrated Data Asset (PLIDA) contains linked individual-level data from the census and several government administrative sources. It was established in 2015 as a collaboration by seven government agencies to “access the substantial value inherent in public data” for policy analysis, research, and statistical purposes (ABS 2023c). PLIDA is maintained by the ABS. The ABS collects, cleans, and links data, ensures data security, and provides access to approved researchers under strict privacy controls. Individual data in PLIDA is de-identified and all output must be approved for release by the ABS to ensure it meets disclosure restrictions.

Datasets in PLIDA are not connected directly, but linked via the Person Linkage Spine, which comprises the unique population covered by three data sources: (1) health records from Medicare, (2) social security data from Centrelink, and (3) Australian Taxation Office data. At the time of writing, the spine covered everyone recorded in these three datasets who was resident in Australia at any time between January 2006 and November 2021. Due to the coverage of the three constituent datasets, there is some undercoverage of temporary visa holders who are not in the labour force.

To examine sub-state overseas migration, we used the Department of Home Affairs’ overseas migration traveller records and address data connected via the Person Linkage Spine. Specifically, the Home Affairs data provides the birth month, birth year, sex, country of origin, visa type, date, immigration or emigration movement, and Australian state or territory of where the record was created. We obtained the SA4 region of departure or arrival from the Combined Location Module, which records the date of all residential address updates since 2006. Residential locations are obtained from the triangulation of the three data sources comprising the Person Linkage Spine, which are then aggregated to the Australian Statistical Geography Standard (ASGS) geographic areas so that individuals cannot be identified by their address description or coordinates.

PLIDA is accessed through the ABS DataLab. DataLab provides a secure and isolated research platform for authorised users to view and analyse microdata with analytical software provided by closed network virtual machines. All analytical output is kept in DataLab and must be approved by the ABS before release.

We extracted from PLIDA annual emigration and immigration flows by State/Territory between 2006 and 2020, and data by sex and five-year age group up to age 65+ for the years 2006-10, 2011-15, and 2016-20, with the data grouped into five-year periods to give counts above 10, which is the minimum cell value permitted by the ABS for release. In addition, we obtained the total number of ‘census returned emigrants’ by SA4 region for the same dates as measured by the 2016

Census (living in Australia 5 years before the census in August 2011, overseas 1 year before in August 2015, and back in Australia on census night in August 2016). This is a different type of migration measure to the annual flow data on migration events. It is strictly a transition-type migration measure, defined as living at an address at a specified point in time which is different from the address at an earlier specified point in time (Rees et al. 2000). It omits counts of multiple migrations within the time interval as well as people who are not part of the Australian population on census night and five years earlier.

2.2. ABS international migration data

International migration estimates are published by the ABS for Australia and the States and Territories for quarterly and annual periods (ABS 2023b, 2022). This data includes immigration, emigration and net overseas migration flows by sex and age group. The estimates are derived primarily from information collected by the Department of Home Affairs' Travel and Immigration Processing System (TRIPS) which is based on data from passports, visa applications, and incoming passenger cards. Individuals' incoming and outgoing border movements are linked. Travellers are deemed international migrants by the ABS if they spend at least one year in or out of Australia. A person is counted as immigrating to Australia if they spend a cumulative total of at least 12 months in the country within a 16-month period. Similarly, a person is counted as emigrating if they spend at least 12 months out of 16 outside Australia. This prevents short visits to and from Australia (e.g., international students returning home for a visit) from distorting the international migration statistics.

The ABS international migration estimates are regarded as high quality because they are based directly on border movements, and have good coverage of migration because almost everyone travelling to or from Australia formally crosses the border. In addition, the intercensal error of closure of population accounts for Australia is usually very small, suggesting that demographic data, including international migration, is accurate. We obtained annual State immigration and emigration flows by sex and age group from the ABS Data Explorer tool (<https://explore.data.abs.gov.au/>).

In addition, we obtained 2016 Census data on 'census returned emigrants'. This data consists of people living in a specified region of Australia 5 years before the 2016 Census, living overseas 1 year before the census (having emigrated), and who were back living somewhere in Australia on census night. This data is captured by the Australian census through its questions on place of residence 5 years ago and place of residence 1 year ago. It is a small sub-set of emigrants, but it does at least enable a direct comparison of emigration between the census and PLIDA to be made at the sub-state scale. We obtained total numbers of census returned emigrants at the SA4 scale. This census data is regarded as reasonably good quality, though the estimated non-response rates

to questions on place of usual residence 1 and 5 years ago were, respectively, 8.6% and 8.0% (ABS 2017).

2.3. Evaluation methods

The evaluation of PLIDA-derived emigration data consisted of two fairly simple tests, comparing PLIDA data against existing good quality emigration data sources. The first test involved comparing State-scale emigration totals from PLIDA for calendar years 2006 to 2020 with published ABS estimates of emigration. Emigration data by sex and five-year age group up to age 65+ for the periods 2006-10, 2011-15, and 2016-20 was also compared with equivalent ABS migration data. Importantly for the purposes of this evaluation, the PLIDA-based data was linked to the Combined Location Module in PLIDA to obtain State/Territory of previous residence, and not State/Territory details in the overseas migration data. We examined both absolute emigration numbers and the geographical distribution of emigration.

Second, we compared data from PLIDA which was equivalent to ‘census returned emigrants’ for sub-state SA4 areas from the 2016 Census. Because this is quite a small sub-set of emigrants, only totals were compared because the numbers by age and sex were often too small in PLIDA to meet ABS disclosure controls. Again, we evaluated both numbers and the geographical distribution of emigration.

In addition, we evaluated PLIDA-derived *immigration* estimates for SA4 areas. Although the coverage of immigration data in PLIDA is not guaranteed to be exactly the same as that of emigration, evaluating immigration should give at least an approximate indication of how well emigration is covered in the dataset. We would expect there to be a reasonable degree of correlation between the coverage of immigration and emigration because the variable coverage of different population groups in the dataset should affect both immigration and emigration. We compared the PLIDA immigration data to our own immigration estimates which were calculated by disaggregating ABS State-level immigration flows to SA4 areas using census data on immigration.

3. Results

3.1. State/Territory emigration

Emigration flows from all States and Territories estimated from PLIDA are noticeably lower than the official ABS estimates. Table 1 presents PLIDA-derived emigration estimates as a percentage of official ABS emigration statistics, showing the PLIDA-based numbers to be generally 60-80% of the ABS emigration flows. There is quite strong clustering evident among these percentages, with most States and Territories being within a few percentage points of each other in any particular year. However, the percentages do fluctuate to some extent from year to year, suggesting varying amounts of coverage by PLIDA over time, with coverage generally improving over the first decade of the study period, and then falling slightly in the most recent years. A likely reason for the undercount of emigration is the underrepresentation of recently arrived migrants, temporary migrants, and very remote populations in the Combined Location Module (Bernard et al. 2024). The Northern Territory stands out for having much lower coverage than other jurisdictions, while Tasmania's coverage of emigration is a little higher than other jurisdictions over the last few years of the study period. For the Northern Territory, the undercount most likely stems from the underrepresentation of very remote populations in the Combined Location Module (author reference). This is associated with lower coverage amongst younger age groups and highly mobile populations (who interact less with government data systems), both of which are common demographic characteristics of very remote areas.

Table 1: PLIDA-derived emigration flows by State/Territory† as a percentage of official ABS emigration estimates, 2006-20

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
NSW	55	61	62	63	70	72	77	75	79	78	78	75	77	69	72
Vic	59	63	63	65	68	72	76	77	78	77	76	72	75	68	68
Qld	54	55	60	66	72	73	77	77	81	77	81	72	78	69	77
SA	58	65	64	66	70	74	78	78	78	80	78	77	79	71	69
WA	54	60	61	66	71	74	78	77	80	81	80	78	80	75	74
Tas	67	73	66	69	67	78	86	79	82	84	87	86	92	77	88
NT	23	30	37	40	46	47	50	58	60	53	55	53	53	53	62
ACT	61	66	69	69	68	74	77	79	82	81	79	79	75	70	73
Total	55	60	61	64	69	72	76	76	79	78	78	74	77	69	71

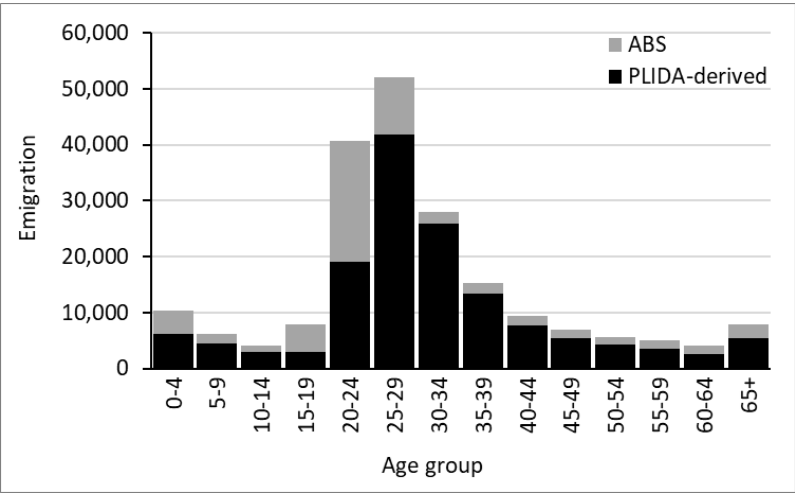
Source: Estimated from data extracted from the Person Level Integrated Data Asset (PLIDA)

† NSW = New South Wales; Vic = Victoria; Qld = Queensland; SA = South Australia; WA = Western Australia; Tas = Tasmania; NT = Northern Territory; ACT = Australian Capital Territory.

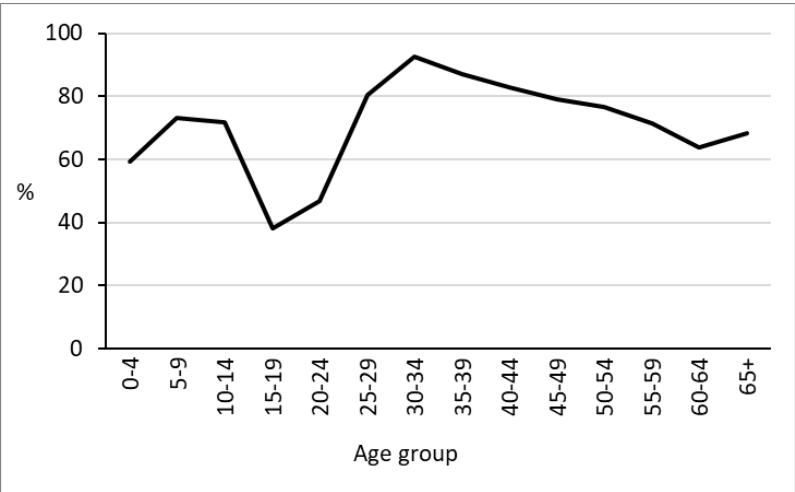
PLIDA-derived emigration estimates by sex and age group were also compared to the official ABS statistics for 2006-10, 2011-15, and 2016-20. Across all three periods, both sexes, and all States/Territories we found a common pattern: lower PLIDA-derived numbers at all ages, with the PLIDA flows as a proportion of the ABS statistics being lowest in the young adult ages. Figure 1, showing male emigration from Victoria for 2016-20, illustrates the typical pattern. It shows

substantial undercoverage in the peak young adult age groups, but much better coverage at ages 30 and above.

While the general relationship between the PLIDA-derived and ABS estimates of emigration by age are similar across sex, time and jurisdiction, the Northern Territory patterns are more dissimilar to other jurisdictions across age and have greater differences between males and females. For Tasmania, the PLIDA-derived emigration estimates are higher than the ABS estimates in 2016-20 for ages 30-39.



(a) Emigration flows, 2016-20



(b) PLIDA-derived emigration as a percentage of ABS emigration statistics

Figure 1: Comparison of PLIDA-derived and official ABS estimates of male emigration by age group from Victoria, 2016-20

Sources: ABS overseas migration statistics; data extracted from the Person Level Integrated Data Asset (PLIDA)

In an evaluation of the coverage of PLIDA, Bernard et al. (2024) found that international students, temporary skilled visa holders, and residents of very remote areas are substantially under-represented in the PLIDA database generally. Could this account for the poor coverage at the young adult ages? To examine this issue, we extracted emigration counts from PLIDA by broad

visa/citizenship categories by state and for the periods 2006-10, 2011-15, and 2016-20. The categories were: permanent visa holders, temporary student visa holders, other temporary visa holders, Australian citizens, and New Zealand citizens. To maintain consistency with the ABS definition of emigration by visa category and citizenship, PLIDA-derived emigration was classified according to “the visa at time of traveller’s overseas migration arrival date” (ABS 2022, Table 3.2, note b). Table 2 presents the PLIDA-derived estimates as a percentage of the equivalent ABS emigration estimates. Unfortunately, we were not able to analyse the coverage of PLIDA-derived emigration flows by visa/citizenship category by age and sex because the ABS does not publish official emigration estimates with this level of detail.

Table 2: PLIDA-derived emigration flows by visa/citizenship status and State/Territory as a percentage of official ABS emigration estimates, 2006-20

	Temporary visa holders			Australian citizens	New Zealand citizens	Total
	Permanent	Student	Other			
2006-10						
NSW	68.6	54.8	38.2	84.8	79.7	66.0
Vic	68.6	50.9	34.5	87.0	81.3	66.5
Qld	71.5	45.1	43.6	84.7	82.5	67.7
SA	70.4	44.6	38.1	91.7	78.0	68.4
WA	73.2	55.3	45.8	86.0	85.1	68.7
Tas	67.8	42.3	35.9	95.9	67.8	71.2
NT	63.5	54.2	35.5	72.6	60.6	54.7
ACT	67.6	46.8	37.0	88.4	64.2	71.3
Total	69.6	51.0	39.4	85.9	81.0	66.9
2011-15						
NSW	71.3	83.2	64.1	85.7	87.7	78.0
Vic	73.1	90.4	53.9	86.4	90.8	77.0
Qld	68.6	74.5	74.0	86.1	92.4	80.6
SA	81.9	77.7	61.7	90.2	90.1	79.0
WA	71.6	92.5	74.9	82.4	96.0	81.4
Tas	68.6	79.0	63.2	100.4	81.2	83.0
NT	58.8	98.1	69.2	79.7	82.8	74.0
ACT	71.6	76.8	62.8	91.2	85.9	81.6
Total	71.8	83.7	65.4	86.0	91.3	78.8
2016-20						
NSW	66.4	78.2	66.5	82.3	89.7	75.2
Vic	72.5	77.3	56.1	82.0	91.7	72.2
Qld	67.5	72.7	74.9	84.8	93.2	79.6
SA	83.0	75.9	60.2	86.7	92.1	76.3
WA	69.9	84.4	75.7	76.6	97.5	79.3
Tas	45.8	101.9	73.8	99.4	95.4	87.1
NT	52.9	116.5	72.6	75.0	87.7	75.3
ACT	70.4	82.4	48.0	89.6	80.3	77.5
Total	69.3	77.7	65.8	82.5	92.6	75.8

Sources: calculated from data extracted from the Person Level Integrated Data Asset (PLIDA) and ABS overseas migration statistics.

Nearly all visa/citizenship-specific emigration flows estimated from PLIDA are lower than official ABS estimates (i.e. the percentages are less than 100%). In the first period, 2006-10, student temporary visa holders were under-represented (51.0%) relative to the population as a whole (66.9%), but this situation changed in the next two periods. The emigration of other temporary visa holders has consistently lower coverage in PLIDA than the population as a whole across all three periods (39.4%, 65.4%, and 65.8% in each period). Australian and New Zealand citizens were relatively well covered in the PLIDA-derived emigration estimates for all three periods. However, the coverage of permanent visa holders was lower. While we cannot provide a definite answer to the young adult under-coverage question, the undercount of other temporary visa holders is likely to be at least part of the answer.

Although PLIDA-derived emigration estimates are lower than the official emigration statistics, one possibility is that they could be used instead to geographically disaggregate the official statistics. We therefore created scaled PLIDA-based emigration estimates for 2006-20 by taking national emigration statistics published by the ABS and disaggregating them to States using the geographical distribution of the original PLIDA-derived emigration estimates. Table 3 presents the percentage errors of these scaled PLIDA estimates, with the final column consisting of the mean absolute percentage error across all years. The errors are mostly quite small, with the exception of the Northern Territory (where the scaled-up PLIDA-derived estimates are too low) and Tasmania (where they are too high).

Table 3: Percentage errors and mean absolute percentage error of scaled PLIDA-derived emigration flows by State, 2006-20

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	MAPE
NSW	-0.2	1.4	0.8	-1.5	0.0	-0.4	0.6	-1.4	-0.3	0.7	0.0	0.9	0.8	-0.3	0.2	0.6
Vic	5.7	5.1	1.8	0.3	-1.6	-0.6	-1.0	0.9	-1.5	-1.5	-2.8	-2.3	-2.7	-1.6	-4.6	2.3
Qld	-2.3	-8.2	-3.0	2.1	2.9	1.5	0.2	1.4	2.8	-0.7	3.6	-2.1	1.7	-0.4	7.3	2.7
SA	4.1	8.3	4.7	2.6	0.7	2.6	2.1	2.2	-1.1	3.0	0.2	4.2	2.9	2.3	-3.2	2.9
WA	-2.7	-0.1	-0.5	2.6	2.4	2.5	2.1	0.7	1.5	4.2	2.4	6.0	3.6	8.3	3.4	2.9
Tas	20.4	21.1	6.7	6.5	-3.0	8.6	12.3	3.7	3.9	8.3	11.5	15.9	19.5	11.7	22.9	11.7
NT	-57.9	-51.0	-40.2	-38.5	-33.8	-35.5	-34.7	-24.2	-24.2	-32.0	-29.4	-27.8	-30.6	-23.4	-12.6	33.0
ACT	9.9	9.7	13.1	7.9	-1.8	2.7	1.1	3.9	3.8	4.1	1.5	7.0	-1.8	0.5	2.6	4.8

Source: Calculated from data extracted from the Person Level Integrated Data Asset (PLIDA) and ABS overseas migration statistics.

Possible explanations for the Northern Territory and Tasmania patterns are not obvious. The Northern Territory has below-average proportions of temporary visa holders among its emigrants (ABS 2022), so undercoverage of temporary visa holders in PLIDA is unlikely to explain much of the undercoverage for this jurisdiction. Table 2 shows that the emigration of permanent residents, Australian citizens, and New Zealand citizens is poorly covered by the PLIDA-derived emigration estimates. It might also be that, because the Northern Territory experiences very high rates of interstate migration, address details in PLIDA are out-of-date for an above-average proportion of

Territory residents. Or it might be due to errors in allocating State/Territory of former residence in the ABS emigration data, especially after the Outgoing Passenger Card, which all travellers leaving Australia had to complete (including address), was discontinued in 2017.

For Tasmania, temporary visa holders make up about the average proportion of the emigration flow, so again, the obvious explanation of temporary visa holder undercoverage seems unlikely to be the main cause. Table 2 shows that Tasmania has very high coverage of Australian citizen emigration and low coverage of permanent residence emigration in PLIDA. Perhaps some Australian citizens shown to be emigrating from Tasmania are really emigrating from other states due to address information being out of date.

We also experimented with distributing ABS national emigration estimates by age and sex across States according to PLIDA emigration numbers by age and sex. This was to determine whether this would produce better PLIDA-derived estimates because of the large variations in PLIDA coverage by age group. Table 4 shows errors of PLIDA-derived estimates of total emigration by State based on the age-sex distribution of emigration on the left side of the table compared to errors of PLIDA-derived estimates of total emigration based simply on PLIDA emigration totals on the right. It shows that distributing national emigration estimates using PLIDA age-sex-specific emigration numbers, unfortunately, does not improve the estimates.

Table 4: Absolute Percentage Errors of PLIDA-derived estimates of State emigration based on distributing ABS national emigration by (a) age and sex and (b) totals

	(a) Estimated using age-sex emigration			(b) Estimated using total emigration		
	2006-10	2011-15	2016-20	2006-10	2011-15	2016-20
NSW	0.8	0.5	0.4	0.1	0.2	0.3
Vic	1.5	1.1	3.1	2.0	0.8	2.8
Qld	0.0	1.8	2.8	1.5	1.0	1.5
SA	4.8	2.8	3.3	4.0	1.7	1.5
WA	2.2	2.4	5.4	0.5	2.2	4.8
Tas	11.4	8.5	15.1	9.7	7.3	15.8
NT	44.3	30.9	27.6	44.4	29.8	26.2
ACT	6.2	3.5	3.1	7.4	3.2	1.8
MAPE	8.9	6.4	7.6	8.7	5.8	6.8

Source: Calculated from data extracted from the Person Level Integrated Data Asset (PLIDA) and ABS overseas migration statistics.

3.2. Sub-state returned emigrants

Figure 2 provides a partial assessment of how well the PLIDA-derived emigration estimates measure sub-state emigration. This shows 2016 Census ‘returned emigrants’ – people who left Australia between 2011 and 2015 from each SA4 region and who had returned by 2016 – and the equivalent PLIDA-derived emigration estimates. We chose 2016 Census data because return migration to Australia was severely disrupted in 2020-21 by COVID border restrictions. As the

graph shows, the two sets of numbers are quite closely aligned ($R^2 = 0.98$), although the PLIDA estimates overall are a little lower than the Census numbers (92% on average).

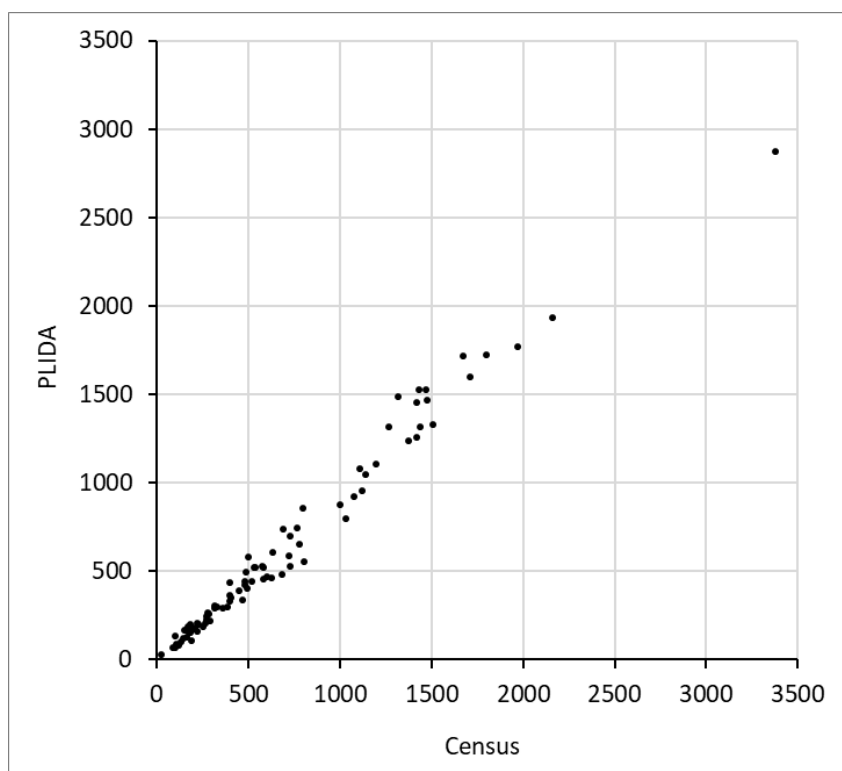
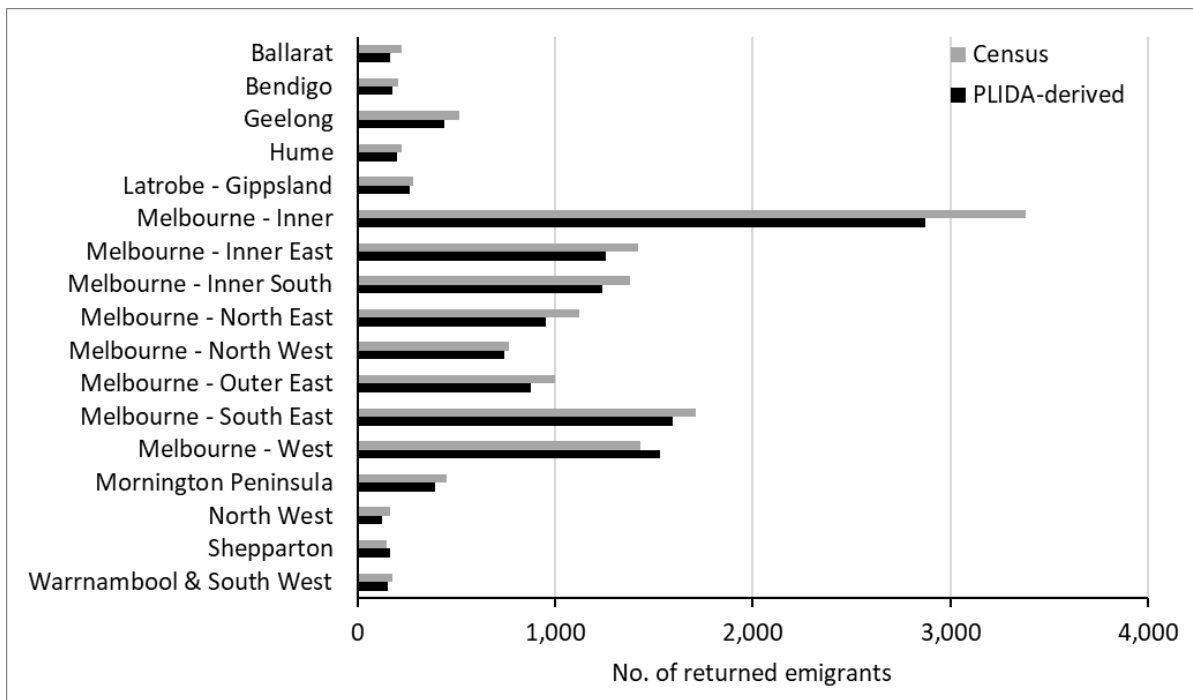


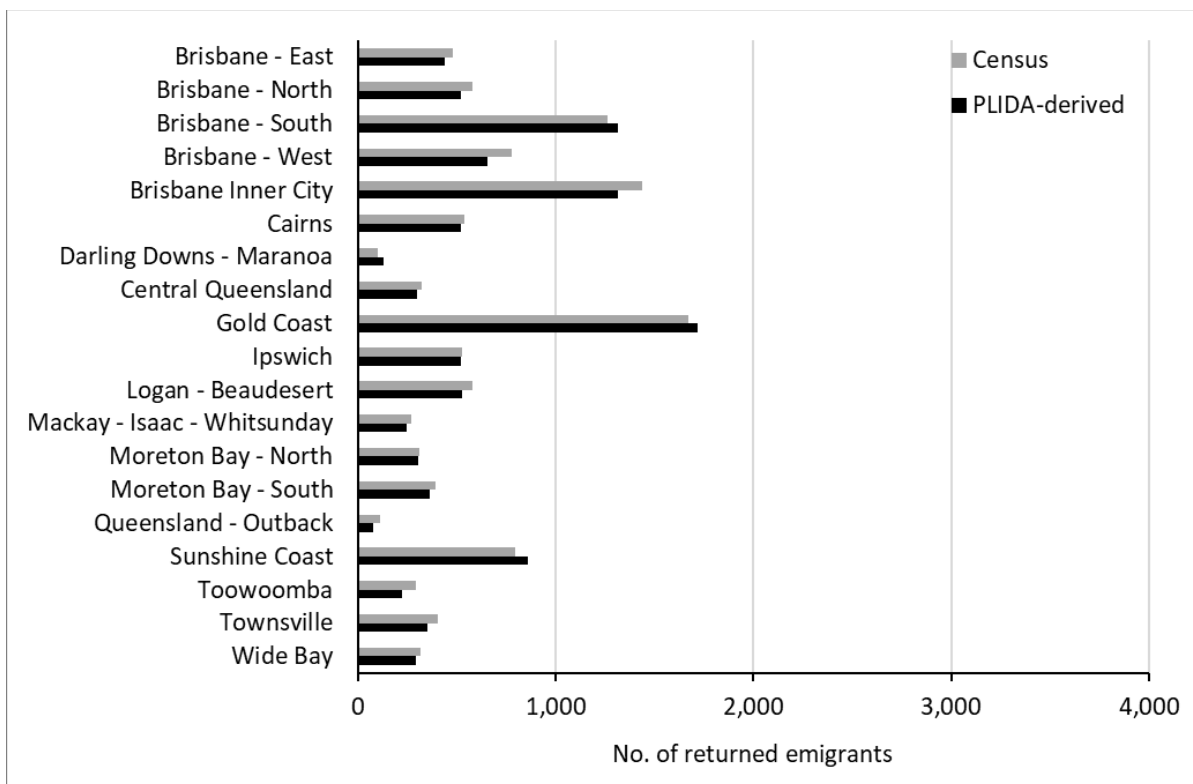
Figure 2: Comparison of PLIDA-derived and 2016 Census counts of 2011-15 emigrants who had returned to Australia by 2016 by SA4 region

Sources: ABS 2016 Census; data extracted from the Person Level Integrated Data Asset (PLIDA)

To illustrate the size of these flows and their differences, Figure 3 shows the returned emigrants estimates from PLIDA alongside equivalent census counts for SA4 regions in the States of Victoria and Queensland. In Victoria, PLIDA-derived returned emigrants as a proportion of the census counts range from 73% in Ballarat to 109% in Shepparton, while in Queensland they vary from 68% in Queensland – Outback to 129% in Darling Downs – Maranoa. The largest numerical difference is for Melbourne – Inner, though the PLIDA-derived estimate is not as low as it may appear, being 85% of the census number. The Appendix contains the equivalent of Figure 3 for the other States and Territories. Both Figures 2 and 3 suggest that the PLIDA-derived estimates do a reasonable job of measuring SA4 scale emigration – at least for those emigrants who later return to Australia.



(a) Victoria



(b) Queensland

Figure 3: Comparison of PLIDA-derived and 2016 Census counts of 2011-15 emigrants who had returned to Australia by 2016 by SA4 region in Victoria and Queensland

Sources: ABS 2016 Census; data extracted from the Person Level Integrated Data Asset (PLIDA)

However, it is important to note that these returned emigrants comprise a small and unrepresentative subset of all emigrants. They are unrepresentative in terms of both geographical

distribution and visa/citizenship category. If the geographical distribution of census returned emigrants was used to distribute national emigration to States, it would result in large errors. For some States the absolute percentage error would exceed 50%. With regards to visa/citizenship, the census does not collect information on visa status, but it does ask about Australian citizenship. Overall, 82% of returned emigrants in PLIDA over the period 2011-15 were Australian citizens (at least at the time of the 2016 Census) compared to 34% Australian citizens in official ABS emigration statistics. In the PLIDA-derived estimates, 68% of returned emigrants were Australian citizens. Given the good coverage of citizens in PLIDA, it is not surprising to see higher coverage of the census returned emigrants than emigration overall.

3.3. Sub-state immigration

PLIDA-based immigration flows for SA4 areas for 2016-20 were compared with our own SA4 area immigration estimates. These were calculated from ABS State-level immigration numbers distributed to SA4 areas according to 2021 Census counts of numbers of people living overseas 5 years ago. Inevitably, there will be some approximation in these disaggregated estimates through the use of census counts because they suffer from a small amount of undercount and consist of a five-year interval transition measure of immigration. Figure 4 shows the comparison of the two sets of data in the black dots. The correlation is high ($r=0.99$) although the PLIDA immigration flows are considerably lower than the distributed ABS immigration estimates, with the PLIDA numbers for SA4 areas being on average 66% of the distributed ABS immigration flows.

If PLIDA immigration numbers by SA4 area are scaled up to sum to ABS State-level immigration estimates, how close do the scaled-up numbers get to our own estimates? Figure 4 shows the answer (using the red + symbols). The numbers are now closer to the immigration estimates with the median absolute percentage error being 7.4% and the interquartile range covering 4.1% to 13.9%. But the deviations from the diagonal shown in Figure 4 demonstrate that the scaled-up numbers remain quite approximate estimates of SA4 scale immigration for many areas.

Would it help if the PLIDA immigration numbers were scaled-up separately by visa/citizenship category? This involves taking PLIDA SA4 area immigration flows by visa/citizenship category, scaling these to ABS State-level estimates of immigration by visa/citizenship category, and then aggregating across visa/citizenship categories to obtain alternative scaled-up PLIDA SA4 immigration numbers. Unfortunately, this results in SA4 immigration numbers which deviate to a greater extent from our disaggregated ABS estimates. The median absolute percentage error from the alternate immigration numbers is 11.7% (with an interquartile range of 5.6-20.3%).

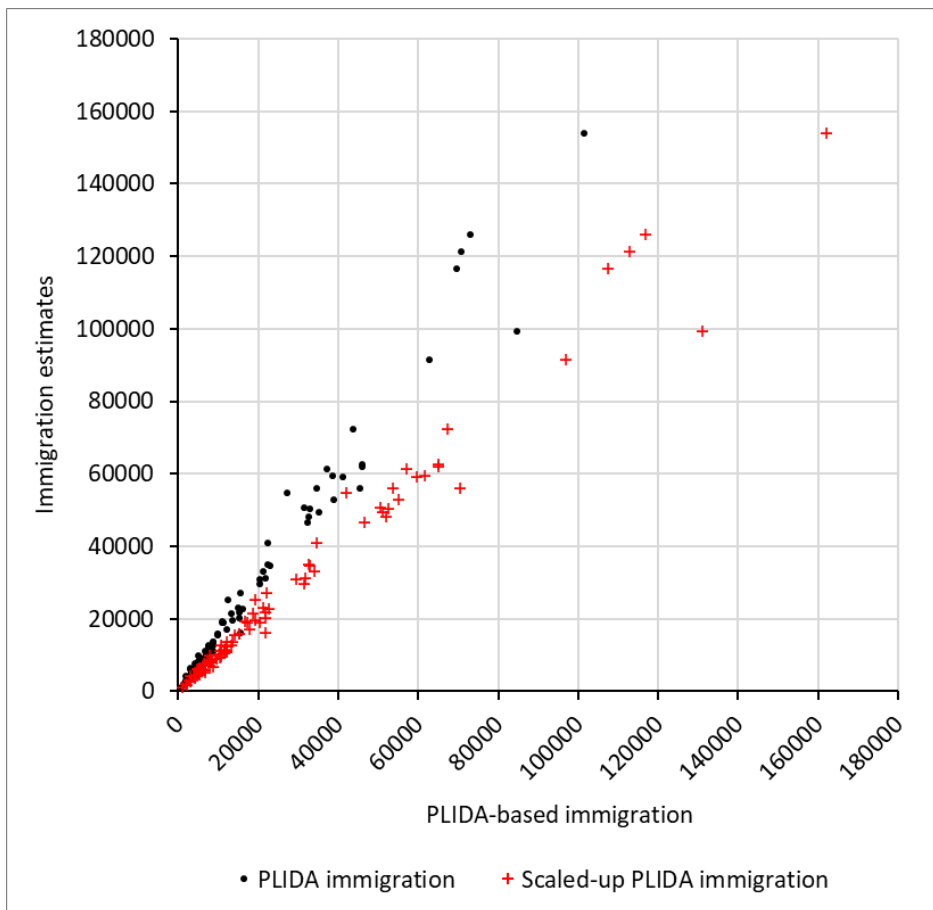


Figure 4: Comparison of PLIDA-based and estimated immigration by SA4 area, 2016-20

Sources: data extracted from the Person Level Integrated Data Asset (PLIDA); own immigration estimates.

Notes: SA4 area immigration estimates were calculated by disaggregating ABS State immigration estimates according to 5-year interval 2021 Census immigration counts.

4. Discussion & Conclusions

The purpose of this study was to discover if SA4 area emigration data calculated from PLIDA is of reasonable quality. Due to limited comparison data and a restriction on cell sizes extracted from PLIDA, it was not possible to undertake a comprehensive evaluation (such as a comparison of PLIDA and official statistics on emigration by visa/citizenship category, State/Territory, and age and sex; or a comparison of PLIDA and official statistics on emigration for a coarse sub-state geography). This is a common situation for many researchers and practising demographers where data assessments must be made even though limited data and information are available to make the assessment. The evaluation we were able to complete revealed non-trivial limitations in the coverage of the PLIDA international migration data, including sizeable geographical, temporal, age, and visa/citizenship variations.

PLIDA-derived emigration at the State scale was found to be substantially lower than the official ABS estimates, especially at the young adult ages. When used to distribute national ABS emigration across States, large errors were found for the Northern Territory and Tasmania. Although geographical and age distribution undercounts at the State scale do not provide definitive evidence of similar limitations at the SA4 scale, some level of undercoverage at the SA4 scale must exist for the State emigration numbers to be too low. The comparison of census returned emigrants, a small subset of all emigration flows, with a similarly defined subset of flows from PLIDA revealed an approximate but imperfect relationship.

PLIDA-based *immigration* numbers for SA4 areas were found to offer approximate immigration estimates when scaled to ABS State-level emigration estimates. It is likely that similarly accurate emigration estimates would be obtained if the PLIDA emigration numbers were scaled up to ABS State-level emigration estimates, though this cannot be stated with absolute certainty. For a rough and ready indication of emigration, consideration could be given to using scaled-up PLIDA emigration numbers, but with the caveat that errors of 20-30% could be possible. For most uses, however, the PLIDA-derived emigration numbers at the sub-state scale are probably not accurate enough. It would be risky to use these data for any analysis involving levels of migration, geographical patterns or temporal trends. Indirect estimation of sub-state migration using a previously described indirect estimation approach is another option (author reference). Alternatively, a simple approach such as taking sub-state immigration estimates and multiplying by State emigration/immigration ratios by age and sex might suffice for some purposes.

However, accurate spatially detailed emigration estimates from a linked administrative dataset such as PLIDA *could* be possible in the future. It would require better coverage of the Australian resident population in the Person Linkage Spine so that it includes all temporary visa holders and better coverage of people living in remote areas. It would also require more timely updates for it to be useful for analyses of emigration trends and use in population forecasting, and the data made

more easily accessible than through DataLab. We encourage ABS to develop PLIDA further so that its coverage and timeliness is sufficient to extract robust sub-state overseas migration estimates. As part of this development, it would be very helpful for users if PLIDA could be made more user-friendly. It is currently a database of very ‘raw’ data which contains many records of those who have died or emigrated. There is limited documentation, and it requires a high-level of expertise to program and extract the data. If PLIDA could be edited and processed to become an annual administrative-based resident population count and made available via TableBuilder, like the census, the uses and value of the dataset would increase enormously. Work by the UK Office for National Statistics in building a dynamic population model based on a variety of data sources and modelling techniques shows that administrative data can be used to produce reasonable quality demographic statistics (ONS 2022).

An alternative approach, of course, would be to measure sub-state emigration directly. Most travellers to and from Australia travel by air, and until 2017 all departing travellers were required to complete an outgoing passenger card which included State of residence/visit. If an electronic version of the outgoing passenger card was to be re-introduced, then the postcode of those departing Australia for 12 months or more could be collected. Postcodes are widely known in Australia and the four digits of a postcode could be easily collected and processed electronically. Realistically, however, it seems unlikely that the Australian Government would be keen on re-introducing paperwork which was removed relatively recently just for statistical purposes.

Emigration data remains the weakest link of sub-state population statistics in Australia (and many other countries). The absence of good quality geographically detailed emigration data has implications for the quality of local population estimates using the demographic balancing equation, for measuring and understanding local population change, for monitoring the effectiveness of regional workforce recruitment and retention policies, and for preparing input data for local population projections. We believe large integrated population datasets such as PLIDA have the *potential* to offer good quality sub-state emigration estimates, though substantial coverage improvements would be needed as the first step in that direction.

Declarations

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

ABS (2017). Census of Population and Housing: Understanding the Census and Census Data, Australia, 2016.

<https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2900.0~2016~Main%20Features~Item%20non-response~10036>. Accessed 18 July 2023.

ABS (2021). Overseas Migration Methodology, 2020-21 financial year.

<https://www.abs.gov.au/methodologies/overseas-migration-methodology/2020-21>. Accessed 10 May 2022.

ABS (2022). Overseas Migration, 2021-22 financial year.

<https://www.abs.gov.au/statistics/people/population/overseas-migration/2021-22-financial-year>. Accessed 18 July 2023.

ABS (2023a). Regional Population Methodology, 2021-22 financial year.

<https://www.abs.gov.au/methodologies/regional-population-methodology/2021-22>. Accessed 17 July 2023.

ABS (2023b). National, State and Territory Population, December 2022.

<https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/dec-2022>. Accessed 17 July 2023.

ABS (2023c). Person Level Integrated Data Asset (PLIDA). <https://www.abs.gov.au/about/data-services/data-integration/integrated-data/person-level-integrated-data-asset-plida>. Accessed 17 July 2023.

Bernard, A., Wu, G., Wilson, T., Argent, N., Zajac, T., & Kimpton, A. (2024). The pitfalls and benefits of using administrative data for internal migration research: An evaluation of Australia's Person Level Integrated Data Asset (PLIDA). *Demographic Research* 51(22): 687-722.

<https://doi.org/10.4054/DemRes.2024.51.22>

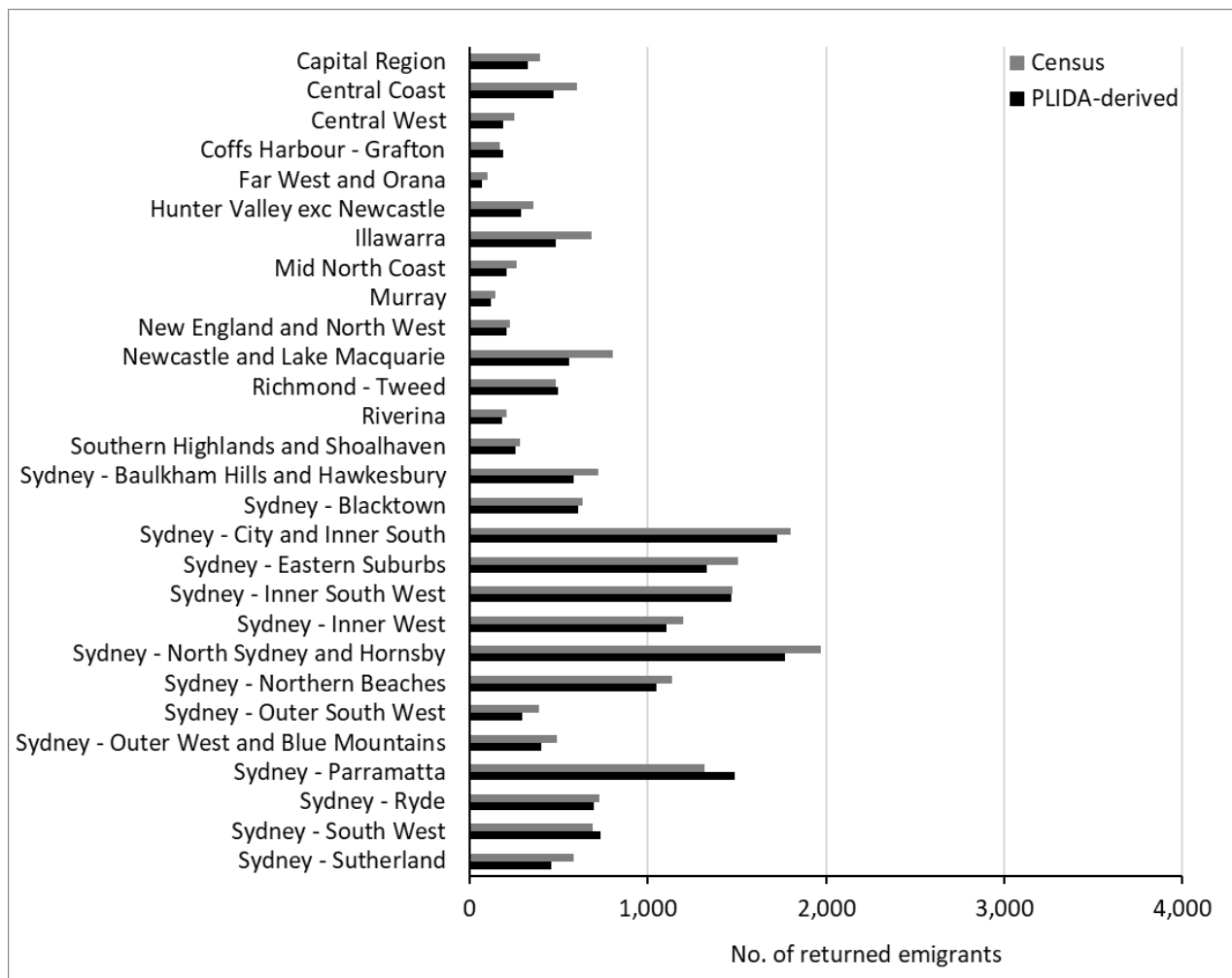
- Boese, M. (2023). Migrant and refugee retention in regional Australia at the intersection of structure and agency. *Journal of International Migration and Integration* 24 (Suppl 6): 1145-1166 (2023). <https://doi.org/10.1007/s12134-023-01022-y>
- Dyrting, S., Taylor, A., & Shalley, F. (2020). A life-stage approach for understanding population retention in sparsely populated areas. *Journal of Rural Studies* 80: 439-451. <https://doi.org/10.1016/j.jrurstud.2020.10.021>.
- Hugo, G. (2008). Australia's state-specific and regional migration scheme: An assessment of its impacts in South Australia. *Journal of International Migration and Integration* 9: 125-145. <https://doi.org/10.1007/s12134-008-0055-y>
- Jensen, E. B. (2013). A review of methods for estimating emigration. U.S. Census Bureau Working Paper No. 101. <https://www.census.gov/content/dam/Census/library/working-papers/2013/demo/POP-twps0101.pdf>. Accessed 10 May 2022.
- ONS (2022). International migration statistical design progress report: July 2022. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/articles/internationalmigrationstatisticaldesignprogressreport/july2022>. Accessed 14 November 2023.
- QGSO (Queensland Government Statistician's Office). (2023). Queensland Government population projections: regions, 2023 edition. <https://www.qgso.qld.gov.au/statistics/theme/population/population-projections/regions>. Accessed 17 July 2023.
- ONS (Office for National Statistics). (2009). Estimating international long-term emigration by local authority. London: ONS.
- Poulain, M., & Herm, A. (2013). Central population registers as a source of demographic statistics in Europe. *Population* (English edition) 68: 183-212. <https://doi.org/10.3917/pope.1302.0183>
- Raymer, J., & Baffour, B. (2018). Subsequent migration of immigrants within Australia, 1981-2016. *Population Research and Policy Review* 37: 1053-1077. <https://doi.org/10.1007/s11113-018-9482-4>
- Raymer, J., Bai, X., Liu, N., & Wilson, T. (2020). Estimating a consistent and detailed time series of immigration and emigration for sub-state regions of Australia. *Applied Spatial Analysis and Policy* 13: 411-439. <https://doi.org/10.1007/s12061-019-09310-w>

Rees, P., Bell, M., Duke-Williams, O., & Blake, M. (2000). Problems and solutions in the measurement of migration intensities: Australia and Britain compared. *Population Studies* 54: 207-222, <https://doi.org/10.1080/713779082>

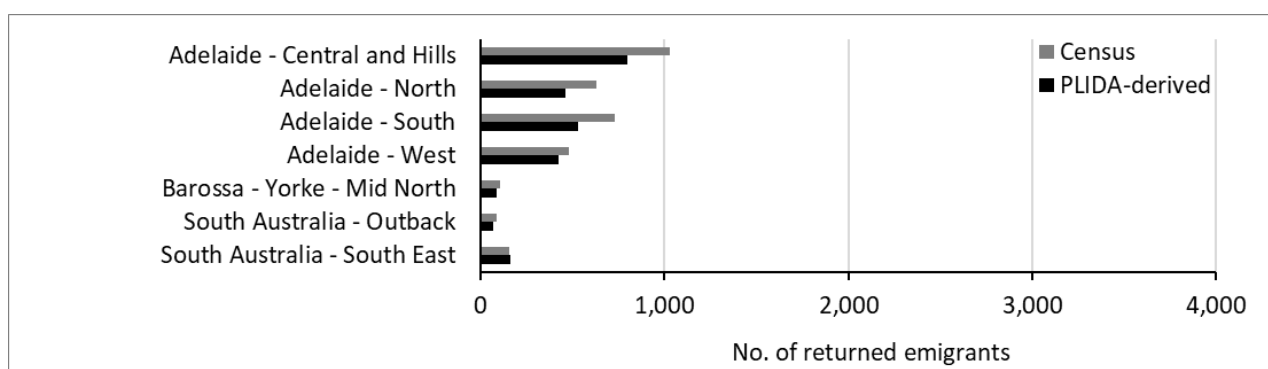
Van Hook, J., Zhang, W., D. Bean, F., & Passel, J.S. (2006). Foreign-born emigration: a new approach and estimates based on matched CPS files. *Demography* 43: 361–382.
<https://doi.org/10.1353/dem.2006.0013>

+ author references excluded for peer review

Appendix



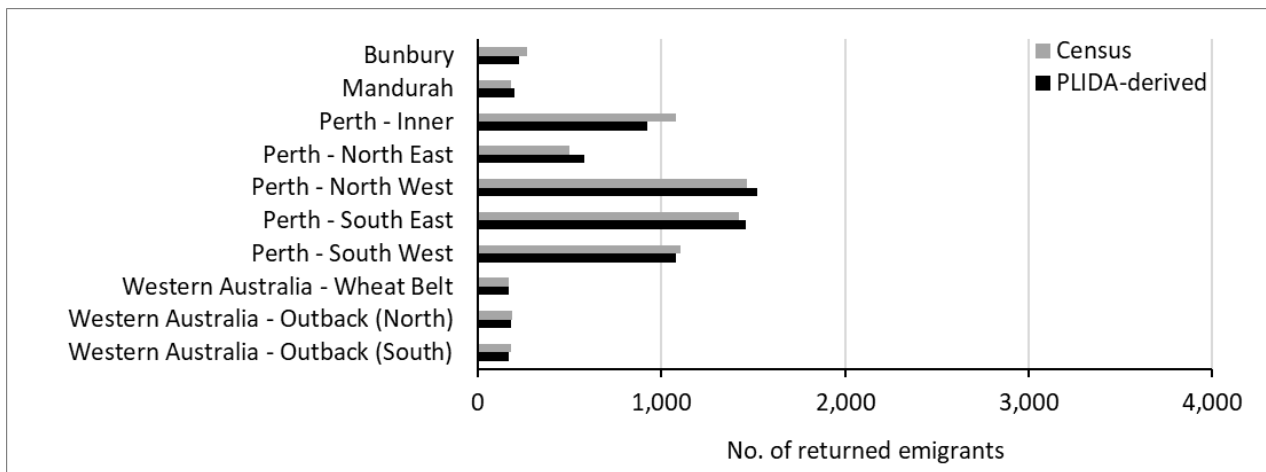
(a) New South Wales



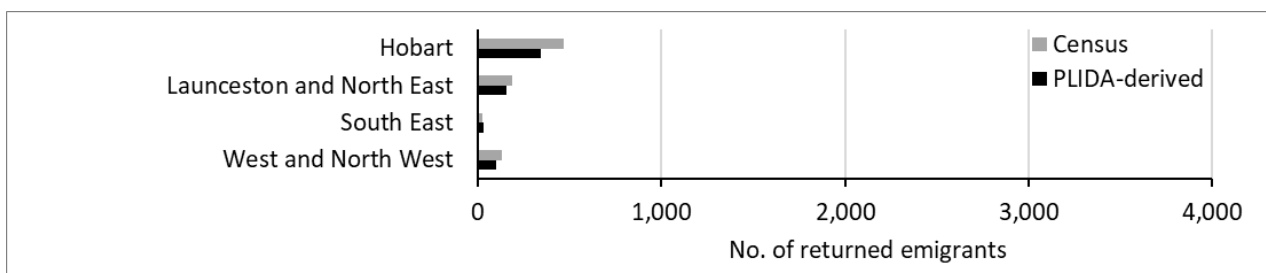
(b) South Australia

Figure A1: Comparison of PLIDA-derived and 2016 Census counts of 2011-15 emigrants who had returned to Australia by 2016 by SA4 region for States and Territories not in Figure 3

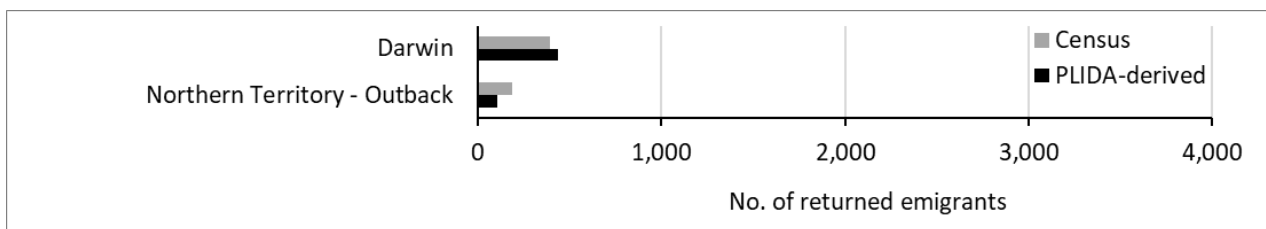
Sources: ABS 2016 Census; data extracted from the Person Level Integrated Data Asset (PLIDA)



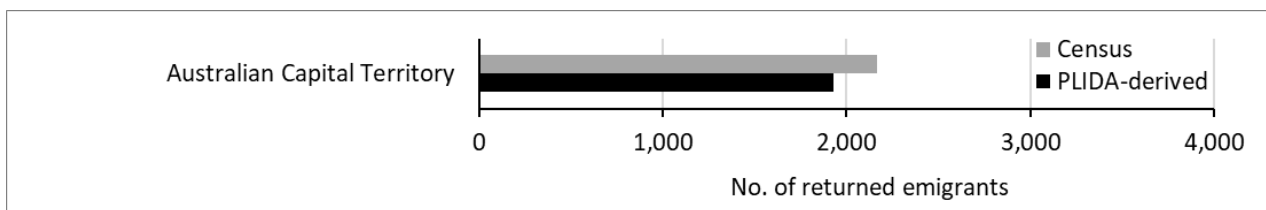
(c) Western Australia



(d) Tasmania



(e) Northern Territory



(f) Australian Capital Territory

Figure A1 continued