DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 51, ARTICLE 22, PAGES 687–722
PUBLISHED 1 OCTOBER 2024
https://www.demographic-research.org/Volumes/Vol51/22
DOI: 10.4054/DemRes.2024.51.22

Research Material

The pitfalls and benefits of using administrative data for internal migration research: An evaluation of Australia's Person Level Integrated Data Asset (PLIDA)

Aude Bernard

Neil Argent

Jing Wu

Tomasz Zając

Tom Wilson

Anthony Kimpton

Contents

Demographic Research: Volume 51, Article 22
Research Material

The pitfalls and benefits of using administrative data for internal migration research: An evaluation of Australia's Person Level Integrated Data Asset (PLIDA)[1]

Aude Bernard[2]
Jing Wu[3]
Tom Wilson[4]
Neil Argent[5]
Tomasz Zając[2]
Anthony Kimpton[6]

Abstract

BACKGROUND
To enhance its data capability, Australia recently set up a longitudinal administrative micro-dataset, the Person Level Integrated Data Asset (PLIDA).

OBJECTIVE
To ensure that users in both scholarly and applied settings understand how PLIDA can be reliably used, we assess its Combined Location Module, which provides place of residence by combining three administrative datasets since 2006.

METHODS
Using descriptive statistics and regression analysis, we compare the population coverage of PLIDA to the census, estimate the incidence of missing values at various spatial scales, quantify spatial mismatch between PLIDA and the census, and compare the intensity, selectivity, and spatial patterns of internal migration between the two datasets.

RESULTS
The PLIDA population is higher than the census population, but very remote populations and recently arrived and temporary migrants are under-represented. We uncover a high mismatch rate between PLIDA and the census for small spatial units, particularly among highly mobile groups. As a result, PLIDA overestimates the level of internal migration and slightly distorts migration age patterns at young ages. Migration patterns are broadly

[1] The authors kindly acknowledge support from the Australian Research Council under its Discovery scheme (DP200100760) and its Linkage scheme (LP210200733).
[2] University of Queensland. Australia. Email: a.bernard@uq.edu.au.
[3] University of Queensland. Australia.
[4] Advanced Demographic Modelling, Melbourne, Australia.
[5] University of New England, New South Wales, Australia.
[6] University of Southern Queensland, Springfield, Queensland, Australia.

comparable at large spatial scales, except for rapidly growing peri-urban regions, inner cities with short-term residents, and regions with a seasonal or temporary workforce.

CONCLUSION
PLIDA can be used in some applied settings outside demography if the spatial scale of analysis is aggregated, the analysis is restricted to census respondents, or the analysis is limited to well-represented groups. However, in its current form PLIDA is not suitable for demographic applications such as internal migration estimates, and we discourage its use for population projections.

CONTRIBUTIONS
Researchers need to be aware of the pitfalls of administrative data to knowingly decide on appropriate use. We recommend researchers to keep abreast of developments by the Australia Bureau of Statistics, which is continuously improving PLIDA.

1. Introduction

Data availability remains a major constraint on migration research (King and Skeldon 2010), including internal migration – that is, a change of region of residence within national borders. Three main types of data are available to internal migration scholars: censuses, surveys, and administrative sources, each with their own strengths and limitations (Bell et al. 2014). The choice of a suitable dataset is constrained not only by availability but also by the theoretical and disciplinary lenses through which researchers analyse migration. The aggregate perspective favoured by demographers and population geographers relies on count data and migration flows, whereas sociologists and economists typically draw on longitudinal microdata (White 2016).

While Australia is well-served in internal migration data, thanks to a quinquennial census and a long-running household survey, both data sources face limitations compared with population register data. In the absence of a population register, Australia has no single administrative dataset that covers the entire population, so coverage is achieved by linking a range of administrative datasets, which are encompassed in the Person Level Integrated Data Asset (PLIDA).[7] Released in 2015 and further developed in 2017 and 2020, PLIDA is yet to be used for internal migration research. This paper takes the first step in assessing its reliability and utility for this purpose. Our findings are also relevant to any geo-spatial research.

Evaluating the strengths and limitations of administrative datasets is essential to ensure that users in both scholarly and applied settings understand the reliability of such

[7] PLIDA was initially named the Multi-Agency Data Integration Project (MADIP).

datasets for research purposes (DeWaard et al. 2022). This is particularly important as national statistical agencies around the world are moving away from traditional censuses (Bell 2015). In some countries, administrative data have replaced census enumeration as the key source of internal migration statistics. In other countries, administrative data are used with a limited census collection to collect data not covered by administrative agencies or for which the quality or coverage is inadequate. At the 2020 census round, two-thirds of European countries used a register-based census, an increase on the 25% that participated in the 2010 census round (Valente 2019). For the 2021 census the Australian Bureau of Statistics (ABS) used administrative data from PLIDA for the first time. For example, PLIDA information was used to identify empty houses on census night, thereby improving the final census count (ABS 2021e), while highlighting the potential use of administrative data sources in future censuses. In that context, there is a pressing need to gauge the reliability and utility of administrative datasets as an alternative source of internal migration data.

In this paper we focus on the Combined Location Module of PLIDA, which combines changes of address recorded by three federal agencies. Section 2 briefly reviews existing sources of internal migration data in Australia and discusses their strengths and limitations. Section 3 introduces PLIDA. It outlines the framework for data linkage and presents the different administrative datasets available in PLIDA. Section 4 introduces our analytical strategy and the rationale for comparing PLIDA to census data. Results are presented in Section 5 sequentially. First, we compare the coverage of PLIDA to that of the census to assess whether some groups are over- or under-represented. Second, we examine missing values and mismatches in region of residence in PLIDA compared with the census. Third, we calculate from both the census and PLIDA a series of internal migration indicators that capture the level, age selectivity, and direction of internal migration and compare findings from both sources. In Section 6 we summarise findings, provide guidance on how to reliably use PLIDA and explore new research avenues opened by PLIDA and the ability to link the Combined Location Module to other administrative datasets.

2. Migration data in Australia: An overview

Three main sources of internal migration data exist in Australia: the census, Medicare, and the Household Income and Labour Dynamics in Australia (HILDA) survey. We briefly introduce each in turn and summarise their strengths and limitations.

2.1 Census

The quinquennial census has been the main source of information on migration since 1981. It has the unique advantage of providing data for the entire population at a fine-grained geographical scale, permitting the construction of detailed origin–destination flow matrices, which are essential for understanding the spatial structure of internal migration systems and their dynamics. The census also contains a wealth of sociodemographic information about migrants and non-migrants that can be used to compare sub-population groups and establish the determinants of internal migration. In recent years the utility of census data has been enhanced, thanks to the probabilistic matching of the last three censuses released at an individual level in a 5% sample (Zhang and Campbell 2012). The Australian Bureau of Statistics (ABS) also recently released two data assets in which visa status from the Department of Home Affairs is linked to census data (Smith and Smith 2014). While this novel dataset has advanced research on the internal migration of international migrants (Laukova, Bernard, and Sigler 2022), it faces the shortcomings of census data.

Regardless of recent advances, the census only runs every five years, and an enduring drawback remains its partial collection of migration. Migration is captured via the comparison of current place of residence with that of 1 and 5 years previously, which means that onward and return moves during the observation period are missed (Rogers, Raymer, and Newbold 2003; Rogerson 1990; Rees and Lomax 2019). This transition measure of migration is particularly problematic in a high-mobility country such as Australia, where over 15% of the population changes address every year (Bell et al. 2015). A facet of this high level of mobility is the repeat migration of a small segment of the population (Bell 1996; Morrison 1971; Bernard et al. 2017) that cannot be identified within census data.

2.2 Medicare

To provide a more timely and regular picture of internal migration, the ABS regularly publishes annual and quarterly internal migration estimates (ABS 2021b). Selected aggregate data are made available at various spatial scales, including migration between states, migration between Greater Capital City Statistical Areas (ABS 2021d), and migration in and out of small areas (ABS 2023b). These migration estimates are based on administrative sources, primarily changes of address notified to Medicare,[8] combined

[8] Medicare is the publicly funded universal insurance scheme in Australia. It is available to all Australian usual residents, including Australian citizens and permanent residents, and non-Australian residents granted

with Department of Defence data on defence force movement plus various modelling adjustments to correct for under-coverage of Medicare (ABS 2021c). It is assumed that the lag between moving and updating an address with Medicare is 3 months.

In summary, while these migration estimates provide up-to-date information, they contain a number of limitations. Extra modelling is required to adjust for the undercount of some groups – particularly young males – and assumed time lags between a permanent shift and the formal changes of address need to be factored in, such as occurred when most of the Australian population was vaccinated against COVID-19. In addition, Medicare-based migration estimates are disaggregated only by sex and age group, which limits their utility, and are not made available at an individual level.

2.3 HILDA

An alternative source of internal migration data is the Household, Income, and Labour Dynamics in Australia (HILDA) survey, which has tracked individuals annually since 2001, with refreshment samples ensuring it remains nationally representative. The HILDA survey has a very high wave-on-wave retention rate by international standards and, thanks to a complex probabilistic sampling design, it produces measures of internal migration broadly comparable to the census (Watson 2020). It has the unique advantage of collecting comprehensive sociodemographic attributes, including personality traits, which have emerged as an important determinant of internal migration (Jokela 2021; Shuttleworth et al. 2020). As in other countries, longitudinal data have become the gold standard in Australian migration research (Vidal and Lersch 2021). As a result, HILDA has been used to examine a broad range of issues, including the selectivity of internal migration (Campbell 2019; Crown, Gheasi, and Faggian 2020), reasons for migrating (Thomas, Gillespie, and Lomax 2019; Bernard and Kalemba 2022; Kalemba et al. 2022), the impact of migration on life outcomes (Clark and Lisowski 2018; Korpi and Clark 2017), and the association between migration and life course transitions (Bernard, Bell, and Charles-Edwards 2016; Vidal et al. 2017; Sander and Bell 2014).

However, as with other longitudinal surveys, HILDA faces three limitations. First, its small sample size means that key sub-population groups such as immigrants and First Nations Australians tend to be under-represented. Detailed analysis of these populations is hence compromised. Second, the coarse scale of geographical resolution does not permit analysis beyond the crude categories of ‘capital city’ and ‘rest of state’. Third, despite a wave-on-wave retention rate between 90% and 95%, the effect of attrition is

temporary protection visas. However, some groups, including defence personnel and First Nations people, are likely to be under-represented because of these groups’ reliance on alternative health services (Cook 2001).

compounded over time, making it difficult to track a large sample of respondents over a sustained period.

In this context, the use of longitudinal administrative micro datasets in the internal migration literature has grown. They are increasingly available not only in Europe (Poulain, Herm, and Depledge 2013) but also in East Asia (Ishikawa 2020) and, more recently, in Australasia, where New Zealand established an integrated Data Infrastructure in 2011 (Poot and Sanderson 2015). Contrary to censuses, administrative sources capture migration as an event. These time-rich and spatially detailed data allow the reconstruction of migration histories over long segments of the life course (Sesma Carlos, Kok, and Oris 2022), permit the distinction between onward and return migration (Mulder, Lundholm, and Malmberg 2020), and enable the disaggregation of migration indicators by move order (Kolk 2019; Bernard and Kolk 2020; Kulu, Lundholm, and Malmberg 2018).

3. The Person Level Integrated Data Asset

3.1 Overview

PLIDA is a longitudinal micro-dataset based on the integration of multiple administrative data sources. It was first established in 2015 and further developed between 2017 and 2020 through the Data Integration Partnership for Australia (DIPA), with the aim of maximising the use and value of the Federal Government's data assets to allow academics and policymakers to gain new insights into health, education, and socioeconomic domains. PLIDA is enabled through a series of partnerships between Federal agencies, including the Australian Taxation Office, the Department of Health, the Department of Social Services, Services Australia, the Department of Home Affairs, the National Disability Insurance Agency, and state agencies including the State and Territory Registrars of Births, Deaths, and Marriage.

The integration of the various datasets is based on a unique identifier called the Person Linkage Spine, hereafter referred to as the 'spine'. To maximise population coverage, the spine is based on the combined population of 3 core datasets: (1) the Medicare Consumer Directory held by the Department of Health, (2) DOMINIO Centrelink Administrative Data from the Department of Social Services, and (3) Personal Income Tax maintained by the Australian Taxation Office. As explained in the previous section, Medicare is available to all Australian usual residents. Social security payments through Centrelink are subject to means testing. Since 2019, immigrants must wait four years before being eligible for most payments (prior to 2019 immigrants had to wait two years before being eligible for social security payments). Conversely, taxation data covers all employed individuals regardless of immigration status. The spine currently covers the period January 2006 to November 2022. It is updated annually and aims to include all people resident in Australia. In this paper we use version 5 of the spine, which is the latest available at the time of writing.

The spine is used by PLIDA users to integrate datasets of interest. At the time of writing, 24 datasets have an enduring link to PLIDA, which means that they are updated on an ongoing basis, with possible future linkages being discussed. Most datasets have been available since 2006, although the exact period available varies from one dataset to another. The full list of datasets with enduring links to PLIDA can be found in Table A-1. Examples of such datasets include: (1) administrative datasets such as personal income tax data from the Australian Tax Office and visa information from Home Affairs; (2) surveys such as the National Health Surveys; and (3) data from the last three censuses in 2011, 2016, and 2021. The link to the census serves to enhance the utility of administrative data, which are attribute-poor (McCollum et al. 2021; Ernsten et al. 2018). However, to reduce the risk of re-identification, administrative datasets in PLIDA can be linked to only one census at a time. An additional 51 datasets have been integrated to PLIDA for once-off purposes for approved projects and these are typically removed after project completion.[9]

The ABS is the Accredited Integrating Authority of PLIDA. It collects, combines, and provides access to integrated administrative data, which are de-identified and confidentialised. The data are made available to approved users located in Australia and analysed in a safe, secured, and controlled environment through the ABS online DataLab platform. Prior to being taken outside of the DataLab environment, all results must be vetted by the ABS to ensure that they comply with the output clearance rules (e.g., minimum cell size requirements) designed to prevent re-identification.

3.2 The Combined Location Module

For ease of analysis, the ABS curates a few bespoke files, including the Combined Location Module, which provides changes of region of residence since 2006 at various spatial scales. This module combines changes of address recorded in the three datasets used for the spine – Medicare, Centrelink, and Income Tax – to account for the varying coverage and quality of location information from the three source datasets. The ABS appends the residential addresses from each dataset and excludes postal and business addresses. When records with the same address have overlapping time periods the records are collapsed to form a single record, with the earliest start date and the latest end date being retained. When the time period overlaps at different addresses, the end date of the

[9] The full list of once-off linkages to PLIDA can be found at https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Statistical+Data+Integration+-+MADIP+data+and+legislation.

older address is adjusted to one day before the start date of the new address. Finally, if multiple addresses are recorded within a month, the address with the longest duration is retained.

For each record, a start and end date are provided. We find that close to 21% of records do not include an end date, so only use start dates in the analysis. For each record, four levels of geography are available, from Statistical Areal level 1 (SA1s), which is the lowest level of geography in the Australian Statistical Geographic Standard (ASGS), to SA2s, SA4s, and states.[10] More information on the ASGS, a social geography classification that divides Australia into a nested hierarchy of statistical areas, can be found in Table A-2.

4. Analytical strategy

To assess the reliability and utility of PLIDA for internal migration and geo-spatial research, we compare it to the census, as the census remains the main source of migration data in Australia. A similar strategy has been used in European studies (Foley, Champion, and Shuttleworth 2017; Ernsten et al. 2018). Table 1 lists in a summary form the characteristics of each data source. We use data from the 2016 census instead of the 2021 census for three main reasons. First, at the time of writing the death module in PLIDA was available only to June 2021, which prevents us from removing individuals in PLIDA who died before census night, which was 10 August 2021. Second, the COVID vaccination campaign that started in 2021 is likely to have resulted in an increase in changes of address reported to Medicare, which may inflate changes of address in PLIDA for that period. Third, COVID-19 disrupted patterns of internal migration in Australia (Borsellino et al. 2022; Perales and Bernard 2023), which means that results from the 2021 census may not reflect long-standing internal migration patterns in Australia.

To ensure comparability between the census and PLIDA we exclude short-term overseas visitors from the census. There were 22,485,854 people in Australia at the 2016 census. However, according to PLIDA, there were 31,756,714 in Australia on census night, 9 August 2016. We use the Combined Demographics module, which includes death records, to remove individuals in PLIDA who died before census night (n = 1,376,827). This number is on par with 1,466,998 deaths recorded by the Australian Bureau of Statistics between mid-2006 and mid-2016 (ABS 2023a). We then use the Travellers module from Home Affairs to remove people who had emigrated (n = 1,184,147). This is lower than the number of international departures reported by the

[10] Since then the data has been made available at the mesh block level, which corresponds to census tracts in the United States.

ABS, which averaged 247,014 per annum between 2006 and 2016 (ABS 2022). This may explain why we are left with 29,195,740 people in PLIDA,[11] which is higher than the number of records in the census. Another possible explanation is the duplication of some records. We return to this limitation in the conclusion and explain how the ABS is planning to address this issue in the future. Here we mitigate the issue by restricting the PLIDA population to people who could be linked to the census to ensure they were in Australia on census night.

Table 1: Characteristics of internal migration data from the census and PLIDA

	Census	PLIDA's Combined Location Module
Data type	Census	Administrative dataset
Period	Quinquennial data collection on the second Tuesday of August 2016	Information recorded at each contact with three federal agencies since 2006
Data collection mode	Form completed online or face-to-face by one or multiple household members. Proxy respondents possible	Indirect data collection via administrative contact
Population universe	All individuals in Australia on census night	All residents of Australia with interactions with at least 1 of 3 government agencies: (1) Medicare, (2) Australian Taxation Office, (3) Social Services
How migration is measured	Migration is recorded as a transition by comparing place of residence at the census to that 1 and 5 years prior	Migration is recorded as an event
Potential sources of error	Undercount Reporting error Proxy reporting error Recall error (place of previous residence)	Delay in reporting a change of address Reporting error Under-representation of some sub-population groups

The analysis proceeds in three complementary steps. The first step consists of comparing the population coverage of both datasets to assess whether some groups are under-represented in PLIDA. This is achieved through logistic regression from a random sample of 10 million observations from each dataset (census and PLIDA), controlling for key sociodemographic characteristics collected at the census, including age, sex, region of birth, duration of residence in Australia, visa, educational attainment, marital status, presence of dependent children, labour force status, household income, indigenous status, state of residence, and remoteness status.

Second, we focus on the region of residence recorded in PLIDA and assess the proportion of missing values and mismatches between PLIDA and the census at a range of spatial scales: SA1, SA2, SA4, and states. Because administrative datasets are known to suffer from delayed reporting (Foley, Champion, and Shuttleworth 2017; McCollum

[11] This means that the PLIDA population we work with is based on the linkage between three modules: the Combined Location Module, the Combined Demographics Module, and the Travellers Module. It is not the raw Combined Location Module.

et al. 2021), we replicate the analysis based on place of residence recorded 1, 2, and 3 months after the census and establish whether spatial matching with the census improves. We then use regression analysis to establish whether places of residence are missing at random using the same control variables as in step 1.

The third step of the analysis consists of comparing a series of internal migration measures derived from both datasets based on place of residence on census day. In line with the census, we construct migration as a transition by comparing the region of residence on census day to that 1 and 5 years prior. We then construct a range of system-wide measures that capture the level of internal migration (crude migration intensity), the age selectivity of migration (age-specific migration intensity), and the spatial structure of migration (migration effectiveness index) (Bell et al. 2002). We calculate these measures based on both 1- and 5-year migration transition data and do so for migration between SA2s, SA4s, and states, which are the most used administrative units for internal migration research in Australia. We complete system-wide measures with region-specific net migration rates to establish any spatial bias.

5. Results

5.1 Population coverage

To establish whether some groups are over- or under-represented, Table 2 reports the results of a logistic regression model in which the dependent variable takes 0 for individuals in the census file and 1 for individuals in PLIDA. For ease of interpretation, results are reported as odds ratios (ORs). ORs greater/smaller than 1 indicate that a given explanatory variable is associated with an increase/decrease in the likelihood of being in PLIDA.

Results in Table 2 indicate that very remote populations are the most under-represented sub-population group, with odds of being in PLIDA rather than in the census 21% lower than the reference group (residents of major cities). With the exception of working holidaymakers, all temporary migrants are under-represented in PLIDA, even if duration of residence is controlled for. International students are particularly affected, having 14% smaller odds of being in PLIDA than the census compared with Australian citizens. By contrast, permanent migrants are adequately represented in PLIDA. Recently arrived migrants are less likely to be in PLIDA, presumably because of limited interactions with government agencies. The odds of being in PLIDA increases with duration of residence and are comparable to those of Australian citizens after 5 years in Australia. Region of origin is not a determinant of under-coverage in PLIDA. We also note the limited association between age and PLIDA coverage. Adolescents and young adults under the age of 25 and older adults aged 50 and over are slightly under-represented, having 1% to 2% lower odds of being in PLIDA than the census compared with individuals aged 40 to 44 years. Collectively, these results suggest that except for very remote populations and recently arrived and temporary immigrants, other sub-groups are adequately represented in PLIDA, including First Nations Australians.

Table 2: Determinants of membership to PLIDA compared with census

	Odds ratio	Standard error
Female	0.999	(0.001)
Age (ref. cat. 40–44)		
0–4	0.995	(0.004)
5–9	0.996	(0.004)
10–14	0.983	(0.004)
15–19	0.984	(0.003)
20–24	0.990	(0.003)
25–29	1.000	(0.003)
30–34	1.001	(0.003)
35–39	1.001	(0.003)
45–49	0.999	(0.003)
50–54	0.994	(0.003)
55–59	0.991	(0.003)
60–64	0.994	(0.003)
65+	0.990	(0.003)
State (ref. cat. New South Wales)		
Victoria	0.997	(0.002)
Queensland	1.003	(0.002)
South Australia	1.006	(0.002)
Western Australia	1.025	(0.007)
Tasmania	1.001	(0.004)
Northern Territory	1.025	(0.007)
Australian Capital Territory	1.000	(0.005)
Remoteness status (ref. cat. major city)		
Inner Regional	1.000	(0.001)
Regional	0.998	(0.002)
Remote	1.000	(0.006)
Very Remote	0.791	(0.006)
Region of birth (ref. cat. Australia)		
Oceania	1.005	(0.007)
North-East Asia	0.992	(0.007)
Southern and Central Asia	1.018	(0.007)
South-East Asia	1.007	(0.006)
North-West Europe	1.007	(0.006)
Southern and Eastern Europe	1.010	(0.006)
North Africa and Middle East	1.001	(0.006)
Sub-Saharan Africa	1.010	(0.007)
Americas	1.008	(0.008)
Duration of residence in Australia (ref. cat. 20 years +)		
Less than a year	0.896	(0.008)
1 to 2 years	0.959	(0.006)
3 to 4 years	0.987	(0.006)
5 to 9 years	1.002	(0.005)
10 to 19 years	1.002	(0.004)

Table 2: (Continued)

	Odds ratio	Standard error
Visa status (ref. cat. Australian citizen)		
Permanent visas		
Skilled	0.990	(0.007)
New Zealand	1.005	(0.007)
Family	1.000	(0.007)
Humanitarian	1.004	(0.010)
Other permanent	0.993	(0.007)
Temporary		
Skilled	0.896	(0.010)
Working holiday	1.060	(0.020)
Student	0.861	(0.008)
Bridging	0.965	(0.006)
Other temporary	0.978	(0.007)
Visa unknown	0.965	(0.014)
Married or partnered (ref. cat. Single or separated)	0.993	(0.006)
Dependent children under 15 years	0.999	(0.002)
First Nations Australians	1.013	(0.004)
Labour force status (ref. cat. Not in the labour force)		
Employed	1.004	(0.003)
Unemployed	1.004	(0.003)
Tertiary educated	1.002	(0.001)
Homeowner	0.999	(0.001)
Constant	2.105	(0.007)

Source: Authors' calculations based on the 2016 census data and PLIDA's Combined Location Module.

5.2 Region of residence: Missing values

We now investigate whether missing values are an issue. These are records for which a change of residence is recorded but the region of residence is missing. Table 3 reports the percentage of records in PLIDA with missing values by geographical units and year. SA4 regions are derived from SA2s, which is why they have the same proportion of missing values. Unsurprisingly, the larger the spatial scale, the lower the number of missing values. The share of missing values is negligible for states and hovers around 1% for SA2 and SA4 regions, although it has increased since 2017. While higher for SA1s, the share of missing values has decreased since 2007 and has remained below 2% since 2009.

Table 3: Percentage of records in PLIDA with missing values by geographical unit and year

Year	SA1	SA2/SA4	State
2006	10.394	3.226	0.011
2007	2.024	0.642	0.003
2008	2.010	0.681	0.003
2009	1.626	0.778	0.002
2010	1.267	0.663	0.002
2011	1.245	0.670	0.002
2012	1.183	0.719	0.002
2013	1.216	0.804	0.002
2014	1.113	0.752	0.002
2015	1.098	0.780	0.002
2016	1.118	0.824	0.013
2017	1.326	1.029	0.003
2018	1.434	1.135	0.004
2019	1.582	1.267	0.003
2020	1.245	1.027	0.002

Source: Authors' calculations based on PLIDA's Combined Location Module.

While the proportion of records with missing values is not an issue, missing values are not randomly distributed in the population. Regression results in Table 4 show that recently arrived and temporary migrants are significantly more likely to have missing records at both the SA1 and SA2 levels, and this is true for adolescents, young adults, older Australians, and First Nations Australians. By contrast, permanent migrants, individuals in the labour force, tertiary-educated individuals, those aged 30 to 44, and individuals with dependent children are significantly less likely to have a missing record.

Table 4: Determinants of missing records

	SA1 units		SA2 units	
	Odds ratio	Standard error	Odds ratio	Standard error
Female	1.009	(0.005)	0.890	(0.025)
Age (ref. cat. 40–44)				
0–4	1.599	(0.030)	1.951	(0.048)
5–9	1.217	(0.023)	1.429	(0.035)
10–14	3.321	(0.055)	2.678	(0.061)
15–19	2.793	(0.045)	2.633	(0.058)
20–24	1.269	(0.021)	1.543	(0.033)
25–29	0.954	(0.016)	1.215	(0.026)
30–34	0.749	(0.013)	0.944	(0.021)
35–39	0.759	(0.015)	0.876	(0.022)
45–49	1.484	(0.027)	1.303	(0.034)
50–54	2.024	(0.035)	1.723	(0.044)
55–59	2.372	(0.041)	1.827	(0.048)
60–64	2.514	(0.044)	1.819	(0.048)
65+	3.102	(0.050)	2.582	(0.059)

Table 4: (Continued)

	SA1 units		SA2 units	
	Odds ratio	**Standard error**	**Odds ratio**	**Standard error**
Region of birth (ref. cat. Australia)				
Oceania	0.933	(0.038)	1.103	(0.053)
North-East Asia	1.574	(0.060)	2.159	(0.095)
Southern and Central Asia	0.862	(0.033)	1.206	(0.053)
South-East Asia	1.203	(0.046)	1.735	(0.076)
North-West Europe	0.743	(0.028)	0.775	(0.035)
Southern and Eastern Europe	1.152	(0.046)	1.892	(0.085)
North Africa and Middle East	1.221	(0.049)	1.791	(0.083)
Sub-Saharan Africa	0.872	(0.036)	1.202	(0.057)
Americas	1.296	(0.052)	1.804	(0.083)
Duration of residence in Australia (ref. cat. 20 years +)				
Less than a year	10.89	(0.233)	10.85	(0.276)
1 to 2 years	5.627	(0.113)	5.808	(0.141)
3 to 4 years	2.626	(0.055)	2.728	(0.069)
5 to 9 years	0.975	(0.021)	1.057	(0.027)
10 to 19 years	0.641	(0.013)	0.661	(0.017)
Visa status (ref. cat. Australian citizens)				
Permanent visas				
Skilled	1.321	(0.045)	2.061	(0.079)
New Zealand	0.791	(0.030)	1.321	(0.058)
Family	0.760	(0.028)	1.143	(0.047)
Humanitarian	0.316	(0.017)	0.337	(0.021)
Other permanent	0.595	(0.002)	0.724	(0.036)
Temporary				
Skilled	11.52	(0.392)	20.02	(0.764)
Working holiday	4.032	(0.159)	6.422	(0.283)
Student	7.210	(0.243)	10.11	(0.384)
Bridging	3.875	(0.017)	6.012	(0.226)
Other temporary	5.729	(0.232)	9.068	(0.408)
Visa unknown	0.741	(0.027)	0.993	(0.042)
Married or partnered (ref. cat. Single or separated)	0.894	(0.005)	0.828	(0.007)
Dependent children under 15 years	0.816	(0.006)	0.943	(0.943)
First Nations Australians	1.297	(0.019)	0.890	(0.025)
Labour force status (ref. cat. Not in the labour force)				
Employed	0.608	(0.004)	0.413	(0.007)
Unemployed	0.726	(0.009)	0.706	(0.009)
Tertiary educated	0.866	(0.006)	0.949	(0.008)
Homeowner	1.149	(0.006)	1.114	(0.008)
Constant	0.005	(0.001)	0.005	(0.001)

Source: Authors' calculations based on the 2016 census data and PLIDA's Combined Location Module.

5.3 Spatial mismatch

We now explore possible spatial mismatch by establishing the share of PLIDA records that do not match those in the census. Unlike for missing values, the results are different for SA2 and SA4 regions, which are reported separately in the remainder of the paper. Results in Table 5 show that spatial mismatch is a problem for both SA1 and SA2 regions, which suffer from mismatch rates of 22% and 18% respectively, compared with 9.5% for

SA4 regions and 2% for states. Because of known delays in reporting changes of address to government agencies, we replicate the analysis based on place of residence in PLIDA 1, 2, and 3 months after census. The mismatch rate decreases only marginally. This suggests that the combination of addresses from multiple government agencies in PLIDA takes care of the issue of delayed address-change registration, which is a well-known limitation of administrative data. However, the high rate of mismatch for finer scale geographies raises questions about the reliability of PLIDA at the SA1 and SA2 levels.

Table 5: Percentage of records with a census–PLIDA mismatch by geographic unit and time lag

	On census day	1 month later	2 months later	3 months later
SA1	22.48	21.61	21.30	21.31
SA2	18.07	17.34	17.07	13.30
SA4	9.50	9.08	8.89	7.34
State	2.18	2.00	1.97	1.96

Source: Authors' calculations based on PLIDA's Combined Location Module and the 2016 census

We explore this further in Table 6 by analysing the determinants of spatial mismatch, focusing on SA1 and SA2 units. Results show a strong age gradient, with children and young adults being more likely to suffer from a spatial mismatch and older adults being less likely. Homeowners and individuals with dependent children are also very unlikely to be affected by a mismatch. However, immigrants who have been in Australia for less than 2 years, international students, working holidaymakers, and immigrants from North-East Asia are more likely to be spatially mismatched, as are First Nations Australians.

Overall, the results suggest that highly mobile groups are more likely to be spatially mismatched, whereas groups with lower levels of mobility are less likely to be affected by this issue. This is unsurprising. Some groups, like young adults, may report a postal address to government agencies (e.g., parental residence) that differs from the usual place of residence because of repeat movement (Foley, Champion, and Shuttleworth 2017). For other groups, such as working holidaymakers and recently arrived migrants, the discrepancy may be the result of frequent movement that is not captured by administrative datasets. For others, it may be residential multilocality (Petzold 2017). Overall, the results suggest that caution is required in the use of PLIDA for analysing the internal migration of the most mobile groups.

Table 6: Determinants of spatial mismatch between PLIDA and census

	SA1		SA2	
	Odds ratio	**Standard error**	**Odds ratio**	**Standard error**
Female	0.9032	(0.001)	0.930	(0.001)
Age (ref. cat. 40–44)				
0–4	1.242	(0.006)	1.350	(0.007)
5–9	1.166	(0.005)	1.196	(0.006)
10–14	1.254	(0.005)	1.270	(0.006)
15–19	1.110	(0.004)	1.171	(0.005)
20–24	1.320	(0.005)	1.393	(0.005)
25–29	1.380	(0.005)	1.432	(0.005)
30–34	1.230	(0.004)	1.267	(0.004)
35–39	1.119	(0.004)	1.137	(0.004)
45–49	0.842	(0.003)	0.836	(0.003)
50–54	0.684	(0.003)	0.684	(0.003)
55–59	0.576	(0.002)	0.580	(0.003)
60–64	0.503	(0.002)	0.512	(0.002)
65+	0.415	(0.002)	0.400	(0.002)
Sate (ref. cat. New South Wales)				
Victoria	0.964	(0.002)	1.003	(0.002)
Qeensland	1.090	(0.002)	1.232	(0.003)
South Australia	0.928	(0.003)	0.958	(0.003)
Western Australia	1.056	(0.003)	1.106	(0.006)
Tasmania	0.860	(0.004)	1.104	(0.006)
Northern Territory	1.462	(0.010)	2.214	(0.016)
Australian Capital Territory	0.936	(0.005)	1.134	(0.006)
Remoteness status (ref. cat. major city)				
Inner Regional	1.138	(0.002)	0.892	(0.002)
Regional	1.212	(0.003)	0.867	(0.003)
Remote	1.558	(0.010)	0.878	(0.006)
Very Remote	1.654	(0.014)	0.971	(0.009)
Region of birth (ref. cat. Australia)				
Oceania	0.985	(0.008)	1.021	(0.009)
North-East Asia	1.295	(0.010)	1.342	(0.012)
Southern and Central Asia	0.940	(0.007)	0.971	(0.008)
South-East Asia	0.915	(0.007)	0.932	(0.008)
North-West Europe	1.009	(0.007)	1.019	(0.008)
Southern and Eastern Europe	0.879	(0.007)	0.913	(0.008)
North Africa and Middle East	1.010	(0.010)	1.063	(0.010)
Sub-Saharan Africa	1.021	(0.009)	1.094	(0.010)
Americas	0.993	(0.009)	1.048	(0.011)
Duration of residence in Australia (ref. cat. 20 years +)				
Less than a year	1.210	(0.012)	1.114	(0.012)
1 to 2 years	1.505	(0.010)	1.330	0.009)
3 to 4 years	1.198	(0.007)	1.096	(0.009)
5 to 9 years	1.076	(0.006)	1.013	(0.006)
10 to 19 years	1.071	(0.005)	1.028	(0.005)
Visa status (ref. cat. Australian citizen)				
Permanent visas				
Skilled	0.936	(0.007)	0.917	(0.008)
New Zealand	1.054	(0.009)	1.015	(0.009)
Family	0.880	(0.008)	0.869	(0.008)
Humanitarian	1.229	(0.014)	1.126	(0.014)
Other permanent	0.994	(0.008)	0.975	(0.009)
Temporary				
Skilled	1.071	(0.013)	0.991	(0.013)
Working holiday	1.754	(0.033)	1.754	(0.007)
Student	1.240	(0.001)	1.102	(0.011)
Bridging	0.978	(0.008)	0.954	(0.009)
Other temporary	1.109	(0.017)	1.074	(0.017)
Visa unknown	0.983	(0.007)	0.971	(0.007)

Table 6: (Continued)

	SA1		SA2	
	Odds ratio	Standard error	Odds ratio	Standard error
Married or partnered (ref. cat. Single or separated)	0.797	(0.001)	0.801	(0.001)
Dependent children under 15 years				
First Nations Australians	1.223	(0.005)	1.201	(0.005)
Labour force status (ref. cat. Not in the labour force)	0.645	(0.001)	0.600	(0.001)
Employed	1.008	(0.004)	0.984	(0.002)
Unemployed	1.088	(0.002)	1.0910	(0.004)
Tertiary educated	1.038	(0.002)	1.091	(0.002)
Homeowner	0.516	(0.000)	0.549	(0.000)
Constant	0.569	(0.002)	0.437	(0.002)

Source: Authors' calculations based on the 2016 census data and PLIDA's Combined Location Module.

5.4 Levels and patterns of internal migration

The final step of the analysis consists of comparing migration indicators obtained from three sources: the census (all observations in the census), PLIDA (all observations in PLIDA excluding those who emigrated or died), and PLIDA linked to census (all observations in PLIDA that could be linked to the census). We do so for migration between SA2 regions, SA4 regions, and states by comparing place of residence 1 and 5 years prior to the census. To that end, Table 7 reports a series of well-established system-wide indicators that capture the intensity or level of internal migration, its age patterns, and spatial impact (Bell et al. 2002). The overall level of migration is conventionally measured by the Crude Migration Intensity (CMI), computed as:

$$CMI = \frac{M}{P} * 100 \quad (1)$$

where M represents the total number of internal migrants during the observation period and P represents the population at risk measured at the beginning of the observation and expressed as a percentage. Results in Table 7 show that PLIDA overestimates the CMI at all spatial scales and for both 1- and 5-year migration data. This is particularly pronounced when PLIDA is restricted to observations that could be linked to the census, which inflates the CMI by 18% to 32% depending on the spatial scale.

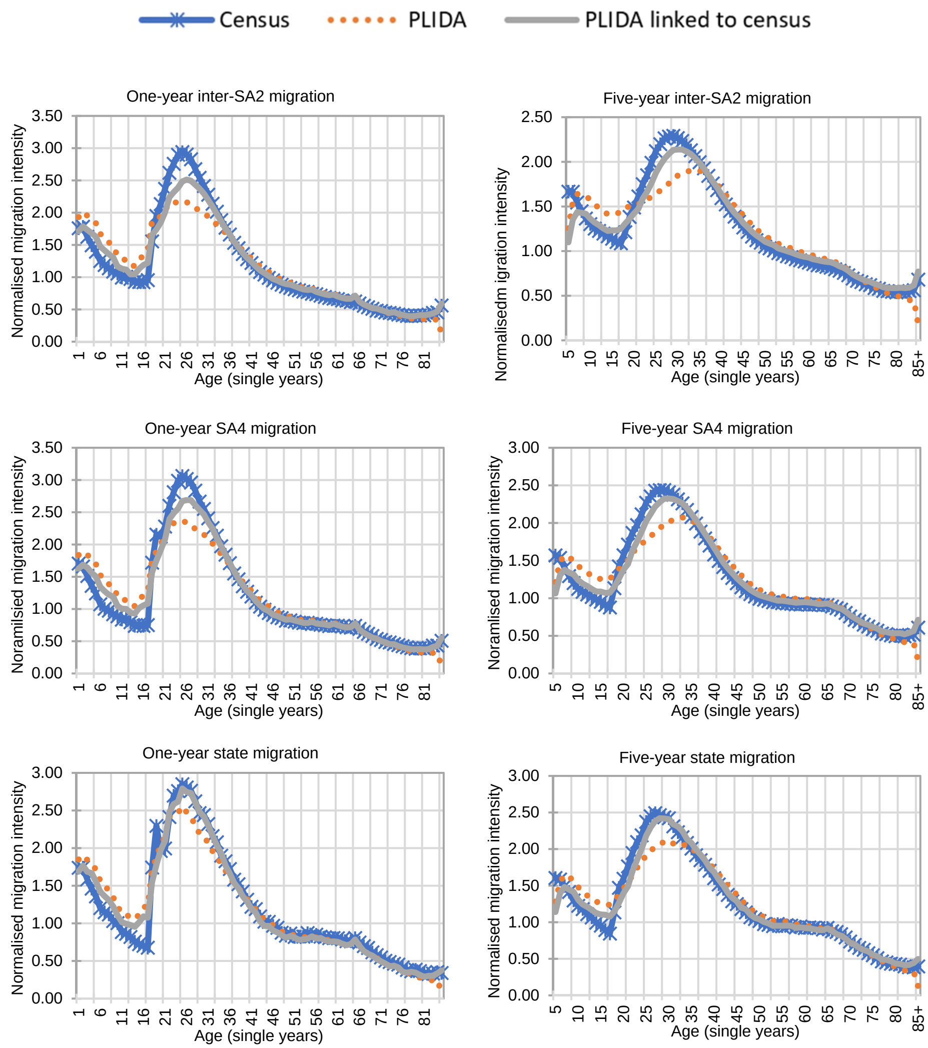
Table 7: Internal migration indicators by data source for 1- year and 5-year migration data

		1-year migration			5-year migration		
		Census	**PLIDA**	**PLIDA linked to census**	**Census**	**PLIDA**	**PLIDA linked to census**
Crude migration intensity	**SA2**	11.29	11.97	13.43	31.0	32.13	36.59
	SA4	5.60	6.35	7.01	16.39	17.90	20.10
	state	1.45	1.78	1.91	4.40	5.07	5.57
Age at peak	**SA2**	25	25	26	29	33	30
	SA4	25	25	26	29	33	30
	State	25	25	25	23	26	24
Normalised intensity at peak	**SA2**	2.94	2.18	2.51	2.29	1.91	2.14
	SA4	3.06	2.37	2.69	2.44	2.07	2.33
	state	2.85	2.55	2.79	2.49	2.12	2.42
Migration effectiveness Index	**SA2**	7.27	5.28	5.76	9.57	7.71	8.23
	SA4	6.23	4.52	4.97	7.48	6.20	6.62
	State	9.52	6.26	7.34	8.80	7.60	8.43

Source: Authors' calculations based on PLIDA's Combined Location Module and data from the 2016 census.

To capture migration age patterns we use two summary metrics, the age and intensity at peak migration, which capture most of the variance in migration age patterns (Rogers and Castro 1981; Bernard, Bell, and Charles-Edwards 2014), accompanied by normalised single-year migration intensity in Figure 1. Migration age patterns are broadly similar across the three data sources, which follow the typical age profile of migration, with migration declining in childhood, peaking in young adulthood, and declining thereafter. Yet there are some significant differences among the three data sources. The peak migration intensity is significantly lower with PLIDA, especially when it is not linked to census data, and the PLIDA-derived age at peak is up to 4 years older than the census measure. In addition, PLIDA slightly overestimates childhood migration.

Figure 1: Age-specific normalised migration intensities by spatial scale for 1- and 5-year migration



Source: Authors' calculations based on PLIDA's Combined Location Module and data from the 2016 census.
Note: Migration age profiles are normalised to unity to compare age patterns independently from differences in migration level.

One of the most striking facets of migration lies in the way it alters settlement patterns by redistributing population between regions. Net migration gains and losses depend on the size of each region's inflows and outflows, but even with large migration flows very little net redistribution can occur if inflows and outflows are closely matched. At a system-wide level, this is captured by the Migration Effectiveness Index (MEI), computed as:

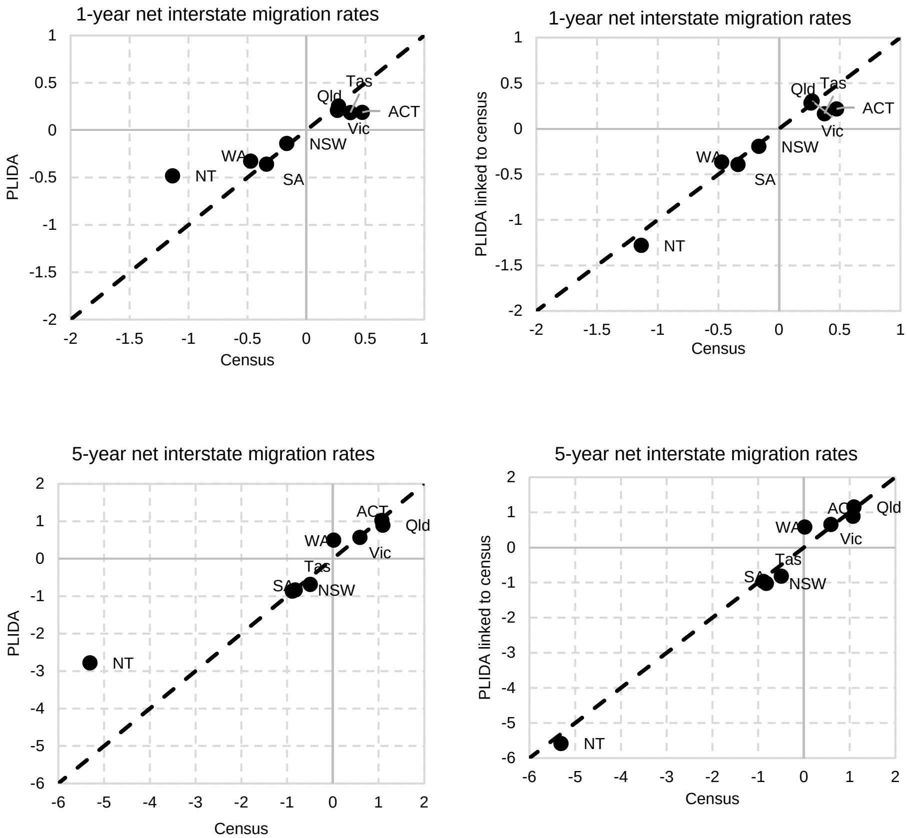
$$MEI = 100 * \sum_i |D_i - O_i| \ / \ \sum_i (D_i + O_i) \qquad (2)$$

where D represents the total inflows to region i and O represents the total outflows from region i. The MEI can assume values between 0 and 100. Higher values indicate that migration flows are asymmetrical, and that migration is highly effective in redistributing population by generating a large net effect for a given volume of migration. The MEI is derived from origin–destination flow matrices.

Results in Table 7 suggest that the effectiveness of migration is lower in PLIDA than the census-derived measure at all spatial scales. This suggests a spatial bias, especially when PLIDA is not linked to the census. We explore this further using net migration rates, which are calculated for each region as the difference between inflows and outflows divided by the population of the region and expressed as a percentage. Figure 2 reports interstate net migration rates obtained from census data (x-axis) against the same net migration rate from PLIDA (y-axis). Results are consistent whether migration is measured over a 1- or 5-year interval. PLIDA thus provides reliable net migration rates, with the notable exception of the Northern Territory: its net migration rate is significantly higher in PLIDA, though not when PLIDA is linked to the census. This discrepancy likely stems from PLIDA's under-coverage of very remote populations, noted earlier.

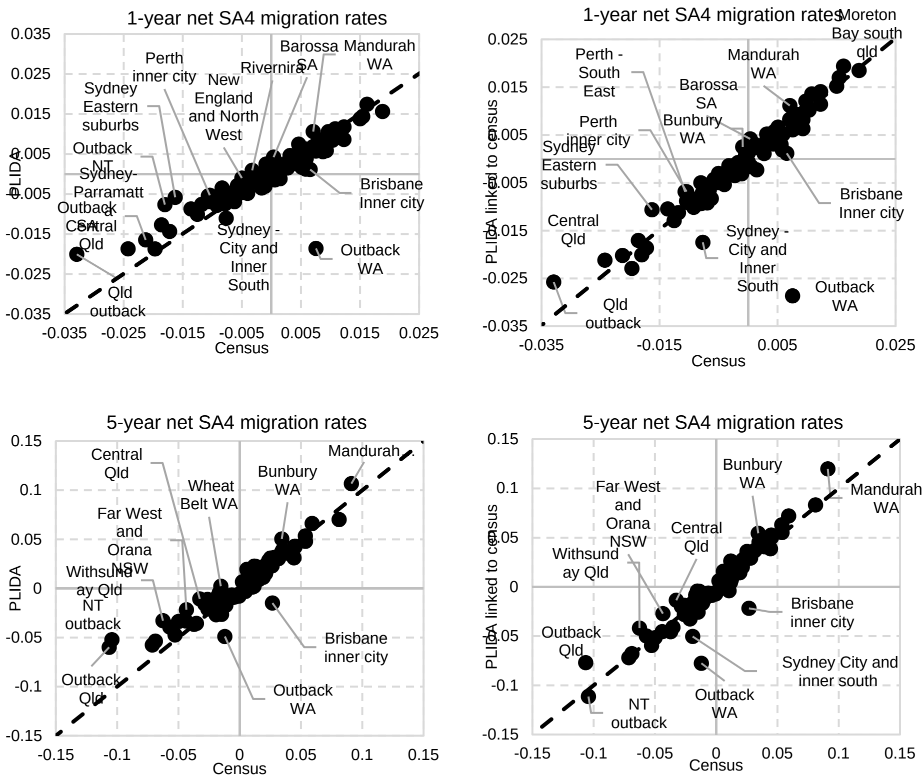
We explore this further by reporting net migration rates at SA4 level in Figure 3. The results reveal 6 types of region for which PLIDA over- or underestimates net migration: (1) very remote regions with small peripatetic populations, particularly in the outback; (2) mining regions such as Far West and Orana in New South Wales that rely on fly-in and fly-out workers; (3) coastal regions such as the Whitsundays in Queensland and Bunbury in Western Australia with a large share of temporary migrants, including working holidaymakers; (4) agricultural regions such as the Barossa in South Australia and the Wheat Belt in Western Australia, which rely on a seasonal workforce; (5) the inner city areas of Brisbane, Sydney, and Perth, with highly mobile and short-term populations such as international students; and (6) rapidly expanding peri-urban regions, such as Moreton Bay (adjacent to Brisbane) and Mandurah (a satellite city of Perth). We conclude that in its current form, PLIDA does not produce population-level internal migration estimates of sufficient quality for the robust analysis of internal migration patterns and trends. This problem is likely exacerbated for lower levels of geography.

Figure 2: Net interstate migration rates by data source for 1- and 5-year migration



Source: Authors' calculations based on PLIDA's Combined Location Module and data from the 2016 census.

Figure 3: Net interstate migration rates by data source for 1- and 5-year migration



Source: Authors' calculations based on PLIDA's Combined Location Module and data from the 2016 census.

6. Discussion and conclusion

6.1 Overview

The establishment of PLIDA represents a major step forward for Australia's sociodemographic data arsenal. It brings Australia's statistical capabilities closer to countries with long-established population registers and administrative datasets such as Norway, Sweden, and the Netherlands. While it opens new research avenues, a thorough and critical examination of the strengths and limitations of PLIDA is an essential prerequisite for ensuring robust empirical research. As migration scholars, we focused our assessment on PLIDA's Combined Location Module, which records changes of address based on the combination of three administrative datasets: Medicare Consumer Directory, Personal Income Tax Office, and DOMINO Centrelink Administrative Data. Our findings provide insights not only for internal migration research but also for any geo-spatial research.

We investigated three inter-related challenges – population coverage, missing values, and spatial mismatch compared with census data – and explored implications for the level and spatial patterns of internal migration. In this section we summarise our findings and provide guidance on how to use PLIDA to avoid the pitfalls we identified. We then discuss the potential benefits of PLIDA and provide examples of research applications. Our overall assessment is that PLIDA can be used in some applied settings outside demography when using the safeguards described below. However, we advise against using it in its current form for demographic applications such as internal migration estimates and population projections where accurate migration data are critical for obtaining robust analyses and reliable population projections.

6.2 Key limitations and potential solutions

Some of the limitations we identified are common to most administrative datasets (McCollum et al. 2021; Ernsten et al. 2018; Foley, Champion, and Shuttleworth 2017; Cook 2001). These include the under-coverage of some groups, particularly temporary and recently arrived immigrants and very remote populations. Researchers interested in these sub-population groups should therefore proceed with care, as further research is needed to understand who within these groups is under-represented. For example, international students who work should be captured by taxation data, so international students out of the labour force are those who are the most likely to be missing in PLIDA. The creation of weighting factors might be a solution to obtain representative results in the future. In the meantime, researchers need to fully understand whether their population of interest is well-captured in PLIDA to make sure they obtain representative results, or be transparent about caveats. For example, PLIDA would not be a useful data source for examining the internal migration of international students after graduation, their labour market outcomes in Australia, or their transition to a skilled visa. This is because it can only provide information on international students who had contact with the Australian government during their stay, which is only a subset of the international student body.

An allied issue is the number of individuals in PLIDA, which significantly exceeds the Australian resident population. After removing deaths and emigration, we obtained a PLIDA population of 29,195,740 people at the 2016 census compared to 22,485,854 census records. We believe that this discrepancy is due to an undercount of emigration and perhaps also a duplication of some records. This a significant issue for any demographic research. We addressed this issue by restricting the analysis to people who are both in PLIDA and the census to ensure that people who had left the country were not counted. We advise researchers to consider a similar approach to obtain results that are representative of the Australian resident population. This approach has important implications for intercensal years that researchers need to consider in order to understand the population they are working with. This finding highlights the importance of censuses to benchmark administrative data and we encourage users to do so with future censuses. More importantly, this shows that censuses continue to play an essential role, even with the advent of administrative data.

However, linking PLIDA to the census is not an ideal solution for longitudinal analysis. To address this issue, the ABS has developed a 'scoping' method to remove individuals who are no longer residing in Australia, based on interactions with 10 government agencies in the previous 12 months. This approach appears to produce more reliable population counts and internal migration estimates for 2021, but at the time of writing this method was not available to PLIDA users outside the ABS or for earlier years (ABS 2021a). We recommend researchers keep abreast of developments and future product releases by the ABS.

The issue of population coverage is compounded by missing values and spatial mismatch, which disproportionally affect highly mobile groups. While the share of records with missing values has been under 2% since 2009, young adults, temporary and recently arrived migrants, renters, individuals who are not in the labour force, and First Nations Australians are more likely to have missing records. This reinforces the need to proceed with care for these sub-population groups and ideally undertake a comparison with the census before analysing PLIDA data.

More concerning is the high rate of spatial mismatch between the census and PLIDA for small spatial units (SA1 and SA2), which is around 20% and is higher among immigrants who have been in Australia for less than 2 years, international students, working holidaymakers, immigrants from North-East Asia, First Nations Australians, and renters. We therefore recommend users not to use PLIDA data on place of residence below the SA4 level. This has important implications for the type of research applications PLIDA can be used for. For example, research on climate-induced migration might consider linking latitude and longitude coordinates of historical cyclone data from the Bureau of Meteorology to PLIDA data at the SA1 level. Our findings suggest that this would result in a high risk of erroneous cyclone exposure.

Overall, internal migration patterns are broadly comparable, but there are some noticeable differences between PLIDA and the census. In particular, PLIDA marginally overestimates the level of internal migration and slightly distorts migration age patterns by underestimating migration among young adults and overestimating childhood migration. Spatial patterns are also broadly comparable, except for rapidly growing peri-urban regions, inner cities with short-term residents, and agricultural, mining, and coastal regions with a seasonal or temporary workforce. Researchers interested in those regions should critically assess any results from PLIDA.

From these results, we conclude that in its current form PLIDA does not produce population-level migration estimates of sufficient quality to robustly analyse internal migration trends and patterns. Consequently, PLIDA estimates should not be used as input for population projections because of the risk of implausible projections, particularly for small areas and rapidly growing regions with high immigration. Weighting factors and a ‘scoping’ method based on recent interactions with multiple government agencies to remove individuals no longer resident in Australia (ABS 2021a) are promising avenues for addressing this issue of representativeness in the future. Again, we recommend PLIDA users to stay informed, as the ABS continuously improves its data products.

One of the unique benefits of PLIDA is its capacity to provide internal migration data in intercensal years. To capitalise on this advantage while addressing the issue of reliability, one way forward might be to apply a modelling framework that combines both the census and PLIDA to produce a reliable time series of internal migration, as done in Europe for international migration estimates (Yildiz et al. 2024).

6.3 Opportunities for internal migration research

In the meantime, PLIDA opens new opportunities for individual-level research by enabling the adoption of a truly longitudinal perspective that dynamically links internal migration to other life course domains thanks to linkage with other administrative datasets, which is where the power of PLIDA lies. While PLIDA does not provide nationally representative internal migration estimates, it should help shed new light on the mechanisms underpinning migration, particularly for migration between SA4 regions and states, if one assumes that the same mechanisms operate for under-represented groups. We outline below two research examples: the determinants and consequences of internal migration.

The Australian government has made increasing efforts to locate permanent immigrants, particularly humanitarian and skilled migrants, in non-metropolitan regions to alleviate population pressure in major cities and address rural labour force shortages (Hugo 2008). This has been achieved through a regional skilled visa scheme that requires immigrants to reside for a minimum of three years to obtain permanent residency. By linking the Combined Location Module to the Visa Applicants and Grants Module, PLIDA should permit, for the first time, reliable measures of the rural retention of this group. Contrary to temporary migrants, this group is unlikely to be missed from PLIDA. This is an example of a research application where underrepresentation is not an issue.

More broadly, understanding the internal migration patterns of permanent immigrants on different visas and whether they concentrate geographically over time is an important avenue for future research. Cross-sectional evidence from Australia suggests that skilled migrants are geographically more mobile than humanitarian migrants, who have less agency (Laukova, Bernard, and Sigler 2022), although Dutch longitudinal administrative data suggest that humanitarian migrants have distinctive internal migration patterns characterised by movement toward ethnically segregated suburbs (Zorlu and Mulder 2008). However, it is unclear how such preferences interplay with the Australian government's effort to settle humanitarian migrants in non-metropolitan regions (Wong, Perales, and Bernard 2023). PLIDA offers an opportunity to generate more systematic evidence and understand how internal migration patterns relate to transitions from temporary to permanent visas and eventually citizenship. Again, one would have to assume that the temporary international migrants who are missed in PLIDA transition to permanent residency similarly to those in PLIDA. Researchers must decide whether such an assumption is tenable regarding their own research subject.

An allied question is the economic benefit of migration, both in terms of wages and occupational mobility. While this question has been extensively investigated in Europe through the escalator regions framework (Fielding 1992; Van Ham et al. 2012), evidence is critically lacking in Australia. This gap can be addressed by linking the Combined Location Modules to employment and income data from the Australian Taxation Office (Biddle and Marasinghe 2021). Further, such explorations using PLIDA would potentially be less sensitive to duplicated records since it would typically entail examining averages rather than population counts. Most studies examine the impact of the last recorded migration on labour market outcomes and assume constant positive economic returns to migration, although the economic benefits of migration may be delayed. For example, in the Netherlands occupational mobility among females is only achieved after three internal migrations (Mulder and Van Ham 2005). PLIDA provides an opportunity to take a long-term perspective in order to understand the cumulative impact of successive migrations on economic outcomes in Australia. It also provides the opportunity to explore migration within data collection periods of the more authoritative data sources, such as the census of population and housing, and provide novel insights such as seasonal variations in migration. However, researchers should proceed with care and bear in mind the limitations we uncovered, including restricting analysis to a large spatial scale and focusing on well-represented groups.

References

ABS, Australian Bureau of Statistics (2021a). Administrative data snapshot of population, methodology. Canberra: ABS.

ABS, Australian Bureau of Statistics (2021b.) Migration, Australia. Canberra: ABS. https://www.abs.gov.au/statistics/people/population/migration-australia/2019-20#regional-internal-migration.

ABS, Australian Bureau of Statistics (2021c). Migration, Australia methodology. Canberra: ABS. https://www.abs.gov.au/methodologies/migration-australia-methodology/2019-20.

ABS, Australian Bureau of Statistics (2021d). Regional internal migration estimates, provisional. Canberra: ABS. https://www.abs.gov.au/statistics/people/population/regional-internal-migration-estimates-provisional/mar-2021.

ABS, Australian Bureau of Statistics (2021e). Using administrative data for the 2021 Census. (Planning the 2021 Census). Canberra: ABS.

ABS, Australian Bureau of Statistics (2022). Overseas migration. Canberra: ABS. https://www.abs.gov.au/statistics/people/population/overseas-migration/latest-release#data-downloads.

ABS, Australian Bureau of Statistics (2023a). Deaths, Australia. Canberra: ABS. https://www.abs.gov.au/statistics/people/population/deaths-australia/latest-release#national.

ABS, Australian Bureau of Statistics (2023b). Regional population. Canberra: ABS. https://www.abs.gov.au/statistics/people/population/regional-population/latest-release.

Bell, M. (1996). How often do Australians move? Alternative measures of population mobility. *Journal of the Australian Population Association* 13: 101–124. doi:10.1007/BF03029490.

Bell, M. (2015). W(h)ither the Census? *Australian Geographer* 46(3): 299–304. doi:10.1080/00049182.2015.1058320.

Bell, M., Blake, M., Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J., and Hugo, G. (2002). Cross-national comparison of internal migration: Issues and measures. *Journal of the Royal Statistical Society: Series A* 165(3): 435–464. doi:10.1111/1467-985X.t01-1-00247.

Bell, M., Charles-Edwards, E., Kupiszewska, D., Kupiszewski, M., Stillwell, J., and Zhu, Y. (2014). Internal migration data around the world: Assessing contemporary practice. *Population, Space and Place* 21(1): 1–17. doi:10.1002/psp.1848.

Bell, M., Charles-Edwards, E., Ueffing, P., Stillwell, J., Kupiszewski, M., and Kupiszewska, D. (2015). Internal migration and development: comparing migration intensities around the world. *Population and Development Review* 41(1): 33–58. doi:10.1111/j.1728-4457.2015.00025.x.

Bernard, A., Bell, M., and Charles-Edwards, E. (2016). Internal migration age patterns and the transition to adulthood: Australia and Great Britain compared. *Journal of Population Research* 33: 123–146. doi:10.1007/s12546-016-9157-0.

Bernard, A., Bell, M., and Charles-Edwards, E. (2014). Life-course transitions and the age profile of internal migration. *Population and Development Review* 40(2): 213–239. doi:10.1111/j.1728-4457.2014.00671.x.

Bernard, A., Forder, P., Kendig, H., and Byles, J. (2017). Residential mobility in Australia and the United States: A retrospective study. *Australian Population Studies* 1(1): 41–54. doi:10.37970/aps.v1i1.11.

Bernard, A. and Kalemba, S. (2022). Internal migration and the de-standardization of the life course: A sequence analysis of reasons for migrating. *Demographic Research* 46(12): 337–354. doi:10.4054/DemRes.2022.46.12.

Bernard, A. and Kolk, M. (2020). Are young Swedes moving more? A cohort analysis of internal migration by move order. *European Journal of Population* 36: 601–615. doi:10.1007/s10680-019-09542-z.

Biddle, N. and Marasinghe, D. (2021). Using census, social security and tax data from the Multi-Agency Data Integration Project (MADIP) to impute the complete Australian income distribution. (Tax and Transfer Policy Institute working paper 8). Canberra: TTPI. doi:10.2139/ssrn.3826001.

Borsellino, R., Bernard, A., Charles-Edwards, E., and Corcoran, J. (2022). A regional renaissance? The shifting geography of internal migration under COVID-19. *Australian Geographer* 53(4) 405–423. doi:10.1080/00049182.2022.2074622.

Campbell, P. (2019). Dispositional traits and internal migration: Personality as a predictor of migration in Australia. *Journal of Research in Personality* 78: 262–267. doi:10.1016/j.jrp.2018.12.009.

Clark, W.A.V. and Lisowski, W. (2018). Examining the life course sequence of intending to move and moving. *Population, Space and Place* 24: e2100. doi:10.1002/psp.2100.

Cook, T. (2001). Evaluation of administrative data sources for use in quarterly estimation of interstate migration between 1996 and 2001. (Demography Working Paper). Canberra: Australian Bureau of Statistics.

Crown, D., Gheasi, M., and Faggian, A. (2020). Interregional mobility and the personality traits of migrants. *Papers in Regional Science* 99(4): 899–914. doi:10.1111/pirs.12516.

DeWaard, J., Hauer, M., Fussell, E., Curtis, K.J., Whitaker, S.D., McConnell, K., Price, K., Egan-Robertson, D., Soto, M., and Anampa Castro, C. (2022). User beware: concerning findings from the post 2011–2012 US Internal Revenue Service migration data. *Population Research and Policy Review* 41: 437–448. doi:10.1007/s11113-021-09663-6.

Ernsten, A., McCollum, D., Feng, Z., Everington, D., and Huang, Z. (2018). Using linked administrative and census data for migration research. *Population Studies* 72(3): 357–367. doi:10.1080/00324728.2018.1502463.

Fielding, A.J. (1992). Migration and social mobility: South East England as an escalator region. *Regional Studies* 26(1): 1–15. doi:10.1080/00343409212331346741.

Foley, B., Champion, T., and Shuttleworth, I. (2017). *Measuring internal migration: Comparing census and administrative data*. Paper presented at UK Administrative Data Research Network Annual Research Conference, Royal College of Surgeons, Edinburgh, UK, 1–2 June 2017 [NILS].

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/Revue de l'intégration et de la migration internationale* 9: 125–145. doi:10.1007/s12134-008-0055-y.

Ishikawa, Y. (2020). Internal migration in Japan. In: Bell, M., Bernard, A., Charles-Edwards, E., and Zhu, Y. (eds.). *Internal migration in the countries of Asia: A cross-national comparison.* Cham: Springer. doi:10.1007/978-3-030-44010-7_7.

Jokela, M. (2021). Personality traits and reasons for residential mobility: Longitudinal data from United Kingdom, Germany, and Australia. *Personality and Individual Differences* 180: 110978. doi:10.1016/j.paid.2021.110978.

Kalemba, S., Bernard, A., Corcoran, J., and Charles-Edwards, E. (2022). Has the decline in the intensity of internal migration been accompanied by changes in reasons for migration? *Journal of Population Research* 39: 279–313. doi:10.1007/s12546-022-09285-5.

King, R. and Skeldon, R. (2010). 'Mind the gap!' Integrating approaches to internal and international migration. *Journal of Ethnic and Migration Studies* 36(10): 1619–1646. doi:10.1080/1369183X.2010.489380.

Kolk, M. (2019). Period and cohort measures of internal migration. *Population* 74(3): 333–350.

Korpi, M. and Clark, W.A.W. (2017). Human capital theory and internal migration: do average outcomes distort our view of migrant motives? *Migration Letters* 14(2): 237–250. doi:10.33182/ml.v14i2.329.

Kulu, H., Lundholm, E., and Malmberg, G. (2018). Is spatial mobility on the rise or in decline? An order-specific analysis of the migration of young adults in Sweden. *Population Studies* 72(3): 323–337. doi:10.1080/00324728.2018.1451554.

Laukova, D., Bernard, A., and Sigler, T. (2022). Settlement and migration patterns of immigrants by visa class in Australia. In: Kawano, M., Kourtit, K., Nijkamp, P., and Higano, Y. (eds). *Theory and history in regional perspective*. Cham: Springer: 165–191. doi:10.1007/978-981-16-6695-7_10.

McCollum, D., Ernsten-Birns, A., Feng, Z., and Everington, D. (2021). Mobile no more? The innovative use of administrative data linked to a census-based longitudinal study to investigate migration within Scotland. *Population, Space and Place* 27(7): e2312. doi:10.1002/psp.2312.

Morrison, P.A. (1971). Chronic movers and the future redistribution of population: A longitudinal analysis. *Demography* 8(2): 171–184. doi:10.2307/2060607.

Mulder, C.H. and Van Ham, M. (2005). Migration histories and occupational achievement. *Population, Space and Place* 11(3): 173–186. doi:10.1002/psp.365.

Mulder, C.H., Lundholm, E., and Malmberg, G. (2020). Young adults' return migration from large cities in Sweden: The role of siblings and parents. *Population, Space and Place* 26(7): e2354. doi:10.1002/psp.2354.

Perales, F. and Bernard, A. (2023). Continuity or change? How the onset of COVID-19 affected internal migration in Australia. *Population, Space and Place* 29(2): e2626. doi:10.1002/psp.2626.

Petzold, K. (2017). Mobility experience and mobility decision-making: An experiment on permanent migration and residential multilocality. *Population, Space and Place* 23(8): e2065. doi:10.1002/psp.2065.

Poot, J. and Sanderson, L. (2015). Researching duration of residence: From census and postal surveys to integrated data infrastructure. *New Zealand Population Review* 41: 69.

Poulain, M., Herm, A., and Depledge, R. (2013). Central population registers as a source of demographic statistics in Europe. *Population* 68(2): 183–212. doi:10.3917/pope.1302.0183.

Rees, P. and Lomax, N. (2019). Ravenstein revisited: The analysis of migration, then and now. *Comparative Population Studies* 44. doi:10.12765/CPoS-2020-10.

Rogers, A. and Castro, L.J. (1981). Model migration schedules. (IIASA research reports RR-81-30). Laxemburg: International Institute for Applied System Analysis (IIASA).

Rogers, A., Raymer, J., and Newbold, K.B. (2003). Reconciling and translating migration data collected over time intervals of differing widths. *The Annals of Regional Science* 37: 581–601. doi:10.1007/s00168-003-0128-y.

Rogerson, P.A. (1990). Migration analysis using data with time intervals of differing widths. *Papers of the Regional Science* 68(1): 97–106. doi:10.1111/j.1435-5597.1990.tb01196.x.

Sander, N. and Bell, M. (2014). Migration and retirement in the life course: An event history approach. *Journal of Population Research* 31(1): 1–27. doi:10.1007/s12546-013-9121-1.

Sesma Carlos, D., Kok, J., and Oris, M. (2022). Internal migrant trajectories within The Netherlands, 1850–1972: Applying cluster analysis and dissimilarity tree methods. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*: 55(3): 1–22. doi:10.1080/01615440.2022.2047852.

Shuttleworth, I., Stevenson, C., Bjarnason, Þ., and Finell, E. (2020). Geography, psychology and the 'Big Five'personality traits: Who moves, and over what distances, in the United Kingdom? *Population, Space and Place*: 27(3): e2418. doi:10.1002/psp.2418.

Smith, D. and Smith, T. (2014). Using linking to sharpen policy-thinking: Early findings from the Australian census and migrants integrated dataset. *Australian Economic Review* 47(2): 220–230. doi:10.1111/1467-8462.12069.

Thomas, M., Gillespie, B., and Lomax, N. (2019). Variations in migration motives over distance. *Demographic Research* 40(38): 1097–1110. doi:10.4054/DemRes.2019.40.38.

Valente, P. (2019). Modernizing the census in Europe: traditional and new methods for the 2020 round. Paris: N-IUSSP. https://www.niussp.org/environment-and-development/modernizing-the-census-in-europe-traditional-and-new-methods-for-the-2020/.

Van Ham, M., Findlay, A., Manley, D., and Feijten, P. (2012). Migration, occupational mobility, and regional escalators in Scotland. *Urban Studies Research* 2012(1): 827171. doi:10.1155/2012/827171.

Vidal, S. and Lersch, P.M. (2021). Panel data in research on mobility and migration: A review of recent advances. *Comparative Population Studies* 46: 121–150. doi:10.12765/CPoS-2021-07.

Vidal, S., Perales, F., Lersch, P.M., and Brandén, M. (2017). Family migration in a cross-national perspective: The importance of within-couple employment arrangements in Australia, Britain, Germany, and Sweden. *Demographic Research* 36(10): 307–338. doi:10.4054/DemRes.2017.36.10.

Watson, N. (2020). Measuring geographic mobility: Comparison of estimates from longitudinal and cross-sectional data. *Survey Research Methods* 14(1): 1–18. doi:10.18148/srm/2020.v14i1.7422.

White, M.J. (ed.) (2016). *International handbook of migration and population distribution.* Dordrecht: Springer. doi:10.1007/978-94-017-7282-2.

Wong, C.Y., Perales, F., and Bernard, A. (2023). Non-metropolitan settlement and integration outcomes of humanitarian migrants in Australia: Combining spatial, multidimensional and longitudinal perspectives. *Journal of Rural Studies* 104: 103151. doi:10.1016/j.jrurstud.2023.103151.

Yildiz, D., Wiśniowski, A., Abel, G.J., Weber, I., Zagheni, E., Gendronneau, C., and Hoorens, S. (2024). Integrating traditional and social media data to predict bilateral migrant stocks in the European Union. *International Migration Review* 0(0). doi:10.1177/01979183241249969.

Zhang, G. and Campbell, P. (2012). Data survey: Developing the statistical longitudinal census dataset and identifying its potential uses. *Australian Economic Review* 45(1): 125–133. doi:10.1111/j.1467-8462.2011.00673.x.

Zorlu, A. and Mulder, C.H. (2008). Initial and subsequent location choices of immigrants to the Netherlands. *Regional Studies* 42(2): 245–264. doi:10.1080/00343400601145210.

Appendix

Table A-1: Datasets with enduring linkages to PLIDA

Dataset	Description
Australian Bureau of Statistics	
Census of Population and Housing 2011, 2016, 2021	Census data provides a rich snapshot of the nation and informs government, communities, and businesses. Current ABS policy restricts the linking of data from more than one census. This means data from multiple censuses cannot be brought together for longitudinal PLIDA analysis. The exception to this is the Australian Census Longitudinal Dataset, which brings together a small sample of census data exploring how Australian society changes over time.
Australian Census Longitudinal Dataset 2016	The Australian Census Longitudinal Dataset brings together a 5% sample of census data from the 2006, 2011, and 2016 censuses. It contains information such as age, sex, country of birth, labour force status, dwelling, and household and family characteristics, across each of these censuses. Researchers can apply to use the ACLD for longitudinal PLIDA analysis for specific projects.
National Health Survey (2014–2015, 2017–2018)	Information on Australian's health and wellbeing such as medical conditions, health and lifestyle risk factors, mental health, and use of health services.
Survey of Disability, Ageing, and Carers (2018)	Information about people with a disability and older people (aged 65 and over), and their carers.
ABS Business Characteristics Survey (2005–2006 to 2017–2018)	An annual survey providing estimates in business use of information technology, innovation, and a broad range of other non-financial business characteristics.
BLADE Core Dataset Indicative data items BLADE Locations data Business Activity Statement Business Income Tax Pay As You Go	An economic data tool combining tax, trade, and intellectual property information with ABS data to provide a better understanding of the Australian economy and business performance over time. A limited set of BLADE information (generally from 2001 onwards) is available in PLIDA.
Australian Taxation Office	
Personal Income Tax data: Client Register (from 2006) Payment Summary (from 2010–2011) Income Tax Return (from 2010–2011)	Information about taxpayers' occupation and income, employment payments and amounts withheld during a financial year, and all persons with a registered tax file number (TFN) for tax and superannuation purposes.
Single Touch Payroll (from 2020–2021)	Information about employees' salaries and wages, pay as you go (PAYG) withholding, and superannuation, as reported through the Single Touch Payroll system.
JobKeeper (2020–2021)	Information about JobKeeper subsidy payments to businesses affected by coronavirus (COVID-19).
Department of Health	
Pharmaceutical Benefits Scheme (from 2006)	Information about the use of prescription medications & services subsidised under the PBS.
Medicare Benefits Schedule (from 2006)	Information on the usage of Medicare-subsidised healthcare services.
Centralised Register of Medical Practitioners (Provider Directory) (from 2011)	Information about registered medical practitioners, including specialties.
Australian Immunisation Register (AIR) (from 2010)	Information about COVID-19 and other vaccination status, including core demographics.

Table A-1: (Continued)

Dataset	Description
Services Australia	
Medicare Consumer Directory (2006–June 2020)	All persons with an active Medicare enrolment. This dataset was previously known as the Medicare Enrolments Database.
Department of Social Services	
DOMINO Centrelink Administrative data (from 2009)	Data Over Multiple Individual Occurrences (DOMINO) contains snapshots of the characteristics of recipients of government payments such as the Age Pension and JobSeeker (formerly Newstart Allowance). This dataset was previously known as Social Security and Related Information (SSRI).
Data Exchange (from July 2015)	Program performance information that contains de-identified data on clients that receive social services including their demographics and services being delivered.
Department of Education, Skills and Employment	
Australian Early Development Census (AEDC) (2009, 2012, 2015, 2018)	Measures how well children in their first year of full time school are developing across 5 important domains.
Australian Apprenticeships Incentives Program (AAIP) and Training Contracts (from 2006)	Information on apprenticeships and traineeships, AAIP payments, and Trade Support Loan (TSL) payments.
Higher Education Information Management System Data (HEIMS) (from 2005)	Information on higher education studies including course type, student status, and study load.
Provider Registration and International Students Management System (PRISMS) (from 2006)	Information on course enrolments of overseas students.
Total VET Activity (TVA) (from 2015)	Information about students and the Vocational Education and Training (VET) delivery activities of registered training organisations. This dataset is supplied by agreement with the Department of Education, Skills, and Employment, the Office of the Student Identifiers Registrar, and the National Centre for Vocational Educational Research.
National Disability Insurance Agency	
National Disability Insurance Scheme data (from June 2019)	Information on primary disability, including services attained, by participants in the National Disability Insurance Scheme
State and Territory Registrars of Births, Deaths, and Marriages	
Death Registrations (from 2005)	Information relating to registered death records from Australian States and Territories.
Department of Home Affairs	
Migration data: Visa information & Citizenship grants (from 2000) Client information (from 1984) Net Overseas Migration (NOM) Traveller data (from 2004) Skilled Migration Points (from 2005) Adult Migrant English Program data (from 2003)	Information about various migrant types, including permanent, skilled, temporary, and other migrant programs, including their demographics and movement over time.

Source: The Australian Bureau of Statistics, https://www.abs.gov.au/about/data-services/data-integration/data-integration-project-register#PLIDA , page accessed on 11/03/2022.

Table A-2: Australian statistical geographic standard

Spatial scale	Definition	Number of units
Statistical Areas Level 1 (SA1)	SA1s generally have a population of 200 to 800 people, and an average population of about 400 people. SA1s are designed to be either urban or rural in character and to represent Aboriginal and Torres Strait Islander communities as accurately as possible, particularly in remote areas. SA1s are generally internally connected by road transport.	61,845
SA2	SA2s generally have a population between 3,000 and 25,000 with an average of about 10,000 people. Their purpose is to represent a community that interacts socially and economically.	2,473
SA3	SA3s are designed to have populations of between 30,000 and 130,000 people. SA3s are often the functional areas of regional towns and cities with a population in excess of 20,000 or clusters of related suburbs around urban commercial and transport hubs within the major urban areas.	359
SA4	SA4s are the largest sub-state regions in the Main Structure of the ASGS. Most SA4s have a population above 100,000 people to provide sufficient sample size for Labour Force estimation. SA4s are designed to incorporate both labour supply (where people live) and demand (where people work).	108
Greater Capital Cities Statistical Areas (GCCSA)	Each state and territory is divided into greater capital city and rest of state, with the expectation of the Australian Capital Territory and other territories	16
States and territories	New South Wales, Victoria, Queensland, South Australia, Western Australia, Tasmania, Northern Territory, Australian Capital Territory and other territories (Jervis Bay Territory, Territory of Christmas Island, Territory of the Cocos (Keeling) Islands and Norfolk Island)	9

Source: The Australian Bureau of Statistics, Main Structure and Greater Capital Statistical Areas (Main Structure and Greater Capital City Statistical Areas | Australian Bureau of Statistics (abs.gov.au)