Author/s:
Andrew, NE;Cadilhac, DA;Sundararajan, V;Thrift, AG;Anderson, P;Lannin, NA;Kilkenny, MF

Title:
Linking Australian Stroke Clinical Registry data with Australian government Medicare and medication dispensing claims data and the potential for bias

Date:
2021-08-01

Citation:
Andrew, N. E., Cadilhac, D. A., Sundararajan, V., Thrift, A. G., Anderson, P., Lannin, N. A. & Kilkenny, M. F. (2021). Linking Australian Stroke Clinical Registry data with Australian government Medicare and medication dispensing claims data and the potential for bias. Australian and New Zealand Journal of Public Health, 45 (4), pp.364-369. https://doi.org/10.1111/1753-6405.13079.

Persistent Link:
https://hdl.handle.net/11343/280404

Linking Australian Stroke Clinical Registry data with Australian government Medicare and medication dispensing claims data and the potential for bias

Nadine E. Andrew,[1,2] Dominique A. Cadilhac,[1,3] Vijaya Sundararajan,[4] Amanda G. Thrift,[1] Phil Anderson,[5,6] Natasha A. Lannin,[7,8] Monique F. Kilkenny[1,3]

Abstract

Objective: We aim to report the accuracy of linking data from a non-government-held clinical quality registry to national claims data and identify associated sources of systematic bias.

Methods: Patients with stroke or transient ischaemic attack admitted to hospitals participating in the Australian Stroke Clinical Registry (AuSCR) were linked with Medicare and medication dispensings through the Australian Medicare enrolment file (MEF). The proportion of registrants in the datasets was calculated and factors associated with a non-merge assessed using multivariable analyses.

Results: A total of 17,980 AuSCR registrants (January 2010 – July 2014) were submitted for linkage (median age 76 years; 46% female; 67% ischaemic stroke); the proportion merged was 97% MEF, 93% Medicare and 95% medication dispensings. Data from registrants born in Asia were less likely to link with the MEF (adjusted Odds Ratio [aOR]: 0.20; 95%Confidence Interval [CI]: 0.15, 0.27). Data for those aged 85-plus compared to those under 65 years were less likely to merge with Medicare (aOR 0.25; 95%CI:0.21, 0.30) but more likely to merge with dispensing claims data (aOR: 2.15 (95%CI:1.71, 2.69).

Implications for public health: Linkage between the AuSCR, a national clinical quality registry and Commonwealth datasets was achieved and potential sources of bias were identified.

Key words: data linkage, record linkage, health data, clinical registry, stroke, primary care, medication, bias

In Australia, healthcare is mainly provided and funded within a public/private federated system with services provided by all levels of government: commonwealth, state and territory, and local as well as private providers. Each routinely collects their own data for different administrative purposes.[1] The Commonwealth Government is responsible for managing and collecting data related to health programs, including those funded under the Medicare Benefits Schedule (MBS or Medicare) and Pharmaceutical Benefits Scheme (PBS or subsidised 'medication dispensing') and collectively termed 'Commonwealth datasets' for this article.[2,3] These data have been used by government for surveillance and quality improvement purposes, and by researchers to supplement research data from clinical trials and cohort studies.[4-6]

Non-government operated clinical quality registries, such as the Australian Stroke Clinical Registry (AuSCR) have been established to provide more detailed disease-specific measures such as clinical care indicators and patient-reported outcomes (e.g. quality of life) on clinical populations, since these are unavailable in government datasets.[7] National linkages between claims data and clinical registries are common in regions with a unique National Health Index (NHI number) such as Scandinavia and New Zealand, but are relatively new in Australia.[8-10] Recent advances in data linkage and data security mean that the infrastructure is now available to enable regular linkage of Commonwealth-held administrative datasets and data routinely collected by clinicians for quality improvement or outcomes research with a waiver of consent in countries that do not have an NHI number.[11,12]

The lack of an NHI number means that probabilistic and/or deterministic methods are required to perform linkages based on the available identifying (personal) information in the datasets to be linked. In the case of

1. *Stroke & Ageing Research, Department of Medicine, School of Clinical Sciences at Monash Health, Monash University, Victoria*
2. *Peninsula Clinical School, Central Clinical School, Monash University, Victoria*
3. *Florey Institute of Neuroscience and Mental Health, Victoria*
4. *Department of Public Health, School of Psychology and Public Health, College of Science, Health and Engineering, La Trobe University, Victoria*
5. *Health Linkage Unit, Australian Institute of Health and Welfare, Australian Capital Territory*
6. *Faculty of Health, University of Canberra, Australian Capital Territory*
7. *Department of Neuroscience, Central Clinical School, Monash University, Victoria*
8. *Alfred Health (Allied Health), Victoria*

Correspondence to: Associate Professor Nadine Andrew, Academic Unit, Peninsula Health, 2 Hastings Road Frankston, Victoria; e-mail: nadine.andrew@monash.edu

Submitted: July 2020; Revision requested: November 2020; Accepted: December 2020

The authors have stated they have no conflict of interest.

Aust NZ J Public Health. 2021; Online; doi: 10.1111/1753-6405.13079

large administrative datasets, the linkages are frequently performed by government data linkage centres and the linked de-identified data is provided to the researchers. Consequently, formally published evaluations of such activities and the potential impact for research are lacking, especially with regards to linkages between clinical quality registries and government-held data. Herein, we report the accuracy of linking data from the AuSCR, a non-government-held clinical quality registry, to the national Australian Medicare enrolment file (MEF). A secondary aim was to compare characteristics of registrants whose data were and were not able to be merged with the MEF and the content data from the associated national datasets to identify potential sources of systematic bias associated with non-linkage.

Methods

Data sources

We used data from four main data sources (Figure 1) as outlined below.

Stroke cohort (AuSCR): The AuSCR is a prospective national clinical quality registry, initiated to monitor, promote and improve the quality of acute stroke care in Australia (www.auscr.com.au).[13] Data are collected on all patients with a clinical diagnosis of stroke or transient ischaemic attack (TIA) admitted to participating hospitals using an opt-out method of consent. The AuSCR provides the systematic collection of data to identify variability in clinical care and health outcomes. A minimum core dataset is collected and includes demographic, clinical and hospital outcome data. Survivors who are discharged from hospital are also contacted at 90–180 days following stroke with a request to complete a follow-up questionnaire. The registry holds patients' details such as full name, address, sex and date of birth to enable routine patient follow-up. AuSCR data are linked annually to the National Death Index (NDI) by the Australian Institute of Health and Welfare (AIHW).[14]

Medicare enrolment file: The Medicare enrolment file contains identifiable information on all Australians enrolled in the Australian Medicare Scheme regardless of whether or not a claim has been made. Australia has universal healthcare whereby all people living in Australia who are Australian or New Zealand citizens or hold a permanent residency visa are covered for basic health care services. However, to claim Medicare-funded healthcare or medication dispensings covered under the PBS those eligible are required to first enrol in Medicare.

Medicare claims data: The MBS database contains transactional data related to all services that are subsidised under the Medicare scheme by the Commonwealth Government such as general practitioners (GPs), also known as primary care physicians, and specialist visits, imaging and pathology. Non-residents from countries with a reciprocal healthcare agreement with Australia (N=11 countries) are also covered for medically necessary care. In addition to the restrictions applied to Medicare enrolment, veterans who hold a Department of Veterans Affairs (DVA) gold card and their families are not included in the MBS database as they are covered under a separate and more comprehensive DVA scheme.[3] However, they are included in the MEF. Following discharge from hospital after a stroke, patients should ideally be returned to the care of their GP for ongoing secondary prevention and disability management. There are a number of items that can be provided and claimed by a GP to support care planning that are relevant to survivors of stroke. For our study, all items related to GP and allied health care as well as relevant specialties (e.g. neurology, cardiology, rehabilitation and geriatrics) were requested. Dates of service and an encrypted service provider number are available and can be used to determine continuity and regularity of GP care.[1,3,15]

Medication dispensing data: The PBS database contains records of medications dispensed that are subsidised by the Australian Commonwealth Government.[1,2] Medications such as over-the-counter medications or those provided while in hospital that are privately purchased or are funded under other specialty schemes (such as for chemotherapy) are not included. All Australian residents who hold a current Medicare card, or are covered under the Medicare reciprocal arrangements, are eligible under the PBS. At the time of this project, the Repatriation Schedule of Pharmaceutical Benefits (RPBS) data were made available to the research team. In line with clinical guidelines for the long-term secondary prevention of stroke,[16] we requested data on all drugs dispensed with World Health Organization (WHO) Anatomical Therapeutic Classification (ATC) codes relevant to antihypertensive, antithrombotic and lipid-lowering medications.

Linkage of datasets

Person-level linkages were requested between all the AuSCR registrants who had been included in the registry between January 2010 and December 2014 as well as their records in the Medicare enrolment file held by the AIHW Data Linkage Unit for claims made between June 2010 to June 2015 (Supplementary Table S1).[2,17] A two-stage separation model of data linkage was used. This means that linkage variables pre-specified by the AIHW data linkage unit (name, address, date of birth, date of death [if applicable], sex) were submitted by the AuSCR data manager with a unique project identifier or linkage key attached. Linkages were performed by staff at the AIHW data linkage unit using a combination of deterministic and probabilistic methods. A deterministic approach, whereby record pairs are considered a match if they exactly agree or disagree on a given set of patient identifiers, was used for the first round of linkages. Probabilistic matching, a weighted statistical approach that considers the degree to which patient identifiers agree and so can account, to some extent, for common typographical errors, was applied to the remaining unlinked records. This was followed by an extensive clerical review by the AIHW staff whereby uncertain matches were checked manually. This approach has been shown to yield high levels of sensitivity and specificity for linkages using the personal identifiers listed above.[14,18]

Approved de-identified content data were then provided to the research team by the two organisations (AIHW and the AuSCR) and merged by the researchers using a unique project-specific, patient-level identification number (or 'linkage key' to ensure the inability to re-identify the data). As this was a complex linkage project containing large and detailed unit record datasets, the research team were required, by the Commonwealth Department of Health, to access and merge the content data via a remote access computing environment known as the Secure Unified Research Environment (SURE), managed by the Sax Institute.

Ethics approvals

Institutional ethics was obtained from Monash University (2017-7864 and 2018-12405). Additional approvals from ethics committees and data custodians were obtained from the AIHW for access to the

MBS and PBS data (EO2017-1-346). A Public Health Act approval was required to allow submission of identifiers from the AuSCR registrants who resided in Queensland to the AIHW Data Linkage Unit. Approval was also sought from the AuSCR Research Task Group and AuSCR Steering Committee. The AuSCR uses an opt-out model of consent (opt-out rate approximately 3%) and a waiver of consent was granted by the reviewing HRECs for data linkage.

Statistical analysis

Descriptive statistics were used to determine the proportion of the AuSCR registrants who were present in the MEF, Medicare dataset or medication dispensing dataset or both. Differences in patient characteristics were compared between registrants who were and were not able to be merged to these datasets. As there were no demographic details provided to the research team from the Commonwealth datasets, we were unable to validate the quality of the matches once they were merged, e.g. compare sex or age matches when contained in both the AuSCR and Medicare or medication dispensings datasets. Instead, we used the date of death obtained from the National Death Index (NDI) that was previously captured within the AuSCR dataset to investigate the proportion of Medicare and medication dispensing claims recorded after death.

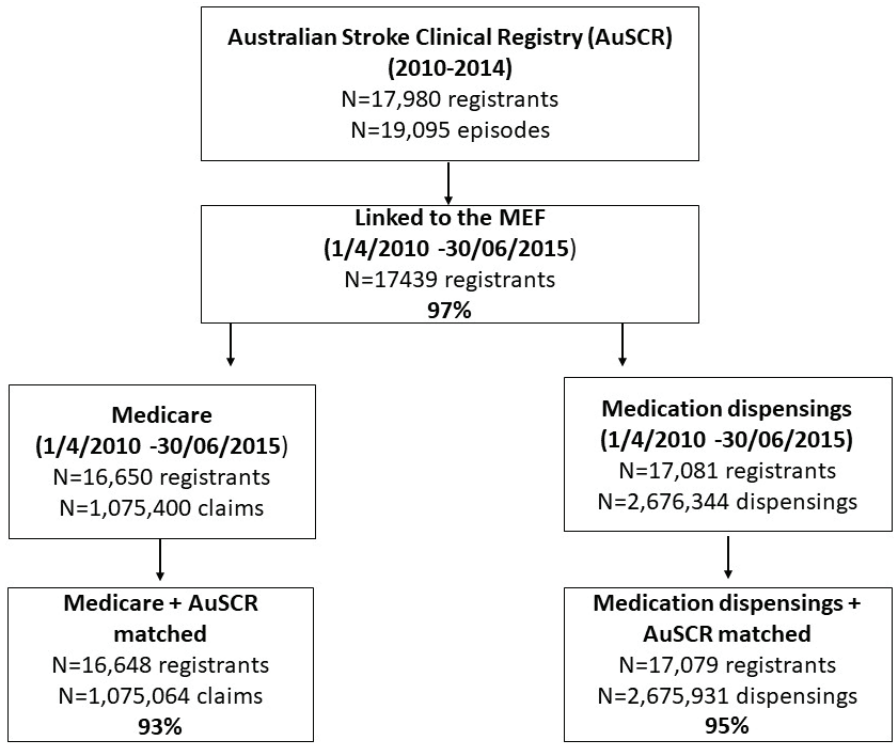
Multivariable multilevel logistic regression, with levels defined as patient and hospital, was used to investigate patient demographic and clinical factors associated with registrant data not being merged with each of the Medicare and medication dispensing datasets (Supplementary Table S1). Covariates included in the model were all of the clinical and demographic variables routinely collected in the AuSCR (Table S1). Data were missing for the following variables: age (0.03%), stroke type (0.05%) and level of social advantage (3.3%). Models were tested for collinearity and odds ratios and corresponding confidence intervals were reported. Sensitivity analyses were performed excluding those registered in the AuSCR in 2010 so that all had a minimum of 12 months of pre-stroke claims data. Prior to April 2012, medications that were under the co-payment threshold were not recorded in the PBS. To understand the potential impact of this on our results, a second sensitivity analysis was performed for those registered in the AuSCR after April 2012. All analyses were performed using Intercooled STATA/SE 15.0 for Windows (Statcorp, College Station, USA, 2017).

Results

Data from 17,980 AuSCR registrants submitted by 26 hospitals in New South Wales, Victoria, Queensland, Tasmania and Western Australia were provided to the AIHW for linking: 46% female, median age (Quartile1 [Q1], Quartile 3 [Q3]) 76 years (65, 84), 67% ischaemic stroke and 16% TIA. Among these, 17,439 registrants (97%) were able to be linked to the MEF, 17,098 (95%) registrants' data were merged with medication dispensing data and 16,648 (93%) were merged with Medicare claims data (Figure 1) resulting in 2,675,931 medication dispensing records and 1,075,064 Medicare claims from the 4.5 years of claims data (Figure 1). Examination of claims that occurred after a death date indicated that 127 (0.8%) of cohort members had one or more (total 1,481) Medicare claims at a median of 5 days (Q1: 1, Q3: 206) after date of death and 411 (2.4%) had one or more (total 5,076) medications dispensed at a median of 6 days (Q1: 2, Q3: 30) after date of death.

Registrants whose data were linked to the MEF differed to those whose data were successfully merged to the Medicare and medication content data, according to the characteristics of registrants (Table 1). Differences in factors associated with being successfully merged with the Medicare data also differed to the medication dispensing data. In multivariable analyses (Table 2), age was not associated with a successful linkage to the MEF. However, registrants aged 85 years or more were less likely to have their registry data merged with the Medicare data than the other age groups (adjusted Odds Ratio [aOR]: 0.25, 95% Confidence Interval [CI]: 0.21, 0.30, reference group <65 years) and those aged <65 years were less likely to have their registry data merged with the medication dispensing data than the other age groups. Compared to registrants born in Australia, those born in Asia were least likely to be linked to the MEF (aOR 0.20, 95%CI: 0.15, 0.27) and least likely to be merged to the Medicare (aOR: 0.45, 95%CI: 0.35, 0.58) and medication dispensing data (aOR: 0.28, 95%CI: 0.22, 0.36). Similar results were observed for those requiring an interpreter. Registrants who died in hospital were less likely to be merged than those who did not die in hospital for both datasets but not the MEF (Table 2). Factors associated with documentation such as missing postcode and undetermined stroke types were also

Figure 1: Merging of the Australian Stroke Clinical Registry with the Medicare Enrolment File, Medicare and medication dispensing datasets.



associated with a non-linkage or non-merge.

Discussion

Results support the linkage of the AuSCR, a national researcher-held clinical quality registry, with Commonwealth datasets through the MEF, where previously we have linked clinical quality registry with hospital data.[11,14] In our study, a large proportion of registrants were linked to the MEF and merged between datasets. The small proportion of claims for medication dispensing or Medicare items made after a date of death were consistent with claim delays, Medicare-funded death confirmation[3] or claim delays or errors.[19] Although these factors indicate overall accuracy for linkages of this kind, we have identified potential sources of systematic bias associated with the small proportion of people whose data were unable to be linked to the MEF or merged with the claims data.

An advantage of using routinely collected data for monitoring and surveillance is a reduced susceptibility to selection bias compared to trials and cohort studies that rely on active recruitment of participants. However, the systematic exclusion of populations due to an inability to link their information or eligibility based exclusions can also result in biases leading to inaccurate conclusions from the data.[20] Unlike clinical trials data, registry data relies on accurate, routine inputting of data by clinical staff who work in the hospital and may have variable training in research. Although training and audits of data are in place to maximise data quality there is still the potential for missing or poorly recorded variables that may impact linkage to other datasets, as noted in the reduced linkage accuracy for those with missing postcode or stroke type. Since data are rarely missing at random, this situation may disproportionately impact certain patient types.

Non-merges as opposed to non-linkages may also be influenced by not having a record in the associated datasets. Patients with stroke require regular contact with their GP and use secondary prevention medications following their stroke.[16] Although all registrants eligible for Medicare should have had one or more claims in these datasets, this may not always be the case. For example, despite recommendations in guidelines for the prescription of secondary prevention medications following a stroke, there is evidence that younger patients are less likely to be prescribed these medications than older patients.[21,22] Non-merges may also be due to claims-based eligibility rather than Medicare eligibility.[23] In Australia, the majority of veterans are elderly; it is likely that the absence of DVA gold card recipients in the Medicare dataset contributed to the reduced ability to merge registrants aged over 85 years to these data. The reduced probability of merging but not linkage for in-hospital death is another example and likely a reflection of survivor bias in which those who survived had an increased opportunity to have a Medicare claim and medication dispensing claims.[24] The impact of this form of bias on study design and outcomes has been well documented in the pharmacoepidemiology literature.[25] Knowledge of these nuances is important for accurate interpretation of these types of linked data.

The greatest potential source of bias is the poor linkage and merging rates for those born in Asian countries compared to registrants of other nationalities. People born in countries such as China and India make up the majority of people residing in Australia on non-resident visa types. These countries do not have reciprocal Medicare arrangements with Australia.[3] Differences in naming conventions for surnames and given names between Asian and Western cultures along with poor English language skills (i.e. requiring an interpreter) can lead to inaccuracies in the recording of details in both administrative and registry datasets, thereby affecting linkage accuracy.[3,25] Given the median age of our cohort, this is likely to account for the majority of the non-merges. These biases have been shown to

Table 1: Proportion of AuSCR registrants in each category merged with Medicare data, medication dispensing data, or either dataset.

AuSCR merged with	AuSCR registrants in each category Total N=17,980	Medicare only	Medication dispensing only	MEF
	n	%	%	%
Female	8,289	92	95	97
Age Groups				
<65 years	4,502	95	93	97
65-74 years	4,038	95	95	96
75- 84 years	5,584	95	96	97
85 years or more	3,850	84	96	97
Country of birth				
Australia	12,597	92	96	98
United Kingdom	1273	94	95	96
Europe	2,480	96	95	97
Asia	721	85	83	87
Other	909	91	90	93
Interpreter required	941	89	89	90
IRSAD				
Most disadvantaged	1,977	94	95	98
Second most disadvantaged	2,827	92	96	98
Third most disadvantaged	3,403	94	97	98
Fourth most disadvantaged	3,968	94	96	98
Least disadvantaged	5,211	92	95	97
Postcode unknown	594	74	76	79
In-hospital stroke	894	94	96	98
Able to walk on admission	6,767	95	96	97
Type of stroke				
Intracerebral haemorrhage	2,452	92	93	97
Ischaemic stroke	12,077	93	95	97
Transient ischaemic attack	2,816	94	96	98
Undetermined	626	88	94	96
Discharged to aged care facility	847	90	96	97
Died in hospital	1,826	86	90	98
Metropolitan residence	14,787	92	97	97

Notes:
AuSCR: Australian Stroke Clinical Registry; IRSAD: Index of Relative Socio-economic Advantage and Disadvantage

be particularly problematic with linkages to death data where a non-linkage would be assumed to mean that the person had not died.[26] Biased assumptions relating to disparities in healthcare access, medication adherence or physician attendances by cultural subgroups are also likely. These biases will be inflated when measuring multiple episodes across longitudinally linked data, thereby creating multiple opportunities for missed linkages. For example, a missed linkage to a medication claim as a result of incorrect recording of personal identifiers can mean registrants may be incorrectly assumed to have poor adherence to medication due to fewer medication claims over a given period.

The research potential of this large linked dataset is great and will enable us to gain novel insights into the primary care management of patients following stroke. This information will help address gaps in recommendations for primary care attendances and enable us to describe changes in the use of subsidised medication dispensing and Medicare claim items as a consequence of stroke. It will also allow a comprehensive assessment of the discontinuation and adherence to medications as well as medical consultations before and after patients experience a stroke/TIA event. Use of data on medication dispensing has been found to be reliable and valid for health services research and monitoring in Australia.[18]

The main limitation of our findings is that our analysis of covariates was limited to those contained within the AuSCR data and there may be other factors associated with absences in the Commonwealth datasets that we were not able to discern. We did not have access to identifiable data and so were not able to perform our own manual review to discern the accuracy of the linkages. As such, we were only able to address the issue of false-negative errors (records that did not find a match) and could not investigate false-positive errors. Nevertheless, we have demonstrated some systematic biases in the characteristics of people contained in the final merged datasets, which may affect the generalisability of future research findings. This knowledge is important for accurate interpretation of analyses when using these and similar linkages for epidemiological research.

Our results confirm that good quality linkages can be achieved between the AuSCR and the Medicare enrolment file, creating a case for future linkages of this kind. We have also identified potential sources of linkage bias that may need to be addressed through sensitivity analyses and other analytic techniques to ensure accurate interpretation of findings to guide policy and practice.

Table 2: Factors associated with successful merging of the AuSCR with Medicare or medication dispensing data.

AuSCR merged with	MEF aOR (95% CI) N=17,974	Medicare aOR (95% CI) N=17,675	Medication dispensing aOR (95% CI) N=17,675
	n	%	%
Patient characteristics			
Female	0.80 (0.67, 0.96)	1.06 (0.94, 1.20)	0.92 (0.79, 1.06)
Age (ref <65 years)	Reference	Reference	Reference
65-74 years	0.80 (0.62, 1.03)	0.87 (0.71, 1.13)	1.51 (1.24, 1.83)
75- 84 years	1.06 (0.82, 1.36)	0.85 (0.70, 1.03)	2.11 (1.74, 2.57)
85 years or more	1.16 (0.86, 1.55)	0.25 (0.21, 0.30)	2.15 (1.71, 2.69)
Country of birth			
Australia	Reference	Reference	Reference
United Kingdom	0.56 (0.40, 0.78)	1.46 (1.12, 1.90)	0.72 (0.55, 0.95)
Europe	0.76 (0.57, 1.02)	2.24 (1.78, 2.83)	0.89 (0.71, 1.13)
Asia	0.20 (0.15, 0.27)	0.45 (0.35, 0.58)	0.28 (0.22, 0.36)
Other	0.35 (0.26, 0.48)	0.74 (0.57, 0.96)	0.45 (0.35, 0.59)
Interpreter required	0.46 (0.34, 0.62)	0.66 (0.51, 0.86)	0.59 (0.45, 0.77)
Able to walk on admission	1.17 (0.95, 1.43)	1.22 (1.06, 1.41)	1.15 (0.97, 1.36)
In-hospital stroke	1.44 (0.88, 2.34)	1.28 (0.96, 1.71)	1.32 (0.92, 1.88)
IRSAD			
Most disadvantaged	Reference	Reference	Reference
Second most disadvantaged	0.81 (0.55, 1.19)	0.81 (0.63, 1.09)	0.88 (0.65, 1.19)
Third most disadvantaged	1.64 (0.98, 2.75)	1.15 (0.84, 1.38)	1.63 (1.10, 2.43)
Fourth most disadvantaged	1.16 (0.83, 1.63)	0.97 (0.77, 1.23)	1.19 (0.92, 1.54)
Least disadvantaged	0.99 (0.66, 1.48)	0.94 (0.72, 1.19)	1.00 (0.73, 1.37)
Postcode unknown	0.09 (0.06, 0.13)	0.16 (0.12, 0.21)	0.14 (0.10, 0.19)
Type of stroke			
Intracerebral haemorrhage	Reference	Reference	Reference
Ischaemic stroke	0.86 (0.65, 1.12)	0.90 (0.75, 1.07)	1.18 (0.97, 1.43)
Transient ischaemic attack	0.95 (0.65, 1.37)	0.90 (0.71, 1.15)	1.24 (0.93, 1.64)
Undetermined	0.42 (0.25, 0.68)	0.53 (0.38, 0.73)	0.55 (0.36, 0.83)
Discharged to aged care facility	1.19 (0.77, 1.84)	1.21 (0.94, 1.57)	0.94 (0.65, 1.35)
Died in hospital	1.34 (0.95, 1.89)	0.57 (0.48, 0.67)	0.34 (0.28, 0.42)
Metropolitan residence	0.69 (0.47, 1.00)	0.95 (0.78, 1.16)	0.59 (0.37, 0.94)

Notes:
AuSCR: Australian Stroke Clinical Registry; IRSAD: Index of Relative Socio-economic Advantage and Disadvantage, also adjusted for year

Acknowledgements

We thank members of the AuSCR Steering Committee and staff from the Florey Institute of Neuroscience and Mental Health who manage the AuSCR. In particular, we wish to acknowledge Mr Sam Shehata, AuSCR data manager, for his assistance with data preparation and submission. We also thank the hospital clinicians and patients who contribute data to the AuSCR. We also acknowledge the staff from the AIHW Data Linkage Unit who undertook the data linkage for this project.

Funding

The AuSCR received funding from a range of government, nongovernment, and industry sources. The following authors receive research fellowship support: Professors Thrift (NHMRC:1042600), Lannin (Heart Foundation: 1112158) and Cadilhac (co-funded Heart Foundation and NHMRC: 1154273); Associate Professors Andrew (NHMRC: 1072053) and Kilkenny (NHMRC: 1109426).

References

1. Australian Institute of Health and Welfare. *Australia's Health 2016*. Canberra (AUST): AIHW; 2016.
2. Australian Department of Health. *Pharmaceutical Benefits Advisory Committee Guidelines*. Canberra (AUST): Government of Australia; 2018.
3. Australian Department of Health. *Medicare Benefits Schedule Book*. Canberra (AUST): Government of Australia; 2018.
4. Sharma Y, Thompson C, Miller M, Shahi R, Hakendorf P, Horwood C, et al. Economic evaluation of an extended nutritional intervention in older Australian hospitalized patients: A randomized controlled trial. *BMC Geriatr*. 2018;18(1):41.
5. Patterson A, Hure A, Burrows T, Jackson J, Collins C. Diet quality and 10-year healthcare costs by BMI categories in the mid-age cohort of the Australian Longitudinal Study on Women's Health. *J Hum Nutr Diet*. 2018;31(4):463-72.
6. Freed GL, Allen AR. Variation in outpatient consultant physician fees in Australia by specialty and state and territory. *Med J Aust*. 206(4):176-80.
7. Cadilhac DA, Andrew NE, Lannin NA, Middleton S, Levi CR, Dewey HM, et al. Quality of acute care and long-term quality of life and survival: The Australian Stroke Clinical Registry. *Stroke*. 2017;48(4):1026-32.
8. Chan WC, Jackson G, Wright CS, Orr-Walker B, Drury PL, Boswell DR, et al. The future of population registers: Linking routine health datasets to assess a population's current glycaemic status for quality improvement. *BMJ Open*. 2014;4(4):e003975.
9. Grey C, Jackson R, Wells S, Thornley S, Marshall R, Crengle S, et al. Maintenance of statin use over 3 years following acute coronary syndromes: A national data linkage study (ANZACS-QI-2). *Heart*. 2014;100(10):770-4.
10. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: Possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol*. 2009;24(11):659-67.
11. Andrew NE, Sundararajan V, Thrift AG, Kilkenny MF, Katzenellenbogen J, Flack F, et al. Addressing the challenges of cross-jurisdictional data linkage between a national clinical quality registry and government-held health data. *Aust N Z J Public Health*. 2016;40(5):436-42.
12. Ung D, Kim J, Thrift AG, Cadilhac DA, Andrew NE, Sundararajan V, et al. Promising use of big data to increase the efficiency and comprehensiveness of stroke outcomes research. *Stroke*. 2019;50(5):1302-9.
13. Cadilhac DA, Lannin NA, Anderson CS, Levi CR, Faux S, Price C, et al. Protocol and pilot data for establishing the Australian Stroke Clinical Registry. *Int J Stroke*. 2010;5(3):217-26.
14. Kilkenny M, Kim J, Andrew N, Sundararajan V, Thrift A, Katzenellenbogen J, et al. Maximising use of data and avoiding data waste: Case study from stroke. *Med J Aust*. 2019;210:27-31.
15. Youens D, Moorin R, Preen D, Harris M, Robinson S. Regular General Practitioner contact - analysis of methods for measurementusing administrative data. *Int J Popul Data Sci*. 2018;3:267.
16. Guidelines Content Working Party Stroke Foundation. *Clinical Guidelines for Stroke Management 2017*. Melbourne (AUST): Stroke Foundation; 2018.
17. Australian Department of Health. *MBS Online - Medicare Benefits Schedule*. Canberra (AUST): Government of Australia; 2018.
18. Mellish L, Karanges EA, Litchfield MJ, Schaffer AL, Blanch B, Daniels BJ, et al. The Australian Pharmaceutical Benefits Scheme data collection: A practical guide for researchers. *BMC Res Notes*. 2015;8:634.
19. Kelman C. The Australian National Death Index: An assessment of accuracy. *Aust N Z J Public Health*. 2000;24(2):201-3.
20. Bohensky MA, Jolley D, Sundararajan V, Evans S, Pilcher DV, Scott I, et al. Data linkage: A powerful research tool with potential problems. *BMC Health Serv Res*. 2010;10:346.
21. Lannin NA, Anderson CS, Kim J, Kilkenny M, Bernhardt J, Levi C, et al. Treatment and outcomes of working aged adults with stroke: Results from a national prospective registry. *Neuroepidemiology*. 2017;49(3-4):113-20.
22. Andrew N, Kim J, Thrift A, Kilkenny M, Lannin N, Anderson C, et al. Prescription of antihypertensive medication at discharge influences survival following stroke. *Neurology*. 2018;90:1-9.
23. Randall S, Brown A, Boyd J, Schnell R, Borgs C, Ferrante A. Sociodemographic differences in linkage error: An examination of four large-scale datasets. *BMC Health Serv Res*. 2018;18(1):678.
24. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol*. 2008;167(4):492-9.
25. Miller EA, McCarty FA, Parker JD. Racial and ethnic differences in a linkage with the national death index. *Ethn Dis*. 2017;27(2):77-84.
26. Lariscy JT. Differential record linkage by Hispanic ethnicity and age in linked mortality studies: Implications for the epidemiologic paradox. *J Aging Health*. 2011;23(8):1263-84.

Supporting Information

Additional supporting information may be found in the online version of this article:

Supplementary Table 1: Project variables and data sources.