Neuroepidemiology

Original Paper

Neuroepidemiology
DOI: 10.1159/000534242

Received: February 20, 2023
Accepted: September 20, 2023
Published online: September 26, 2023

Differing Definitions of First-Ever Stroke Influence Incidence Estimates More than Trends: A Study Using Linked Administrative Data

David Youens[a, b] Judith Katzenellenbogen[a] Rathina Srinivasa Ragavan[c]
Nita Sodhi-Berry[a] Jennie Carson[a] Dawit Zemedikun[a] Amanda G. Thrift[c]
Valery Feigin[d] Lee Nedkoff[a, e]

[a]Cardiovascular Epidemiology Research Centre, School of Population and Global Health, The University of Western Australia, Crawley, WA, Australia; [b]School of Population Health, Curtin University, Bentley, WA, Australia; [c]Department of Medicine, School of Clinical Sciences at Monash Health, Monash University, Melbourne, VIC, Australia; [d]National Institute for Stroke and Applied Neuroscience, Auckland University of Technology, Auckland, New Zealand; [e]Victor Chang Cardiac Research Institute, Darlinghurst, NSW, Australia

Keywords

Incidence · Methodology · Data-linkage · Rates · Epidemiology

Abstract

Introduction: Researchers apply varying definitions when measuring stroke incidence using administrative data. We aimed to investigate the sensitivity of incidence estimates to varying definitions of stroke and lookback periods and to provide updated incidence rates and trends for Western Australia (WA). ***Methods:*** We used linked state-wide hospital and death data from 1985 to 2017 to identify incident strokes from 2005 to 2017. A standard definition was applied which included strokes coded as the principal hospital diagnosis or the underlying cause of death, with a 10-year lookback used to clear prevalent cases. Alternative definitions were compared against the standard definition by percentage difference in case numbers. Age-standardised incidence rates were calculated, and age- and sex-adjusted Poisson regression models were used to estimate incidence trends. ***Results:*** The standard definition with a 10-year lookback period captured 31,274 incident strokes. Capture increased by 19.3% when including secondary diagnoses, 4.1% when including nontraumatic subdural and extradural haemorrhage, and 8.1% when including associated causes of death. Excluding death records reduced capture by 11.1%. A 20-year lookback reduced over-ascertainment by 2.0%, and a 1-year lookback increased capture by 13.3%. Incidence declined 0.6% annually (95% confidence interval −0.9, −0.3). Annual reductions were similar for most definitions except when death records were excluded (−0.1%, CI: −0.4, 0.2) and with the shortest lookback periods (greatest annual reduction). ***Conclusion:*** Stroke incidence has declined in WA. Differing methods of identifying stroke influence estimates of incidence to a greater extent than estimates of trends. Reductions in stroke incidence over time are primarily driven by declines in fatal strokes.

Introduction

Stroke remains a leading cause of mortality globally [1], accounting for 2.4% of the burden of disease in Australia [2]. Accurate estimates of stroke incidence inform resource allocation for prevention and treatment.

karger@karger.com
www.karger.com/ned

Correspondence to:
David Youens, david.youens@uwa.edu.au

Administrative data are appealing for assessing incidence due to their capture of complete populations and low cost compared to prospective data collection [3]. Linking data on hospitalisations and deaths enables the capture of first-ever strokes, with a lookback or clearance period typically applied to differentiate first-ever from recurrent strokes [4].

Methods used to capture incident strokes in administrative data vary widely. Differences exist in the diagnoses considered to reflect stroke and the number of diagnosis fields reviewed, while some investigators consider care type or length of stay (LOS) [5–8]. Additional diagnosis fields may be used to identify historical strokes [5, 8, 9], while some investigators do not differentiate between potential incident strokes and prevalent strokes [7, 10, 11]. Some investigators include deaths occurring outside hospital [12], while others exclude these so do not capture total incidence [5, 6, 9, 11, 13]. The availability of historical linked data varies between jurisdictions and data sources within jurisdictions, resulting in varying lookback periods. Finally, some studies include transient ischaemic attack (TIA), while others do not. Changes in diagnostic criteria [14] mean that many events previously considered TIA would now be diagnosed as stroke [15].

Differences in stroke capture methods can affect estimates of incidence and trends, potentially biasing comparisons between populations or studies. Trend estimates may be particularly impacted where changes have occurred over time in data coding or diagnostic practices as for stroke/TIA [14].

Using data from Western Australia (WA), a state with a population of 2.6 million in 2017 and Australia's longest-running collection of linked health data, we aimed to investigate the sensitivity of estimates and trends of stroke incidence from administrative data according to (1) the definition of the potential incident strokes captured, (2) the characteristics used to assess historical strokes, and (3) the duration of the lookback period. We also provide contemporary estimates of stroke incidence and trends for WA (2005–2017).

Materials and Methods

Data Sources

The source dataset included all WA residents hospitalised with a principal or secondary diagnosis of cardiovascular disease or TIA or dying from cardiovascular disease or TIA, coded as underlying or additional cause of death from 1985 onwards (International Classification of Diseases, version 10, Australian Modification (ICD-10-AM) codes I00-I99 plus G45.x, ICD-9-Clinical Modification [CM]/ICD-9 390–459). Hospitalisation data were from the Hospital Morbidity Data Collection (HMDC), which includes all inpatient episodes in all public and private hospitals in WA (with principal and 20 secondary diagnoses provided to researchers). ICD codes are used in Australian hospitals for financial and reporting purposes; these are entered by staff trained in clinical coding or health information management [16]. Death data were from the WA Death Registrations dataset which captures all deaths in WA, including in-hospital and emergency department (ED) deaths. Data were linked and extracted by the WA Data Linkage Branch.

Incident Stroke Capture

Potential incident strokes from 2005 to 2017 were identified using hospitalisation and death records and classified as incident when no historical stroke was recorded within the lookback period. Admissions recorded during or ≤1 day after a prior hospitalisation were combined to form a continuous episode of care. Potential incident strokes were flagged regardless of sequence within an episode of care. The admission date for the first hospitalisation in the episode was usually defined as the stroke date. When a non-stroke admission >2 days in length was followed by a transfer with a principal stroke diagnosis, the transfer date was used as the date of stroke, as these were assumed to be in-hospital strokes.

Stroke Definitions

Our definition of stroke reflects the most common methods in existing literature (the "standard definition") (Table 1). We also applied alternative definitions, reflecting the heterogeneity reported in previous literature. Each component of the alternative definitions was tested separately to estimate the influence of each element on stroke incidence and trends, in comparison with the standard definition.

Methods applied in previous studies vary widely. In capturing potential incident strokes, most studies incorporate the principal hospital diagnosis field only [6, 8, 17], although some also incorporate secondary diagnoses [9, 19]. Most studies include subarachnoid and intracerebral haemorrhage, cerebral infarction, nontraumatic intracranial haemorrhage, and unspecified stroke codes [7, 8, 13, 17]. However, some exclude unspecified strokes [6], while others include nontraumatic subdural and extradural haemorrhages [6, 7, 11], retinal vascular occlusions [5, 8], all cerebrovascular disease [10], or TIA [11]. Some studies exclude non-acute care type [5–8] or short LOS followed by discharge home (assumed to indicate non-acute admissions) [8]. When flagging historical strokes, some authors adopt a broader set of diagnosis fields [8] or diagnoses [5, 8, 9], or apply the same criteria used for incident strokes [7, 10, 11, 13, 17]. Table 1 outlines definitions used in our study, including the lookback periods. Lookback periods were of fixed length, beginning on the date of the potential incident stroke and capturing all admissions within the prior 10 years (or alternative duration).

Statistical Analysis

For each alternative definition, we recorded the number of incident strokes captured and the percentage difference in ascertainment compared with the standard definition, calculated as follows: ([count with alternative definition – count with standard definition]/count with standard definition)*100.

This was repeated for stratifications of age, sex, Indigenous status, socioeconomic status, remoteness, and study year.

Table 1. Definitions of potential incident and historical stroke applied

Parameter	Standard definition	Alternative definitions
Potential incident strokes		
Data source	Hospitalisations and deaths [12]	Exclude deaths [5, 6, 9, 11, 13], i.e., hospital (and in some cases ED) records
Diagnoses[a]	Subarachnoid haemorrhage, intracerebral haemorrhage, nontraumatic intracranial haemorrhage, cerebral infarction, unspecified stroke type [4, 7, 8, 17], and sequelae of cerebrovascular disease (in death data[c])	Include nontraumatic subdural and extradural haemorrhage[d] [6, 7, 11]
Diagnosis fields[b]	Principal diagnosis only [6, 8, 16]	Include all secondary diagnoses [9, 17]
Cause of death fields	Underlying cause	Include additional causes
Length of stay	No restrictions	Exclude separations to home on same date as admission [8]
Admission type	No restrictions	Exclude non-acute, non-emergency records [5–8]
Historical strokes		
Data source	Hospitalisations only	N/A
Diagnoses	As for potential incident strokes (above), plus sequelae of cerebrovascular disease in hospitalisation data	Nontraumatic subdural and extradural haemorrhage included when part of alternative potential incident stroke definition
Diagnosis fields	Principal and all secondary	N/A
Length of stay	No restrictions	N/A
Admission type	No restrictions	N/A
Lookback period length	10 years	20, 15, 8, 5, 3, 1 years, and no lookback

[a] International Classification of Diseases codes for each diagnosis (ICD-10-AM/ICD-9-CM) are as follows: subarachnoid haemorrhage, I60.x/430.x; intracerebral haemorrhage, I61.x/431.x; nontraumatic intracranial haemorrhage, I62.9/432.9; unspecified stroke, I64/436; sequelae of cerebrovascular disease, I69/438; subdural haemorrhage, I62.0/432.1; nontraumatic extradural haemorrhage, I62.1/432.0; cerebral infarction, I63.x/433.x1+434.x1. Prior to July 1995, the 5th digit on codes 433 and 434 did not exist. Through this period, 434.x was included and 433.x was excluded, in line with definitions applied by the Australian Institute of Health and Welfare [18]. Codes I65-I68 (non-stroke cerebrovascular diseases) did not contribute to definitions. [b] Includes cases identified from ICD-10-AM codes used for the standard definition only. [c] These represent strokes which were not hospitalised, were hospitalised prior to the lookback period, or were hospitalised outside WA. Although the date of these strokes is unknown, omitting these would result in under-ascertainment; hence, these have been included with date of death assigned as the stroke date. These represent 2.1% of strokes captured with the standard definition. [d] In principal diagnosis field.

Remoteness was based on the Accessibility/Remoteness Index of Australia [20], which classifies areas as major cities, inner regional, outer regional, remote and very remote, reflecting accessibility to services. Socioeconomic status was based on the area-level Socio-Economic Indices for Areas – Index of Relative Socio-Economic Disadvantage (SEIFA-IRSD) [21]. Statistical area level 2 areas were stratified into quintiles from least to greatest disadvantage. Crude stroke incidence rates were produced for each stratification (online suppl. File 1; for all online suppl. material, see https://doi.org/10.1159/000534242).

We computed age-standardised stroke incidence rates for each study year using the direct method [22], with the WA population [23] as the denominator and the 2016 Australian population as the standard [24]. Denominators were adjusted to remove prevalent individuals [25], separately for each definition. We estimated trends in stroke incidence using Poisson regression models in STATA [22]. The average annual percentage change in stroke incidence was calculated from the exponential of the beta-coefficient for the calendar year as a continuous variable adjusting for 5-year age group and sex. Finally, for each definition by year, we report the percentage of incident strokes with a hospitalisation for TIA (ICD-10-AM G45/ICD-9-CM 435) during the lookback period. This study was approved by the WA Department of Health Human Research Ethics Committee (#2014/55).

Results

Incident Stroke Events

The standard definition captured 31,274 incident strokes during the study period (Table 2). Approximately half were women, and 53.2% within those aged ≥75 years,

Table 2. Incident (first-ever) strokes captured by the standard definition[1] and percentage differences[2] in ascertainment with each alternative definition

	Standard definition, *n*	Independent changes to potential incident stroke definition							Changes to length of lookback period in years						
		additions to standard definition				exclusions from standard definition									
		all I62[6]	2nd diag.	I62[6] and 2nd diag.	all COD fields	same-day discharges home	non-acute non-emergency records	death records (underlying cause only)	20	15	8	5	3	1	0
Overall	31,274	4.1	19.3	24.9	8.1	−1.1	−1.0	−11.1	−2.0	−1.4	1.2	3.9	7.2	13.3	38.7
Sex															
Female	15,450	2.7	18.1	21.8	9.0	−0.9	−1.1	−14.9	−2.1	−1.5	1.3	4.0	7.6	13.9	40.3
Male	15,824	5.5	20.4	27.9	7.2	−1.2	−0.9	−7.3	−2.0	−1.4	1.2	3.8	6.8	12.8	37.1
Age, years															
Below 55	4,459	3.9	24.3	29.6	10.2	−2.6	−0.5	−2.8	−0.6	−0.4	0.4	1.1	2.1	4.2	21.4
55–74	10,175	4.7	19.2	25.3	3.7	−1.3	−0.7	−3.5	−1.5	−1.0	0.8	2.6	4.8	9.2	27.2
75 or above	16,640	3.8	18.0	23.3	10.3	−0.5	−1.3	−17.9	−2.7	−2.0	1.8	5.5	10.1	18.3	50.3
Indigenous status[3]															
Indigenous	949	6.6	22.8	31.4	8.9	−0.6	−0.2	−5.7	−1.8	−1.4	0.8	4.5	8.2	17.9	48.2
Non-indigenous	30,318	4.0	19.2	24.7	8.1	−1.1	−1.0	−11.2	−2.0	−1.4	1.3	3.9	7.2	13.2	38.4
Accessibility[3, 4]															
Major cities	22,880	4.2	19.3	25.2	8.1	−1.2	−1.0	−11.4	−2.1	−1.4	1.4	4.2	7.7	13.9	39.6
Inner regional	3,253	3.7	20.2	25.0	8.9	−1.0	−1.3	−11.6	−2.0	−1.4	1.1	3.7	6.6	12.8	39.8
Outer regional	2,802	3.4	19.0	23.3	9.0	−0.6	−1.0	−11.8	−2.2	−1.7	0.9	3.5	6.7	12.3	38.1
Remote	996	3.9	20.8	25.8	8.2	−0.6	−0.4	−6.9	−1.9	−1.2	0.7	3.1	5.9	12.3	33.4
Very remote	524	6.1	20.8	27.9	4.8	−0.8	−0.4	−4.2	−1.5	−1.3	0.8	3.6	5.9	13.2	35.5
SES[3, 5]															
Q1	8,344	3.6	20.0	24.8	8.7	−0.8	−1.1	−11.4	−2.3	−1.6	1.4	4.9	8.4	15.2	42.0
Q2	6,703	4.2	19.4	25.3	8.2	−0.9	−1.0	−11.9	−2.2	−1.6	1.3	3.8	7.6	13.9	40.1
Q3	5,831	4.1	19.5	25.2	8.9	−1.1	−0.9	−11.7	−2.1	−1.5	1.3	4.0	7.5	13.9	39.6
Q4	4,849	4.4	19.6	25.6	7.5	−1.2	−1.1	−10.1	−1.8	−1.1	1.4	3.4	6.7	12.4	37.3
Q5	4,668	4.5	18.4	24.7	7.0	−1.9	−1.0	−9.9	−1.6	−1.2	1.0	3.3	5.9	10.9	34.3

Accessibility, accessibility to services; COD, cause of death; 2nd diag., secondary diagnoses; SES, socioeconomic status. [1]2.5% of strokes were flagged subsequent to an initial non-stroke admission >2 days in length and had the transfer date assigned as the stroke date. [2]Percentage differences were calculated as ([count with alternative definition – count with standard definition]/count with standard definition)*100. [3]Does not sum to total due to missing information for some individuals. [4]Based on the Accessibility and Remoteness Index of Australia [18]. [5]Based on the Socio-Economic Index For Areas – Index of Relative Socio-Economic Disadvantage [19]. Q1 indicates greatest disadvantage. [6]ICD-10-AM code I62, i.e., nontraumatic subdural or extradural haemorrhage.

Table 3. Percentage change[1] in stroke ascertainment for each definition in comparison to the standard definition by calendar year (WA, 2005–2017)

Year of incident stroke	Standard definition, *n*	Independent changes to potential incident stroke definition							Changes to length of lookback period in years						
		additions to standard definition				exclusions from standard definition									
		all I62[2]	2nd diag.	all I62[2] & 2nd diag.	all COD fields	same-day discharge home	non-acute non-emergency records	death records (underlying cause only)	20	15	8	5	3	1	0
2005	2,117	4.3	21.4	27.1	9.5	−0.6	−0.3	−10.5	−1.8	−1.5	1.5	5.1	8.9	16.6	45.9
2006	2,166	4.8	21.7	27.0	6.5	−1.4	−0.3	−13.1	−1.6	−1.2	0.8	4.0	7.0	14.0	43.1
2007	2,187	3.1	19.0	23.7	8.0	−0.6	−0.4	−12.7	−2.1	−1.6	1.2	3.6	7.0	13.5	39.2
2008	2,282	3.8	17.7	22.6	7.2	−0.8	−0.3	−14.3	−2.0	−1.5	1.2	3.3	7.1	15.0	44.1
2009	2,172	4.4	19.5	25.1	9.3	−0.8	−0.6	−13.4	−2.0	−1.5	1.7	4.7	8.5	14.5	40.1
2010	2,208	4.8	18.1	24.2	10.2	−1.2	−0.4	−13.0	−1.9	−1.3	1.4	4.4	8.7	15.7	41.7
2011	2,262	4.6	19.6	25.7	7.9	−1.4	−0.4	−11.4	−2.2	−1.6	1.3	4.4	7.9	14.0	38.8
2012	2,429	4.7	19.7	25.9	8.6	−1.3	−0.3	−11.4	−2.3	−1.4	1.2	4.0	6.8	11.8	35.5
2013	2,503	3.6	19.8	25.7	8.2	−1.0	−1.0	−10.7	−2.1	−1.6	1.3	4.0	7.3	12.7	34.5
2014	2,661	4.2	19.5	25.5	8.1	−0.9	−1.0	−10.0	−2.2	−1.4	1.2	3.7	6.6	11.5	34.1
2015	2,660	4.7	20.5	27.2	7.4	−1.0	−1.7	−8.6	−1.7	−1.1	1.4	3.4	6.6	12.4	37.8
2016	2,847	2.7	16.6	21.1	7.4	−1.1	−2.8	−8.3	−2.2	−1.4	1.1	3.6	6.7	12.3	36.6
2017	2,780	4.0	18.1	23.6	7.5	−1.4	−2.5	−8.5	−2.0	−1.4	0.9	3.1	5.8	11.2	35.1
Total	31,274	4.1	19.3	24.89	8.1	−1.0	−1.0	−11.01	−2.0	−1.4	1.2	3.9	7.2	13.3	38.7

COD, cause of death; 2nd diag., secondary diagnoses. [1] Percentage differences were calculated as ([count with alternative definition – count with standard definition]/count with standard definition)*100. [2] ICD-10-AM code I62, i.e., nontraumatic subdural or extradural haemorrhage.

with a greater proportion among those living in areas of greatest disadvantage (27.5%) relative to least disadvantage (15.4%).

Compared with the standard definition, capture increased by 4.1% when including nontraumatic subdural and extradural haemorrhage (ICD-10-AM codes I62.x) as the principal diagnosis, 19.3% when including diagnoses from all secondary diagnosis fields, and 24.9% with a combination of these approaches. This combined approach includes nontraumatic subdural and extradural haemorrhages in secondary diagnosis fields that are not included in either of the single definitions (Table 2; online suppl. Table S1). Incident strokes captured via the HMDC using the standard definition had a median of five secondary diagnoses recorded (mean 6.0). Using all causes of death compared with underlying cause alone increased capture by 8.1%. Exclusions of same-day discharges home and non-acute non-emergency records each made minimal difference. When mortality data were excluded, stroke capture was reduced by 11.1%. Use of a longer 20-year lookback period resulted in a 2.0% reduction in stroke capture, representing over-ascertainment using the 10-year lookback from prevalent strokes incorrectly classified as incidents. When compared to the 10-year lookback, shorter lookbacks resulted in over-ascertainment of cases by 3.9% for 5 years, 13.3% for 1 year, and 38.7% for no lookback period.

The effect of alternative definitions differed substantially between subgroups (Table 2). A 1-year lookback resulted in an over-ascertainment of 18.3% in those aged ≥75 years versus 4.2% in those aged <55 years. The exclusion of mortality data resulted in a larger reduction in those aged ≥75 years (−17.9%) than those aged <55 (−2.9%) and caused a larger reduction in females (−14.9%) than males (−7.4%). The effects of changes in definition were relatively constant over time (Table 3), except for lookback period length, where the over-ascertainment observed with shorter lookback periods reduced over time.

The effect of lookback period duration also differed when using only hospital data (excluding death data). A 1-year lookback resulted in 9.9% over-ascertainment compared to 13.3% when using both data sources, and no lookback resulted in 19.7% over-

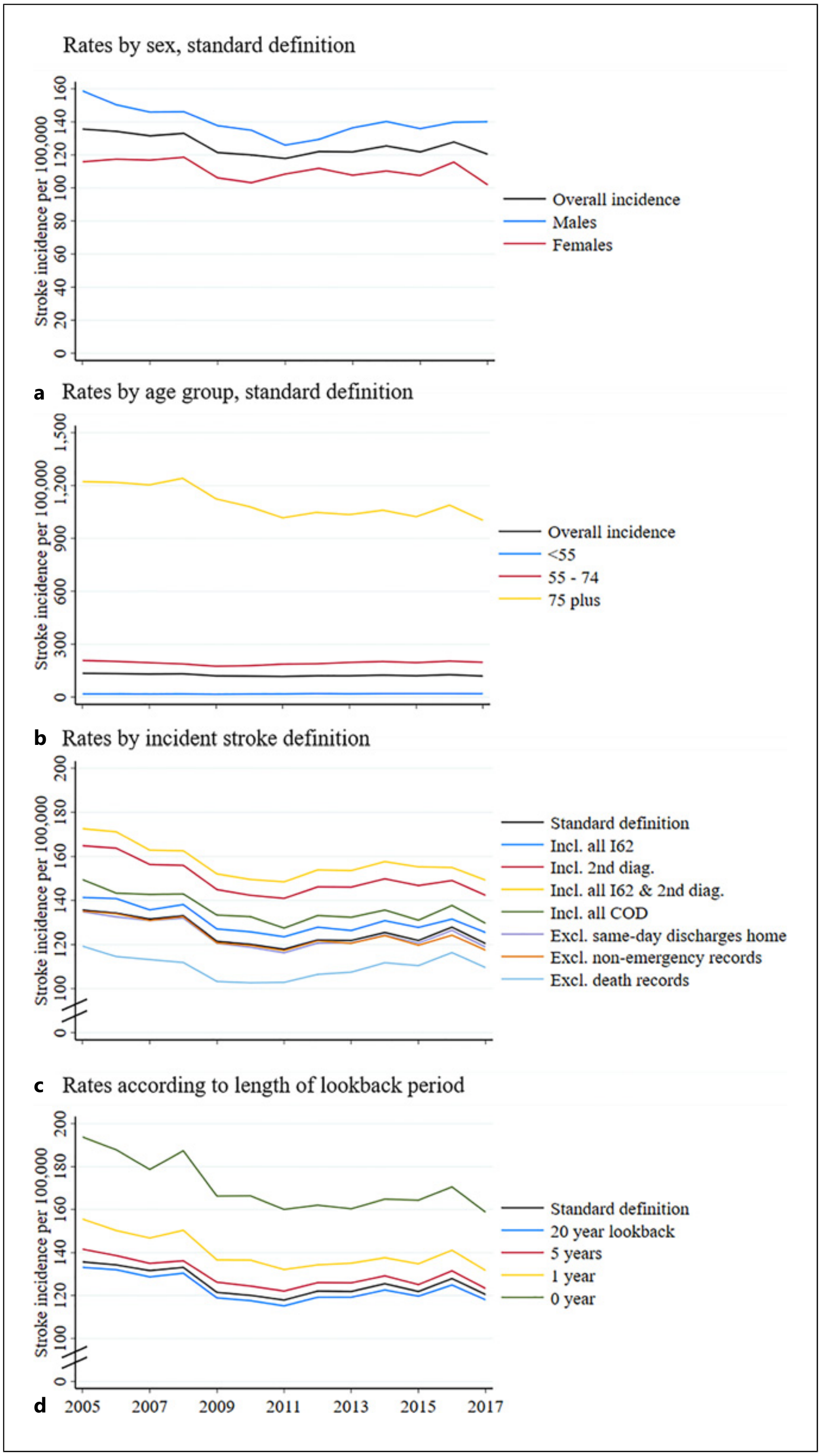

Fig. 1. Stroke incidence per 100,000 person-years at risk, age-standardised to the 2016 Australian population (2005–2017, WA). COD, cause of death; 2nd diag., secondary diagnoses.

ascertainment when using only hospital data, compared to 38.7% with both datasets (online suppl. Table S2).

Age-Standardised Incidence Rates

Age-standardised incidence rates (ASIRs) were greater in males (139.3/100,000) than females (110.6/100,000), and in people aged ≥75 years (1,096.1/100,000) than <55 years (19.1/100,000) (Fig. 1, online suppl. Table S3). Crude rates by age reflected ASIRs, though crude rates by sex differed from ASIRs (online suppl. Table S4). ASIRs declined by 0.61%/year (95% CI: −0.91, −0.32) using the standard definition, although the decline may have attenuated since 2011. This trend was similar between males and females but differed substantially between those aged ≥75 years (−1.66%, 95% CI: −2.06, −1.26) and <55 years (+1.44%/year, 0.64, 2.25).

ASIRs differed according to the definition used (Fig. 1; online suppl. Table S5), although the annual change was similar irrespective of diagnostic codes, number of diagnosis fields, LOS, and admission type. However, excluding mortality data resulted in no significant annual reduction in incidence (−0.1%, 95% CI: −0.4%, 0.2%). The average annual decline in stroke incidence for lookback periods of >5 years were similar to the standard definition, but declines were greater with shorter lookback periods. The proportion of incident strokes with a TIA admission during the lookback period declined from 10.3% in 2005 to 6.3% in 2017 (online suppl. Table S6); this was similar across definitions.

Discussion

These findings lead to several important conclusions. First, the stroke definition applied to administrative data can substantially affect case capture, with the greatest increase in capture when including secondary diagnoses and additional causes of death and the greatest decrease when excluding death records. Second, the longest-available 20-year lookback period led to only marginal improvements in ascertainment over the 10-year lookback through a reduction of prevalence pooling/over-ascertainment. Lookback lengths of <5 years resulted in substantial over-ascertainment. Third, definitions of stroke incidence generally have little effect on incidence trends, except when mortality data are excluded. Finally, stroke incidence appears to be gradually declining in WA overall but increasing in those aged <55 years.

ASIRs for 2005 were below rates in WA previously reported for 2000 [26], consistent with downward trends observed. The magnitude of decline between males and females was similar, although trends differed substantially between age groups. While the annual reduction in stroke incidence was small, this trend appeared to flatten more recently, as observed elsewhere [5]. The relatively small magnitude of the overall trend was driven by trends in fatal strokes, with minimal change in hospitalised incidence. Additionally, we observed a decrease over time in the percentage of patients with incident strokes with prior TIA. The transition from clinical to tissue-based definitions of stroke [14], where many events previously counted as TIA are now considered stroke [15], may contribute to both trends. Despite the overall reduction in incidence, the increase among the youngest group is concerning and reflects trends elsewhere [5].

We add to the limited literature on algorithms developed for stroke incidence by reporting in detail on a wider range of parameters and by assessing subgroup differences and trends. In a linked data study from New South Wales (NSW), the exclusion of cases discharged alive within 48 h (presumed non-acute strokes) led to a 5.7% reduction in capture [8]. We excluded only same-day discharges home, consistent with another recent study [27], which resulted in a smaller reduction in capture than the NSW study but ensured inclusion of patients with ischaemic stroke discharged home the day following stroke [28]. We observed a similar increase in ascertainment from including secondary diagnoses, as reported in a Northern Territory study [19]; some of these likely reflect in-hospital strokes captured as secondary diagnoses. Stroke monitoring in Australia typically uses principal diagnoses only because these are considered more accurate as they are typically acute strokes unrelated to in-hospital complications. While we demonstrate the upper level of possible under-ascertainment when using this approach, trends remain unaffected. The suitability of secondary diagnosis fields will depend on their accuracy. A recent study reported that, in a large linked database of stroke/TIA covering multiple hospitals in Australia, 90.7% of patients with principal stroke/TIA codes were registered as stroke/TIA patients in a linked clinical registry. The addition of secondary diagnoses improved sensitivity but reduced positive predictive value [16]. Inclusion of secondary diagnoses may depend on whether sensitivity or positive predictive value is a greater concern for a given study.

We have uniquely demonstrated that a 10-year lookback results in minimal over-ascertainment of cases relative to 20 years. Previous linked data studies in WA and NSW demonstrate similar minimal overestimation at 10 years, although with a reduced reference of 14 years (WA) and 12 years (NSW) [8, 29]. Our observation that over-ascertainment resulting from the use of short lookback lengths reduced over time may be explained by a decline in recurrence or an increase in time to recurrence following incident stroke [30, 31].

Significant under-ascertainment exists when using only hospitalisation data for incidence, so prior work based on inpatient data only [5, 6, 9, 11, 13] (or inpatient plus ED records, in particular in studies capturing TIA [5, 9, 13]) will provide underestimates of incidence. Furthermore, the use of hospital data only resulted in no decline in incidence over time, suggesting that overall declines in incidence (when including both hospital and mortality data) are driven by declines in fatal stroke and/or decreasing stroke severity. Omission of death data caused a larger reduction in capture among women and older people, possibly due to these groups making up the majority of nursing home residents in Australia [32]. They may be more likely to be discharged from ED or treated at their residence, hence not captured in hospitalisation data (personal communication, A. Thrift, October 17, 2022). Case-fatality from stroke is also greater among older people [33]. The over-ascertainment that occurred with shorter lookback durations was smaller when only hospital data were used compared with including death data; researchers deciding on their study lookback length may need to consider the inclusion or exclusion of death data to minimise over-ascertainment. An additional point to consider is the validity of cause-of-death data. In a previous Australian comparison of stroke codes from death registry data to neurologist classification based on patient medical histories, the sensitivity was 53.1% and the specificity 99.7% [34], suggesting that death registry data may not capture all out-of-hospital stroke deaths, though it does not result in the addition of false positives, i.e., non-stroke deaths.

Strengths and Limitations

Strengths of this study include comprehensive examination of the effect of several stroke definitions and lookback periods, whole-population capture of stroke cases, and our ability to investigate lookback periods up to 20 years in length. While we pragmatically selected the most pertinent definitions, we did not assess the effect of all variations reported in the literature. We do not report results by stroke type, since many strokes have an unspecified type in mortality data; results observed here may differ by subtype. Exclusion of same-day discharges home may have differing effects between ischaemic and haemorrhagic strokes, as the likelihood of a brief admission for an acute event may differ between these types. Patterns described may differ outside of Australia because of differences in factors such as demographics, coding, and administrative practices. We did not assess ED data, so we likely excluded the capture of minor strokes not requiring admission. However, use of ED data may also risk the incorrect capture of non-stroke events, considering the false positive rate of ED stroke codes [35]. This study was conducted within the confines of linked administrative data. Our case capture using the standard definition was 8% less in those aged ≥65 years than for the Perth Community Stroke Study, conducted using an "ideal" prospective stroke study methodology [36], though it was equivalent at younger ages. The discrepancy in those aged ≥65 years may result from under-representation of aged care residents in hospitalisation data (personal communication, A. Thrift, October 17, 2022). Finally, where a non-admitted stroke resulted in death during the post-acute period, the death record will have been treated as the incident event due to the stroke itself being unobserved, though if death data were excluded, these strokes would have been missed entirely.

Given that 35% of West Australians are born overseas [37], people who migrated to WA may have experienced strokes elsewhere before arriving in WA, so these strokes would not be captured. Unfortunately, capture of data on duration of residence in WA is unavailable. However, when allowing for the healthy migrant effect [38], prior strokes before arriving in WA may be less likely in recent migrants.

Conclusion

Methods of stroke capture applied to administrative data affect estimates of incidence to a greater extent than estimates of trends. Our study assists with understanding the effect of different stroke definitions, including their influence on specific population subgroups, and will facilitate comparisons between studies employing differing methods. The definitions selected will ultimately depend on the area of epidemiological enquiry; our findings provide researchers with guidance as to which aspects of stroke definitions are most important in addressing their aims.

Acknowledgments

We acknowledge the WA Data Linkage Branch and the Data Custodians of the WA Department of Health Hospital Morbidity Data Collection and the Death Registrations for access to and provision of data, and the Victorian Department of Justice and Community Safety, the State and Territory Coroners, and the National Coronial Information System for enabling Cause of Death Unit Record File data to be used.

Statement of Ethics

This study was reviewed and approved by the WA Department of Health Human Research Ethics Committee (#2014/55). A waiver of consent was granted due to the size of the population involved, the proportion likely to have moved or died since health information was collected, and the risk of introducing bias into the research.

Conflict of Interest Statement

LN reports consultancy support unrelated to the topic of this paper from CSL Behring. VLF reports being an Executive Committee member of WSO, Honorary Medical Director of Stroke Central New Zealand, and CEO of the New Zealand Stroke Education (charitable) Trust, outside of the submitted work. Authors have received funding unrelated to this paper from the Brain Research New Zealand Centre of Research Excellence, National Health and Medical Research Council, Heart Foundation, and World Stroke Organization.

Funding Sources

The study was funded via a National Health and Medical Research Council Synergy Grant, GNT1182071. The funders had no input into study design, analysis or interpretation of data, the preparation of the current paper, or the decision to publish.

Author Contributions

Study design was performed by D.Y., J.K., R.S.R., A.G.T., and L.N. L.N. arranged data access and managed study data. Analysis was performed by D.Y., J.K., and L.N. D.Y., J.K., R.S.R., N.S.B., A.G.T., V.F., and L.N. provided interpretation of the analysis results. The preparation of the initial manuscript was led by D.Y., J.K., and L.N. Critical review of multiple drafts and approval of final manuscript was given by DY, J.K., R.S.R., N.S.B., J.C., D.Z., A.G.T., V.F., and L.N.

Data Availability Statement

Due to ethical and legal restrictions, patient-level data from this study cannot be shared. However, aggregated data and coding that support the findings of this study are available from the corresponding author on reasonable request.

References

1 Avan A, Digaleh H, Di Napoli M, Stranges S, Behrouz R, Shojaeianbabaei G, et al. Socioeconomic status and stroke incidence, prevalence, mortality, and worldwide burden: an ecological analysis from the Global Burden of Disease Study 2017. BMC Med. 2019;17(1):191.

2 Australian Institute of Health and Welfare. Heart, stroke and vascular disease – Australian facts. Canberra: Australian Government; 2021.

3 Olaiya MT, Sodhi-Berry N, Dalli LL, Bam K, Thrift AG, Katzenellenbogen JM, et al. The allure of big data to improve stroke outcomes: review of current literature. Curr Neurol Neurosci Rep. 2022;22(3):151–60.

4 Epping J, Geyer S, Tetzlaff J. The effects of different lookback periods on the sociodemographic structure of the study population and on the estimation of incidence rates: analyses with German claims data. BMC Med Res Methodol. 2020;20(1):229.

5 Joundi RA, Smith EE, Yu AYX, Rashid M, Fang J, Kapral MK. Temporal and age-specific trends in acute stroke incidence: a 15-year population-based study of administrative data in Ontario, Canada. Can J Neurol Sci. 2021;48(5):685–9.

6 Kim RB, Kim BG, Kim YM, Seo JW, Lim YS, Kim HS, et al. Trends in the incidence of hospitalized acute myocardial infarction and stroke in Korea, 2006–2010. J Korean Med Sci. 2013;28(1):16–24.

7 Lee HJ, Jang SI, Park EC. Effect of adherence to antihypertensive medication on stroke incidence in patients with hypertension: a population-based retrospective cohort study. BMJ Open. 2017;7(6):e014486.

8 Worthington JM, Gattellari M, Goumas C, Jalaludin B. Differentiating incident from recurrent stroke using administrative data: the impact of varying lengths of look-back periods on the risk of misclassification. Neuroepidemiology. 2017;48(3–4):111–8.

9 Sundararajan V, Thrift AG, Phan TG, Choi PM, Clissold B, Srikanth VK. Trends over time in the risk of stroke after an incident transient ischemic attack. Stroke. 2014;45(11):3214–8.

10 Bhopal RS, Bansal N, Fischbacher CM, Brown H, Capewell S, Scottish Health and Ethnic Linkage Study. Ethnic variations in the incidence and mortality of stroke in the Scottish health and ethnicity linkage study of 4.65 million people. Eur J Prev Cardiol. 2012;19(6):1503–8.

11 Clissold BB, Sundararajan V, Cameron P, McNeil J. Stroke incidence in Victoria, Australia – emerging improvements. Front Neurol. 2017;8:180.

12 Katzenellenbogen JM, Vos T, Somerford P, Begg S, Semmens JB, Codde JP. Burden of stroke in Indigenous Western Australians: a study using data linkage. Stroke. 2011;42(6):1515–21.

13 Vyas MV, Silver FL, Austin PC, Yu AYX, Pequeno P, Fang J, et al. Stroke incidence by sex across the lifespan. Stroke. 2021;52(2):447–51.

14 Easton JD, Saver JL, Albers GW, Alberts MJ, Chaturvedi S, Feldmann E, et al. Definition and evaluation of transient ischemic attack: a scientific statement for healthcare professionals from the American heart association/American stroke association stroke Council; Council on cardiovascular surgery and anesthesia; Council on cardiovascular radiology and intervention; Council on cardiovascular nursing; and the interdisciplinary Council on peripheral vascular disease. The American academy of neurology affirms the value of this statement as an educational tool for neurologists. Stroke. 2009;40(6):2276–93.

15 Shahjouei S, Sadighi A, Chaudhary D, Li J, Abedi V, Holland N, et al. A 5-decade analysis of incidence trends of ischemic stroke after transient ischemic attack: a systematic review and meta-analysis. JAMA Neurol. 2021;78(1):77–87.

16 Ryan OF, Riley M, Cadilhac DA, Andrew NE, Breen S, Paice K, et al. Factors associated with stroke coding quality: a comparison of registry and administrative data. J Stroke Cerebrovasc Dis. 2021;30(2):105469.

17 Sposato LA, Kapral MK, Fang J, Gill SS, Hackam DG, Cipriano LE, et al. Declining incidence of stroke and dementia: coincidence or prevention opportunity? JAMA Neurol. 2015;72(12):1529–31.

18 Jamrozik K, Dobson A, Hobbs M, McElduff P, Ring I, D'Este K, et al. Monitoring the incidence of cardiovascular disease in Australia. Australian Institute of Health and Welfare. Canberra: AIHW; 2001.

19 You J, Condon JR, Zhao Y, Guthridge SL. Stroke incidence and case-fatality among Indigenous and non-Indigenous populations in the Northern Territory of Australia, 1999–2011. Int J Stroke. 2015;10(5):716–22.

20 Australian Bureau of Statistics. The Australian statistical geography standard (ASGS) remoteness structure canberra. ABS. [Available from: https://www.abs.gov.au/websitedbs/d3310114.nsf/home/remoteness+structure.

21 Australian Bureau of Statistics. Socio-economic indexes for areas, Canberra. ABS; 2022. Available from: https://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa.

22 StataCorp. Stata. 17.0 ed. College Station, Texas: StataCorp; 2021.

23 Australian Bureau of Statistics. Estimated resident population by single year of age, Western Australia. Canberra; 2022.

24 Australian Bureau of Statistics. Estimated resident population by single year of age, Australia. In: ABS, editor. Canberra; 2022.

25 Katzenellenbogen JM, Sanfilippo FM, Hobbs MS, Briffa TG, Ridout SC, Knuiman MW, et al. Variable effects of prevalence correction of population denominators on differentials in myocardial infarction incidence: a record linkage study in Aboriginal and non-Aboriginal Western Australians. J Clin Epidemiol. 2011;64(6):658–66.

26 Katzenellenbogen JM, Begg S, Somerford P, Anderson CS, Semmens JB, Codde JP, et al. Disability burden due to stroke in Western Australia: new insights from linked data sources. Int J Stroke. 2010;5(4):269–77.

27 Seminog OO, Scarborough P, Wright FL, Rayner M, Goldacre MJ. Determinants of the decline in mortality from acute stroke in England: linked national database study of 795,869 adults. BMJ. 2019;365:l1778.

28 Labán-Seminario LM, Carrillo-Larco RM, Bernabé-Ortiz A. Stroke-related length of hospitalization trends and in-hospital mortality in Peru. PeerJ. 2022;10:e14467.

29 Katzenellenbogen JM, Somerford P, Semmens JB, Codde JP. Effect of clearance periods on hospital stroke incidence using linked administrative data. Int J Stroke. 2010;5(4):336–7.

30 Lewsey J, Jhund PS, Gillies M, Chalmers JW, Redpath A, Briggs A, et al. Temporal trends in hospitalisation for stroke recurrence following incident hospitalisation for stroke in Scotland. BMC Med. 2010;8:23.

31 Bernegger A, Mikšová D, Posekany A, Krebs S, Ferrari J, Greisenegger S, et al. Time trends in stroke severity in the years 2005 to 2020: results from the Austrian Stroke Unit Registry. J Neurol. 2022;269(8):4396–403.

32 Australian Institute of Health and Welfare. Aged care canberra: AIHW; 2021. Available from: https://www.aihw.gov.au/reports/australias-welfare/aged-care.

33 Kashida YT, Lillicrap T, Walker R, Holliday E, Hasnain MG, Tomari S, et al. Transition in incidence rate of hospitalised stroke and case fatality rate in the Hunter region, Australia, 2001-2019: a prospective hospital-based study. J Stroke Cerebrovasc Dis. 2022;31(4):106266.

34 Harriss LR, Ajani AE, Hunt D, Shaw J, Chambers B, Dewey H, et al. Accuracy of national mortality codes in identifying adjudicated cardiovascular deaths. Aust N Z J Public Health. 2011;35(5):466–76.

35 Hall R, Mondor L, Porter J, Fang J, Kapral MK. Accuracy of administrative data for the coding of acute stroke and TIAs. Can J Neurol Sci. 2016;43(6):765–73.

36 Islam MS, Anderson CS, Hankey GJ, Hardie K, Carter K, Broadhurst R, et al. Trends in incidence and outcome of stroke in Perth, Western Australia during 1989 to 2001: the Perth community stroke study. Stroke. 2008;39(3):776–82.

37 Australian Bureau of Statistics. Australia's Population by Country of Birth. Canberra; 2022. Available from: https://www.abs.gov.au/statistics/people/population/australias-population-country-birth/latest-release.

38 Helgesson M, Johansson B, Nordquist T, Vingård E, Svartengren M. Healthy migrant effect in the Swedish context: a register-based, longitudinal cohort study. BMJ Open. 2019;9(3):e026972.