Stroke surveillance in Manitoba, Canada: Estimates from administrative databases

DF Moore, MD, PhD (1); LM Lix, PhD (2); MS Yogendran, MSc (3); P Martens, PhD (3); A Tamayo, MD (4)

Abstract

This study investigated the use of population-based administrative databases for stroke surveillance. First, a meta-analysis was conducted of four studies, identified via a PubMed search, which estimated the sensitivity and specificity of hospital data for ascertaining cases of stroke when clinical registries or medical charts were the gold standard. Subsequently, case-ascertainment algorithms based on hospital, physician and prescription drug records were developed and applied to Manitoba's administrative data, and prevalence estimates were obtained for fiscal years 1995/96 to 2003/04 by age group, sex, region of residence and income quintile. The meta-analysis results revealed some over-ascertainment of stroke cases from hospital data when the algorithm was based on diagnosis codes for any type of cerebrovascular disease (Mantel-Haenszel Odds-Ratio [OR] – 1.70 [95% confidence interval (CI): 1.53 – 1.88]). Analyses of Manitoba administrative data revealed that while the total number of stroke cases varied substantially across the algorithms, the trend in prevalence was stable regardless of the algorithm adopted.

Key words: *administrative data, surveillance, population health, stroke, longitudinal, diagnoses*

Introduction

Death due to stroke is ranked third in Canada and in other developed countries after heart diseases and cancer, while the stroke burden and case fatality rate is estimated to eclipse that of other chronic diseases.[1-3] Despite advances in acute stroke care, prevention of stroke-related risk factors is likely to remain the most effective mechanism to reduce the disease burden.[4] Population-based surveillance allows researchers and policy analysts to describe the disease burden for population groups defined by such characteristics as age, sex, region of residence and income level. Surveillance can also facilitate assessments of the effectiveness of risk prevention strategies over time. Large-scale administrative databases have been used for population-based surveillance of a number of chronic conditions, including stroke.[5, 6] The advantages of administrative data include: (a) ability to generalize prevalence estimates to the whole population rather than just to specific sub-populations, (b) lower costs associated with establishment and maintenance of a surveillance system,[7] (c) ability to monitor trends in prevalence, and (d) opportunity to investigate associated co-morbidities.[7]

Hospital administrative data have often been used to identify stroke cases in the population, but there has been only limited investigation of physician billing claims for this purpose.[8-12] The choice of diagnosis codes to identify stroke cases has been a critical issue in the development of case-ascertainment algorithms using administrative data. An algorithm based on diagnostic codes for all forms of cerebrovascular disease will have improved sensitivity, but could result in possible over-ascertainment of stroke cases when compared with a clinical data source. An algorithm based on a narrower set of diagnostic codes will have improved specificity, but may result in conservative estimates of the number of stroke cases.[13,14] In the medical literature, ischemic stroke is often classified according to the TOAST (Trial of ORG 10172 in Acute Stroke Treatment) criteria.[15-17] This classification better defines the aetiology of ischemic stroke by focusing on clinical treatment strategies such as the use of warfarin for cardio-embolic stroke or anti-platelet therapy for large and small vessel disease.

At present, no studies have extended the methodology of stroke case ascertainment to include data from population-based prescription drug dispensation records,[7] which are now routinely maintained in a number of jurisdictions. It is possible that the combination of three data sources – hospital separations, physician billing claims, and prescription drug records – may improve stroke case ascertainment.[18,19] In fact, in one study of six chronic diseases (hypertension, heart failure, chronic lung disease, arthritis, glaucoma and diabetes), the use of multiple administrative databases to ascertain disease cases resulted in specificity greater than 0.95 and sensitivity greater than 0.90 when compared with an independent validation data source.[20]

In the current study we begin by using meta-analysis techniques to assess the validity of diagnoses in administrative data for stroke case ascertainment. We then apply the meta-analysis results to develop case-ascertainment algorithms based on

Author References

1 Walter Reed Army Medical Center, Washington, DC
2 School of Public Health, University of Saskatchewan, Saskatoon, SK
3 Manitoba Centre for Health Policy, University of Manitoba, Winnipeg, MB
4 Brandon Regional Health Authority, Brandon, MB
Correspondence: Lisa M. Lix, School of Public Health, University of Saskatchewan, Health Sciences Building, 107 Wiggins Road, Saskatoon SK S7N 5E5, Tel.: 306-966-1617, Email: lisa.lix@usask.ca

diagnoses in hospital and physician data as well as records of dispensations for prescription drugs used in the treatment of stroke. The algorithms are applied to administrative data from Manitoba, Canada, to estimate the prevalence of stroke by age, sex, region of residence and income group over time.

Methods

Meta-analysis of stroke case ascertainment

A comprehensive PubMed search on the terms administrative database **AND** stroke **OR** cerebrovascular disease identified a total of 28 references for the period 1965-2005. After full review of these articles by the first author, four were selected for investigation using meta-analytic techniques. All of the selected studies used a "gold standard", that is, an independently maintained stroke registry or a prospective or retrospective chart review to validate administrative data for stroke case ascertainment. The studies selected for the meta-analysis relied on hospital data in which diagnoses were coded using the International Classification of Diseases, 9th revision (ICD-9). Studies excluded from the meta-analysis did not validate the administrative data, did not use ICD-9 codes to ascertain disease cases, and/or did not disclose the information necessary to construct a 2 x 2 classification table composed of the number of stroke cases and non-cases in the administrative and validation datasets.

The studies selected for the meta-analysis identified acute stroke cases using a "sensitive" algorithm based on all diagnosis codes for cerebrovascular disease (i.e. ICD-9-CM 430 to 438) and/or a "specific" algorithm based on a subset of ICD-9-CM codes most likely to identify only acute stroke cases in administrative data.

Odds ratios (ORs) were calculated for each of the studies included in the meta-analysis for the sensitive and specific algorithms. Pooled odds ratios were calculated for the sensitive and specific algorithms using the Mantel-Haenszel method.[21,22] The pooled ORs were based on three datasets for the sensitive algorithm and four datasets for the specific algorithm. An OR = 1.0 indicates that the probability of the event (i.e. stroke case ascertainment) is equally likely in both the administrative data and validation data source. An OR > 1.0 indicates an overestimate of stroke ascertainment by the administrative data compared to the validation data source while an OR < 1.0 indicates that stroke case ascertainment is lower in administrative data than in the validation data source. The meta-analysis was conducted using SAS software.[23]

Stroke case ascertainment in Manitoba's administrative data

The Research Data Repository housed at the Manitoba Centre for Health Policy (MCHP) was used to estimate stroke prevalence for sensitive and specific case-ascertainment algorithms. The Repository has been used in many studies of population health and health services use.[24,25]

MCHP maintains comprehensive population-based administrative data, including hospital separations, physician billing claims, and out-patient prescription drug dispensation records, for all health insurance registrants.[5] The Manitoba population is approximately 1.2 million according to Statistics Canada Census figures.[26] Nonparticipation in the provincial health insurance program is minimal since no premium payment requirement exists.[5] Administrative data files in the Repository can be linked over time via a unique anonymized personal health identification number (PHIN). Demographic information for health insurance registrants, including age, sex and geographic location of residence is available in the Repository by linking to the population registry. As well, income groups have been derived by linking the Repository data to data for dissemination areas. These are the smallest geographic areas for which Statistics Canada Census data are provided.

A hospital separation abstract is completed at the point of discharge from an acute care facility; each abstract contains up to 16 diagnosis codes. Physician billing claims contain a single diagnosis code. A small number of physicians in Manitoba are salaried; however, most of them submit "shadow billing claims" for billing purposes. It has been estimated that shadow billing results in at least 80% capture of services (Katz A., personal communication, February 2007). Diagnoses in hospital and physician data are recorded using ICD-9-CM codes up to fiscal year 2003/04, but commencing in fiscal year 2004/05, ICD-10-CA coding was introduced in hospital separations.

The Drug Programs Information Network (DPIN) is an on-line point-of-sale prescription drug database linking all retail pharmacies in Manitoba. DPIN captures prescription drug dispensations for all Manitoban residents regardless of coverage mechanism. Prescription drugs are identified via drug identification numbers (DINs) which are linked to the Drug Product Database maintained by Health Canada. Anatomic Therapeutic Chemical (ATC) codes[27] are added to allow categorization of drugs into appropriate therapeutic and pharmacological subgroups.

All 16 diagnosis fields in a hospital separation abstract were searched to identify stroke cases. Table 1 lists ICD-9-CM codes for cerebrovascular disease, the stroke type to which each code corresponds, and its relationship to the TOAST criteria. Using TOAST, ischemic stroke is categorized as stroke related to (a) large artery atherosclerosis (including large artery thrombosis and artery-to-artery embolism), (b) cardio-embolism, (c) small artery occlusion, (d) stroke of other determined cause, and/or (e) stroke of undetermined cause. The TOAST cardio-embolic stroke category cannot be identified from ICD-9-CM codes.

After reviewing the literature and consulting with clinical experts, the authors selected the following drug categories for identification of stroke cases from the DPIN data: (a) *anti-platelet agents* such as aspirin (ASA) at 81 or 325 mg once a day, clopidogrel, ticlopidine, dipyridamole and combination agents such as Aggrenox (ASA 25 mg dipyridamole 200 mg slow release) and (b) *oral anti-coagulants* such as warfarin, phenindione, and nicoumalone. The ATC codes (fifth level) were B01AA02, B01AA03, B01AA07, B01AC07, B01AB01, B01AC30, B01AC05, B01AC06, B01AC04, B01AB09, B01AB04, B01AB10. Thrombolytic agents such as rt-PA (recombinant tissue plasminogen activator) and intravenous anti-platelet agents (anti GP 2b/3a) such as

abciximab, tirofiban and eptifibatide cannot be identified in DPIN data.

Stroke cases were identified by the following rules: at least one hospital separation in one fiscal year (i.e. 1 + H), or at least two ICD-9-CM physician billing claims in one fiscal year (i.e. 2 + P), or at least one physician billing claim in one fiscal year together with at least two prescription drug records in one fiscal year (i.e. 1 + P and 2 + Rx). Case counts were derived for each of the fiscal years 1995/96 to 2003/04. The fiscal year extends from April 1 to March 31 of the following year. This time period was chosen because in 1994/95 the DPIN system originated, while in 2004/05 ICD-10-CA coding was introduced. Consistency was therefore maintained by the yearly application of a case-ascertainment algorithm based on ICD-9-CM codes.

Frequencies of stroke cases were compiled for both sensitive and specific algorithms by study year, age group (19 to 44, 45 to 54, 55 to 64, 65 to 74, 75 to 84, 85 years and older), sex, region of residence (Northern Regional Health Authorities, Southern Regional Health Authorities and Winnipeg Regional Health Authority, with the Regional Health Authority representing the health administrative unit of the province) and income quintile (Q1 to Q5, with Q1 representing the lowest income group). The geographic areas broadly correspond to sparsely populated rural, rural, and urban communities. Individuals were assigned to income quintiles using average household income data for dissemination areas, and then ranked according to these areas. The quintiles are defined so that approximately 20% of the total population is assigned to each group.[24] Income quintiles are defined separately for urban and rural areas. Prevalence estimates were calculated using data from the provincial registry to compute the denominator of the estimate.

Regression analyses were conducted to test for differences in the relative rate (RR) of stroke for different population sub-groups and over time. The data for each study year were first analyzed using generalized linear models[28] to relate stroke counts to the main effects of age group, sex, region of residence, income quintile and algorithm (i.e. sensitive, specific), as well as selected two-way interactions among these variables. To ensure a parsimonious model, only those interactions that resulted in a significant improvement in model fit, as evaluated with a likelihood ratio test, were retained. An offset, the log of the total population, was included in all models. The data were initially parameterized using Poisson, negative binomial and gamma distributions. Goodness-of-fit statistics were compared, and the distribution resulting in the best fitting model was selected. The longitudinal data were analyzed using a generalized linear model with generalized estimating equations (GEEs) to account for correlation among the stroke case counts over time.[29] The main effects of age group, sex, region of residence, income quintile, algorithm and year/time were included in the model. Selected two-way interactions were included, but these model effects were only retained if they resulted in a significant improvement in model fit. Again, the data were initially parameterized using Poisson, negative binomial and gamma distributions, and goodness-of-fit statistics were compared. The correlation structure was chosen to be exchangeable after examination of the sample correlation matrix. The regression analyses were conducted using SAS software.[23]

TABLE 1
ICD-9-CM codes for ascertaining cases of stroke in administrative data, relationship to TOAST criteria and frequency in Manitoba hospital separations, 1995/96 to 2003/04

ICD-9-CM Code	Stroke Type	TOAST Criterion	Freq[b]	%
430	Subarachnoid hemorrhage	N/A[a]	901	1.7
431	Cerebral hemorrhage	N/A	2038	4.0
432	Other and unspecified intracranial hemorrhage	N/A	951	1.8
433	Occlusion and stenosis of pre-cerebral arteries	Large vessel disease	5957	11.9
434	Occlusion of cerebral arteries	Large and small vessel disease	5968	11.9
435	Transient cerebral ischemia	N/A	8189	16.3
436	Acute but ill-defined cerebrovascular disease	Stroke of other determined cause	12 061	24.0
437	Other and ill-defined cerebrovascular disease	Stroke of undetermined cause	2667	6.5
438	Late effects of cerebrovascular disease	N/A	12 266	22.4
All			50 098	100.0

[a] N/A = not applicable

[b] Frequencies are the number of individuals (19 years and older) with at least hospital separation having the identified ICD-9-CM code

Results

Meta-analysis of stroke case ascertainment

Table 2 reports the study-specific ORs for stroke case ascertainment from hospital administrative data. The specific algorithm based on the restricted set of ICD-9-CM codes, for which there were three datasets available for analysis, resulted in ORs which were smaller in magnitude and closer to 1.0 than the ORs for the sensitive algorithm based on all diagnosis codes for cerebrovascular disease, for which there were four datasets available for analysis. As Table 2 reveals, the pooled OR for the datasets that used a sensitive algorithm was 1.70 (95% CI: 1.53, 1.88) while the pooled OR for the datasets that used a specific algorithm was 1.02 (95% CI: 0.93, 1.13), indicating some over-ascertainment for the former but not for the latter.

TABLE 2
Odds ratios (ORs) for the meta-analysis of stroke case ascertainment in hospital data for sensitive and specific sets of diagnosis codes

Study	Sensitive Algorithm OR (95% CI)[a]	Specific Algorithm OR (95% CI)[b]
Ellekjaer et al. [32]	1.76 (1.51, 2.05)	1.17 (1.00, 1.38)
Leibson et al. [18]	1.47 (1.17, 1.86)	1.20 (0.95, 1.53)
Reker et al.[33]	1.76 (1.46, 2.10)	0.72 (0.58, 0.89)
Tirschwell et al. [13]	--	1.02 (0.80, 1.31)
Pooled OR	1.70 (1.53, 1.88)	1.02 (0.93, 1.13)

[a] The sensitive algorithm is based on ICD-9-CM codes 430 to 438 for all studies except for Reker et al. 33 who excluded 437 and 438

[b] The specific algorithm is based on ICD-9-CM codes 430, 431, 434 and 436 for all studies except Reker et al. 33 (excluded 436), Leibson et al. 18 (included 437), and Tirschwell et al. 13 (included 435)

Stroke case ascertainment in Manitoba's administrative data

Table 1 shows the frequency of stroke cases identified from hospital data for each of ICD-9-CM codes 430 to 438 for the period 1995/96 to 2003/04. More than half (54.7%) of cases had non-specific diagnostic codes of 432, 436, 437 and 438.

The frequency of stroke cases from hospital, physician, and pharmacy administrative data is reported next. The number of cases satisfying the 1 + H rule was identified first, followed by the number of cases identified with the 2 + P rule, and then the number of additional cases identified with the 1 + P and 2 + Rx rule. The results are reported separately for the sensitive algorithm based on all diagnoses for cerebrovascular disease in hospital and physician administrative data and the specific algorithm based on a subset of diagnoses most likely to identify acute stroke cases. For the latter, we initially reported the results for two different subsets of ICD-9-CM codes: one set included transient ischemic attacks while the other did not. Only the results for the first specific algorithm are included in subsequent regression analyses.

As Table 3 reveals for the sensitive algorithm, 49.9% of stroke cases were identified from hospital data at the beginning of the study period (i.e. 1995/96); this percentage dropped substantially to 38.5% by the end of the study period. However, the percentage of stroke cases identified solely from physician data remained relatively constant over time. The percentage of cases identified from a combination of physician data and prescription drug data increased over time, from 5.6% in 1995/96 to 15.8% in 2003/04. The same trend of decreasing numbers of stroke cases identified from hospital data and increasing numbers identified from physician and prescription drug data was also observed when the more specific algorithm was adopted.

The crude provincial prevalence estimates (Table 3) are relatively unchanged across time regardless of the algorithm used. However, the rate based on the smallest set of ICD-9-CM codes is approximately half the value of the rate derived using the full set of diagnostic codes for cerebrovascular disease.

Next, the number of stroke cases in each year was analyzed using generalized linear models. The negative binomial distribution provided a better fit to the data than either the Poisson or gamma distributions as judged by the ratio of the residual deviance to the model degrees of freedom. The likelihood ratio test showed that models containing two-way interactions were not a significantly better fit to the data than a simpler model containing main effects only ($p > .05$), thus the latter was retained. The model results for fiscal year 1998/99 are reported in Table 4; similar results were observed for all other years of data and are therefore not reported here. The relative rate (RR) of stroke was significantly lower in both southern rural and urban regions than in northern Manitoba, and significantly higher in older groups. An income gradient was observed, such that the RR of stroke was lower in higher income quintiles. The rate was higher in males than in females, and the rate for the specific algorithm was significantly lower than for the sensitive algorithm.

The longitudinal prevalence data were also modeled. A negative binomial distribution was again selected because it resulted in a better fit to these data than either the gamma or Poisson distributions. The inclusion of year x region and year x age group interaction terms resulted in a significant improvement in model fit; other two-way interaction terms were not retained in the model because they did not significantly improve model fit. For the year x region interaction, the analyses revealed that after adjusting for other model effects, the RR of change in stroke prevalence was greater for southern rural (RR = 1.01, 95% CI = 1.00, 1.02) and for urban (RR = 1.02, 95% CI = 1.01, 1.04) regions than for the northern regions of Manitoba. For the latter, the analyses showed that compared to the 19 to 44 years age group, the RR of change in stroke prevalence was lower for 55 to 64 years (RR = 0.96; 95% CI = 0.95, 0.99), 65 to 74 years (RR = 0.96; 95% CI = 0.94, 0.97), 75 to 84 years (RR = 0.95; 95% CI = 0.94, 0.97), and 85+ years (RR = 0.96; 95% CI = 0.94, 0.98). Figures 1 and 2 illustrate the nature of the trends in prevalence estimates for age groups and regions. Main effects of sex, income quintile and algorithm were also significant ($p < 0.05$), and the RR estimates were similar to those reported in Table 4.

Discussion

This study builds on previous research that has explored the potential role of administrative data for stroke surveillance. A number of case-ascertainment algorithms were applied to administrative databases available in Canadian provinces and territories. The stroke prevalence estimates were indirectly validated via a meta-analysis of previous studies that compared hospital data to medical chart or registry data. The results of the meta-analysis revealed that the odds of agreement between administrative data and chart or registry data were better when a specific set of diagnostic codes was used instead of a sensitive set of codes.

TABLE 3
Frequency (%) of stroke cases by data source and ICD-9-CM codes and crude prevalence of stroke, 1995/96 – 2003/04

	Hospital[a] (1 + H)		Physician (2 + P)		Physician + Drug (1 + P & 2 + Rx)		Total	
Fiscal Year	Freq	%	Freq	%	Freq	%	Freq	Prev (%)
Sensitive Algorithm (ICD-9-CM 430 to 438)								
1995/96	4882	49.9	4349	44.5	551	5.6	9782	1.16
1996/97	5053	51.5	4203	42.9	547	5.6	9803	1.16
1997/98	4790	48.7	4339	44.1	701	7.1	9830	1.16
1998/99	4777	48.6	4357	44.3	702	7.1	9836	1.16
1999/00	4488	45.8	4398	44.9	920	9.4	9806	1.16
2000/01	4585	44.3	4587	44.3	1176	11.4	10 348	1.21
2001/02	4276	41.5	4557	44.3	1462	14.2	10 295	1.20
2002/03	3948	39.9	4447	45.0	1488	15.1	9883	1.14
2003/04	3993	38.5	4746	45.7	1635	15.8	10 374	1.19
Specific Algorithm #1 (ICD-9-CM 430, 431, 434, 435, 436)								
1995/96	3283	43.8	3704	49.4	517	6.9	7504	0.89
1996/97	3239	43.8	3573	48.3	584	7.9	7396	0.88
1997/98	3166	42.3	3656	48.9	662	8.8	7484	0.89
1998/99	3234	43.4	3554	47.7	656	8.8	7444	0.88
1999/00	2956	39.9	3603	48.6	855	11.5	7414	0.87
2000/01	2955	37.5	3836	48.7	1084	13.8	7875	0.92
2001/02	2831	36.1	3726	47.5	1281	16.3	7838	0.91
2002/03	2666	35.6	3546	47.4	1275	17.0	7487	0.87
2003/04	2705	33.8	3869	48.4	1422	17.8	7996	0.92
Specific Algorithm #2 (ICD-9-CM 430, 431, 434, 436)								
1995/96	2431	45.1	2518	46.7	439	8.1	5388	0.64
1996/97	2406	45.7	2433	46.2	426	8.1	5265	0.62
1997/98	2281	43.2	2525	47.9	468	8.9	5274	0.62
1998/99	2328	44.1	2478	46.9	475	9.0	5281	0.62
1999/00	2091	40.1	2507	48.1	614	11.8	5212	0.61
2000/01	2197	39.5	2599	46.8	761	13.7	5557	0.65
2001/02	2113	37.7	2601	46.3	898	16.0	5612	0.65
2002/03	2035	37.8	2468	45.9	875	16.3	5378	0.62
2003/04	2023	35.3	2683	46.8	1024	17.9	5730	0.66

[a] Hospital separations (1 + H) have precedence over physician billing claims (2 + P), which in turn have precedence over combined physician and prescription drug data (1 + P and 2 + Rx) for identification of stroke cases

Examination of the distribution of stroke cases across the diagnosis codes can yield insights into health-care resource requirements. In Manitoba we found that in hospital data, almost 1000 stroke cases per year (i.e. about 16%) are identified as transient ischemic attacks with, by definition, no extended functional neurological deficits. Patients with such events are good candidates for secondary prevention through the correct assessment of cerebrovascular risk factors together with subsequent treatment involving surgery or medical therapy. The number of subarachnoid and cerebral hemorrhages was seen to represent approximately 10% of all stroke cases in Manitoba compared to the usually quoted 20% rate for hemorrhagic stroke in most population studies. While it is likely that coding for hemorrhagic stroke is reliable due to the nature of the clinical encounter, such discrepancies in ascertainment can only really be resolved by validation studies involving direct chart review or by development of an inclusive stroke registry.

The use of physician billing claims and prescription drug records in addition to hospital separations increased the total number of stroke cases identified with Manitoba's administrative data. In fact, the number of identified cases nearly doubled merely by inclusion of the physician billing data. The use of prescription drug records in combination with physician billing claims resulted in identification of a small but increasing number of cases over time.

TABLE 4
Regression analyses of the relative rate (RR) of stroke in Manitoba, 1998/99

Model Effect	Estimate (se)[a,b]	RR (95% CI)[c]
South Rural	-0.27 (0.05)	0.77 (0.70, 0.84)
Winnipeg	-0.25 (0.05)	0.78 (0.71, 0.85)
North Rural	Ref	–
85 years and older	4.54 (0.05)	93.41 (84.92, 102.76)
75-84 years	4.24 (0.05)	69.61 (63.67, 76.10)
65-74 years	3.55 (0.05)	34.97 (31.95, 38.28)
55-64 years	2.72 (0.05)	15.20 (13.83, 16.72)
45-54 years	1.70 (0.05)	4.47 (4.93, 6.06)
19-44 years	Ref	–
Quintile 5	-0.30 (0.03)	0.74 (0.69, 0.79)
Quintile 4	-0.20 (0.03)	0.82 (0.49, 0.62)
Quintile 3	-0.16 (0.03)	0.86 (0.81, 0.91)
Quintile 2	-0.13 (0.03)	0.88 (0.83, 0.93)
Quintile 1	Ref	–
Male	0.23 (0.02)	1.26 (1.21, 1.31)
Female	Ref	–
Specific[d] algorithm	-0.27 (0.02)	0.77 (0.74, 0.80)
Sensitive algorithm	Ref	–

[a] se = standard error

[b] Parameter estimates were obtained using a generalized linear model with a negative binomial distribution

[c] CI = confidence interval

[d] The specific algorithm was based on the following ICD-9-CM codes: 430, 431, 434, 435 and 436

The total number of stroke cases identified with Manitoba's administrative data changed very little over time, although the total number of cases identified with hospital data decreased. This has important implications for future stroke surveillance studies. Use of a single administrative data source could give a misleading picture of changes in stroke prevalence over time.

Analyses of the annual stroke data revealed significant variations across both income groups and geographic areas of Manitoba, even after adjusting for differences in age and sex. The trend analyses showed that the prevalence of stroke decreased slightly across older age groups relative to the youngest age group, but increased slightly across urban and southern rural regions relative to the northern region. Our study results are largely consistent with other epidemiological studies, such as the Framingham study, which shows a greater prevalence of stroke in males compared to females and a stroke burden that is largely unchanged over time.[4] These results have important implications for stroke therapy, post stroke dependency, rehabilitation and the development of targeted stroke prevention programs. For example, thrombolytic therapy in the older age group results in a higher mortality, although this age group has higher mortality than younger age groups even without rt-PA.[30,31]

This study has some limitations. We have estimated the annual prevalence of stroke, but stroke incidence is not investigated. However, the stroke incidence rate can be estimated using the slope of the prevalence trend. The stability of the prevalence estimates over time suggests that incidence remains relatively constant in Manitoba. A complete picture of incidence would, however, use provincial vital statistics data in addition to hospital, physician, and prescription drug data. Like all epidemiological studies, investigations of stroke case ascertainment from administrative databases represent a snapshot of events with an implied estimation error. While estimation error was not quantified in the current study, it was approximated by reporting results for both sensitive and specific algorithms, which likely represent the upper and lower bounds of stroke prevalence.

The data capture all residents of Manitoba having health registration coverage from Manitoba Health in any one year. No attempt was made to eliminate residents who were only partially covered during the fiscal year due to in-province or out-of-province migration, and therefore may have had a lower probability of meeting the criteria required for identification of stroke cases. At the same time, the inclusion of residents with only partial coverage during the fiscal year because of death means that the data represent a better estimate of all stroke cases in Manitoba, not just stroke survivors.

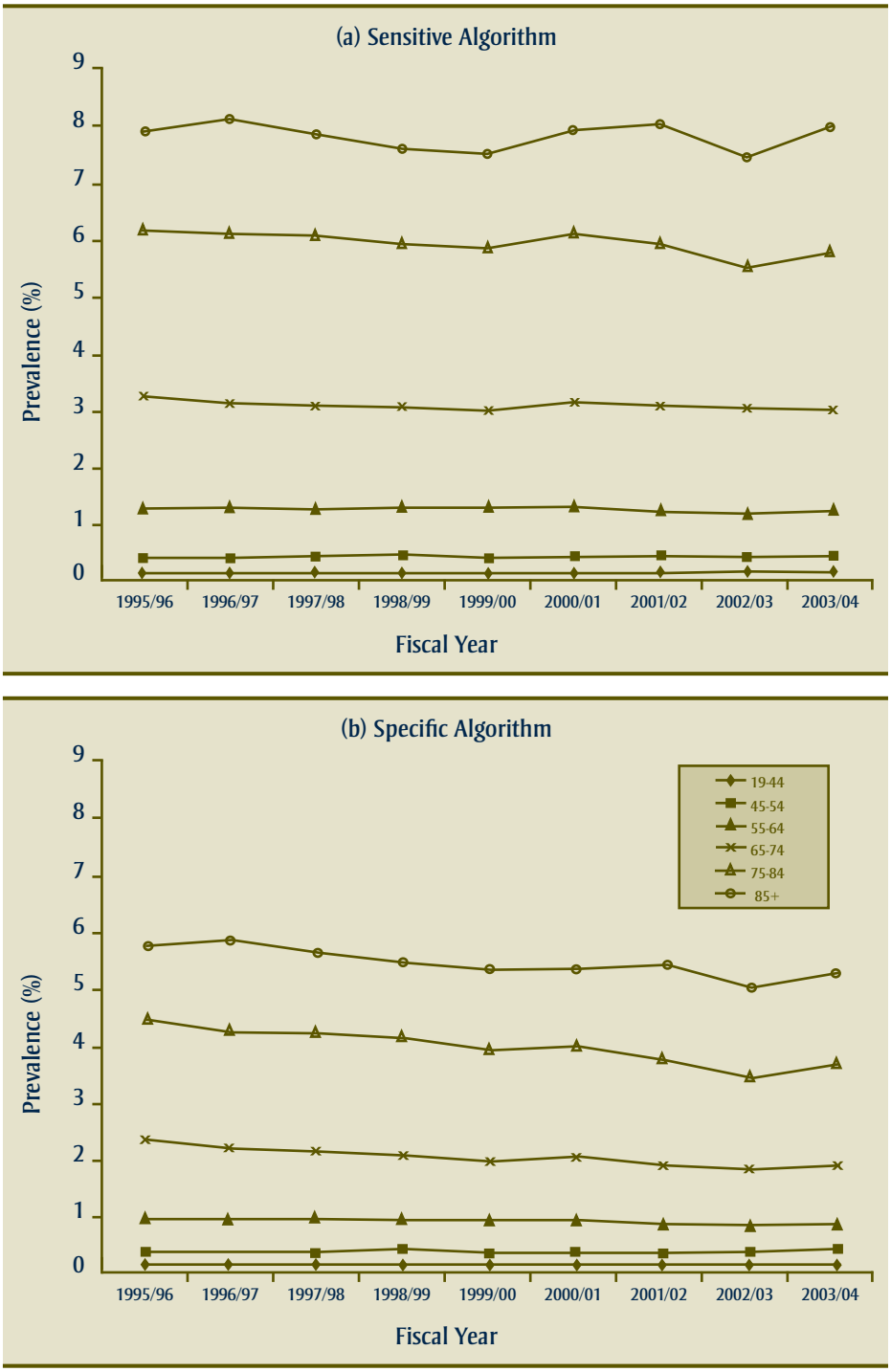
In a recent review of the transition from ICD-9 to ICD-10 coding, no significant difference was found between the two classification systems for stroke case ascertainment or risk factors.[6] However, the effect of the change in coding on stroke case ascertainment in hospital separations warrants further investigation using Manitoba data.

In conclusion, administrative data can be used for population-based surveillance of a variety of chronic conditions, including stroke. Administrative data can be used to describe socio-demographic variations in the population-prevalence of stroke and to conduct retrospective studies of change over time. These data represent a cost-effective tool for providing information about the burden of stroke on the population and for informing health policy decisions.

Acknowledgements

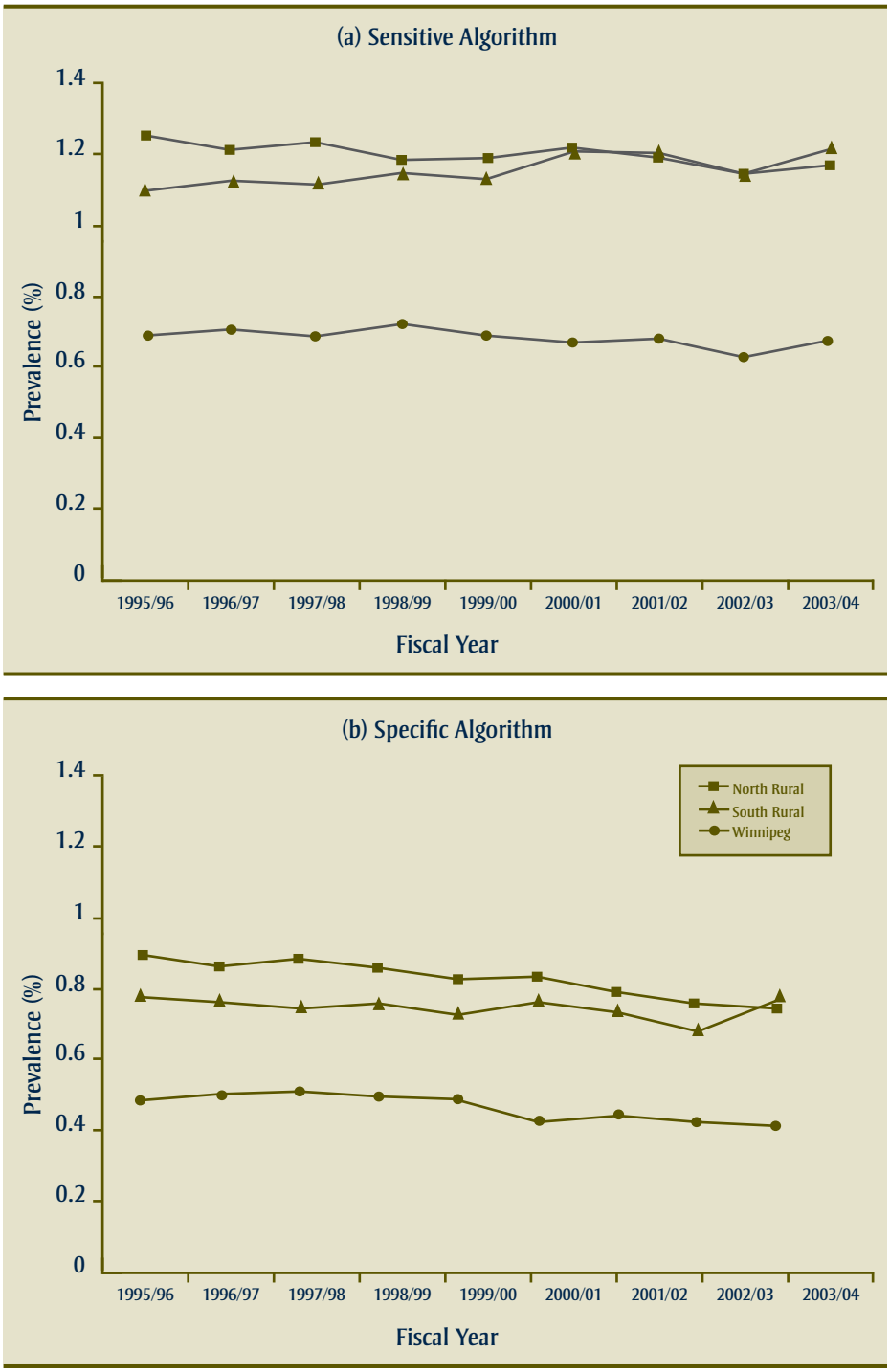
The authors are indebted to Manitoba Health & Healthy Living for the provision of data under project #2004/05-01. This research was supported, in part, by a CIHR New Investigator Award to the second author. The results and conclusions are those of the authors, and no official endorsement by Manitoba Health & Healthy Living is intended or should be inferred. The authors have no competing interests to declare.

FIGURE 1
Trends in crude prevalence of stroke by age group for (a) the sensitive algorithm and (b) the specific algorithm,[a] 1995/96 to 2003/04



[a] The specific algorithm was based on the following ICD-9-CM codes: 430, 431, 434, 435 and 436

FIGURE 2
Trends in crude prevalence of stroke by region of residence for (a) the sensitive algorithm and (b) the specific algorithm,[a] 1995/96 to 2003/04



[a] The specific algorithm was based on the following ICD-9-CM codes: 430, 431, 434, 435 and 436

References

1. Statistics Canada. Selected Leading Causes of Death, By Sex. Ottawa: Statistics Canada, 1997.
2. Verbrugge L, Lepkowski JM, Imanaka Y. Comorbidity and its impact on disability. Milbank Q. 1989;67:450-484.
3. Dobkin B. The Clinical Science of Neurologic Rehabilitation. Second ed. New York: Oxford, 2003:375-376.
4. Wolf P. Epidemiology of stroke. In: Mohr J, Choi DW, Grotta JC, Weir B, Wolf PA, eds. Stroke Pathophysiology, Diagnosis, and Management. Philadelphia: Churchill Livingstone, 2004:13-34.
5. Roos L, Walld R, Uhanova J, Bond R. Physician visits, hospitalizations, and socioeconomic status: Ambulatory care sensitive conditions in a Canadian setting. Health Research and Educational Trust. 2005;10:1167-1185.
6. Kokotailo R, Hill MD. Coding of stroke and stroke risk factors using International Classification of Disease, Revision 9 and 10. Stroke. 2005;36:1776-1781.
7. Lix L, Yogendran M, Burchill C, et al. Defining and Validating Chronic Disease: An Administrative Data Approach. Winnipeg: Manitoba Center for Health Policy, 2006.
8. Yiannakoulias N, Svenson LW, Hill MD, et al. Regional comparison of inpatient and outpatient patterns of cerebrovascular disease diagnosis in the province of Alberta. Chronic Diseases in Canada 2003;24:9-16.
9. Ostbye T, Levy AR, Mayo NE. Hospitalization and case fatality rates for subarachnoid hemorrhage in Canada from 1982 through 1991. Stroke. 1997;28:793-798.
10. Mayo N, Chockalingam A, Reeder BA, et al. Surveillance for stroke in Canada. Health Reports. 1994;6:62-72.
11. Lappala J, Virtamo J, Heinonen OP. Validation of stroke diagnosis in the National Hospital Discharge Register and the Register of Causes of Death in Finland. Euro J Epi. 1999;15:155-160.

12. Benesch C, Witter DM, Wilder MA, et al. Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. Neurology. 1997;49:660-664.

13. Tirschwell D, Longstreth WT. Validating administrative data in stroke research. Stroke. 2002;33:2465-2470.

14. Rhys Willams G, Jiang JG, Matchar DB, et al. Incidence and occurrence of total (first-ever and recurrent) stroke. Stroke. 1999;30:2423-2528.

15. The Publication Committee for the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) Investigators. Low molecular weight heparinoid, ORG 1072 (danaparoid) and outcome after acute stroke. JAMA. 1998;279:1265.

16. Adams H, Bendixen BH, Kapele LJ, et al. Classification of subtype of acute ischemic stroke: Definitions for use in a multicenter clinical trial. Stroke. 1993;24:35-41.

17. Kolominisky-Rabas P, Weber M, Gefeller O, et al. Epidemiology of ischemic stroke subtypes according to TOAST criteria, Incidence, recurrence, and long-term survival in ischemic stroke subtypes: A population-based study. Stroke. 2001;32:2735-2740.

18. Stroup D, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology – A proposal for reporting. JAMA. 2000;283:2008-2012.

19. Friedman L, Furberg CD, DeMets DL. Fundamentals of Clinical Trials. 3rd ed. New York: Springer, 1998:313-317.

20. SAS Institute Inc. SAS/STAT User's Guide, version 9.1. Cary, NC: SAS Institute Inc., 2004.

21. Roos N, Mustard CA. Variation in health and health care use by socioeconomic status in Winnipeg, Canada: Does the system work well? Yes and No. Milbank Q. 1997;75:89-111.

22. Robinson R, Young KT, Roos L, et al. Estimating the burden of disease: Comparing administrative data and self-reports. Med Care. 1997;35:932-947.

23. Statistics Canada. Provincial and Territorial Profiles, Manitoba. Ottawa, ON: Statistics Canada, 2001.

24. World Health Organization. WHO Collaborating Centre for Drug Statistics Methodology: ATC Classification Index with DDDs and Guidelines for ATC Classification and DDD Assignment. Oslo, Norway: Norwegian Institute of Public Health, 2006.

25. McCulloch CE, Searle SR. Generalized, Linear, and Mixed Models. New York: Wiley, 2001.

26. Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. Hoboken, NJ: Wiley, 2004.

27. Engelter S, Reichart M, Sekoranja L, et al. Thrombolysis in stroke patients aged 80 years and older: Swiss survey of IV thrombolysis. Neurology. 2005;65:1795-1798.

28. Hemphill J, Lyden P. Stroke thrombolysis in the elderly: Risk or benefit. Neurology. 2005;65:1690-1691.

29. Ellekjaer H, Holmein J, Oystein K, et al. Identification of incident stroke in Norway hospital discharge data compared with a population-based stroke register. Stroke. 1999;30:56-60.

30. Reker DM, Hamilton BB, Duncan PW, et al. Stroke: Who's counting what? J Rehabil Res Devel. 2001;38:281-289.

31. Leibson C, Naessens JM, Brown RD, et al. Accuracy of hospital discharge abstracts for identifying stroke. Stroke. 1994;25:2348-2355.

32. Madans J, Reubens C, Rothwell S, et al. Differences in morbidity measures and risk factor identification using multiple data sources: The case of stroke. J Epidemiol Biostat. 1999;4:37-43.

33. Rector T, Wickstrom SL, Shah M, et al. Specificity and sensitivity of claims-based algorithms for identifying members of Medicare plus Choice health plans that have chronic medical conditions. Health Serv Resear. 2004;39:1839-1861.