



## UWA Research Publication

Jamrozik, Z, Hyde, Z, Alfonso, H, Flicker, L, Almeida, O, Yeap, B, Norman, P, Hankey, G & Jamrozik, K 2014, 'Validity of Self-Reported versus Hospital-Coded Diagnosis of Stroke: A Cross-Sectional and Longitudinal Study' *Cerebrovascular Diseases*, vol 37, no. 4, pp. 256-262

© 2014 S. Karger AG, Basel

---

This is pre-copy-editing, author-produced version of an article accepted for publication in *Cerebrovascular Diseases* following peer review. The definitive published version (see citation above) is located on the journal home page of the publisher, Karger.

This version was made available in the UWA Research Repository on the 6<sup>th</sup> of May 2015, in compliance with the publisher's policies on archiving in institutional repositories.

Use of the article is subject to copyright law.

## **The validity of self-reported versus hospital-coded diagnosis of stroke – a cross sectional and longitudinal study**

Euzebiusz Jamrozik<sup>1,2</sup>, Zoë Hyde<sup>1,3</sup>, Helman Alfonso<sup>4</sup>, Leon Flicker<sup>1,3</sup>, Osvaldo P. Almeida<sup>4</sup>, Bu B. Yeap<sup>3</sup>, Paul Norman<sup>5</sup>, Graeme J. Hankey<sup>3</sup>, Konrad Jamrozik<sup>6</sup>

1. Western Australian Centre for Healthy Ageing (WACHA), Centre for Medical Research, WA Institute for Medical Research, Perth, Australia

2. Centre for Human Bioethics, Monash University, Melbourne, Australia

3. School of Medicine and Pharmacology, University of Western Australia, Perth, Australia

4. School of Psychiatry and Clinical Neurosciences, University of Western Australia, Perth, Australia

5. School of Surgery, University of Western Australia, Perth, Australia

6. (Deceased) Previously School of Population Health and Clinical Practice, University of Adelaide, Adelaide, Australia

Corresponding author:

Dr Euzebiusz Jamrozik. C/O WA Centre for Healthy Ageing. Phone: +61 9224 2750.  
Email: zeb.jamrozik@gmail.com

Key words: Stroke, Self Report, Prognosis, Longitudinal Survey

Cover Title: Prognosis of Self-reported and Hospital-coded Stroke

Word Count: 2,379

### Tables and Figures:

- Table 1: Cross-tabulation of Self-reported History of Stroke with Hospital-coded Stroke
- Table 2: Comparison of baseline characteristics of participants with no stroke, self-reported (SR) stroke, hospital-coded (HC) stroke, and both self reported and hospital-coded stroke (both).
- Table 3: Risk of incident or recurrent stroke or incident cardiovascular events after the baseline assessment in 1996-1999 according to the coding of stroke at baseline (self-reported, hospital-coded, or both)
- Figure 1: Adjusted Kaplan Meier Curve For Time to Incident or Recurrent Stroke
- Figure 2. Comparison, by means of the log rank test, of the survival curves for recurrent stroke among men with a diagnosis of stroke by self-report (SR), Hospital-coded (HC), or both self report and hospital-coded stroke.

## Abstract

**Background** Population-based studies, as well as clinicians, often rely on self-report and hospital records to obtain a history of stroke. This study aimed to compare the validity of the diagnosis of stroke by self-report and by hospital coding according to their cross-sectional association with prevalent vascular risk factors, and longitudinal association with recurrent stroke and major cardiovascular outcomes in a large cohort of older Australian men.

**Methods** Between 1996 and 1999, 11,745 older men were surveyed for a self-reported history of stroke as part of the Health in Men Study (HIMS). Previous hospitalization for stroke was obtained with consent from linked medical records via the Western Australian Data Linkage System (WADLS). Subjects were followed by WADLS until December 31, 2010 for hospitalization for stroke, cardiovascular events, and all-cause mortality. The primary outcome was hospitalisation for stroke during follow-up. Secondary outcomes included incident vascular events and composite vascular endpoints.

**Results:** At baseline, a history of stroke was reported by 903 men (7.7%), previous hospitalisation for stroke was recorded in 717 (6.1%), both self-report and hospitalisation in 467 (4.0%), and no history of stroke in 10,696 men (91.1%). Prevalent cardiovascular disease and peripheral arterial disease were more common among men with previous hospitalisation for stroke than a history of self reported stroke ( $p < 0.001$ ). In longitudinal analyses, incident aortic aneurysm was also more common among men with baseline history of hospitalization for stroke (adjusted hazard ratio [HR]: 1.71 95% CI 1.12-2.60) than among men with self-reported stroke (HR: 0.88 95% CI 0.56-1.36) compared to men with no history of stroke. With regards to the primary outcome, the rate of hospitalisation for stroke during follow-up was significantly higher among men with self-reported stroke (HR: 2.44, 95% CI: 2.03 to 2.94), hospital-coded stroke (adjusted HR 3.02, 2.42 to 3.78) and both self-reported and hospital-coded stroke (adjusted HR 3.33; 2.82 to 3.92) compared to participants with no previous stroke. Time to recurrent stroke was similar among different methods of initial stroke diagnosis ( $p = 0.067$ ).

**Conclusions:** Self-reported stroke and hospital-coded stroke have similar prognostic value for predicting the risk of recurrent stroke. This supports the use of these ways of assessing a history of stroke for the clinical purposes of secondary prevention and for further epidemiological studies.

Word Count (abstract): 366

## Introduction

Stroke is defined by the World Health Organisation as a syndrome of “rapidly developing clinical evidence of focal (or global) disturbance of cerebral function, with symptoms lasting 24 hours or longer or leading to death, with no apparent cause other than of vascular origin”<sup>1</sup>. However, the clinical diagnosis of stroke is not always valid or reliable<sup>2, 3</sup> and there is no standardized “gold standard” test. Such uncertainty creates problems for clinicians as well as for population-based studies of stroke that do not include neurological examination or brain imaging as part of case ascertainment. Population studies therefore often rely on: (i) self-report, (ii) diagnosis based on data linkage records of hospitalisation, or (iii) diagnosis based on death certificates. The validity of each of these methods is uncertain, and likely to be suboptimal compared with clinical and radiological assessment.

A recent study reported in a population of 14,490 New Zealanders found that a self-reported history of stroke was present in 3% (n= 385) whereas only 1% (n=150) were hospitalised with a diagnosis of stroke<sup>4</sup>. It is not clear whether the higher prevalence of self-reported history of stroke reflected false-positive diagnoses by self-report, false-negative diagnoses by hospitalization (i.e. strokes that occurred during a hospital admission for another diagnosis may not have been fully documented), or failure of data linkage records to register cases not admitted to hospital.

An alternative method of defining the validity of a diagnosis of stroke may be by its prognosis, because individuals with a previous stroke have substantially higher rate of recurrent stroke, cardiovascular events, vascular dementia and death than individuals of the same age and sex who have not had a stroke<sup>5-9</sup>.

The REasons for Geographic And Racial Differences in Stroke (REGARDS) study has reported that among participants without self-reported stroke or transient ischemic attack, prior stroke symptoms are highly predictive of future stroke events; almost as large as the impact of smoking and hypertension and larger than the impact of diabetes and heart disease<sup>10</sup>. The REGARDS study also suggests that there is a spectrum of risk from stroke symptoms (SS) only, to TIA, to distant stroke (DS), and recent stroke (RS)<sup>11</sup>. Among 30 239 black and white Americans  $\geq 45$  years of age who were followed over  $5.0 \pm 1.72$  years for the occurrence of 737 strokes, there was 1.20-fold increased stroke risk for SS (95% CI, 0.96-1.51), 1.73-fold for TIA (95% CI, 1.27-2.36), 2.23-fold for DS (95% CI, 1.61- 3.09), and 2.85-fold for RS (95% CI, 2.16-3.76) compared with those who were asymptomatic.

In this study we aimed to: (i) compare the cross-sectional agreement between the diagnosis of stroke made by self-report and the diagnosis of stroke made by hospital coding, (ii) investigate the association between stroke ascertained by either method and the prevalence of cardiovascular risk factors and events and (iii) prospectively investigate the association between stroke (by either method) and subsequent hospitalisation for stroke, vascular events and death in a large cohort of community-based elderly men.

## **Methods**

### *Ethics*

This study followed the principles of the Helsinki Declaration and all participants provided written informed consent, including for access to data-linkage for study purposes. The Human Research Ethics Committee of the University of Western Australia approved the study.

### *Setting, Design and Participants*

The Health in Men Study (HIMS) is an observational study of community-dwelling men living in Perth, Western Australia, which was originally designed as a trial of screening for abdominal aortic aneurysm (AAA)<sup>5</sup>. In 1995, 19,352 men aged 65 years or older were randomly selected from the electoral roll (enrollment to vote is compulsory in Australia). Between 1996 and 1999 (wave 1: W1), 12,203 men attended a clinic and completed a health questionnaire, providing a range of demographic and clinical data, including self-reported prevalence of vascular risk factors and diseases. We have published previously a detailed account of the design and procedures for the assessment of HIMS participants<sup>5</sup>.

For hospital coded stroke, this study used the Western Australian Data Linkage System (WADLS), which provides electronic linkage to the state's population health data collections, including information from the death registry, cancer registry, mental health registry and hospital morbidity data system (which includes all separations (i.e. diagnoses) for public and private hospitals in the state of Western Australia dating back to 1971).

### *Definitions and diagnosis of prevalent stroke*

Prevalent stroke at the time of the initial assessment (wave 1) in 1996-1999 was ascertained by two methods: (i) self-report by questionnaire, and (ii) hospital coding detected by data linkage (see above). Self-reported prevalent stroke was lifetime prevalent stroke as defined by an answer of yes to the question: "Have you ever been told that you have had a stroke?" Transient ischaemic attack (TIA) was not measured separately by the questionnaire. Hospital coded stroke was defined according to ICD-9 codes 362.3, 433.x1, 434.x1, 430, 431, 436 or ICD-10 codes H34.1, I60, I61, I63, I64<sup>12</sup> and was retrieved from the Western Australian Data Linkage System (WADLS)<sup>13</sup> between 1971 and the time of assessment between 1996 and 1999.

#### *Follow-up*

Participants were followed up by means of the WADLS to 31 December 2010.

#### *Outcome events in the longitudinal study*

The primary outcome measure was incident (first-ever stroke for participants with no prior stroke) or recurrent stroke (for participants with history of prevalent stroke). Incident stroke was defined by hospital coded stroke as recorded by WADLS any time after the baseline assessment (1996-1999) and 31 December 2010.

Secondary outcomes included death, myocardial infarction (ICD-9 code 410 or ICD-10 code I21) and a composite measure of cardiovascular disease endpoints (CVD;

including coronary heart disease (ICD-9 codes 410-414; ICD-10 codes I20-I25), aortic aneurysm (ICD-9 codes 441.3-4; ICD-10 codes I71.3-4), and peripheral arterial disease (PAD; ICD-9 codes 440.2-4, 443, 444.2, 445.0; ICD-10 codes I70.2, I73, I74.2-4)).

### *Statistical methods*

Data from 458 of the 12,203 men assessed in 1996-1999 (W1) were excluded because they completed an early version of the questionnaire that did not ask about a previous stroke, resulting in a study population of 11,745. Data were analyzed using the statistical software package Stata version 11.1 (StataCorp).

The agreement between the diagnosis of a history of prevalent stroke by self-report and by data linkage were compared using Kappa statistics. The sensitivity, specificity and predictive values of diagnosis of stroke by self-report were assessed by taking the diagnosis by hospital record as the reference group. The prevalence of vascular risk factors and diseases among men with and without a diagnosis of stroke was compared by means of Pearson chi-square tests.

Cox regression (proportional hazards model) was performed to analyze the association between self-report or hospital record of prevalent stroke and time to incident stroke, cardiovascular event or death, after adjusting for age. The Schoenfeld residuals were examined to assess the proportional hazards assumption. All *p*-values were two-tailed, with a level of less than 0.05 considered statistically significant.

## Results

### *Cross sectional analyses*

Table 1 shows that among the 11,745 men included in the study, self-reported history of stroke was reported by 903 (7.7%), while hospital coded stroke between 1971 and 1996-9 was identified in 717 men (6.1%). Both self-reported and hospital-coded strokes were present in 467 men (4.0%), and no history of stroke, by either self-report or hospital coding, was present in 10,592 men (91.1%).

Using hospital-coded stroke as the reference group, the specificity (96%) and negative predictive value (98%) of self-reported stroke were high, while the sensitivity (65%) and positive predictive value (52%) were comparatively lower (Table 1). The kappa statistic for agreement between self-reported stroke and hospital-coded stroke beyond chance was 0.55 (95% CI 0.52 to 0.58; moderate agreement).

Compared with participants with no stroke (n=10,592), those who had a diagnosis only by self-report (n=436), only by hospital coding (n=250), or by both self-reported and hospital coded stroke (n=467) were older and had a higher prevalence of CVD risk factors and peripheral arterial disease (PAD) (Table 2). However, after adjusting for multiple variables, only prevalent CVD and PAD were more frequent in those with hospital-coded compared with self-reported stroke.

### *Longitudinal analyses*

Table 3 shows the absolute and relative rates of subsequent stroke, other vascular events, all cause mortality, and composite cardiovascular endpoints after the baseline assessment in 1996-1999 in men who had never had a stroke, and men who had self-reported stroke only, hospital-coded stroke only or both self-reported stroke and hospital-coded stroke. Men with self-reported stroke, hospital-coded stroke and both had higher rates of recurrent stroke, major vascular events and death than men with no past stroke, with the exception of incident aortic aneurysm, which was higher uniquely among those with hospital coded stroke only (HR 1.71 (1.12-2.60)).

Among the 10,597 men (91.1%) with neither a self-reported nor hospital-coded previous stroke, the rate of incident stroke was 12.0 per 1000 person-years (95% CI: 11.4 to 12.6). The rate of recurrent stroke was higher among men with self-reported stroke (adjusted hazard ratio [HR] 2.44, 95% CI: 2.03 to 2.94), with hospital-coded stroke (adjusted HR 3.02, 2.42 to 3.78), and with both self-report and hospital coding (adjusted HR 3.33; 2.82 to 3.92), with statistically similar risks among the three groups (Table 3).

Longitudinal survival curves for time to incident or recurrent stroke, incident myocardial infarction, all cause mortality, and composite measures of stroke or MI, and stroke or MI or death are shown in Figure 1. Comparison of survival curves by the log rank test (Figure 2) demonstrated that time to incident or recurrent stroke

was similar among different measures of stroke (self-report, hospital coding, or both;  $p=0.067$ ). Similar results were found for time to incident MI ( $p=0.963$ ), and composite vascular endpoints ( $p>0.1$ ). There was a marginally significant difference for all cause mortality, with higher risk of death among men with both measures of stroke ( $p=0.047$ ).

## Discussion

This study in a large population of Australian men has demonstrated that self reported stroke and hospital-coded stroke are both independently associated with a significantly increased risk of recurrent stroke, cardiovascular events, and death over more than a decade after the initial stroke. A self-reported history of stroke and a diagnosis of stroke by means of hospital coding conferred a similar risk of recurrent stroke, as did the presence of both a diagnosis of stroke by self-report and hospital coding. Time to recurrent stroke by log rank testing of longitudinal event curves was similar for either of the two methods of diagnosing stroke, or both ( $p=0.067$ ).

A gold standard for the diagnosis of stroke is the opinion of a specialist stroke clinician based on clinical information and imaging of the brain. Yet even in these circumstances inter-observer agreement may only be about 90%<sup>14, 15</sup>. In many other circumstances, such as in the primary care setting or in population-based research, this gold standard is not available or cannot always be readily confirmed, thus general practitioners and stroke epidemiology researchers alike rely on patients' self report of their history of stroke, hospital record, or death certificate documentation<sup>16, 17</sup>. It is therefore essential to ascertain to what degree such methods may be relied upon for the purposes of prognostication, secondary prevention and clinical epidemiology.

In the present study, similar to previously reported New Zealand data<sup>4</sup>, only approximately one third of the men who self reported a stroke had a hospital coded diagnosis of stroke. Self-reported stroke had a high specificity and negative predictive value for hospital-coded stroke but only modest sensitivity and positive predictive value. This apparent mismatch may reflect men who had attended hospitals outside of Western Australia, those who had had a stroke or TIA and not attended hospital, incomplete data linkage records or misattribution of symptoms caused by non-stroke conditions by patients or their physicians. However, hospital morbidity data obtained via WADLS possesses accuracy comparable to adjudication of medical records<sup>18</sup>. It may also reflect, to some extent, the different time periods over which a history of stroke was ascertained by the two methods. Self-reported stroke could include any stroke between birth (1913 to 1933 in the cohort) and the initial assessment in 1996-1999 whereas hospital-coded stroke was only recorded between 1971 (the commencement of WADLS, see Methods) and 1996-9. We did not ask our participants for the date of the self-reported stroke, so we could not verify if self-reported stroke occurred during or after 1971. However, most self-reported strokes would have occurred after 1971 when the age range of the cohort was 38 to 58 years. It is also possible that a small number of men admitted with TIA recalled this event as a stroke, as the self-reported questionnaire did not differentiate between stroke or TIA.

We found that self-report and hospital-coded stroke conferred similar risks of recurrent stroke, myocardial infarction, all cause mortality and composite vascular

events, suggesting that both measures have prognostic value for these events. Neither measure demonstrated a statistically significant advantage as a prognostic marker of recurrent stroke. Log rank analysis of event curves for recurrent stroke suggested that time to recurrent stroke was similar for men with either diagnostic measure of stroke, or both ( $p=0.067$ )

We demonstrated similar findings for other incident cardiovascular events in longitudinal analyses, with overlap of adjusted HR estimates for all outcome measures, with one exception. Men with hospital-coded stroke, whether or not self-report of stroke was present, had a significantly higher incidence of AAA (HR 1.71 95% CI 1.12-2.60 for hospital coded only; HR 1.98 95% CI 1.53-2.56 for both) during follow-up, whereas men with self-reported stroke only did not have an increased risk for AAA (HR 0.88 95% CI 0.56-1.36). This would be consistent with hospital-coded stroke being a more robust predictor of subsequent vascular risk; perhaps because those admitted with a stroke had more co-morbidities or more severe vascular disease at baseline. Other workers have identified peripheral arterial disease (PAD) as an independent prognostic risk factor for future morbidity and mortality after stroke<sup>19</sup>. Similarly, atherothrombotic burden (i.e. vascular disease in multiple organs) has been shown to be an adverse prognostic marker after stroke<sup>20</sup>.

In summary, this study demonstrates that self-reported stroke and hospital-coded stroke are both reliable prognostic measures of subsequent stroke with recurrent

stroke between 2.5 and 3.5 times more likely over more than a decade among those with either diagnostic measure compared to men with no stroke by either measure at baseline. Both measures were also predictive of incident myocardial infarction and other vascular events, with neither measure outperforming the other.

These results may be useful in at least two ways. Firstly, they add to evidence of the validity of self-reported stroke and hospital-coded stroke and their use in population health studies. Secondly, they confirm other evidence that a history of stroke may be used in a clinical context to inform patients' and physicians' decisions about secondary prevention.<sup>9-11, 16, 17</sup>; Whether or not a history of stroke is confirmed by radiological imaging, the risk of incident stroke, vascular events and death is significantly increased in those who report or have a documented hospital diagnosis of history of a prior stroke event.

## **Acknowledgements**

Special thanks to all men who participated in the Western Australian Abdominal Aortic Aneurysm Programme. Thanks to the staff and investigators of the original screening trial. The authors are also grateful for assistance received from the State Electoral Commission, the Australian Bureau of Statistics, the Registrar General of Births, Deaths and Marriages, and the Data Linkage Unit of the Health Department of Western Australia, and to hospitals in Perth for providing space in which to conduct screening.

## **Sources of Funding**

HIMS was supported by NHMRC Project Grants 964145, 139093, 403963 and 455811 with additional funding from the National Heart Foundation and the Western Australian Health Promotion Foundation (Healthway).

## **Conflicts of Interest**

None declared.

## References:

1. Monica W. The world health organization monica project (monitoring trends and determinants in cardiovascular disease): A major international collaboration: Who monica project principal investigators. *J Clin Epidemiol.* 1988;41:105-114
2. Nor AM, Davis J, Sen B, Shipsey D, Louw SJ, Dyker AG, Davis M, Ford GA. The recognition of stroke in the emergency room (rosier) scale: Development and validation of a stroke recognition instrument. *Lancet Neurol.* 2005;4:727-734
3. Hand PJ, Kwan J, Lindley RI, Dennis MS, Wardlaw JM. Distinguishing between stroke and mimic at the bedside: The brain attack study. *Stroke.* 2006;37:769-775
4. Carter K, Barber PA, Shaw C. How does self-reported history of stroke compare to hospitalization data in a population-based survey in new zealand? *Stroke.* 2010;41:2678-2680
5. Norman PE, Flicker L, Almeida OP, Hankey GJ, Hyde Z, Jamrozik K. Cohort profile: The health in men study (hims). *Int J Epidemiol.* 2009;38:48-52
6. Hardie K, Hankey GJ, Jamrozik K, Broadhurst RJ, Anderson C. Ten-year survival after first-ever stroke in the perth community stroke study. *Stroke.* 2003;34:1842-1846
7. Hankey GJ. Long-term outcome after ischaemic stroke/transient ischaemic attack. *Cerebrovasc Dis.* 2003;16 Suppl 1:14-19
8. Pendlebury S, Rothwell P. Risk of recurrent stroke, other vascular events and dementia after transient ischaemic attack and stroke. *Cerebrovascular diseases.* 2009;27:1-11
9. Kuwashiro T, Sugimori H, Ago T, Kamouchi M, Kitazono T. Risk factors predisposing to stroke recurrence within one year of non-cardioembolic stroke onset: The fukuoka stroke registry. *Cerebrovascular diseases.* 2011;33:141-149
10. Kleindorfer D, Judd S, Howard VJ, McClure L, Safford MM, Cushman M, Rhodes D, Howard G. Self-reported stroke symptoms without a prior diagnosis of stroke or transient ischemic attack: A powerful new risk factor for stroke. *Stroke.* 2011;42:3122-3126
11. Judd SE, Kleindorfer DO, McClure LA, Rhodes JD, Howard G, Cushman M, Howard VJ. Self-report of stroke, transient ischemic attack, or stroke symptoms and risk of future stroke in the reasons for geographic and racial differences in stroke (regards) study. *Stroke.* 2013;44:55-60
12. Kokotailo RA, Hill MD. Coding of stroke and stroke risk factors using international classification of diseases, revisions 9 and 10. *Stroke.* 2005;36:1776-1781
13. Holman CD, Bass AJ, Rouse IL, Hobbs MS. Population-based linkage of health records in western australia: Development of a health services research linked database. *Aust N Z J Public Health.* 1999;23:453-459
14. Allder SJ, Moody AR, Martel AL, Morgan PS, Delay GS, Gladman JR, Fentem P, Lennox GG. Limitations of clinical diagnosis in acute stroke. *Lancet.* 1999;354:1523
15. Norris JW, Hachinski VC. Misdiagnosis of stroke. *Lancet.* 1982;1:328-331

16. Engstad T, Bonna KH, Viitanen M. Validity of self-reported stroke: The tromso study. *Stroke*. 2000;31:1602-1607
17. Reitz C, Schupf N, Luchsinger JA, Brickman AM, Manly JJ, Andrews H, Tang MX, DeCarli C, Brown TR, Mayeux R. Validity of self-reported stroke in elderly african americans, caribbean hispanics, and whites. *Arch Neurol*. 2009;66:834-840
18. Barr E, Tonkin A, Welborn T, Shaw J. Validity of self-reported cardiovascular disease events in comparison to medical record adjudication and a statewide hospital morbidity database: The ausdiab study. *Internal medicine journal*. 2009;39:49-53
19. Meves SH, Diehm C, Berger K, Pittrow D, Trampisch H-J, Burghaus I, Tepohl G, Allenberg J-R, Endres HG, Schwertfeger M. Peripheral arterial disease as an independent predictor for excess stroke morbidity and mortality in primary-care patients: 5-year results of the getabi study. *Cerebrovascular diseases*. 2010;29:546-554
20. Cimminiello C, Zaninelli A, Carolei A, Sacco S, Toni D, Gensini G. Atherothrombotic burden and medium-term prognosis in patients with acute ischemic stroke: Findings of the sirio study. *Cerebrovascular diseases*. 2012;33:341-347

**Table 1: Cross-tabulation of Self-reported History of Stroke with Hospital-coded Stroke**

		Hospital-coded Stroke				Total
		Yes		No		
		n	%	n	%	
<b>Self-reported Stroke</b> (n, %)	<b>Yes</b>	467	4.0%	436	3.7%	903
	<b>No</b>	250	2.1%	10,592	91.1%	10,842
	<b>Total</b>	717		11,028		11,745
		Sensitivity*	65.1%	PPV*	51.7%	
		Specificity*	96.0%	NPV*	97.7%	

\*Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) refer to self-reported stroke against hospital-coded stroke.

**Table 2: Comparison of baseline characteristics of participants with no stroke, self-reported (SR) stroke, hospital-coded (HC) stroke, and both self reported and hospital-coded stroke (both).**

	None N=10,592	SR Only N=436	HC Only N=250	Both N=467	Adjusted P-value (all groups)	Adjusted P-value (SR vs. HC. vs. both)
Age, mean (SD)	72.0 (4.4)	73.2 (4.4)	73.4 (4.6)	73.4 (4.3)	<0.001	0.6824
Hypertension	4381 (41.4)	270 (61.9)	155 (62.0)	310 (66.4)	<0.001	0.308
Diabetes	1216 (11.5)	82 (18.8)	36 (14.4)	85 (18.2)	<0.001	0.314
Dyslipidaemia	3671 (34.7)	188 (43.1)	123 (49.2)	209 (44.8)	<0.001	0.300
Smoking status					<0.001	0.345
Never	3136 (29.6)	96 (22.0)	58 (23.2)	116 (24.8)		
Ex-smoker	6264 (59.2)	286 (65.6)	164 (65.6)	312 (66.8)		
Current	1188 (11.2)	54 (12.4)	28 (11.2)	39 (8.4)		
Prevalent CVD	2562 (24.2)	170 (39.0)	129 (51.6)	211 (45.2)	<0.001	0.005
Prevalent PAD	303 (2.9)	28 (6.4)	40 (16.0)	54 (11.6)	<0.001	<0.001

\* *p*-value testing by means of chi square for between-group differences by oneway analysis of variance (age) and Pearson chi-square statistic (all variables but age).

- Lipid Disorder, defined as high cholesterol, high triglycerides, or medication prescribed by a doctor to lower blood cholesterol
- Hypertension, self-reported or medication
- Diabetes, self-reported or medication
- CVD: cardiovascular diseases by self-reported myocardial infarction (MI), angina, aneurysm, coronary artery bypass graft (cabg) surgery, coronary artery angioplasty
- PAD: peripheral arterial disease by data linkage, using codes provided in the method section

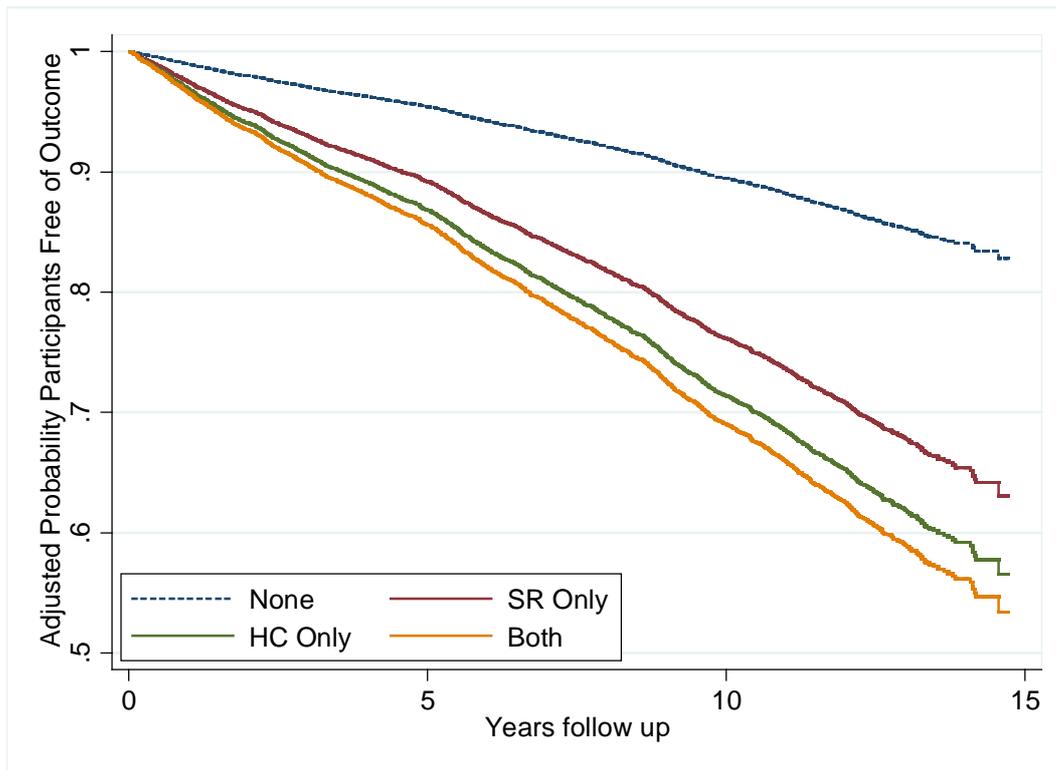
**Table 3: Risk of incident or recurrent stroke or incident cardiovascular events after the baseline assessment in 1996-1999 according to the coding of stroke at baseline (self-reported, hospital-coded, or both)**

Outcome	Stroke Coding at Baseline	Total N	Event N (person-years)	Rate per 1,000-py	Crude HR (95%CI)	Adj* HR (95%CI)
Incident or recurrent stroke	None	10592	1311 (109418)	12.0 (11.4,12.6)	1	1
	Self-reported only	436	128 (3642)	35.1 (29.6,41.8)	3.00 (2.50-3.59)	2.44 (2.03-2.94)
	Hospital-coded only	250	83 (1953)	42.5 (34.3,52.7)	3.63 (2.90-4.53)	3.02 (2.42-3.78)
	Both	467	166 (3600)	46.1 (39.6,53.7)	3.96 (3.37-4.66)	3.33 (2.82-3.92)
Incident Myocardial infarction	None	9657	1366 (100297)	13.6 (12.9,14.4)	1	1
	Self-reported only	372	82 (3299)	24.9 (20.0,30.9)	1.86 (1.49-2.32)	1.48 (1.18-1.85)
	Hospital-coded only	195	42 (1632)	25.7 (19.0,34.8)	1.94 (1.42-2.63)	1.62 (1.19-2.20)
	Both	379	82 (3319)	24.7 (19.9,30.7)	1.85 (1.48-2.32)	1.52 (1.21-1.90)
Incident Aortic aneurysm	None	10541	509 (110481)	4.6 (4.2,5.0)		
	Self-reported only	431	21 (3899)	5.4 (3.5,8.3)	1.12 (0.73-1.74)	0.88 (0.56-1.36)
	Hospital-coded only	247	23 (2055)	11.2 (7.4,16.8)	2.27 (1.50-3.45)	1.71 (1.12-2.60)
	Both	413	63 (3383)	18.6 (14.5,23.8)	2.58 (1.99-3.33)	1.98 (1.53-2.56)
Stroke or myocardial infarction	None	9657	2321 (96726)	24.0 (23.0,25.0)		
	Self-reported only	372	165 (3016)	54.7 (47.0,63.7)	2.33 (1.99-2.73)	1.87 (1.59-2.19)
	Hospital-coded only	195	92 (1452)	63.4 (51.7,77.8)	2.71 (2.20-3.34)	2.29 (1.85-2.82)

	Both	379	194 (2929)	66.2 (57.5,76.2)	2.85 (2.46-3.30)	2.40 (2.07-2.78)
Stroke, myocardial infarction or death	None	9657	5056 (96726)	52.3 (50.9,53.7)		
	Self-reported only	372	267 (3016)	88.5 (78.5,99.8)	1.77 (1.57-2.01)	1.43 (1.27-1.62)
	Hospital-coded only	195	139 (1452)	95.8 (81.1,113.1)	1.93 (1.63-2.28)	1.66 (1.40-1.97)
	Both	379	303 (2929)	103.4 (92.4,115.8)	2.10 (1.87-2.36)	1.82 (1.62-2.04)
All-cause mortality	None	10592	4692 (113865)	41.2(40.0-42.4)		
	Self-reported only	436	268 (4038)	66.4(58.9-74.8)	1.69 (1.49-1.91)	1.38 (1.22-1.56)
	Hospital-coded only	250	161 (2214)	72.7(62.3-84.9)	1.87 (1.60-2.19)	1.57 (1.34-1.84)
	Both	467	328 (4100)	80.0(71.8-89.1)	2.08 (1.86-2.33)	1.73 (1.54-1.94)

*\*Adjusted for age, prevalent hypertension, prevalent diabetes, prevalent lipid disorder and smoking status  
HR, hazard ratio obtained using standard Cox Regression Models*

Figure 1: Adjusted Kaplan Meier Curve For Time to Incident or Recurrent Stroke



**Figure 2. Comparison, by means of the log rank test, of the survival curves for recurrent stroke among men with a diagnosis of stroke by self-report (SR), Hospital-coded (HC), or both self report and hospital-coded stroke.**

