



# Geographic Variations of Stroke Incidence in Chinese Communities: An 18-Year Prospective Cohort Study from 1997 to 2015

Fan Xia,<sup>a,\*</sup> Xuexin Yu,<sup>b,\*</sup> Yunke Li,<sup>a,\*</sup> Yuqi Chen,<sup>a</sup> Wei Zhang,<sup>b</sup> Chao You,<sup>a</sup> Xin Hu<sup>a</sup>

<sup>a</sup>Department of Neurosurgery, West China Hospital, Sichuan University, Chengdu, China

<sup>b</sup>West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China

**Background and Purpose** As a leading cause of disability and death in China, stroke as well as its epidemiologic features have gained increasing attention. Prior studies, however, have overgeneralized the north-to-south gradient in China. Whether the differences exist across urban and rural areas remains unexplored. This study therefore aims to investigate the north-to-south gradient in stroke incidence across urban and rural China.

**Methods** The present prospective cohort study analyzed data from the China Health and Nutrition Survey 1997 to 2015. By including 16,917 individuals from diverse social contexts, we calculated the age-standardized incidence of stroke across regions and the age-adjusted risk ratio (aRR). Cox proportional hazards models with time-varying covariates were employed to analyze variations in incident stroke.

**Results** During the follow-up, age-standardized incidence of stroke ranged from 4.17 per 1,000 person-years (95% confidence interval [CI], 3.38 to 4.96) in the north region to 1.95 (95% CI, 1.60 to 2.30) in the south region (aRR, 2.04; 95% CI, 1.58 to 2.64;  $P < 0.001$ ). The north-to-south gradient of stroke incidence was observed only in rural areas, but not in urban areas. Hierarchical modelling analyses further indicated that the regional differences could be mostly explained by the disparities in the prevalence of hypertension.

**Conclusions** The present study extends the current evidence on the north-to-south gradient by demonstrating that the difference varied across urban and rural China. Our findings highlight the importance of hypertension management as the measure for alleviating regional differences in stroke incidence.

**Keywords** Stroke; Incidence; Geography, medical; China

## Introduction

Although geographic variations have been widely recognized as a challenging public health concern,<sup>1-3</sup> the variations in stroke burden have not been fully clarified. As a leading cause of disability and death in China,<sup>4,5</sup> stroke as well as its epidemi-

ologic features have gained increasing attention. Previous studies have reported a north-to-south gradient in the prevalence of stroke in China.<sup>4,6,7</sup> The prevalence measure, however, could not fully reflect the epidemiologic transition in current China, as it is sensitive to changes in the life expectancy.<sup>7-9</sup> Along with the rapid socio-economic development over the

**Correspondence:** Xin Hu  
Department of Neurosurgery, West China Hospital, Sichuan University, 37 Guoxue Alley, Chengdu 610041, China  
Tel: +86-28-85422972  
Fax: +86-28-85422972  
E-mail: [huxingxy@gmail.com](mailto:huxingxy@gmail.com)  
<https://orcid.org/0000-0002-6084-1341>

Received: June 17, 2020

Revised: July 20, 2020

Accepted: July 24, 2020

\*These authors contributed equally to the manuscript as first author.

past two decades, the average lifespan in China has increased from 70.43 years in 1997 to 75.93 in 2015, with 10.92% population aged over 65.<sup>10,11</sup> Therefore, up-to-date information regarding geographic variations in stroke incidence is warranted.

Earlier studies, however, have limitations and knowledge gaps that should be addressed and bridged in a comprehensive way. First, prior investigations directed insufficient attention on the interaction between regional variations and urban vs. rural differences in the epidemiologic features of stroke, although empirical evidence on population health indicated that regional disparities in urban China tended to be marginal due to rapid urbanization.<sup>12</sup> Whether the north-to-south gradient in stroke incidence varies across urban and rural areas in China is unknown. Furthermore, prior research had a major limitation of retrospective cohort or cross-sectional study-designs,<sup>4,6,7,13</sup> which were not able to provide robust estimates of the attributable risk factors of the disparities.

Overall, this study aims to (1) examine the north-to-south gradient in stroke incidence and their interaction with urban versus rural settings. Meanwhile, this project has an advantage of prospective cohort study-design to (2) analyze the potential attributable risk factors of the geographic variations in stroke incidence.

## Methods

All data and materials have been made publicly available at the website of the China Health and Nutrition Survey (CHNS) and can be accessed at <https://www.cpc.unc.edu/projects/china>.

### Data source

The present study analyzed data from the CHNS. CHNS included over 30,000 individuals from twelve provinces and three provincial-level cities in China. Launched in 1989, this ongoing open cohort has been jointly conducted by the Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute for Nutrition and Health at the Chinese Center for Disease Control and Prevention. CHNS employed a multistage random cluster process to draw the study sample and included individuals from diverse social contexts (e.g., low-, middle-, and high-income counties). As of May 2020, CHNS has released data of ten wave periods of 1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009, 2011, and 2015. Details of the survey design, sampling procedures, data collection, and quality control have been published previously.<sup>14</sup>

### Study design and sample collection

This study employed a prospective cohort study-design. Be-

cause CHNS did not record self-reported disease outcomes until 1997, the present study included data from 1997 to 2015. Because of the low prevalence of stroke among neonates and children,<sup>15</sup> we included individuals aged over 18 years old. Focusing on stroke incidence, we enrolled stroke-free individuals at his/her entry point and excluded individuals with only one observation. Last, observations after the diagnosis of stroke were excluded. Details regarding the sample collection are provided in Supplementary Figure 1.

### Measures

We employed the self-reported diagnosis of stroke as the health outcome. As three provinces were not enrolled in CHNS until 2015, the present study included individuals from nine provinces and three provincial-level cities. According to the latitude, provinces and provincial-level cities were clustered as the south region, the central region, and the north region to explore the north-to-south differences (Figure 1A). Each region consisted of several provinces as well as one provincial-level city, which represented a highly urbanized metropolitan area. Urban-rural settings were derived from CHNS.

Furthermore, we included multiple potential risk factors of stroke, including socio-demographic status (age, sex, race, marital status, employment, and educational level), lifestyle behaviors (body mass index [BMI] and smoking history), and self-reported disease history (hypertension, diabetes mellitus, and myocardial infarction).<sup>4,5,7,13,16</sup>

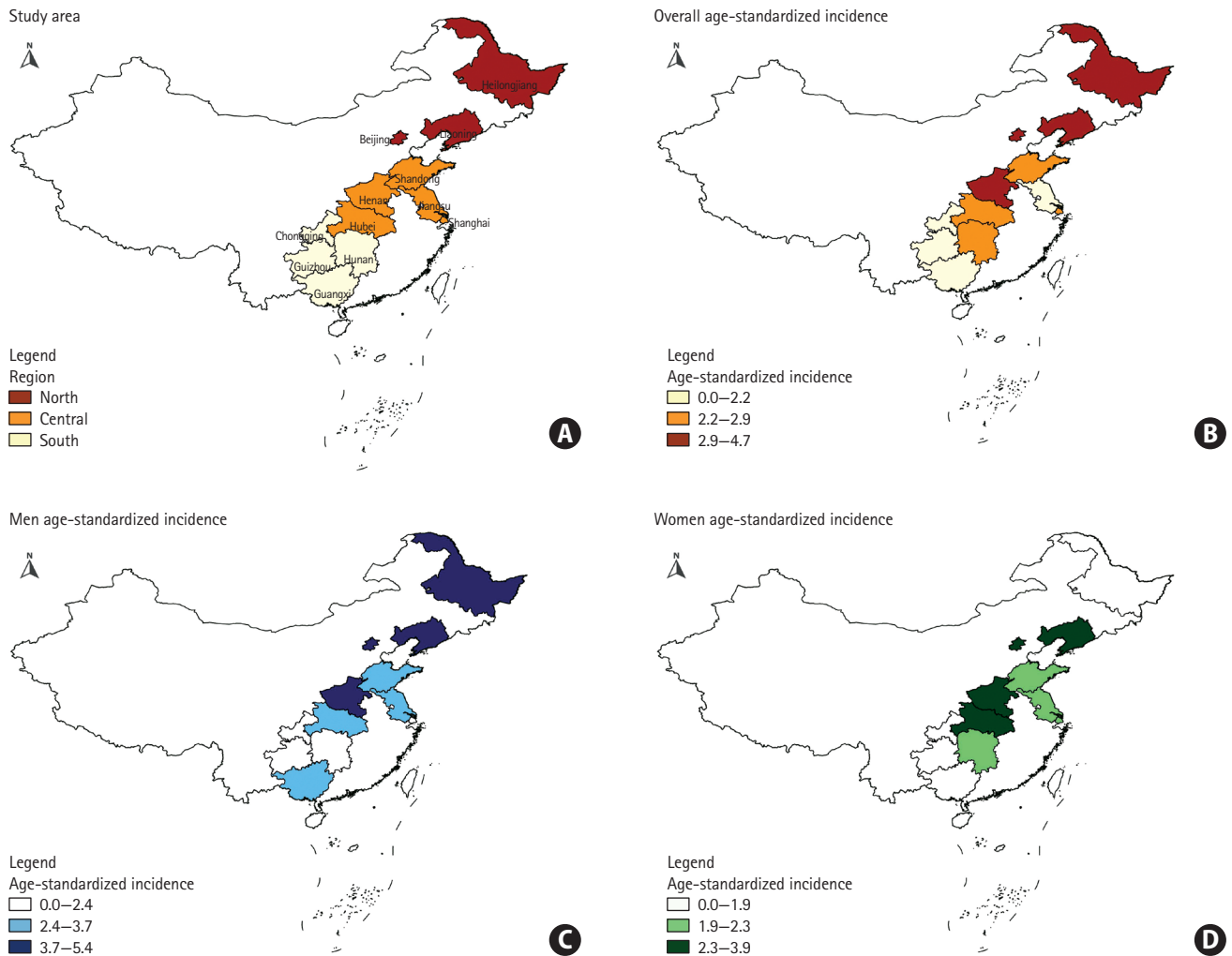
### Data analyses

To compare baseline characteristics across regions, we conducted multiple significance tests (Pearson chi-square tests, analysis of variance [ANOVA], and Kruskal-Wallis rank-sum tests) according to data characteristics. To depict the geographic variations in the incidence of stroke, we calculated the crude incidence as described below:

$$\text{Incidence} = \frac{\text{Number of newly diagnosed stroke}}{\text{Total person-years at risk}}$$

We then calculated the age-standardized incidence of stroke in each region, using the south region study sample as the standard population. To compare the age-standardized incidence across regions, log-binomial regression was performed to calculate age-adjusted risk ratios (aRRs).

To further investigate the geographic variations in stroke incidence, we employed the extended Cox proportional hazards models. Hierarchical modelling analyses were performed to investigate potential attributable risk factors of the geographic variations. In addition, we designated time-varying covariates to



**Figure 1.** Study area and age-standardized incidence across provinces. The age-standardized incidence was calculated by using the study sample in the Guangxi province as the standard population. Details about age-standardized incidence are provided in Supplementary Table 1. (A) The study area included nine provinces and three provincial-level cities: Beijing, Chongqing, Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning, Shandong, and Shanghai. (B) Overall age-standardized stroke incidence, per 1,000 person-years. (C) Age-standardized stroke incidence among men, per 1,000 person-years. (D) Age-standardized stroke incidence among women, per 1,000 person-years.

control for changes in potential risk factors over time. We clustered observations from the same individual to generate robust standard errors. The interaction between urban versus rural settings and geographic regions was included in the models to estimate regional differences across urban and rural areas. As for subgroup analyses, we repeated the modeling analyses by sex.

### Missing values and sensitivity analyses

The percent of missing values ranged from 0.8% for marital status to 7% for BMI. Missing covariates were imputed with data from prior observations. As covariates, such as hypertension, smoking history, and employment status, may change over time, we conducted additional modelling analyses by using the dataset with missing values. In addition, as differences remained in

various statistical techniques when estimating rate ratios,<sup>17</sup> we estimated incidence rate ratios and odds ratios respectively from Poisson regression and logistic regression as well as Mantel-Haenszel rate ratios stratified by age groups. Furthermore, the main outcome was a self-reported stroke occurrence, which was likely to be influenced by educational level and age. To reduce the recall bias and underestimation, we repeated the modeling analyses restricted to individuals with at least middle school education and aged under 65 years old at baseline. The rationale was that better educated and younger individuals might have higher disease awareness and memory function.

Statistical analyses were performed with Stata/SE 15.0 (StataCorp., College Station, TX, USA). A two-tailed *P*-value of less than 0.05 was considered statistically significant.

### Ethics approval

CHNS was approved by the ethics committee of Carolina Population Center at the University of North Carolina at Chapel Hill and the NINH at the CCDC. Informed consent was obtained from all subjects before the investigation. The present study derived data from the public domain, and therefore ethics statement and informed consents were not applicable.

## Results

### Study sample and baseline characteristics

Overall, this study included 66,380 observations from 16,917 individuals. The cohort accrued 169,004 person-years with an average follow-up of 9.99 years. During the follow-up, 442 stroke cases were identified with a crude incidence rate of 2.61 per 1,000 person-years (95% confidence interval [CI], 2.38 to 2.87). The mean age at entry into the cohort was 42.74±15.40 (range, 18 to 100) with no significant difference across the

south, central, and north regions ( $P=0.06$ ). However, there were differences in several baseline characteristics across regions. In general, individuals in the north region were more likely to be a smoker ( $P<0.001$ ) and obese ( $P<0.001$ ) and to have hypertension ( $P<0.001$ ) and diabetes ( $P<0.01$ ) as compared to those in the south and central regions (Table 1).

In addition, we examined the differences in baseline characteristics between the study sample and excluded individuals, i.e., those without stroke records or with only one observation. Excluded individuals were younger (40.41 years old vs. 42.74 years old,  $P<0.001$ ) than the study sample. In comparison with those in the study, individuals not in the cohort were more likely to be obese ( $P<0.001$ ), well-educated ( $P<0.001$ ), and to live in urban rather than rural areas ( $P<0.001$ ). Urban areas in the north region respectively had a 3.16% and 0.94% higher rate of missing stroke record as compared to their counterparts in the central and south regions ( $P<0.05$ ). Likewise, in rural areas, the north region respectively had a 4.28% and 3.98%

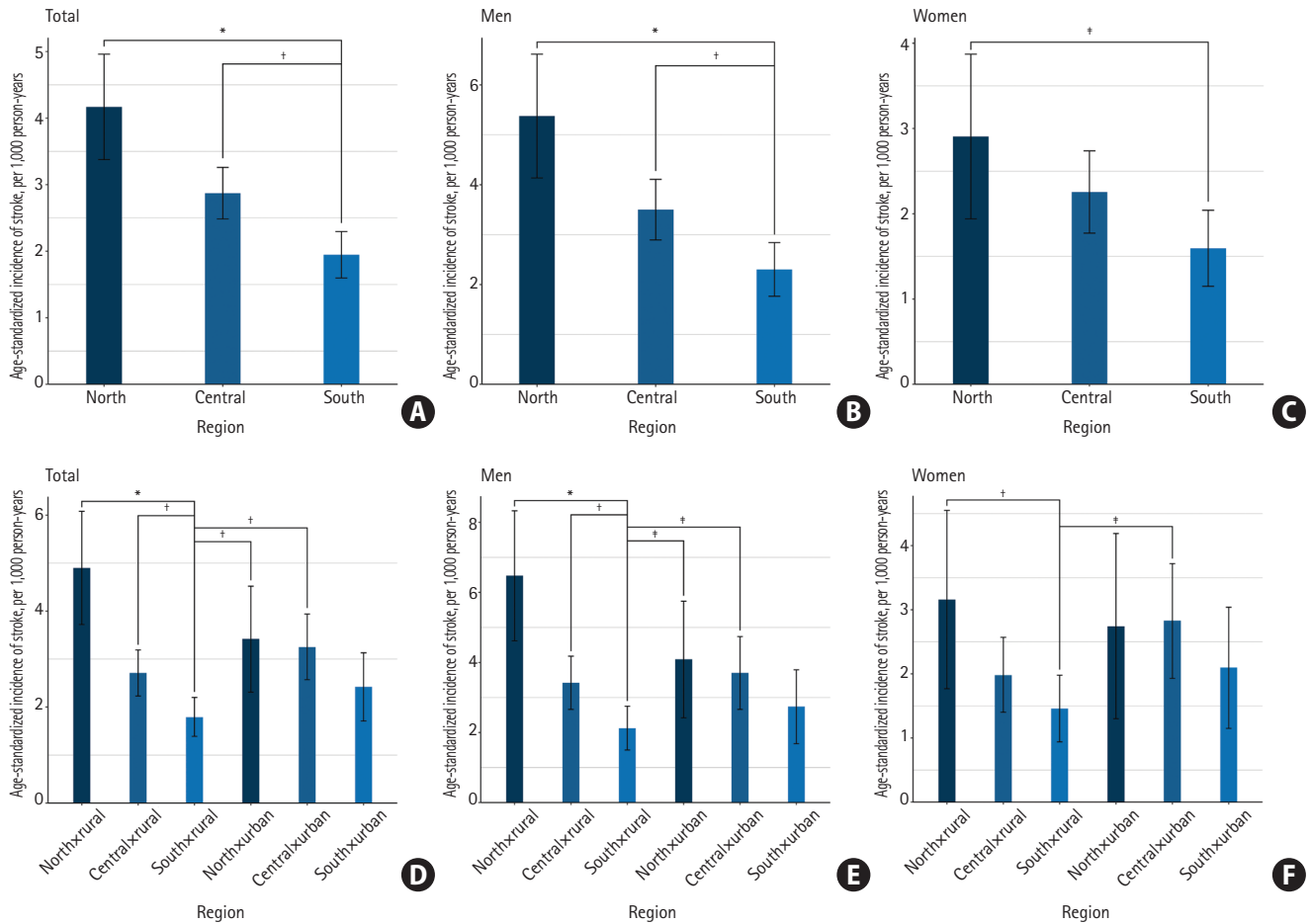
**Table 1.** Baseline characteristics across regions

Variable	South (n=5,861)	Central (n=7,449)	North (n=3,607)	Total (n=16,917)	P
Urban (vs. rural)	2,122 (36.21)	2,830 (37.99)	1,500 (41.59)	6,452 (38.14)	<0.001*
Age	42.54±15.84	43.05±15.54	42.42±14.36	42.74±15.40	0.06 <sup>†</sup>
Men (vs. women)	2,878 (49.10)	3,597 (48.29)	1,755 (48.66)	8,230 (48.65)	0.65*
Race (Han vs. others)	4,589 (78.47)	7,331 (99.19)	3,132 (86.88)	15,05 (89.36)	<0.001*
Married (vs. others)	4,510 (77.21)	6,028 (81.21)	3,095 (85.90)	13,63 (80.83)	<0.001*
Employed (vs. others)	4,072 (69.55)	5,053 (67.90)	2,492 (69.22)	11,617 (68.75)	0.10*
Educational level					<0.001 <sup>†</sup>
Illiteracy	1,397 (24.17)	1,610 (21.92)	363 (10.27)	3,370 (20.23)	
Primary school	1,277 (22.09)	1,269 (17.28)	697 (19.72)	3,243 (19.47)	
Middle school	1,829 (31.64)	2,367 (32.23)	1,135 (32.12)	5,331 (32.00)	
High school	1,278 (22.11)	2,099 (28.58)	1,339 (37.89)	4,716 (28.31)	
BMI (kg/m <sup>2</sup> )					<0.001 <sup>†</sup>
<18.5	610 (10.71)	407 (5.71)	168 (4.80)	1,185 (7.26)	
18.5–24.9	4,195 (73.69)	4,772 (66.93)	2,181 (62.30)	11,148 (68.29)	
24.9–29.9	803 (14.11)	1,696 (23.79)	968 (27.65)	3,467 (21.24)	
>29.9	85 (1.49)	255 (3.58)	184 (5.26)	524 (3.21)	
Smoking history					<0.001 <sup>†</sup>
Non-smoker	3,958 (67.72)	5,239 (70.40)	2,404 (66.87)	11,601 (68.72)	
Smoking cessation	79 (1.35)	109 (1.46)	70 (1.95)	258 (1.53)	
Current smoker	1,808 (30.93)	2,094 (28.14)	1,121 (31.18)	5,023 (29.75)	
Hypertension	248 (4.27)	623 (8.39)	339 (9.44)	1,210 (7.19)	<0.001*
Diabetes	104 (1.81)	152 (2.07)	100 (2.81)	356 (2.14)	<0.01*
Myocardial infarction	13 (0.22)	24 (0.32)	18 (0.50)	55 (0.33)	0.07*

Values are presented as number (%) or mean±standard deviation. Missing rate: race (0.43%), marital status (0.30%), employment status (0.12%), educational level (1.52%), BMI (3.51%), smoking history (0.21%), hypertension (0.49%), diabetes (1.66%), myocardial infarction (0.18%).

BMI, body mass index; ANOVA, analysis of variance.

\*Pearson chi-square test; <sup>†</sup>ANOVA; \*Kruskal-Wallis rank-sum test.



**Figure 2.** Age-standardized incidence across regions and urban vs. rural settings. Details and age-adjusted risk ratios are provided in Tables 2 and 3. (A–C) Age-standardized stroke incidence across regions, per 1,000 person-years. (A) In total; (B) among men; (C) among women. (D–F) Age-standardized stroke incidence across regions and urban vs. rural settings, per 1,000 person-years. (D) In total; (E) among men; (F) among women. Southxurban represents rural areas in the south region. \* $P < 0.001$ ; † $P < 0.01$ ; ‡ $P < 0.05$ .

**Table 2.** Crude and age-standardized incidence across regions, per 1,000 person-years

Region	Person-year	Stroke case	Crude incidence	Age-standardized incidence (95% CI)	aRR (95% CI)*
<b>Overall</b>					
South	61,132	119	1.95	1.95 (1.60–2.30)	-
Central	75,542	212	2.81	2.87 (2.49–3.26)	1.46 (1.17–1.82) <sup>†</sup>
North	32,370	111	3.43	4.17 (3.38–4.96)	2.04 (1.58–2.64) <sup>‡</sup>
<b>Men</b>					
South	30,415	70	2.30	2.30 (1.76–2.84)	-
Central	37,073	128	3.45	3.50 (2.89–4.11)	1.52 (1.14–2.03) <sup>†</sup>
North	15,821	34	4.68	5.38 (4.14–6.62)	2.23 (1.62–3.08) <sup>‡</sup>
<b>Women</b>					
South	30,717	49	1.60	1.60 (1.15–2.04)	-
Central	38,469	84	2.18	2.26 (1.77–2.74)	1.38 (0.97–1.96)
North	16,549	37	2.24	2.91 (1.94–3.87)	1.72 (1.12–2.63) <sup>§</sup>

The age-standardized incidence was calculated by using the south region study sample as the standard population.

CI, confidence interval; aRR, adjusted risk ratio.

\*The age-adjusted risk ratio was derived from log-binomial models with age groups introduced as confounders; † $P < 0.01$ ; ‡ $P < 0.001$ ; § $P < 0.05$ .

higher rate of missing stroke record in comparison with the central and south regions ( $P<0.001$ ).

### Age-standardized incidence and risk ratios

Figures 1 and 2 visualize the age-standardized incidence across regions (Supplementary Table 1). Table 2 provides additional information about the CIs and age-aRRs. During the study period, the age-standardized incidence of stroke ranged from 4.17 per 1,000 person-years (95% CI, 3.38% to 4.96) in the north region to 1.95 (95% CI, 1.60 to 2.30) in the south region (aRR, 2.04; 95% CI, 1.58 to 2.64;  $P<0.001$ ). A similar pattern was observed in both men (central vs. south [aRR, 1.52; 95% CI, 1.14 to 2.03;  $P<0.01$ ], north vs. south [aRR, 2.23; 95% CI, 1.62 to 3.08;  $P<0.05$ ]) and women (north vs. south: aRR, 1.71; 95% CI, 1.12 to 2.63;  $P<0.05$ ).

Table 3 and Figure 2 provide additional information about the north-to-south differences by urban-rural settings. Specifically, the age-standardized incidence among rural individuals in the north region (4.90; 95% CI, 3.72 to 6.08) was higher than their counterparts in the south region (1.79; 95% CI, 1.39 to 2.20;

aRR, 2.59, 95% CI, 1.88 to 3.58,  $P<0.001$ ). The north-to-south differences were observed in both rural men (central vs. south [aRR, 1.66; 95% CI, 1.16 to 2.39;  $P<0.01$ ], north vs. south [aRR, 2.87; 95% CI, 1.93 to 4.27;  $P<0.05$ ]) and women (north vs. south: aRR, 2.09; 95% CI, 1.22 to 3.59;  $P<0.01$ ), whereas urban residents appeared to have similar age-standardized incidence across regions, as indicated by the overlapped 95% CIs of aRRs.

### Extended Cox proportional hazards models

Table 4 provides results from hierarchical extended Cox proportional hazards analyses. In general, Models 1 and 2 were adjusted for age and sex; Model 3 was further adjusted for socio-economic covariates; Model 4 was additionally adjusted for lifestyle attributes; and Model 5 was fully adjusted and included age, sex, socio-economic covariates, lifestyle attributes, and disease history.

Results from Model 1 in Table 4 indicated that individuals in the north (hazard ratio [HR], 2.19; 95% CI, 1.68 to 2.86) and central regions (HR, 1.52; 95% CI, 1.21 to 1.90) respectively had a higher risk of incident stroke in comparison with their counter-

**Table 3.** Crude and age-standardized incidence across regions and urban vs. rural areas, per 1,000 person-years

Region	Person-year	Stroke case	Crude incidence	Age-standardized incidence (95% CI)	aRR (95% CI)*
<b>Overall</b>					
Southxrural	41,272	74	1.79	1.79 (1.39–2.20)	-
Centralxrural	51,371	126	2.45	2.71 (2.23–3.19)	1.55 (1.16–2.06) <sup>†</sup>
Northxrural	20,990	74	3.53	4.90 (3.72–6.08)	2.59 (1.88–3.57) <sup>‡</sup>
Southxurban	19,860	45	2.27	2.42 (1.71–3.13)	1.28 (0.89–1.85)
Centralxurban	24,171	86	3.56	3.25 (2.57–3.94)	1.66 (1.22–2.25) <sup>†</sup>
Northxurban	11,380	37	3.25	3.42 (2.31–4.52)	1.74 (1.18–2.57) <sup>†</sup>
<b>Men</b>					
Southxrural	20,715	44	2.12	2.12 (1.50–2.75)	-
Centralxrural	25,415	80	3.15	3.42 (2.66–4.18)	1.66 (1.16–2.39) <sup>†</sup>
Northxrural	10,361	51	4.92	6.48 (4.62–8.33)	2.87 (1.93–4.27) <sup>‡</sup>
Southxurban	9,700	26	2.68	2.74 (1.68–3.79)	1.22 (0.76–1.98)
Centralxurban	11,658	48	4.12	3.70 (2.66–4.74)	1.59 (1.06–2.38) <sup>§</sup>
Northxurban	5,460	23	4.21	4.09 (2.42–5.75)	1.76 (1.07–2.89) <sup>§</sup>
<b>Women</b>					
Southxrural	20,557	30	1.46	1.46 (0.94–1.98)	-
Centralxrural	25,956	46	1.77	1.98 (1.40–2.57)	1.39 (0.88–2.20)
Northxrural	10,629	23	2.16	3.16 (1.77–4.55)	2.09 (1.22–3.59) <sup>†</sup>
Southxurban	10,160	19	1.87	2.10 (1.15–3.04)	1.37 (0.77–2.42)
Centralxurban	12,513	38	3.04	2.83 (1.93–3.72)	1.77 (1.10–2.84) <sup>§</sup>
Northxurban	5,920	14	2.37	2.74 (1.30–4.19)	1.69 (0.90–3.16)

Southxrural represents rural areas in the south region. The age-standardized incidence was calculated by using study sample in rural areas in the south region as the standard population.

CI, confidence interval; aRR, adjusted risk ratio.

\*The age-adjusted risk ratio was derived from log-binomial models with age groups introduced as confounders; <sup>†</sup> $P<0.01$ ; <sup>‡</sup> $P<0.001$ ; <sup>§</sup> $P<0.05$ .

**Table 4.** Extended Cox proportional hazards analyses of stroke occurrence, HR (95% CI)

Independent variable	Model 1	Model 2	Model 3	Model 4	Model 5
Urban (vs. rural)	1.06 (0.87–1.29)				
<b>Region</b>					
South	Reference				
Central	1.52 (1.21–1.90)*				
North	2.19 (1.68–2.86)*				
<b>Region×Urban vs. rural settings</b>					
South×rural		Reference	Reference	Reference	Reference
Central×rural		1.57 (1.18–2.09)*	1.58 (1.18–2.13)*	1.46 (1.08–1.96) <sup>†</sup>	1.26 (0.93–1.69)
North×rural		2.69 (1.95–3.72)*	2.69 (1.94–3.73)*	2.48 (1.78–3.44) <sup>†</sup>	2.08 (1.50–2.90)*
South×urban		1.28 (0.88–1.85)	1.16 (0.80–1.68)	1.13 (0.78–1.64)	1.00 (0.69–1.46)
Central×urban		1.82 (1.33–2.48)*	1.62 (1.17–2.24)*	1.46 (1.05–2.03) <sup>†</sup>	1.00 (0.72–1.40)
North×urban		1.97 (1.32–2.93)*	1.85 (1.22–2.79)*	1.70 (1.12–2.58) <sup>†</sup>	1.35 (0.88–2.05)
<b>Socio-demographic</b>					
Age (in a unit of 10 years)	2.16 (2.04–2.29)*	2.17 (2.05–2.29)*	1.79 (1.64–1.94)*	1.77 (1.62–1.93)*	1.51 (1.37–1.66)*
Women (vs. men)	0.53 (0.44–0.65)*	0.53 (0.44–0.65)*	0.46 (0.37–0.56)*	0.44 (0.35–0.55)*	0.42 (0.33–0.52)*
Race (Han vs. others)			1.04 (0.75–1.45)	1.01 (0.73–1.41)	0.94 (0.68–1.31)
Married (vs. others)			1.16 (0.90–1.49)	1.14 (0.89–1.46)	1.17 (0.91–1.50)
Employed (vs. others)			0.32 (0.24–0.42)*	0.34 (0.25–0.45)*	0.39 (0.29–0.52)*
<b>Educational level</b>					
Illiteracy			Reference	Reference	Reference
Primary school			1.07 (0.83–1.39)	1.03 (0.80–1.33)	0.97 (0.75–1.26)
Middle school			0.82 (0.61–1.10)	0.80 (0.59–1.07)	0.75 (0.56–1.02)
High school or above			0.79 (0.58–1.07)	0.74 (0.55–1.01)	0.73 (0.54–0.99) <sup>†</sup>
<b>Lifestyle</b>					
<b>BMI (kg/m<sup>2</sup>)</b>					
<18.5				Reference	Reference
18.5–24.9				1.44 (0.93–2.23)	1.15 (0.74–1.79)
24.9–29.9				1.63 (1.03–2.57) <sup>†</sup>	0.97 (0.61–1.55)
>29.9				2.69 (1.58–4.59)*	1.33 (0.77–2.31)
<b>Smoking history</b>					
Non-smoker				Reference	Reference
Smoking cessation				0.66 (0.29–1.50)	0.58 (0.25–1.34)
Current smoker				0.49 (0.27–0.89) <sup>†</sup>	0.54 (0.30–1.00) <sup>†</sup>
<b>Disease history</b>					
Hypertension (yes vs. no)					10.96 (6.44–18.65)*
Diabetes (yes vs. no)					1.40 (1.01–1.94) <sup>†</sup>
Myocardial infarction (yes vs. no)					2.48 (1.68–3.68)*
<b>Time-varying covariates</b>					
<b>Smoking history</b>					
Non-smoker				Reference	Reference
Smoking cessation				1.08 (1.01–1.14) <sup>†</sup>	1.08 (1.01–1.15) <sup>†</sup>
Current smoker				1.04 (0.99–1.09)	1.03 (0.98–1.08)
Hypertension (yes vs. no)					0.94 (0.91–0.98) <sup>†</sup>



**Table 4.** Continued

Independent variable	Model 1	Model 2	Model 3	Model 4	Model 5
No. individuals	16,917	16,917	16,835	16,773	16,700
No. stroke case	442	442	440	439	434

HR, hazard ratio; CI, confidence interval; BMI, body mass index.

\* $P < 0.001$ ; † $P < 0.01$ ; ‡ $P < 0.05$ .

parts in the south region. By comparing 95% CIs of the six combinations between regions and urban vs. rural settings, we found that the north-to-south differences did not exist in urban areas (Model 2 to 5). Results here were in line with those in Table 3.

In rural areas, the differences between the north and south regions remained consistent even in the fully adjusted model (Model 5, HR, 2.08; 95% CI, 1.50 to 2.90). In contrast, results from Models 4 and 5 indicated that in rural areas, the differences between the central and south regions (Model 4 [HR, 1.46; 95% CI, 1.08 to 1.96], Model 5 [HR, 1.26; 95% CI, 0.93 to 1.69]) could be fully explained by the disparities in the prevalence of hypertension and myocardial infarction. Having hypertension (Model 5, main effect, HR, 10.96; 95% CI, 6.44 to 18.65) and myocardial infarction (HR, 2.48, 95% CI, 1.68 to 3.68) were positively associated with incident stroke, although the effect of hypertension appeared to decrease over time (interaction effect, HR, 0.94; 95% CI, 0.91 to 0.98). Further analyses suggested that the disparities between the central and south regions were fully explained by the disparities in hypertension but not myocardial infarction (results not presented here).

Subgroup analyses by sex (Table 5) shared the same hierarchical modelling process with pooled analyses. Consistent with the results in Table 3 and Figure 2, the differences between the central and south regions were only observed in men. The differences could be fully explained by the disparities in disease history, which were in line with the results in Table 4. Additionally, the north-to-south disparities among women were not statistically significant once the disease history was controlled. Likewise, further analyses among women suggested that the geographic variations were fully explained by the disparities in hypertension but not myocardial infarction (results not presented here).

### Sensitivity analyses

Rate ratios from Poisson regression, odds ratios from logistics regression, and Mantel-Haenszel rate ratios were qualitatively similar to the age-aRRs from log-binomial regression. Results from sensitivity analyses that used dataset before imputation were consistent with those from our main analyses. Last, results from sensitivity analyses restricted to individuals with at least middle school education and aged under 65 years old

(Supplementary Tables 2 and 3) were similar to the findings from the primary analyses.

## Discussion

By employing a community-based cohort with 16,917 individuals, the present study extends the prior understanding of the regional disparities in stroke burden in China by focusing on the stroke incidence measure.<sup>4,6,7</sup> We found the north-to-south gradient only existed in rural areas but not in urban areas. In addition, our hierarchical modeling analyses based on a prospective cohort study-design indicated that the disparities in the prevalence of hypertension might account for the regional disparities. These findings help guide nationwide and region-specific strategies for stroke prevention in China.

### Comparison with existing research

Prior studies provided insufficient evidence on the interaction between the geographic gradient and urban-rural settings, even though the cross-sectional study based on a National Epidemiological Survey of Stroke in China (NESS-China) indicated a general north-to-south geographical gradient in stroke prevalence in conjunction with the heavy disease burdens in rural areas.<sup>7</sup> As urbanization has been accelerated over the past few decades, the regional differences among urban areas may have diminished over time.<sup>12</sup> Therefore, more attention should be directed to geographic disparities in rural areas when considering the priority of public health intervention.

Among rural areas between the north and south regions, we found that the disparities remained consistent even after controlling for potential stroke risk factors, including hypertension and myocardial infarction. In contrast to our findings, a higher stroke incidence was found in the southeastern United States, which has been widely acknowledged as 'the Stroke Belt' due to higher mortality of stroke in comparison with other regions.<sup>18-20</sup> Prior studies also indicated "the Stroke Belt" was a result of a significant number of rural residents, African American, residents with a higher prevalence of traditional stroke risk factors, inflammation and infection, as well as socio-economically deprived individuals.<sup>18,19,21-23</sup> However, it should be noted that findings from western nations may not be general-



**Table 5.** Subgroup extended Cox proportional hazards analyses of stroke occurrence by sex, HR (95% CI)

Independent variable	Model 1	Model 2	Model 3	Model 4	Model 5
<b>Men</b>					
Urban (vs. rural)	0.95 (0.74 to1.23)				
<b>Region</b>					
South	Reference				
Central	1.59 (1.19–2.13)*				
North	2.37 (1.70–3.31)*				
<b>Region×Urban vs. rural settings</b>					
South×rural		Reference	Reference	Reference	Reference
Central×rural		1.69 (1.17–2.43) <sup>†</sup>	1.73 (1.19–2.53)*	1.58 (1.08–2.32) <sup>†</sup>	1.41 (0.96–2.07)
North×rural		2.94 (1.96–4.39)*	2.92 (1.95–4.38)*	2.67 (1.77–4.05)*	2.53 (1.68–3.81)*
South×urban		1.22 (0.75 to1.98)	1.09 (0.67–1.77)	1.04 (0.63–1.70)	1.03 (0.63–1.67)
Central×urban		1.73 (1.14–2.60) <sup>†</sup>	1.50 (0.98–2.30)	1.32 (0.85–2.04)	0.96 (0.62–1.49)
North×urban		1.93 (1.16–3.21) <sup>†</sup>	1.75 (1.03–2.97) <sup>†</sup>	1.56 (0.92–2.64)	1.28 (0.75–2.19)
<b>Women</b>					
Urban (vs. rural)	1.26 (0.92–1.71)				
<b>Region</b>					
South	Reference				
Central	1.41 (0.99–2.01)				
North	1.91 (1.23–2.96)*				
<b>Region×Urban vs. rural settings</b>					
South×rural		Reference	Reference	Reference	Reference
Central×rural		1.39 (0.88–2.21)	1.36 (0.85–2.17)	1.28 (0.79–2.06)	1.05 (0.66–1.67)
North×rural		2.28 (1.31–3.95)*	2.28 (1.31–3.98)*	2.21 (1.28–3.82) <sup>†</sup>	1.55 (0.87–2.74)
South×urban		1.36 (0.77–2.42)	1.28 (0.72–2.25)	1.27 (0.72–2.24)	0.97 (0.54–1.74)
Central×urban		1.94 (1.20–3.14) <sup>†</sup>	1.79 (1.10–2.92) <sup>†</sup>	1.67 (1.02–2.75) <sup>†</sup>	1.09 (0.66–1.80)
North×urban		2.03 (1.07–3.85) <sup>†</sup>	2.07 (1.08–3.98) <sup>†</sup>	2.00 (1.01–3.97) <sup>†</sup>	1.45 (0.73–2.88)

Model 1 and Model 2 adjusted for age; Model 3 adjusted for socio-economic covariates; Model 4 adjusted for socio-economic and lifestyle covariates. Model 5 adjusted for socio-economic and lifestyle covariates as well as disease history. Details were presented in Supplementary Tables 4 and 5.

HR, hazard ratio; CI, confidence interval.

\* $P < 0.001$ ; <sup>†</sup> $P < 0.01$ ; <sup>‡</sup> $P < 0.05$ .

ized to eastern countries since culture and social structure differ substantially. For instance, in our study, racial compositions (e.g., Han vs. others) were not associated with stroke incidence. Additionally, in comparison with those in the south region, individuals in high-risk north and central regions had a higher educational level, which, to some extent, represented higher socio-economic status (Table 1). These factors may explain the discrepancy between our results and the 'Stroke Belt' results in the United States. We further hypothesize that the disparities in the present study could be somewhat attributed to differences in alcohol consumption and dietary pattern. Earlier Chinese studies showed that in comparison with those in other regions, individuals in the northeastern region had a higher intake of alcohol and salt, which may be associated with an increased risk of stroke incidence in the northeastern region.<sup>7,24–26</sup>

Among rural areas, we found that the disparities of stroke incidence between the central and south regions could be explained by disparities in hypertension. Our findings here were supported by an existing study, results of which indicated that the variations in stroke incidence were mainly yielded from the differences in the prevalence of hypertension.<sup>6</sup> Hypertension is well recognized as the most important and modifiable stroke risk factor.<sup>21</sup> In addition, we observed an interesting interaction between time and hypertension, indicating that the risk of hypertension declined during the study period. This finding appeared to suggest an improvement of hypertension management among the Chinese population, while subgroup analyses by sex (Supplementary Tables 4 and 5) indicated that the improvement only existed among women. Although public awareness and standardized treatment have improved quickly

in China, current evidence has shown that the proportion of controlled hypertension (<20%) is still lower in comparison with developed countries.<sup>5,27</sup> Besides, many studies have pointed out that awareness, treatment, and control of hypertension are lower in rural areas,<sup>5-7,28</sup> which suggests that the improvement in hypertension management in rural areas holds promise for stroke prevention in the future.

Our study confirmed that myocardial infarction history was significantly associated with incident stroke.<sup>4-7,13</sup> Notably, both acute myocardial infarction (e.g., left ventricular thrombus in the specific setting of an acute myocardial infarction) and chronic myocardial infarction together with reduced ejection fraction are considered to be high-risk cardiac sources of embolic strokes,<sup>29-31</sup> which may account for the association between myocardial infarction and stroke in our study. In addition to management of hypertension as discussed above, coronary heart diseases, e.g., myocardial infarction, were poorly managed in China as well.<sup>5,7</sup> The burdens were also observed by studies from other countries.<sup>4,21</sup> As such, one should direct more attention among patients with myocardial infarction. Intervention strategies like anticoagulant therapy should be timely and appropriately applied for cardioembolic stroke prevention practice.<sup>32</sup>

### Strengths and limitations

Based on a prospective cohort study–design, the present study depicted the north-to-south gradient in stroke incidence across urban and rural China and provided robust estimates of the attributable risk factors that could account for regional disparities. The observed variations and potential attributable risk factors may provide clear and actionable implications for region-specific resources allocation.

Nonetheless, this study is subject to several limitations. First, our study did not identify the subtypes of stroke, and therefore our findings could not accurately reflect the conditions of ischemic and hemorrhagic stroke, respectively. Second, although we included individuals from China's diverse communities, our findings could not reflect the general conditions among Chinese as CHNS did not include all provinces in China. Particularly, our findings could not reflect the conditions of provinces such as Inner Mongolia, Xinjiang, and Tibet, where the geographic contexts, culture, and socio-economic development differed from the provinces enrolled in the present study.

Meanwhile, non-ignorable bias exists because the stroke record was not missing completely at random. However, the small differences in missing rates across regions (<5%), would not substantially affect our main findings. Fourth, as a community-based survey, CHNS did not include institutionalized

individuals, which not only diminished the representation of our study but also resulted in the potential underestimation of the incidence of stroke. Furthermore, self-reported outcome measures may lead to recall bias and could be affected by individuals' educational level and disease awareness, while our sensitivity analyses restricted to younger individuals and those with a higher educational level accounted for the bias to some extent (Supplementary Tables 2 and 3).

Last, the present study is subject to the attrition bias as we are unable to analyze the deaths by deriving data from CHNS. In this regard, we may underestimate the incidence of stroke, particularly in underserved areas. Due to residents' low health literacy and limited access to health care as well as untimely treatment, the mortality-to-incidence ratio of the first-ever stroke in underserved areas could be higher than that in economically developed areas.<sup>5,33</sup> However, by employing the national health claim data, the NESS-China study found that the mortality-to-incidence ratios were generally consistent in the vast majority of regions (ranging from 0.42 to 0.47), except the southwest region (0.68), where most of the provinces were not included in the present study.<sup>5,7</sup> Taken into account the small regional differences in mortality-to-incidence ratios, the attrition bias derived from deaths would not significantly affect the observed regional disparities in stroke incidence.

### Conclusions

The present 18-year period prospective cohort study extends the current literature of the north-to-south gradient in stroke burden by focusing on the incidence measure. Higher risks were observed among rural residents in the north region in comparison with their counterparts in the south region. Focusing on the management of hypertension could greatly alleviate regional differences.

### Supplementary materials

Supplementary materials related to this article can be found online at <https://doi.org/10.5853/jos.2020.02383>.

### Disclosure

The authors have no financial conflicts of interest.

### Acknowledgments

This work is supported by the 1·3·5 project for disciplines of excellence–Clinical Research Incubation Project, West China Hos-

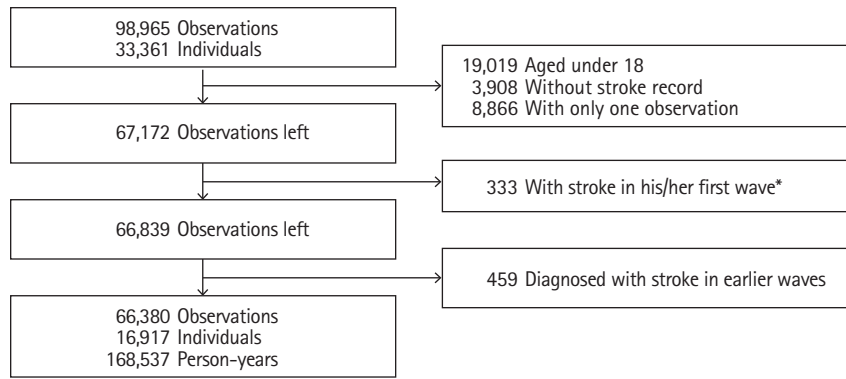
pital, Sichuan University (grant number ZY2016102), and the project of the Sichuan Science and Technology Bureau (grant number 2017SZ0142).

We would like to thank Dr. Yi Li and Dr. Jian Kang from the Department of Biostatistics, School of Public Health, the University of Michigan, for suggestions in data analyses. We thank the China Center for Disease Control and Prevention and Carolina Population Center at the University of North Carolina at Chapel Hill for providing the CHNS data.

## References

- Zhou M, Wang H, Zeng X, Yin P, Zhu J, Chen W, et al. Mortality, morbidity, and risk factors in China and its provinces, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2019;394:1145–1158.
- Boot E, Ekker MS, Putaala J, Kittner S, De Leeuw FE, Tuladhar AM. Ischaemic stroke in young adults: a global perspective. *J Neurol Neurosurg Psychiatry* 2020;91:411–417.
- Venketasubramanian N, Yoon BW, Pandian J, Navarro JC. Stroke epidemiology in South, East, and South-East Asia: a review. *J Stroke* 2017;19:286–294.
- Feigin VL, Forouzanfar MH, Krishnamurthi R, Mensah GA, Connor M, Bennett DA, et al. Global and regional burden of stroke during 1990–2010: findings from the Global Burden of Disease Study 2010. *Lancet* 2014;383:245–254.
- Wu S, Wu B, Liu M, Chen Z, Wang W, Anderson CS, et al. Stroke in China: advances and challenges in epidemiology, prevention, and management. *Lancet Neurol* 2019;18:394–405.
- He J, Klag MJ, Wu Z, Whelton PK. Stroke in the People's Republic of China. I. Geographic variations in incidence and risk factors. *Stroke* 1995;26:2222–2227.
- Wang W, Jiang B, Sun H, Ru X, Sun D, Wang L, et al. Prevalence, incidence, and mortality of stroke in China: results from a nationwide population-based survey of 480687 adults. *Circulation* 2017;135:759–771.
- Geiss LS, Wang J, Cheng YJ, Thompson TJ, Barker L, Li Y, et al. Prevalence and incidence trends for diagnosed diabetes among adults aged 20 to 79 years, United States, 1980–2012. *JAMA* 2014;312:1218–1226.
- Magliano DJ, Islam RM, Barr ELM, Gregg EW, Pavkov ME, Harding JL, et al. Trends in incidence of total or type 2 diabetes: systematic review. *BMJ* 2019;366:l5003.
- Life expectancy at birth, total (years): China. The World Bank. <https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=CN>. Accessed August 17, 2020.
- Population ages 65 and above (% of total population): China. The World Bank. <https://data.worldbank.org/indicator/SP.POP65UP.TO.ZS?locations=CN>. Accessed August 17, 2020.
- Yu X, Zhang W. All-cause mortality rate in China: do residents in economically developed regions have better health? *Int J Equity Health* 2020;19:12.
- Guan T, Ma J, Li M, Xue T, Lan Z, Guo J, et al. Rapid transitions in the epidemiology of stroke and its risk factors in China from 2002 to 2013. *Neurology* 2017;89:53–61.
- Popkin BM, Du S, Zhai F, Zhang B. Cohort profile: The China Health and Nutrition Survey: monitoring and understanding socio-economic and health change in China, 1989–2011. *Int J Epidemiol* 2010;39:1435–1440.
- Ferriero DM, Fullerton HJ, Bernard TJ, Billinghamurst L, Daniels SR, DeBaun MR, et al. Management of stroke in neonates and children: a scientific statement from the American Heart Association/American Stroke Association. *Stroke* 2019;50:e51–e96.
- Wang CC, Cheng XM, Li SC, Wang WZ, Wu SP, Wang KX, et al. Epidemiological survey of neurological disorders in six urban areas of the People's Republic of China [In Chinese]. *Chin J Neurosurg* 1985;1:2–7.
- McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol* 2003;157:940–943.
- Howard G, Howard VJ. Twenty years of progress toward understanding the stroke belt. *Stroke* 2020;51:742–750.
- Lanska DJ, Kuller LH. The geography of stroke mortality in the United States and the concept of a stroke belt. *Stroke* 1995;26:1145–1149.
- Labarthe DR, Howard G, Safford MM, Howard VJ, Judd SE, Cushman M, et al. Incidence and case fatality at the county level as contributors to geographic disparities in stroke mortality. *Neuroepidemiology* 2016;47:96–102.
- Koton S, Schneider AL, Rosamond WD, Shahar E, Sang Y, Gottesman RF, et al. Stroke incidence and mortality trends in US communities, 1987 to 2011. *JAMA* 2014;312:259–268.
- El-Saed A, Kuller LH, Newman AB, Lopez O, Costantino J, McTigue K, et al. Geographic variations in stroke incidence and mortality among older populations in four US communities. *Stroke* 2006;37:1975–1979.
- Howard VJ, Cushman M, Pulley L, Gomez CR, Go RC, Prineas RJ, et al. The reasons for geographic and racial differences in stroke study: objectives and design. *Neuroepidemiology* 2005;25:135–143.
- Anderson CA, Appel LJ, Okuda N, Brown IJ, Chan Q, Zhao L, et al. Dietary sources of sodium in China, Japan, the United Kingdom, and the United States, women and men aged 40 to 59 years: the INTERMAP study. *J Am Diet Assoc* 2010;110:736–

- 745.
25. Li Y, He Y, Lai J, Wang D, Zhang J, Fu P, et al. Dietary patterns are associated with stroke in Chinese adults. *J Nutr* 2011;141:1834-1839.
  26. Xing L, Jing L, Tian Y, Wang W, Sun J, Jiang C, et al. Epidemiology of stroke in urban northeast China: a population-based study 2018-2019. *Int J Stroke* 2020 Jan 20 [Epub]. <https://doi.org/10.1177/1747493019897841>.
  27. Kim YD, Jung YH, Saposnik G. Traditional risk factors for stroke in East Asia. *J Stroke* 2016;18:273-285.
  28. Xing L, Jing L, Tian Y, Liu S, Lin M, Du Z, et al. High prevalence of stroke and uncontrolled associated risk factors are major public health challenges in rural northeast China: a population-based study. *Int J Stroke* 2020;15:399-411.
  29. Ay H, Furie KL, Singhal A, Smith WS, Sorensen AG, Koroshetz WJ. An evidence-based causative classification system for acute ischemic stroke. *Ann Neurol* 2005;58:688-697.
  30. Doufekias E, Segal AZ, Kizer JR. Cardiogenic and aortogenic brain embolism. *J Am Coll Cardiol* 2008;51:1049-1059.
  31. Arsava EM, Ballabio E, Benner T, Cole JW, Delgado-Martinez MP, Dichgans M, et al. The causative classification of stroke system: an international reliability and optimization study. *Neurology* 2010;75:1277-1284.
  32. Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, et al. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2014;45:2160-2236.
  33. GBD 2016 Healthcare Access and Quality Collaborators. Measuring performance on the Healthcare Access and Quality Index for 195 countries and territories and selected subnational locations: a systematic analysis from the Global Burden of Disease Study 2016. *Lancet* 2018;391:2236-2271.



**Supplementary Figure 1.** Study design and sample collection. \*Three hundred and thirty-three observations from 118 individuals.

**Supplementary Table 1.** Crude and age-standardized incidence by provinces, per 1,000 person-years

Region	Province	Person-year	Stroke case	Crude incidence	Age-standardized incidence (95% CI)
<b>Overall</b>					
South	Guangxi	20,900	46	2.20	2.20 (1.57–2.84)
	Guizhou	19,438	28	1.44	1.32 (0.83–1.81)
	Hunan	17,790	40	2.25	2.29 (1.58–3.00)
Central	Chongqing	3,004	5	1.66	1.39 (0.09–2.69)
	Hubei	18,510	48	2.59	2.82 (2.01–3.64)
	Henan	17,796	63	3.54	3.72 (2.80–4.65)
	Shanghai	3,316	9	2.71	2.23 (0.70–3.75)
	Jiangsu	18,983	44	2.32	2.16 (1.51–2.80)
North	Shandong	16,937	48	2.83	2.72 (1.95–3.49)
	Beijing	3,156	9	2.85	4.12 (1.36–6.89)
	Liaoning	13,300	61	4.59	4.67 (3.48–5.86)
	Heilongjiang	15,914	41	2.58	3.13 (2.11–4.16)
<b>Men</b>					
South	Guangxi	10,666	27	2.53	2.53 (1.58–3.48)
	Guizhou	9,656	18	1.86	1.70 (0.91–2.49)
	Hunan	8,709	22	2.53	2.42 (1.41–3.43)
	Chongqing	1,384	3	2.17	1.93 (0.00–4.12)
Central	Hubei	9,359	26	2.78	2.91 (1.78–4.03)
	Henan	8,774	42	4.79	4.84 (3.37–6.30)
	Shanghai	1,560	5	3.21	2.14 (0.21–4.07)
	Jiangsu	9,203	27	2.93	2.44 (1.51–3.36)
North	Shandong	8,177	28	3.42	3.27 (2.06–4.48)
	Beijing	1,492	6	4.02	4.70 (0.95–8.45)
	Liaoning	6,519	39	5.98	5.43 (3.72–7.15)
	Heilongjiang	7,810	29	3.71	4.56 (2.83–6.29)
<b>Women</b>					
South	Guangxi	10,234	19	1.86	1.86 (1.02–2.69)
	Guizhou	9,782	10	1.02	0.94 (0.35–1.53)
	Hunan	9,081	18	1.98	2.23 (1.18–3.28)
	Chongqing	1,620	2	1.24	0.64 (0.00–1.51)
Central	Hubei	9,151	22	2.40	2.69 (1.53–3.85)
	Henan	9,022	21	2.33	2.58 (1.46–3.71)
	Shanghai	1,756	4	2.28	1.94 (0.00–3.97)
	Jiangsu	9,780	17	1.74	1.89 (0.98–2.81)
North	Shandong	8,760	20	2.28	2.20 (1.23–3.18)
	Beijing	1,664	3	1.80	3.20 (0.00–7.03)
	Liaoning	6,781	22	3.24	3.88 (2.20–5.56)
	Heilongjiang	8,104	12	1.48	1.53 (0.66–2.41)

Age-standardized incidence was derived by using study sample in Guangxi province as the standard population. CI, confidence interval.



**Supplementary Table 2.** Extended Cox proportional hazards analyses among individuals with at least middle school education, HR (95% CI)

Independent variable	Model 1	Model 2	Model 3	Model 4	Model 5
Urban (vs. rural)	0.99 (0.72–1.37)				
Region					
South	Reference				
Central	1.45 (0.97–2.18)				
North	2.13 (1.38–3.29)*				
Region×urban vs. rural settings					
South×rural		Reference	Reference	Reference	Reference
Central×rural		1.93 (1.08–3.44) <sup>†</sup>	1.83 (1.02–3.30) <sup>†</sup>	1.63 (0.90–2.94)	1.49 (0.85–2.64)
North×rural		2.72 (1.47–5.05) <sup>*</sup>	2.75 (1.48–5.09) <sup>†</sup>	2.35 (1.26–4.37) <sup>†</sup>	2.19 (1.19–4.06) <sup>†</sup>
South×urban		1.52 (0.78–2.98)	1.33 (0.68–2.62)	1.23 (0.62–2.45)	1.16 (0.59–2.29)
Central×urban		1.64 (0.90–2.99)	1.45 (0.79–2.68)	1.27 (0.69–2.35)	0.97 (0.53–1.76)
North×urban		2.53 (1.36–4.72) <sup>*</sup>	2.36 (1.24–4.47) <sup>†</sup>	2.04 (1.08–3.87) <sup>*</sup>	1.65 (0.87–3.11)

Model 1 and Model 2 adjusted for age; Model 3 adjusted for socio-economic covariates; Model 4 adjusted for socio-economic and lifestyle covariates; Model 5 adjusted for socio-economic and lifestyle covariates as well as disease history.

HR, hazard ratio; CI, confidence interval.

\* $P < 0.001$ ; <sup>†</sup> $P < 0.01$ ; <sup>‡</sup> $P < 0.05$ .

**Supplementary Table 3.** Extended Cox proportional hazards analyses among individuals aged under 65, HR (95% CI)

Independent variable	Model 1	Model 2	Model 3	Model 4	Model 5
Urban (vs. rural)	1.07 (0.80–1.45)				
Region					
South	Reference				
Central	1.36 (0.96–1.91)				
North	1.93 (1.31–2.84) <sup>*</sup>				
Region×urban vs. rural settings					
South×rural		Reference	Reference	Reference	Reference
Central×rural		1.56 (1.00–2.45) <sup>†</sup>	1.65 (1.04–2.62) <sup>†</sup>	1.54 (0.97–2.45)	1.35 (0.86–2.14)
North×rural		2.67 (1.65–4.33) <sup>*</sup>	2.87 (1.76–4.66) <sup>*</sup>	2.61 (1.61–4.23) <sup>*</sup>	2.04 (1.24–3.35) <sup>†</sup>
South×urban		1.62 (0.93–2.83)	1.42 (0.81–2.48)	1.40 (0.80–2.46)	1.33 (0.75–2.33)
Central×urban		1.77 (1.06–2.97) <sup>†</sup>	1.65 (0.97–2.81)	1.51 (0.88–2.57)	1.11 (0.65–1.88)
North×urban		1.71 (0.90–3.26)	1.87 (0.98–3.59)	1.68 (0.87–3.21)	1.26 (0.62–2.53)

Model 1 and Model 2 adjusted for age; Model 3 adjusted for socio-economic covariates; Model 4 adjusted for socio-economic and lifestyle covariates; Model 5 adjusted for socio-economic and lifestyle covariates as well as disease history.

HR, hazard ratio; CI, confidence interval.

\* $P < 0.001$ ; <sup>†</sup> $P < 0.01$ ; <sup>‡</sup> $P < 0.05$ .

**Supplementary Table 4.** Extended Cox proportional hazards analyses of stroke occurrence among men, HR (95% CI)

Independent variable	Model 1	Model 2	Model 3	Model 4	Model 5
Urban (vs. rural)	0.95 (0.74–1.23)				
<b>Region</b>					
South	Reference				
Central	1.59 (1.19–2.13)*				
North	2.37 (1.70–3.31)*				
<b>Region×urban vs. rural settings</b>					
South×rural		Reference	Reference	Reference	Reference
Central×rural		1.69 (1.17–2.43) <sup>†</sup>	1.73 (1.19–2.53)*	1.58 (1.08–2.32) <sup>†</sup>	1.41 (0.96–2.07)
North×rural		2.94 (1.96–4.39)*	2.92 (1.95–4.38)*	2.67 (1.77–4.05)*	2.53 (1.68–3.81)*
South×urban		1.22 (0.75 to 1.98)	1.09 (0.67–1.77)	1.04 (0.63–1.70)	1.03 (0.63–1.67)
Central×urban		1.73 (1.14–2.60) <sup>†</sup>	1.50 (0.98–2.30)	1.32 (0.85–2.04)	0.96 (0.62–1.49)
North×urban		1.93 (1.16–3.21) <sup>†</sup>	1.75 (1.03–2.97) <sup>†</sup>	1.56 (0.92–2.64)	1.28 (0.75–2.19)
<b>Socio-demographic</b>					
Age (in a unit of 10 years)	2.18 (2.01–2.35)*	2.18 (2.02–2.35)*	1.77 (1.60–1.97)*	1.71 (1.53–1.91)*	1.49 (1.32–1.68)*
Race (Han vs. others)			0.99 (0.65–1.50)	0.94 (0.61–1.43)	0.85 (0.57–1.29)
Married (vs. others)			1.50 (1.02–2.20) <sup>†</sup>	1.50 (1.02–2.20) <sup>†</sup>	1.56 (1.05–2.32) <sup>†</sup>
Employed (vs. others)			0.32 (0.23–0.46)*	0.34 (0.24–0.48)*	0.38 (0.27–0.55)*
<b>Educational level</b>					
Illiteracy			Reference	Reference	Reference
Primary school			1.20 (0.86–1.68)	1.15 (0.82–1.60)	1.05 (0.75–1.48)
Middle school			0.86 (0.59–1.24)	0.82 (0.57–1.17)	0.75 (0.52–1.08)
High school or above			0.91 (0.62–1.32)	0.84 (0.58–1.23)	0.79 (0.55–1.15)
<b>Lifestyle</b>					
<b>BMI (kg/m<sup>2</sup>)</b>					
<18.5				Reference	Reference
18.5–24.9				1.51 (0.85–2.68)	1.28 (0.71–2.32)
24.9–29.9				1.66 (0.91–3.01)	1.07 (0.58–1.99)
>29.9				2.19 (1.04–4.63) <sup>†</sup>	1.23 (0.56–2.70)
<b>Smoking history</b>					
Non-smoker				Reference	Reference
Smoking cessation				1.71 (1.25–2.35)*	1.57 (1.14–2.16) <sup>†</sup>
Current smoker				0.74 (0.56–0.98) <sup>†</sup>	0.75 (0.56–0.99) <sup>†</sup>
<b>Disease history</b>					
Hypertension (yes vs. no)					4.85 (3.66–6.43)*
Diabetes (yes vs. no)					1.39 (0.90–2.15)
Myocardial infarction (yes vs. no)					2.33 (1.41–3.86)*

No covariates interacted with time in the model.  
 HR, hazard ratio; CI, confidence interval; BMI, body mass index.  
 \* $P < 0.001$ ; <sup>†</sup> $P < 0.01$ ; <sup>‡</sup> $P < 0.05$ .

**Supplementary Table 5.** Extended Cox proportional hazards analyses of stroke occurrence among women, HR (95% CI)

Independent variable	Model 1	Model 2	Model 3	Model 4	Model 5
Urban (vs. rural)	1.26 (0.92–1.71)				
<b>Region</b>					
South	Reference				
Central	1.41 (0.99–2.01)				
North	1.91 (1.23–2.96)*				
<b>Region×urban vs. rural settings</b>					
South×rural		Reference	Reference	Reference	Reference
Central×rural		1.39 (0.88–2.21)	1.36 (0.85–2.17)	1.28 (0.79–2.06)	1.05 (0.66–1.67)
North×rural		2.28 (1.31–3.95)*	2.28 (1.31–3.98)*	2.21 (1.28–3.82) <sup>†</sup>	1.55 (0.87–2.74)
South×urban		1.36 (0.77–2.42)	1.28 (0.72–2.25)	1.27 (0.72–2.24)	0.97 (0.54–1.74)
Central×urban		1.94 (1.20–3.14) <sup>†</sup>	1.79 (1.10–2.92) <sup>†</sup>	1.67 (1.02–2.75) <sup>†</sup>	1.09 (0.66–1.80)
North×urban		2.03 (1.07–3.85) <sup>†</sup>	2.07 (1.08–3.98) <sup>†</sup>	2.00 (1.01–3.97) <sup>†</sup>	1.45 (0.73–2.88)
<b>Socio-demographic</b>					
Age (in a unit of 10 years)	2.14 (1.96–2.33)*	2.14 (1.97–2.33)*	1.74 (1.49–2.02)*	1.79 (1.53–2.10)*	1.48 (1.24–1.77)*
Race (Han vs. others)			1.15 (0.66–2.02)	1.16 (0.67–2.03)	1.09 (0.62–1.92)
Married (vs. others)			0.97 (0.66–1.43)	0.96 (0.65–1.40)	0.96 (0.66–1.40)
Employed (vs. others)			0.27 (0.15–0.47)*	0.29 (0.16–0.50)*	0.34 (0.19–0.59)*
<b>Educational level</b>					
Illiteracy			Reference	Reference	Reference
Primary school			0.93 (0.60–1.45)	0.90 (0.58–1.40)	0.88 (0.57–1.36)
Middle school			0.81 (0.48–1.38)	0.83 (0.49–1.41)	0.84 (0.50–1.41)
High school or above			0.57 (0.30–1.08)	0.58 (0.31–1.10)	0.61 (0.33–1.13)
<b>Lifestyle</b>					
BMI (kg/m <sup>2</sup> )				Reference	Reference
<18.5				1.36 (0.69–2.67)	0.99 (0.51–1.94)
18.5–24.9				1.60 (0.78–3.28)	0.87 (0.42 to 1.78)
24.9–29.9				3.33 (1.52–7.32)*	1.38 (0.63–3.02)
Current smoker (vs. others) <sup>§</sup>				0.72 (0.37–1.42)	0.84 (0.44–1.63)
Hypertension (yes vs. no)					17.63 (7.27–42.80)*
Diabetes (yes vs. no)					1.30 (0.80–2.13)
Myocardial infarction (yes vs. no)					2.83 (1.53–5.21)*
<b>Time-varying covariates</b>					
Hypertension (yes vs. no)					0.92 (0.86–0.99) <sup>†</sup>

HR, hazard ratio; CI, confidence interval; BMI, body mass index.

\* $P<0.001$ ; <sup>†</sup> $P<0.01$ ; <sup>‡</sup> $P<0.05$ ; <sup>§</sup>Non-smoker and smoking cessation were combined due to limited number of individuals with smoking cessation.