Supplemental material

Expanded Materials & Methods:

Section A. Case identification and ascertainment;

Section B. Validation study;

Section C. Strategy used in the estimation of numerator.

Online Table S1 – S5

Sun Y, *et al. Stroke Vasc Neurol* 2021;0:1–9. doi: 10.1136/svn-2021-000909

Section A. Case identification and ascertainment

In the first step, diagnostic codes and diagnostic texts related to moyamoya disease (MMD) were extracted from the medical insurance databases. To avoid missing patients when using the medical terms in Chinese, we constructed a relatively loose algorithm using fuzzy string matching to extract all potential MMD patients in the database. Keywords were defined according to ICD-10 code (I67.5) and medical terms in Chinese and English, shown in **Table A1**. The search algorithm is therefore defined as:

**(I67.5 OR I67.500 OR I67.501).* OR .*(烟雾病).* OR .*(自发性脑底/基底/颅底动脉闭塞).* OR .* (脑底/颅底异常血管网).* OR .*(烟雾状/样血管).* OR .*(Moyamoya).* OR .*(MMD).*

Using the searching algorithm, we extracted 58,373 records of 2,262 patients from UEBMI and URBMI databases. Each record contained a diagnostic expression combination, which was combined by six variables of diagnostic text or code (i.e., primary diagnosis, secondary diagnosis 1, secondary diagnosis 2) and separated by commas.

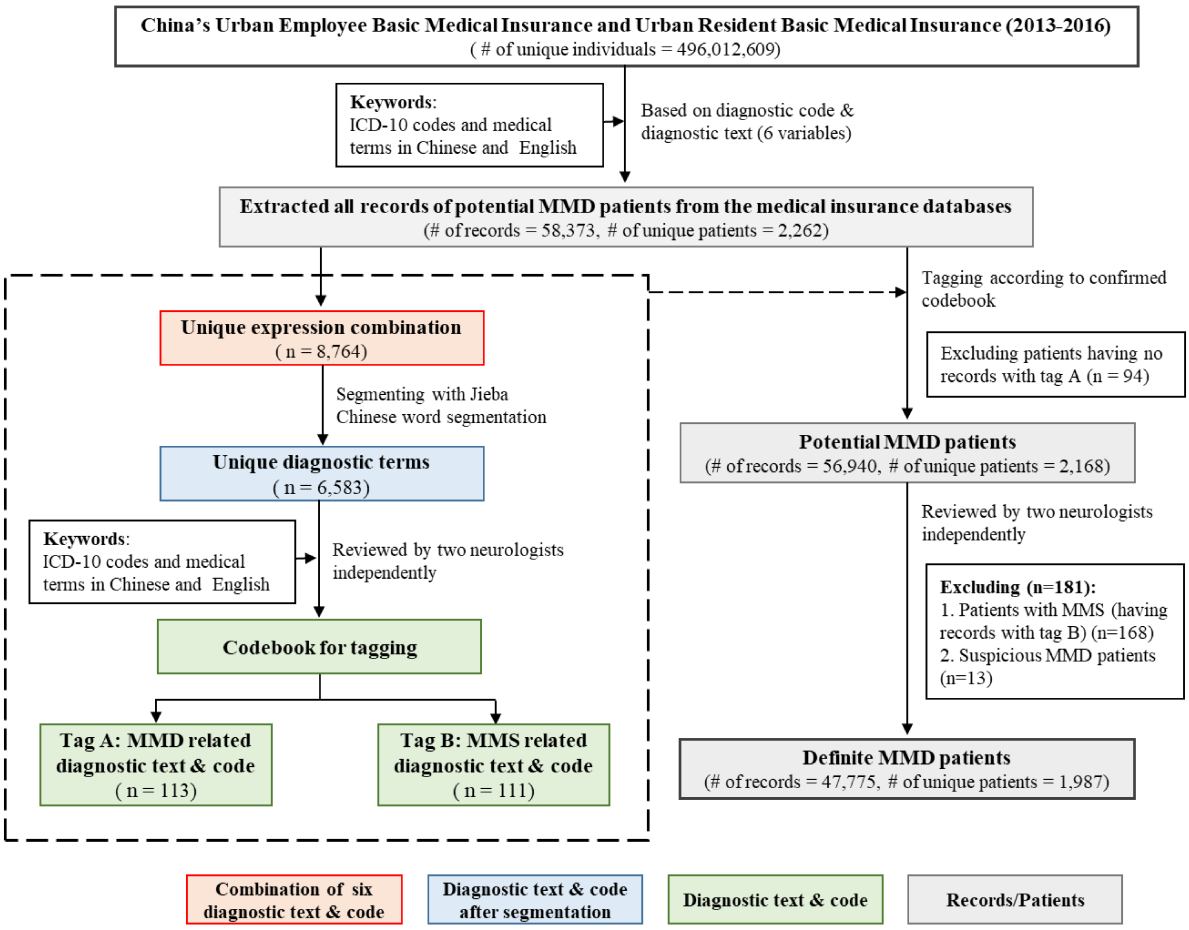
Table A1. Keywords used to search for extracting MMD cases.

Diagnostic text	Corresponding English translation	ICD-10
烟雾病/症	Moyamoya disease	I67.5
烟雾氏病	Moyamoya disease	I67.5
云雾病	Moyamoya disease	I67.5
自发性脑底动脉闭塞	Spontaneous occlusion of the circle of Willis	I67.5
自发性基底动脉闭塞	Spontaneous occlusion of the circle of Willis	I67.5
自发性颅底动脉闭塞	Spontaneous occlusion of the circle of Willis	I67.5
脑底异常血管网	Abnormal vascular network at the base of the brain	I67.5
颅底异常血管网	Abnormal vascular network at the base of the brain	I67.5
烟雾状/样血管	Moyamoya vessels	I67.5
Moyamoya	Moyamoya disease	I67.5
MMD	Moyamoya disease	I67.5

In the second step, two neurologists performed manual verification of the diagnostic codes or texts with the assistance of the computer algorithm. The flow-chart for the case ascertainment was shown in **Figure A1.** In the verification stage, first of all, a total of 8,764 unique diagnostic expression combinations (i.e., six variables of diagnostic text or code were combined and separated by commas) were obtained, from the 58,373 records extracted by the above algorithm. These expressions were then segmented using Jieba Chinese word segmentation tool to gain unique diagnostic terms, according to a clinical terminology database. Overall, 6,583 terms were obtained. Then, these diagnostic terms were reviewed by two neurologists independently to create a codebook of MMD related terms for tagging. Diagnostic terms associated with MMD, including various Chinese transliteration names of MMD as well as their non-standard writing, were labeled as "Tag A". And diagnostic terms associated with moyamoya syndrome (MMS) were labeled as "Tag B", including arteriosclerosis, autoimmune disease (systemic lupus erythematosus, sicca syndrome, antiphospholipid antibody syndrome, arteritis nodosa, etc.), meningitis, brain neoplasm, von Recklinghausen disease, Down syndrome, head trauma, irradiation to the head, hyperthyroidism, Turner syndrome, Alagille syndrome, Williams syndrome, Noonan syndrome, Marfan syndrome, nodular sclerosis, Hirschsprung disease, diabetes mellitus-type IA, Prader-Willi syndrome, Wilms tumor, primary oxalosis, sickle cell anemia, Fanconi anemia, spherocytosis, eosinophilic granuloma, plasminogen abnormality II, leptospirosis, pyruvate kinase deficiency, protein S deficiency, fibromuscular dysplasia, osteogenesis imperfecta, polycystic kidney.

In the third step, we used the summarized tagging codebook to mark the records extracted in the first step from both UEBMI and URBMI, i.e. individual records were categorized into those with or without tagging text. For records with tagging text, two neurologists reviewed them independently to further ascertain MMD patients. Patients who had records with Tag A were included (i.e., MMD patients after the verification of their diagnostic codes or texts). Patients were excluded if (1) they had any records with Tag B (i.e., MMS patients); or (2) the diagnostic text associated with MMD contained words such as "uncertainty", "undetermined", "suspicious", "?" and other synonyms (i.e., suspicious MMD patients). Finally, 47,775 records from 1,987 patients were reviewed and included as definite MMD patients in our study.

Figure A1. Flow chart of case ascertainment.



Section B. Validation study

The objective of this validation study was to evaluate the accuracy of the diagnosis-based searching algorithm identifying MMD patients in the UEBMI and URBMI databases. Considering the feasibility, we chose a high-ranked tertiary hospital to conduct the validation study.

1. Case identification

All patient-related data in the UEBMI and URBMI databases are anonymous. Therefore, we used indirectly identifying data (identifier of the hospital, MD5 encrypted ID number, dates of start of hospital stay and of discharge, age and gender) to identify the cases at the hospital level. We identified 149 potential MMD cases in this tertiary hospital from 2013 to 2016.

2. CRF development

According to the diagnostic criteria in “Chinese guidelines for the diagnosis and treatment of Moyamoya disease (2017)”, a case report form (CRF) (Table B1) was designed to define the information list which is needed to validate the diagnosis, including imaging examination, diagnostic information and so on.

3. Medical records extraction and chart review

Only local doctors can have access to patients' private information and original medical records during this process. Two trained neurologists independently reviewed the medical records of each potential MMD case to extract information, fill out the anonymous CRF, and confirm the diagnosis status of cases. Any disagreements were resolved by discussion with a third expert.

Of these 149 patients, 19 had findings consistent with MMS, including 17 patients with arteriosclerosis, 1 patient with hyperthyroidism and 1 patient having irradiation to the head. Besides, one patient was probably coding errors because there were no clinical features suggestive of MMD. Ultimately, 129 patients were confirmed as having true MMD by chart review, based on the diagnostic criteria in “Chinese guidelines for the diagnosis and treatment of Moyamoya disease (2017)”.

4. PPV calculation

We calculated the positive predictive value (PPV) of the confirmed MMD cases as the number

of true positives after medical chart review divided by the number of cases identified by the algorithm.

		Chart review (Gold standard)		
		Disease	No disease	Total
Algorithm	Disease	A (True cases correctly identified by algorithm)	B (Non-cases wrongly identified as cases by algorithm)	A+B
	No disease	C (True cases not identified by algorithm)	D (true non-cases correctly identified by algorithm)	C+D
	Total	A+C	B+D	A+B+C+D

$$\mathrm{PPV} = \mathrm{A}/(\mathrm{A} + \mathrm{B})$$

Of the 149 algorithm-identified cases, 129 were confirmed as true MMD by chart review, yielding an acceptable PPV of 87%. The result was similar to a Danish population-based study on MMD, which validated the diagnosis of ICD-10 code (I67.5) and calculated the PPV of 86% (***Eur J Neurol.*** 2020;27:2446-2452.). The validation study suggests that MMD patients can be identified in the UEBMI and URBMI databases using the algorithms of ICD-10 code (I67.5) or MMD related medical terms in Chinese and English. The accuracy of MMD diagnoses in both UEBMI and URBMI is acceptable, supporting their use in epidemiologic studies.

Table B1. The case report form (CRF).

Chart Review		
Reviewer：	The date for chart review：□□/□□/□□□□	Signature：
Moyamoya disease?	□ Yes □ No □ Unclear_____	
Moyamoya syndrome?	□ Yes □ No □ Unclear_____	
Diagnostic basis? (Please briefly describe, including differential diagnosis)		
1. Basic information		
1.1 Anonymous ID: □□□□□□□□	1.2 Age: □□□years old	
1.3 Date of visit: □□/□□/□□□□	1.4 Sex: □Male □Female	
1.5 Type: □Outpatient □Emergency □Inpatient		
2. Diagnostic information		
2.1 Primary diagnosis	Diagnostic text_____ICD-10 code_____	
2.2 Secondary diagnosis 1	Diagnostic text_____ICD-10 code_____	
2.3 Secondary diagnosis 2	Diagnostic text_____ICD-10 code_____	
2.4 Chief complaint	_____	
2.5 Operation/Surgery	Operation term_____ Operation note_____	
3. Imaging examination		
3.1 Digital Subtraction Angiography (DSA)	Examined: □ Yes, Date: □□/□□/□□□□ □ No	
	Any indications of MMD? □ Yes □ No	
	Results (Please briefly describe)_____	
3.2 Magnetic Resonance Imaging/Angiography (MRI/MRA)	Examined: □ Yes, Date: □□/□□/□□□□ □ No	
	Any indications of MMD? □ Yes □ No	
	Results (Please briefly describe)_____	

Section C. Strategy used in the estimation of numerator

Initially, Poisson regression model was established based on M_1 and N_1, considering covariates including age, sex, and insurance type. $M_1 + M_2$ was then estimated by replacing N_1 with $N_1 + N_2$ in the Poisson regression model. Based on the normal distribution, additional nine estimates of each province were selected from the 95% CI of the estimate from Poisson regression. According to Rubin's Rule, the ten estimates were combined to calculate the pooled prevalence or incidence for each province.

Table S1. Crude incidence of moyamoya disease in urban China in 2016, grouped by sex, age group, and area (Units: /100 000 Person-year).

		Incidence (95% CI)	
		Male	**Female**
Total		0.52 (0.40, 0.65)	0.66 (0.51, 0.81)
Age group			
	0~9	0.27 (0.15, 0.40)	0.37 (0.24, 0.51)
	10~19	0.14 (0.08, 0.21)	0.16 (0.08, 0.23)
	20~29	0.25 (0.17, 0.34)	0.44 (0.30, 0.58)
	30~39	0.42 (0.27, 0.57)	0.62 (0.43, 0.81)
	40~49	0.45 (0.29, 0.60)	0.54 (0.37, 0.72)
	50~59	0.50 (0.33, 0.66)	0.62 (0.42, 0.83)
	60~69	0.52 (0.34, 0.70)	0.60 (0.41, 0.80)
	70~79	0.46 (0.29, 0.63)	0.50 (0.30, 0.70)
	>=80	0.40 (0.16, 0.63)	0.26 (0.10, 0.43)
Region [a]			
	East	0.66 (0.43, 0.90)	0.93 (0.65, 1.22)
	North	0.35 (0.19, 0.51)	0.45 (0.17, 0.73)
	North-East	0.34 (0.23, 0.44)	0.37 (0.17, 0.58)
	North-West	0.30 (0.00, 0.67)	0.98 (0.12, 1.83)
	South-Central	0.73 (0.37, 1.09)	0.82 (0.45, 1.19)
	South-West	0.45 (0.05, 0.85)	0.39 (0.13, 0.65)

a. East area included Jiangsu, Zhejiang, Anhui and Shandong provinces; North area included Shanxi and Inner Mongolia provinces; North-East area included Liaoning, Jilin and Heilongjiang provinces; North-West area included Shaanxi, Qinghai and Xinjiang provinces; South-Central area included Henan, Hubei, Hunan, Guangdong, Guangxi and Hainan provinces; South-West area included Chongqing, Guizhou and Yunnan provinces.

Table S2. Prevalence and incidence of moyamoya disease in different region of China in 2016 (Units: /100 000 Person-year).

Region [a]	Prevalence (95% CI)		Incidence (95% CI) [b]	
	Crude rate	Standardized rate [a]	Crude rate	Standardized rate [c]
East	1.60 (1.10, 2.10)	1.27 (0.73, 1.81)	0.84 (0.60, 1.09)	0.69 (0.41, 0.96)
North	0.50 (0.36, 0.64)	0.51 (0.10, 0.91)	0.39 (0.27, 0.52)	0.39 (0.02, 0.76)
North-East	0.62 (0.40, 0.84)	0.54 (0.26, 0.82)	0.38 (0.21, 0.55)	0.33 (0.12, 0.54)
North-West	0.97 (0.59, 1.34)	0.70 (0.26, 1.15)	0.75 (0.10, 1.39)	0.39 (0.00, 0.85)
South-Central	1.17 (0.68, 1.66)	0.97 (0.42, 1.52)	0.80 (0.47, 1.14)	0.63 (0.24, 1.02)
South-West	0.79 (0.18, 1.41)	0.58 (0.06, 1.09)	0.48 (0.15, 0.80)	0.30 (0.03, 0.56)

a. East area included Jiangsu, Zhejiang, Anhui and Shandong provinces; North area included Shanxi and Inner Mongolia provinces; North-East area included Liaoning, Jilin and Heilongjiang provinces; North-West area included Shaanxi, Gansu, Qinghai and Xinjiang provinces; South-Central area included Henan, Hubei, Hunan, Guangdong, Guangxi and Hainan provinces; South-West area included Chongqing, Guizhou and Yunnan provinces.

b. The new-onset moyamoya disease was defined using a 3-year wash-out period. One province (Gansu) was excluded due to the time ranges <4 years.

c. Standardized by 2010 Chinese census data.

Table S3. Prevalence and incidence of moyamoya disease in urban China during 2013~2016 (Units: /100 000 Person-year).

Year	Prevalence (95% CI)		Incidence (95% CI) [a]	
	Crude rate	Standardized rate [a]	Crude rate	Standardized rate [b]
2013	0.62 (0.48, 0.77)	0.37 (0.22, 0.52)	-	-
2014	0.99 (0.77, 1.20)	0.66 (0.44, 0.89)		
2015	1.06 (0.86, 1.27)	0.69 (0.49, 0.91)	-	-
2016	1.01 (0.81, 1.21)	0.72 (0.50, 0.95)	0.59 (0.49, 0.68)	0.42 (0.27, 0.56)

a. The new-onset moyamoya disease was defined using a 3-year wash-out period. One province (Gansu) was excluded due to the time ranges <4 years.
b. Standardized by 2010 Chinese census data.

Table S4. Crude prevalence of moyamoya disease in urban China in 2016, grouped by sex, age group, and area (Units: /100 000 Person-year).

		Prevalence (95% CI)	
		Male	**Female**
Total		0.90 (0.70, 1.09)	1.05 (0.83, 1.27)
Age group			
	0~9	0.38 (0.25, 0.51)	0.45 (0.31, 0.60)
	10~19	0.13 (0.07, 0.19)	0.16 (0.08, 0.23)
	20~29	0.44 (0.31, 0.58)	0.76 (0.55, 0.98)
	30~39	0.83 (0.58, 1.07)	1.12 (0.80, 1.44)
	40~49	0.87 (0.60, 1.13)	0.99 (0.70, 1.27)
	50~59	0.92 (0.65, 1.20)	1.03 (0.73, 1.34)
	60~69	0.99 (0.68, 1.29)	1.01 (0.71, 1.32)
	70~79	0.96 (0.63, 1.28)	0.89 (0.56, 1.22)
	>=80	0.67 (0.37, 0.97)	0.45 (0.23, 0.67)
Region [a]			
	East	1.28 (0.86, 1.70)	1.75 (1.14, 2.37)
	North	0.38 (0.20, 0.56)	0.60 (0.31, 0.90)
	North-East	0.60 (0.38, 0.82)	0.60 (0.39, 0.80)
	North-West	0.66 (0.46, 0.86)	1.14 (0.61, 1.68)
	South-Central	1.12 (0.58, 1.66)	1.15 (0.63, 1.67)
	South-West	0.80 (0.04, 1.55)	0.72 (0.26, 1.17)

a. East area included Jiangsu, Zhejiang, Anhui and Shandong provinces; North area included Shanxi and Inner Mongolia provinces; North-East area included Liaoning, Jilin and Heilongjiang provinces; North-West area included Shaanxi, Gansu, Qinghai and Xinjiang provinces; South-Central area included Henan, Hubei, Hunan, Guangdong, Guangxi and Hainan provinces; South-West area included Chongqing, Guizhou and Yunnan provinces.

Table S5. The results of sensitivity analysis (Units: /100 000 Person-year).

	Prevalence (95% CI)	Incidence (95% CI)
Main analysis	1.01 (0.81, 1.21)	0.59 (0.49, 0.68)
Excluding the top 10% of provinces with missing diagnostic information [a]	1.04 (0.83, 1.26)	0.57 (0.47, 0.67)
Excluding the top 20% of provinces with missing diagnostic information [b]	0.94 (0.74, 1.15)	0.55 (0.45, 0.65)
Using only observed cases [c]	0.67 (0.64, 0.69)	0.41 (0.40, 0.43)

Note:
a. the results were calculated by using data of 20 provinces, excluding Shandong and Shanxi.
b. the results were calculated by using data of 18 provinces, excluding Shandong, Shanxi, Jilin, and Anhui.
c. known to be an underestimation of rates.