Article

Stroke to Dementia Associated with Environmental Risks—A Semi-Markov Model

Kung-Jeng Wang [1],*, Chia-Min Lee [1], Gwo-Chi Hu [2] and Kung-Min Wang [3]

[1] Department of Industrial Management, National Taiwan University of Science and Technology, Taipei 106, Taiwan; m10501003@mail.ntust.edu.tw

[2] Department of Rehabilitation Medicine, Mackay Memorial Hospital, Number 92, Section 2, Zhongshan North Road, Zhongshan District, Taipei City 10449, Taiwan; kung527@yahoo.com.tw

[3] Department of Surgery, Shin-Kong Wu Ho-Su Memorial Hospital, Shilin District, Taipei 111, Taiwan; albert.hua@msa.hinet.net

* Correspondence: kjwang@mail.ntust.edu.tw

check for updates

Abstract: Background: Most stroke cases lead to serious mental and physical disabilities, such as dementia and sensory impairment. Chronic diseases are contributory risk factors for stroke. However, few studies considered the transition behaviors of stroke to dementia associated with chronic diseases and environmental risks. Objective: This study aims to develop a prognosis model to address the issue of stroke transitioning to dementia associated with environmental risks. Design: This cohort study used the data from the National Health Insurance Research Database in Taiwan. Setting: Healthcare data were obtained from more than 25 million enrollees and covered over 99% of Taiwan's entire population. Participants: In this study, 10,627 stroke patients diagnosed from 2000 to 2010 in Taiwan were surveyed. Methods: A Cox regression model and corresponding semi-Markov process were constructed to evaluate the influence of risk factors on stroke, corresponding dementia, and their transition behaviors. Main Outcome Measure: Relative risk and sojourn time were the main outcome measure. Results: Multivariate analysis showed that certain environmental risks, medication, and rehabilitation factors highly influenced the transition of stroke from a chronic disease to dementia. This study also highlighted the high-risk populations of stroke patients against the environmental risk factors; the males below 65 years old were the most sensitive population. Conclusion: Experiments showed that the proposed semi-Markovian model outperformed other benchmark diagnosis algorithms (i.e., linear regression, decision tree, random forest, and support vector machine), with a high R^2 of 90%. The proposed model also facilitated an accurate prognosis on the transition time of stroke from chronic diseases to dementias against environmental risks and rehabilitation factors.

Keywords: chronic disease; dementia; environmental risk; multivariate analysis; rehabilitation; stroke; semi-Markov process

1. Introduction

Stroke results from the sudden occlusion or rupture of a blood vessel that supplies blood to the brain. This condition is categorized as either ischemic or hemorrhagic, where 87% of such cases are ischemic [1]. Approximately 15 million people suffer from stroke each year, and the number of stroke deaths increases annually [2]. In Taiwan, stroke is the leading cause of death, which places a substantial burden on the national healthcare system [3]. Stroke is also one of the largest causes of serious long-term mental and physical disabilities [1,4,5] and a relevant factor in dementia development [6]. Previous studies proposed that poststroke dementia is a designation for dementia following stroke over time [7–10].

An estimated 12.6 million people died as a result of living or working in an unhealthy environment. This number is nearly one-fourth of the total global deaths. Among the top causes of environment-related deaths, stroke causes 2.5 million deaths annually [2]. In addition, chronic diseases, stroke, and dementia are closely related. Mortality and readmission of stroke patients are related to multiple chronic diseases [11]. Chronic hypertension, diabetes, and hyperlipidemia are important risk factors for stroke [3].

Stroke can be initially prevented by identifying risk factors. This condition is directly related to chronic diseases, specifically diabetes mellitus [12,13], hypertension [14], and hyperlipidemia [15].

In terms of environmental interfering factors, the trends in stroke rates are associated with changing exposures to population risk factors [16]. Considering the development of the country, its exposure to high levels of air pollutants related to combustion increases the risk of stroke in people [17,18]. Some studies proclaimed that seasonal variation and temperature trigger the incidence of stroke [19], and cold temperature is positively correlated with stroke admissions [20,21].

The risk interfering factors in stroke may also be related to socioeconomic factors, such as marital status [22], low socioeconomic status, and elderly living alone [23].

Dementia is a frequent outcome of stroke, which increases the probability of long-term disability and mortality [24]. The presumed etiology of poststroke dementia indicates that two-thirds of the patients have vascular dementia and one-third of those have Alzheimer's disease [7,25–27].

Few studies considered the transition behavior in stroke to dementia, which is associated with chronic diseases, environmental risks, medication, and rehabilitation factors due to high complexity with multivariates and multilayers [28–30]. Therefore, the present study aims to address this issue. We developed a semi-Markovian process (SMP) to represent multiple states with time-based transition behavior of stroke to dementia. The objectives of this study were as follows: (i) to identify the environmental risk factors associated with stroke and (ii) to predict the transition of stroke from a chronic disease to dementia associated with environmental risks and rehabilitation via SMP.

2. Materials and Methods

2.1. Population

The population surveyed in this study was from National Health Insurance Research Database (NHIRD) in Taiwan and consisted of the detailed healthcare data from more than 25 million enrollees and covered over 99% of Taiwan's entire population [31]. This article does not contain any studies with human participants or animals performed by any of the authors. The accuracy of stroke diagnosis in NHIRD has been validated in previous research [32]. A total of 10,627 stroke patients were selected, and the selection procedure is depicted in Figure 1. Stroke patients diagnosed with stroke between 2000 and 2010 during at least three consecutive clinic visits or hospitalization were screened. Patients younger than 30 years old and those diagnosed with hemorrhagic and ischemic stroke on the same day were excluded. This study did not explore the cases of stroke in youth because stroke often occurs at middle age. Furthermore, the rare case of simultaneous occurrence of the two types of stroke was excluded.
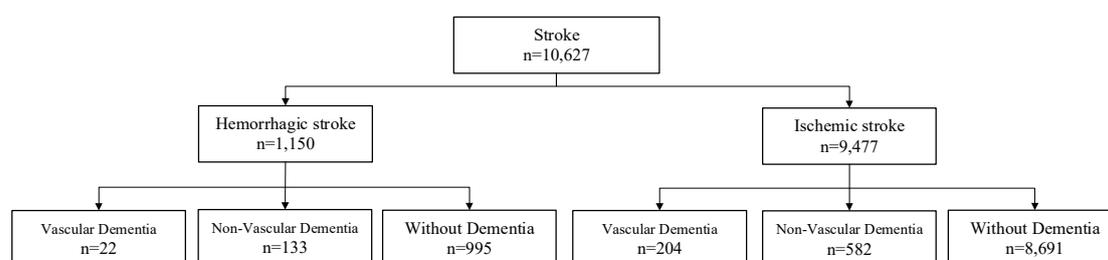
Figure 1. Stroke population in the study.

The criteria for enrollment in NHIRD included two types of stroke identified from January 1, 2000 to December 31, 2010. In International Classification of Diseases (ICD) ICD-9-CM coding system, stroke patients were defined by ICD-9-CM code of 430,432 for hemorrhagic stroke and ICD-9-CM code of 433,437 for ischemic stroke. The ICD-9-CM code for relative chronic diseases and dementia is listed in Table 1. For the modeling, the date of stroke diagnosis must be beyond the date of chronic disease diagnosis, and the date of dementia diagnosis must be beyond the date of stroke diagnosis.

Table 1. ICD-9-CM code definition of all diseases.

Disease Type	Disease Name	ICD-9-CM
Chronic disease	Hypertension	401–405
	Diabetes mellitus	250
	Hyperlipidemia	272
	Chronic lung disease	490–496
	Hyperthyroidism	242
	Chronic kidney disease	585
	Heart failure	428
	Atrial fibrillation	42731
	Sleep apnea	780.51, 780.53, 780.57
	Gout	274
	Peripheral artery disease	437.3, 440, 441, 443.1, 443.2, 443.8, 443.9, 447.1, 447.8, 447.9, 445
Stroke	Hemorrhagic stroke	430, 432
	Ischemic stroke	433, 437
Dementia	Vascular dementia	2904
	Non-vascular dementia	294, 337

Most of the stroke patients in Taiwan were diagnosed with hypertension, diabetes mellitus, and hyperlipidemia. This study used these three chronic diseases for the basic classification of stroke patients. Eleven types of chronic diseases were divided into 5 categories: newly diagnosed with hypertension, newly diagnosed with diabetes mellitus, newly diagnosed with hyperlipidemia, newly diagnosed with other chronic disease (including chronic lung disease, hyperthyroidism, chronic kidney disease, heart failure, atrial fibrillation, sleep apnea, gout, and peripheral artery disease), and stroke patients without any of the 11 chronic diseases before stroke. Furthermore, stroke patients usually suffer from multiple chronic diseases after being diagnosed with a chronic disease. This phenomenon is called the metabolic risk factor. Therefore, this study used the main types of chronic diseases including metabolic risk factor for re-classification. We used chi-square significance test for the selected chronic diseases and related metabolic risk factors to build an SMP model.

2.2. Markovian-Based Modeling

Various approaches have been applied to explore the influence of key factors on stroke patients for occurrences over time; however, only a few achieved high prediction [33]. In addition to standard statistical analyses, such as correlation analysis, covariate analysis, and subgroup analysis, the transition probability in Markovian process is preferable to analyze the risk factor effect on stroke [33]. Among the variants, the SMP provides a flexible approach to characterize the distribution of the sojourn times between states.

Multi-state models are often used to model the development of a disease in medical research, wherein the different levels of the disease can be seen as the states of the model [34]. The parametric SMP models allow the incorporation of covariates in the distribution of sojourn times to investigate the effect on the transition risk by using a proportional-hazard regression model [35,36].

The processes and the calculation of the transition probabilities of a SMP can be described as follows. First consider a model with k states belonging to finite state space $E = \{1, 2, \ldots, k\}$, and consider

$X_0, X_1, X_2, \ldots, X_n \in E$ be the successive states in n visits by a random process, in which $0 = T_0 < T_1 < \ldots < T_n$ are the consecutive time of entrance into each of these states. And the probability of the n transitions jumping from state i to j, embedded into model can be written as:

$$P_{ij} = P(X_{n+1} = j | X_n = i) \tag{1}$$

As the Markov process does not deal with the sojourn time of the state transitions, the random process regards the transition sojourn time $(T_{n+1} - T_n)$ in an SMP and its distribution that satisfies:

$$Q_{ij}(t) = P(X_{n+1} = j, T_{n+1} - T_n \leq t | X_n = i) \tag{2}$$

The probability density function of sojourn time in state i before passing to state j is given by:

$$f_{ij}(t) = \lim_{\Delta t \to 0} \frac{P(t < T_{n+1} - T_n < t + \Delta t | X_{n+1} = j, X_n = i)}{\Delta t} \tag{3}$$

The cumulative probability function, $F_{ij}(t)$ and the corresponding survival function of waiting time in state i, $S_i(t)$ are defined by:

$$F_{ij}(t) = P(T_{n+1} - T_n \leq t | X_{n+1} = j, X_n = i) = \frac{Q_{ij}(t)}{P_{ij}} \tag{4}$$

$$S_i(t) = 1 - P(T_{n+1} - T_n \leq t | X_n = i) = \sum_{j \in E} P_{ij}(1 - F_{ij}(t)) \tag{5}$$

The hazard function of an SMP model, which represents the probability of transition towards state j between time t and $t + \Delta t$, given that the process is in state i for a duration t can be drawn as follows:

$$\begin{aligned} \lambda_{ij}(t) &= \lim_{\Delta t \to 0} \frac{P(X_{n+1}=j, t<T_{n+1}-T_n<t+\Delta t | X_n=i, T_{n+1}-T_n>t)}{\Delta t} \\ &= \frac{P_{ij} f_{ij}(t)}{S_i(t)} = \frac{P_{ij} S_{ij}(t) \alpha_{ij}(t)}{S_i(t)}, \qquad \begin{array}{l} i, j \in K \\ i \neq j \end{array} \\ &\quad \lambda_{ii}(t) = -\sum_{i \neq j} \lambda_{ij}(t) \end{aligned} \tag{6}$$

Let m_{ij} be the covariate specific states transition i to j and the vector is represented by $z_{ij} = (z_{ij}^1, z_{ij}^2, \ldots, z_{ij}^{m_{ij}})$, and the baseline hazard function of the transition i to j is $h_{0,ij}(x)$, the parameter vector in a Cox model is β'_{ij}. The hazard function with covariates is defined by:

$$h_{ij}(x, z_{ij}) = h_{0,ij}(x) \cdot \exp(\beta'_{ij} \cdot z_{ij}) \tag{7}$$

Incorporating the Cox model can not only deal with multivariate models, but investigate the effect upon the relative risk (RR) of transferred stroke. Further, under the transition-specific strategy for covariates, applying Weibull distributions as the sojourn time distribution is more flexible to fit the hazard function, also generalizes the exponential distribution by two parameters $W(\sigma_{ij}, v_{ij})$, making the model well suit for various shapes, especially for medical research [35]. The Weibull function of the sojourn time distribution is defined by:

$$\alpha_{ij}(t) = v_{ij} \left(\frac{1}{\sigma_{ij}}\right)^{v_{ij}} t^{v_{ij}-1}, \forall t \geq 0, \forall \sigma_{ij} \geq 0, \forall v_{ij} \geq 0 \tag{8}$$

where σ_{ij}, v_{ij} are scale and shape parameters.

Covariate analysis and Markovian-based prediction modeling were used in this study. The covariates included in the SMP model must be of categorical parameters as limitation (Table 2). The environmental variables are defined in Table 3. In this study, a value beyond one standard deviation of the mean was considered as high incidence group.

Table 2. The covariate in the proposed SMP model.

	0 (Base Group)	1
Gender	Female	Male
Age	Otherwise	>65 years old
Divorce rate	Otherwise	High divorce rate
Unemployment rate	Otherwise	High unemployment rate
Elderly living alone rate	Otherwise	High elderly living alone rate
Temperature	Otherwise	Low winter temperature
Air pollution	Otherwise	High rate of PSI > 100

Note: The definition of environmental factors.

Table 3. Variable definition.

Environmental Type	Definition
Divorce rate	$\frac{\text{The number of divorced population over 15 years old}}{\text{Total population over 15 years old}} \times 100\%$
Unemployment rate	$\frac{\text{The number of unemployed population}}{\text{Labor force}} \times 100\%$
Air pollution	$\frac{\text{Daily PSI} > 100}{\text{Total monitoring days of PSI}} \times 100\%$
Elderly living alone rate	$\frac{\text{The number of elderly people living alone}}{\text{Total population over 65 years old}} \times 100\%$
Temperature	Average of temperature in January, February and December

The divorce and unemployment rates used in this study were from Socio-Economic Geographic Information System (SEGIS) [37]. The temperature data were from Taiwan Typhoon and Flood Research Institute (TTFRI) [38]. The air pollution index was from the environmental resource database of Environmental Protection Administration (EPA) [39]. The rate of elderly people living alone was from Ministry of Health and Welfare [40]. Each patient was characterized by one set of covariates to assess the influence of the environmental risk factors on the stroke patients. The connection between the individual patient and the environmental data was examined using the registered residence of the stroke patients recorded in the NHIRD. The environmental risk factors were collected from 22 administrative districts of Taiwan.

The stroke patients were identified into nine states in the SMP model, including five states of chronic disease, two types of stroke, and two types of dementia, to explore the phenomenon among state transitions in SMP (Figure 2). This study combined hypertension and hyperlipidemia into one state to investigate the relationship between chronic disease and stroke. The state is defined in Table 4. Chronic disease states included hypertension and with one metabolic risk factor after diagnosed with hypertension, diabetes mellitus, hyperlipidemia and with one metabolic risk factor diagnosed with hyperlipidemia, with more than four metabolic risk factors after having another chronic disease, and without any of the 11 types of chronic disease. These chronic states represent the initial state in the SMP model. Hemorrhagic and ischemic stroke are the stroke states, whereas vascular dementia and non-vascular dementia are the dementia states.

The state diagram of the SMP model is depicted in Figure 2. The arrows indicate transitions among states, and the probability transition matrix was computed from our stroke patient population. This process used Weibull distribution to construct the time delay between different transitions. This study focused on the transferred stroke cases, such as chronic disease transfer to stroke state and stroke transfer to dementia state or remain in the stroke state. Such multi-state SMP model can be resolved via a software package called "semi-Markov" in R [33].

(a)

State	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	0.103	0.897	0	0
2	0	0	0	0	0	0.094	0.906	0	0
3	0	0	0	0	0	0.07	0.93	0	0
4	0	0	0	0	0	0.041	0.959	0	0
5	0	0	0	0	0	0.243	0.757	0	0
6	0	0	0	0	0	0	0	0.142	0.858
7	0	0	0	0	0	0	0	0.26	0.74
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0

(b)
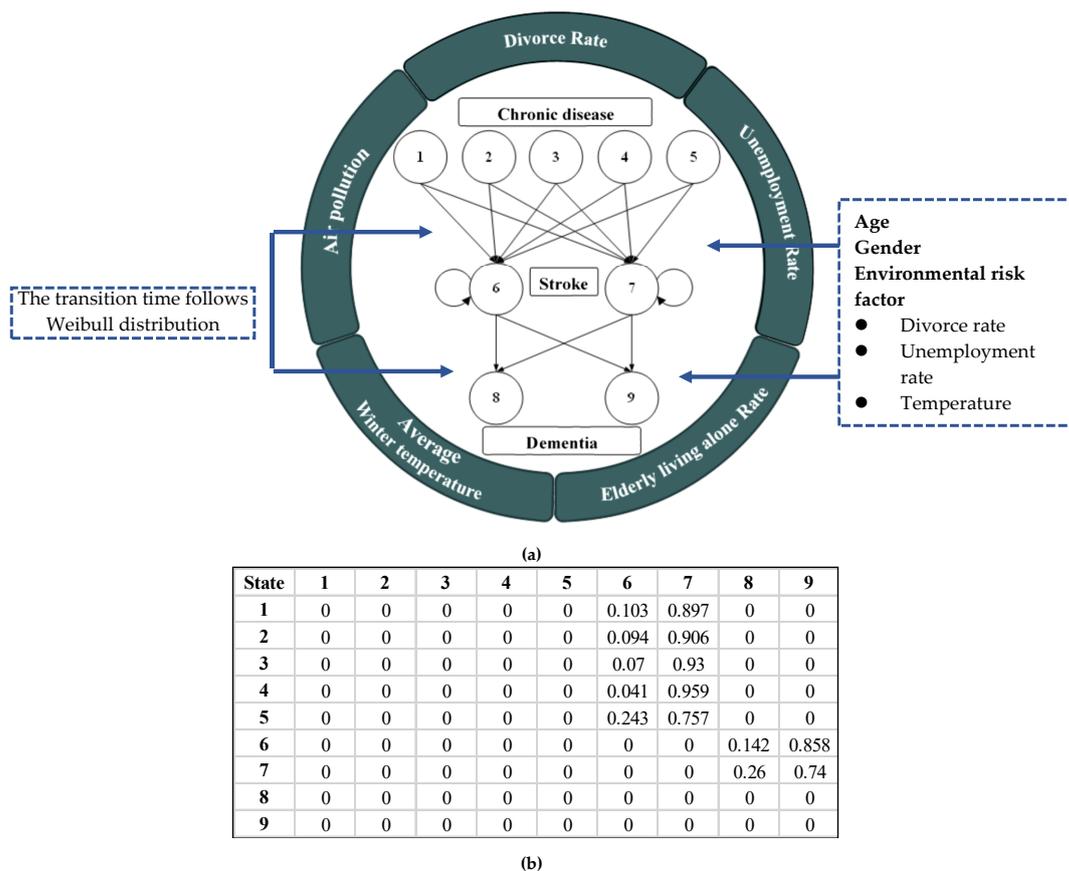
Figure 2. (**a**) Probability transition matrix; (**b**) The transition probability matrix in the SMP model.

Table 4. The number of states and descriptions in the SMP model.

State	Type	Description
1	Chronic disease	Diagnosed only with hypertension or with one metabolic risk factor after having hypertension
2		Diagnosed only with diabetes mellitus
3		Diagnosed only with hyperlipidemia or with one metabolic risk factor after having hyperlipidemia
4		Diagnosed with more than four metabolic risk factors after having other 8 chronic diseases
5		Diagnosed other than 11 types of chronic disease or without chronic
6	Stroke	Hemorrhagic stroke
7		Ischemic stroke
8	Dementia	Vascular dementia
9		Non-vascular dementia

3. Results and Discussion

3.1. Multivariate Analysis

In the covariate model, we obtained 98 transitions, consisting of 10 transitions in the first layer (5 types of chronic diseases multiplied by 2 types of stroke) and 4 transitions in the second layer (2 types of stroke multiplied by 2 types of dementia) for the 7 covariates (i.e., gender, age, and the five types of environmental risk factors). A Wald test (H0: $\beta ij = 0$; H1: $\beta ij \neq 0$) with *p*-value < 0.05 was given.

In the multivariate model, all covariates selected from the statistically significant transitions from univariate analysis were included. The results of multivariate analysis are presented in Table 5. The three sets of models were as follows: A. environmental risk factors, B. medication and rehabilitation factors, and C. all factors. The negative estimation β for the Cox regression model indicated that the transition risk was higher for the base group than for the other groups.

Int. J. Environ. Res. Public Health **2020**, 17, 1944

7 of 13

Table 5. The multivariate analysis by different covariates models.

	Model A: Environmental Risk				Model B: Medication and Rehabilitation				Model C: All Factors			
	Transition	Estimation β	Relative Risk	*p*-Value	Transition	Estimation β	Relative Risk	*p*-Value	Transition	Estimation β	Relative Risk	*p*-Value
Gender	β_{16} (1→6)	0.214	1.24	0.0042 *	β_{16} (1→6)	0.183	1.20	0.0143 *	β_{16} (1→6)	0.096	1.10	0.0108
	β_{36} (3→6)	0.493	1.64	0.02 *	β_{36} (3→6)	0.509	1.66	0.0164 *	β_{36} (3→6)	0.490	1.63	0.0206
Age	β_{17} (1→7)	−0.117	0.89	<0.0001 *	β_{17} (1→7)	−0.097	0.91	0.0002 *	β_{17} (1→7)	−0.124	0.88	<0.0001
	β_{56} (5→6)	−2.366	0.09	<0.0001 *	β_{56} (5→6)	−2.382	0.09	<0.0001 *	β_{56} (5→6)	−2.374	0.09	<0.0001
	β_{57} (5→7)	−2.538	0.08	<0.0001 *	β_{57} (5→7)	−2.538	0.08	<0.0001 *	β_{57} (5→7)	−2.541	0.08	<0.0001
	β_{69} (6→9)	−0.394	0.67	0.0259 *	β_{69} (6→9)	−0.394	0.67	0.0259 *	β_{69} (6→9)	−0.393	0.68	0.0259
	β_{79} (7→9)	−0.372	0.69	0.0008 *	β_{79} (7→9)	−0.389	0.68	0.0005 *	β_{79} (7→9)	−0.366	0.69	0.001
Divorce rate	β_{26} (2→6)	1.364	3.91	0.0122 *					β_{26} (2→6)	1.364	3.91	0.0123
Unemployment rate	β_{16} (1→6)	−0.434	0.65	0.0004 *					β_{16} (1→6)	−0.447	0.64	0.0003
	β_{17} (1→7)	−0.369	0.69	<0.0001 *					β_{17} (1→7)	−0.365	0.69	<0.0001
	β_{36} (3→6)	−0.673	0.51	0.0379 *					β_{36} (3→6)	−0.673	0.51	0.0377
	β_{37} (3→7)	−0.486	0.62	<0.0001 *					β_{37} (3→7)	−0.493	0.61	<0.0001
	β_{47} (4→7)	−0.904	0.40	<0.0001 *					β_{47} (4→7)	−0.915	0.40	<0.0001
Temperature	β_{16} (1→6)	0.237	1.27	0.0088 *					β_{16} (1→6)	0.222	1.25	0.0142
	β_{17} (1→7)	0.408	1.50	<0.0001 *					β_{17} (1→7)	0.397	1.49	<0.0001
	β_{27} (2→7)	0.540	1.72	<0.0001 *					β_{27} (2→7)	0.625	1.87	<0.0001
	β_{37} (3→7)	0.237	1.27	0.0015 *					β_{37} (3→7)	0.227	1.25	0.0023
	β_{78} (7→8)	−0.316	0.73	0.0411 *					β_{78} (7→8)	−0.316	0.73	0.0409
	β_{79} (7→9)	−0.444	0.64	<0.0001 *					β_{79} (7→9)	−0.446	0.64	<0.0001
Air pollution	β_{17} (1→7)	0.071	1.07	0.0057 *					β_{17} (1→7)	0.079	1.08	0.0035
	β_{47} (4→7)	0.272	1.31	0.0059 *					β_{47} (4→7)	0.260	1.30	0.0085
Blood pressure lowing drug					β_{16} (1→6)	−0.275	0.76	0.0042 *	β_{16} (1→6)	−0.318	0.73	0.0009
					β_{17} (1→7)	−0.458	0.63	<0.0001 *	β_{17} (1→7)	−0.469	0.63	<0.0001
					β_{37} (3→7)	−0.218	0.80	0.0006 *	β_{37} (3→7)	−0.215	0.81	0.0008
					β_{47} (4→7)	−0.331	0.72	0.0309 *	β_{47} (4→7)	−0.351	0.70	0.0222
					β_{68} (6→8)	−0.822	0.44	<0.0001 *	-	-	-	-
Blood lipid lowing drug					β_{17} (1→7)	−0.261	0.77	<0.0001 *	β_{17} (1→7)	−0.250	0.78	<0.0001
Glycemic lowing drug					β_{17} (1→7)	−0.132	0.88	0.0168 *	β_{17} (1→7)	−0.150	0.89	0.0068
					β_{27} (2→7)	−0.254	0.78	0.0119 *	β_{27} (2→7)	−0.349	0.71	0.0007
Thrombus prevention drug					β_{16} (1→6)	−0.237	0.79	0.0053 *	β_{16} (1→6)	0.096	1.10	0.0108
Rehabilitation					β_{79} (7→9)	−0.187	0.83	0.0243 *	-	-	-	-

In the transition of chronic diseases to stroke, in model A, age, unemployment rate, and temperature showed high effect occurrences. The risk to transfer from hypertension (RR = 1.24) and hyperlipidemia (RR = 1.64) to hemorrhagic stroke was higher in males than in females as the base group. For the age factor, the patient under 65 years old has higher risk. The divorce rate indicated that high divorce rate was 3.91 times risk in diabetes mellitus transfer to hemorrhagic stroke group than in the base group. Conversely, the unemployment rate factor indicated that the base group had higher risk compared with the high unemployment rate. The temperature factor revealed that low temperature had 1.27–1.72 times risk compared with normal temperature. The air pollution revealed that high air pollution had 1.07 and 1.31 times risk to be transferred from hypertension and with other eight chronic diseases to hemorrhagic stroke.

In the transition of stroke to dementia, only the normal temperature (the base group) had high risk, such as in ischemic stroke transfer to vascular dementia (RR = 0.73) and to non-vascular dementia (RR = 0.64).

To validate the effect of environmental risk factors, this study compared with the model on medication and rehabilitation of stroke patients, as shown in model B. Moreover, this study built an overall multivariate model in which covariates were selected from models A and B. Model C revealed that the environmental risk factors retained the risk effect on transitions. However, two transitions were removed from model C.

3.2. Findings in SMP

The predictive model and corresponding result, which includes mean time, standard deviation, and adjusted R^2, are shown in Table 6. The prediction result on overall sojourn time indicated that the transition of diabetes mellitus to ischemic stroke had a sojourn time of 1.96 years. For the transition of stroke to dementia, hemorrhagic stroke to non-vascular dementia had a sojourn time of 1.29 years. The R^2 values in the transition were all greater than 0.8, indicating that the overall prediction model had 80% prediction performance.

The results of sojourn time prediction by covariates indicated that different models showed their prediction merit with respect to different chronic diseases. In general, the adjusted R^2 of model A outperformed the others in most transition cases from chronic to both hemorrhagic and ischemic strokes. This result indicated that the environmental risk has high explanatory power in the prediction of transition time. For the transition of stroke to dementia, model C had higher prediction performance than the other models did, except in hemorrhagic stroke to vascular dementia that was preferred by model A. As shown in Figure 3, the transition time prediction was remarkably improved in the proposed model compared with only considering medication and rehabilitation.

Table 6. Sojourn time prediction.

Overall Sojourn Time Prediction						
Chronic to Stroke	Hemorrhagic Stroke			Ischemic Stroke		
	Mean (Year)	SD	R^2 *	Mean	SD	R^2
State 1: Hypertension	2.94	2.94	92.3%	2.72	2.89	87.0%
State 2: Diabetes mellitus	2.48	2.50	90.2%	1.96	2.36	86.1%
State 3: Hyperlipidemia	2.93	2.94	95.5%	2.63	2.80	92.9%
State 4: Other 8 chronic diseases	4.67	3.74	92.3%	4.77	3.71	97.5%
State 5: Other than above 11 types of chronic disease or without chronic	25.53	20.24	90.3%	29.03	22.65	91.3%
Stroke to Dementia	State 8: Vascular dementia			State 9: Non-vascular dementia		
	Mean	SD	R^2	Mean	SD	R^2
State 6: Hemorrhagic stroke	1.94	2.23	77.3%	1.29	2.02	80.0%
State 7: Ischemic stroke	2.61	2.85	83.4%	2.71	2.79	88.6%

Table 6. *Cont.*

Covariate Transition		A: Environmental Risk			B: Medication and Rehabilitation			C: All Factors		
		Mean (Year)	SD	R^2	Mean (Year)	SD	R^2	Mean (Year)	SD	R^2
Hemorrhagic stroke	β_{16} (1→6)	2.79	2.86	95.34%	3.19	3.06	93.11%	3.12	3.05	94.07%
	β_{26} (2→6)	2.36	2.45	90.92%	2.37	2.43	91.81%	2.24	2.36	89.23%
	β_{36} (3→6)	2.80	2.88	95.06%	2.68	2.79	94.28%	2.97	3.01	93.61%
	β_{46} (4→6)	7.03	3.56	91.73%	6.83	3.68	91.49%	7.04	3.58	90.82%
	β_{56} (5→6)	30.22	22.70	90.33%	29.82	22.62	90.35%	30.19	22.68	92.31%
		Mean (yr)	SD	R^2	Mean (yr)	SD	R^2	Mean (yr)	SD	R^2
Ischemic stroke	β_{17} (1→7)	2.71	2.90	92.63%	3.15	3.10	87.83%	3.11	3.09	92.30%
	β_{27} (2→7)	1.86	2.30	88.28%	1.97	2.37	87.47%	1.93	2.35	86.08%
	β_{37} (3→7)	2.55	2.76	93.98%	2.65	2.81	93.26%	2.64	2.82	93.06%
	β_{47} (4→7)	5.03	3.88	96.04%	5.56	3.93	97.07%	5.69	4.02	96.34%
	β_{57} (5→7)	37.15	25.55	91.22%	36.97	25.51	91.23%	37.15	25.55	91.57%
		Mean (yr)	SD	R^2	Mean (yr)	SD	R^2	Mean (yr)	SD	R^2
Vascular dementia	β_{68} (6→8)	1.68	2.15	91.79%	2.25	2.36	80.85%	2.39	2.38	86.29%
	β_{69} (6→9)	3.01	3.05	87.23%	3.10	3.07	85.12%	3.16	3.07	92.63%
		Mean (yr)	SD	R^2	Mean (yr)	SD	R^2	Mean (yr)	SD	R^2
Non-vascular dementia	β_{78} (7→8)	1.44	2.13	78.96%	1.42	2.12	81.19%	1.44	2.13	85.49%
	β_{79} (7→9)	3.07	2.99	90.07%	3.11	3.00	89.37%	3.17	3.03	93.54%

* Note: Adjusted $R^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-p-1}$, where p is the number of regressors, and n is the sample size.
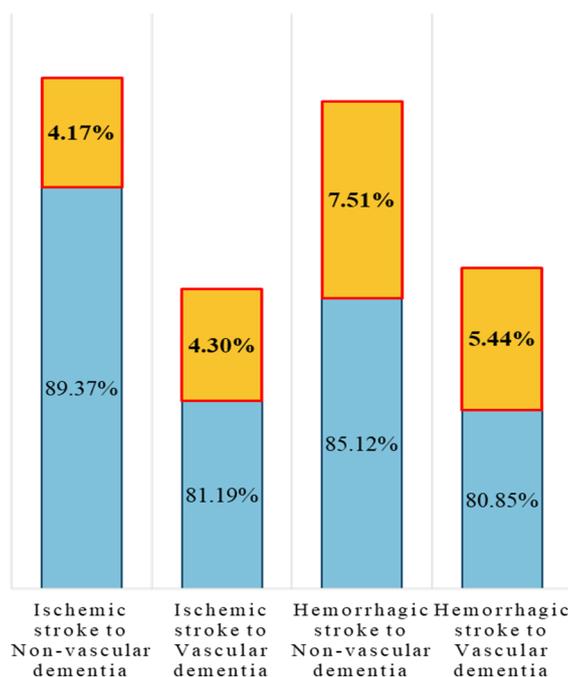
Figure 3. Improvement by the proposed models as compared to only considering medication and rehabilitation.

Table 7 reveals that the SMP prediction model has high R^2, which indicated that the proposed transition time prediction model has a good explanatory power compared with linear regression, decision tree, random forest, and support vector machine (SVM) [29].

Table 7. R^2 of SMP comparison with benchmark models.

Hemorrhagic Stroke	Linear Regression	Decision Tree	Random Forest	SVM	SMP
Environmental risk	88.52%	91.35%	88.84%	91.93%	92.66%
Medication and Rehabilitation	88.41%	91.87%	88.49%	90.22%	92.22%
All factors	88.50%	92.00%	91.60%	90.19%	92.01%
Ischemic stroke					
Environmental risk	85.63%	88.57%	85.09%	89.40%	92.42%
Medication and Rehabilitation	85.36%	90.06%	85.81%	90.00%	91.38%
All factors	85.60%	90.35%	91.57%	90.56%	91.88%
Vascular dementia	Linear Regression	Decision Tree	Random Forest	SVM	SMP Model
Environmental risk	75.74%	79.61%	81.18%	73.28%	89.51%
Medication and Rehabilitation	77.46%	82.61%	82.88%	75.54%	83.00%
All factors	78.76%	85.27%	86.91%	78.52%	89.46%
Non-vascular dementia					
Environmental risk	75.80%	76.44%	77.70%	73.60%	84.52%
Medication and Rehabilitation	77.87%	81.11%	81.29%	76.77%	85.28%
All factors	75.66%	82.96%	84.45%	81.67%	89.52%

Model C considered age, gender, environmental risk factors, and medication and rehabilitation of the stroke patient as covariates. This model indicated that the drugs for lowering blood pressure, blood lipid, and blood sugar decreased the risk of most transitions of chronic disease to stroke and the transition of hypertension to ischemic stroke. In addition, drug interaction data revealed that patients medicated with these three drugs in the same period have low transition risk. Only the blood pressure lowering drug influenced the transition of stroke to dementia (RR = 0.44). For rehabilitation, the results suggested that for people who suffer from ischemic stroke, being active on rehabilitation could reduce their risk of transition to non-vascular dementia (RR = 0.83).

Multivariate analysis results indicated that divorce rate (RR = 3.91), temperature (RR = 1.27, 1.5, 1.72) and air pollution (RR = 1.07, 1.31) have positive signs on chronic disease transfer to stroke. By contrast, the risk of transfer decreases when the unemployment rate shows the negative sign, which indicated high unemployment rate. In the transition of stroke to dementia, the temperature indicated that the risk of transfer to dementia decreased with the temperature (RR = 0.73, 0.64).

After all the environmental factors in multivariate analysis were considered, two transitions were removed from the medication and rehabilitation. These two transitions do not substantially influence the stroke transfer to dementia due to the explanation power of environmental factors.

4. Conclusions

This study facilitates accurate prognosis on the transition behavior and time of stroke from chronic diseases to dementias against environmental risks, medication, and rehabilitation factors using a multilayer and multistate SMP model. Results indicated that the covariates incorporated in the study had different effects on sojourn time in stroke transition. Multivariate analysis results revealed that the high divorce rate, high air pollution, and low temperature were high-risk factors that triggered the transition of a chronic disease to stroke. Previous researchers associated cold and hot temperatures with an increased risk of stroke mortality [19,41]. Their study investigated the mortality owing to stroke, whereas our present study confirmed the effect of temperature on the transition of a chronic disease to stroke. Hyperthermia in acute stroke victims caused a poor prognostic parameter on a short-term basis. Roy and Ray [42] focused on body temperature, whereas our present study investigated environmental temperature. Lim et al. [43] addressed that diurnal temperature change over the preceding 24 h is associated with daily stroke incidence. Their study analyzed the short-term effect of environmental temperature, whereas our present study addressed a long-term effect of environmental temperature annually. Age and temperature are high-risk factors for the transition of

stroke to dementia. Compared with a covariate model of medication and rehabilitation from a previous study, the proposed environmental factor-based prognosis model had a good explanation power of the risk effects on state transitions. The present study highlighted the high-sensitive population through subgroup analysis. Our result pointed out that the population of males below 65 years old was the most sensitive one against the environmental risk factors. Among all factors, air pollution had the highest influence in the transition of a chronic disease to stroke, and the unemployment rate had the highest influence in the transition of stroke to dementia.

The constructed SMP model predicts the transition time the stroke patients stay in different states (i.e., from chronic to stroke, and from stroke to dementia). By comparing the adjusted R^2 with other benchmark algorithms (linear regression, decision tree, random forest, and SVM), we found that the proposed prediction model has the highest prediction performance. In addition, we also confirmed that the Weibull distribution has a good flexible fitting ability to deal with medical survival situations.

Future research may incorporate additional environmental risk factors, such as marital status, socioeconomic status, dietary habits, employment types, and water quality, which may lead to a high prediction on the transition behaviors of stroke. Given the lack of information integration between environmental and medical data, this study connected these data only through the patient's registered residence. However, integrating these heterogeneous data precisely and individually is still a challenge worthy to be further investigated. Environmental risk factors served as covariates in the proposed model in a binary form due to computational constraints. Future studies may investigate the influence of the numerical types of covariates. The current SMP model may also be extended by adding death state to investigate the survival time of stroke patients.

References

1. Arslan, A.K.; Colak, C.; Sarihan, M.E. Different medical data mining approaches based prediction of ischemic stroke. *Comput. Methods Programs Biomed.* **2016**, *130*, 87–92. [CrossRef] [PubMed]
2. Arboix, A. Cardiovascular risk factors for acute stroke: Risk profiles in the different subtypes of ischemic stroke. *World J. Clin. Cases* **2015**, *3*, 418. [CrossRef] [PubMed]
3. Benjamin, E.J.; Virani, S.S.; Callaway, C.W.; Chamberlain, A.M.; Chang, A.R.; Cheng, S.; Chiuve, S.E.; Cushman, M.; Delling, F.N.; Deo, R.; et al. Heart disease and stroke statistics—2018 update: A report from the American Heart Association. *Circulation* **2018**, *137*, e67–e492. [CrossRef] [PubMed]
4. Blank, R.H. Alzheimer's Disease and Other Dementias: An Introduction. In *Social & Public Policy of Alzheimer's Disease in the United States*; Palgrave Pivot: Singapore, 2019; pp. 1–26.
5. Bordone, M.P.; Salman, M.M.; Titus, H.E.; Amini, E.; Andersen, J.V.; Chakraborti, B.; Diuba, A.V.; Dubouskaya, T.G.; Ehrke, E.; de Freitas, A.E.; et al. The energetic brain—A review from students to students. *J. Neurochem.* **2019**, *151*, 139–165. [CrossRef]
6. Cao, Q.; Buskens, E.; Feenstra, T.; Jaarsma, T.; Hillege, H.; Postmus, D. Continuous-time semi-Markov models in health economic decision making: An illustrative example in heart failure disease management. *Med. Decis. Mak.* **2016**, *36*, 59–71. [CrossRef]
7. Chen, R.; Wang, C.; Meng, X.; Chen, H.I.; Thach, T.Q.; Wong, C.M.; Kan, H. Both low and high temperature may increase the risk of stroke mortality. *Neurology* **2013**, *81*, 1064–1070. [CrossRef]

8. Cheng, C.L.; Kao, Y.H.Y.; Lin, S.J.; Lee, C.H.; Lai, M.L. Validation of the national health insurance research database with ischemic stroke cases in Taiwan. *Pharmacoepidemiol. Drug Saf.* **2011**, *20*, 236–242. [CrossRef]

9. Cox, D.R. Regression models and life-tables. *J. R. Stat. Soc. Ser. B* **1972**, *34*, 187–220. [CrossRef]

10. Desmond, D.W.; Moroney, J.T.; Sano, M.; Stern, Y. Incidence of dementia after ischemic stroke: Results of a longitudinal study. *Stroke* **2002**, *33*, 2254–2262. [CrossRef]

11. Dichgans, M. Dementia risk after transient ischaemic attack and stroke. *Lancet Neurol.* **2019**, *18*, 223–225. [CrossRef]

12. Eichler, T.S.; Hoffmann, W.; Hertel, J.; Richter, S.; Wucherer, D.; Michalowsky, B.; Dreier, A.; Thyrian, J.R. Living alone with dementia: Prevalence, correlates and the utilization of health and nursing care services. *J. Alzheimers Dis.* **2016**, *52*, 619–629. [CrossRef]

13. Emerging Risk Factors Collaboration. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: A collaborative meta-analysis of 102 prospective studies. *Lancet* **2010**, *375*, 2215–2222. [CrossRef]

14. Feigin, V.L.; Wiebers, D.O. Environmental factors and stroke: A selective review. *J. Stroke Cerebrovasc. Dis.* **1997**, *6*, 108–113. [CrossRef]

15. Feigin, V.L.; Forouzanfar, M.H.; Krishnamurthi, R.; Mensah, G.A.; Connor, M.; Bennett, D.A.; Moran, A.E.; Sacco, R.L.; Anderson, L.; Truelsen, T.; et al. Global and regional burden of stroke during 1990–2010: Findings from the global burden of disease study 2010. *Lancet* **2014**, *383*, 245–255. [CrossRef]

16. Franklin, S.S.; Wong, N.D. Hypertension and cardiovascular disease: Contributions of the Framingham Heart Study. *Glob. Heart* **2013**, *8*, 49–57. [CrossRef] [PubMed]

17. Guo, P.; Zheng, M.; Feng, W.; Wu, J.; Deng, C.; Luo, G.; Wang, L.; Pan, B.; Liu, H. Effects of ambient temperature on stroke hospital admissions: Results from a time-series analysis of 104,432 strokes in Guangzhou, China. *Sci. Total Environ.* **2017**, *580*, 307–315. [CrossRef] [PubMed]

18. Hachinski, V.; Einhäupl, K.; Ganten, D.; Alladi, S.; Brayne, C.; Stephan, B.C.; Sweeney, M.D.; Zlokovic, B.; Iturria-Medina, Y.; Iadecola, C.; et al. Preventing dementia by preventing stroke: The Berlin Manifesto. *Alzheimers Dement.* **2019**, *15*, 961–984. [CrossRef] [PubMed]

19. Henon, H.; Durieu, I.; Guerouaou, D.; Lebert, F.; Pasquier, F.; Leys, D. Poststroke dementia incidence and relationship to prestroke cognitive decline. *Neurology* **2001**, *57*, 1216–1222. [CrossRef]

20. Hu, G.C.; Chen, Y.M. Post-stroke dementia: Epidemiology, mechanisms and management. *Int. J. Gerontol.* **2017**, *11*, 210–214. [CrossRef]

21. Hsieh, F.I.; Chiou, H.Y. Stroke: Morbidity, risk factors, and care in taiwan. *J. Stroke* **2014**, *16*, 59. [CrossRef]

22. Ivan, C.S.; Seshadri, S.; Beiser, A.; Au, R.; Kase, C.S.; Kelly-Hayes, M.; Wolf, P.A. Dementia after stroke: The Framingham Study. *Stroke* **2004**, *35*, 1264–1268. [CrossRef] [PubMed]

23. Kapetanakis, V.; Matthews, F.E.; Hout, A. A semi-Markov model for stroke with piecewise-constant hazards in the presence of left, right and interval censoring. *Stat. Med.* **2013**, *32*, 697–713. [CrossRef] [PubMed]

24. Kurichi, J.E.; Kwong, P.L.; Xie, D.; Bogner, H.R. Predictive indices for functional improvement and deterioration, institutionalization, and death among elderly medicare beneficiaries. *PmR* **2017**, *9*, 1065–1076. [CrossRef]

25. Lee, J.W.; Lim, H.S.; Kim, D.W.; Shin, S.A.; Kim, J.; Yoo, B.; Cho, K.H. The development and implementation of stroke risk prediction model in National Health Insurance Service's personal health record. *Comput. Methods Programs Biomed.* **2018**, *153*, 253–257. [CrossRef] [PubMed]

26. Li, T.; Horton, R.M.; Bader, D.A.; Liu, F.; Sun, Q.; Kinney, P.L. Long-term projections of temperature-related mortality risks for ischemic stroke, hemorrhagic stroke, and acute ischemic heart disease under changing climate in Beijing, China. *Environ. Int.* **2018**, *112*, 1–9. [CrossRef]

27. Lim, J.S.; Kwon, H.M.; Kim, S.E.; Lee, J.; Lee, Y.S.; Yoon, B.W. Effects of temperature and pressure on acute stroke incidence assessed using a Korean nationwide insurance database. *J. Stroke* **2017**, *19*, 295. [CrossRef]

28. Listwon, A.; Saint-Pierre, P. Semimarkov: An R package for parametric estimation in multi-state semi-markov models. *J. Stat. Softw.* **2015**, *66*, 784.

29. Listwon, A.; Saint-Pierre, P.; Listwon, M.A. Package 'SemiMarkov' 2013. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.408.2595&rep=rep1&type=pdf (accessed on 1 December 2019).

30. Ljungman, P.L.; Mittleman, M.A. Ambient air pollution and stroke. *Stroke* **2014**, *45*, 3734–3741. [CrossRef]

Int. J. Environ. Res. Public Health **2020**, *17*, 1944

13 of 13

31. Madsen, T.E.; Roberts, E.T.; Kuczynski, H.; Goldmann, E.; Parikh, N.S.; Boden-Albala, B. Gender, social networks, and stroke preparedness in the stroke warning information and faster treatment Study. *J. Stroke Cerebrovasc. Dis.* **2017**, *26*, 2734–2741. [CrossRef]

32. Makond, B.; Wang, K.J.; Wang, K.M. Probabilistic modeling of short survival in patients with brain metastasis from lung cancer. *Comput. Methods Programs Biomed.* **2015**, *119*, 142–162. [CrossRef]

33. Mijajlović, M.D.; Pavlović, A.; Brainin, M.; Heiss, W.D.; Quinn, T.J.; Ihle-Hansen, H.B.; Hermann, D.M.; Assayag, E.B.; Richard, E.; Thiel, A.; et al. Post-stroke dementia—A comprehensive review. *BMC Med.* **2017**, *15*, 11. [CrossRef] [PubMed]

34. Moulin, S.; Labreuche, J.; Bombois, S.; Rossi, C.; Boulouis, G.; Hénon, H.; Duhamel, A.; Leys, D.; Cordonnier, C. Dementia risk after spontaneous intracerebral haemorrhage: A prospective cohort study. *Lancet Neurol.* **2016**, *15*, 820–829. [CrossRef]

35. National Health Insurance Administration. 2019. Available online: https://www.nhi.gov.tw/english/ (accessed on 1 December 2019).

36. Rothwell, P.; Slattery, J.; Warlow, C.P.; Wroe, S.J. Is stroke incidence related to season or temperature? *Lancet* **1996**, *347*, 934–936. [CrossRef]

37. Roy, M.K.; Ray, A. Effect of body temperature on mortality of acute stroke. *J. Assoc. Physicians India* **2004**, *52*, 959–961.

38. Wang, K.J.; Adrian, A.M.; Chen, K.-H.; Wang, K.M. A hybrid classifier combining Borderline-SMOTE with AIRS algorithm for estimating brain metastasis from lung cancer: A case study in Taiwan. *Comput. Methods Programs Biomed.* **2015**, *119*, 63–76. [CrossRef]

39. Wang, Q.; Gao, C.; Liu, H.; Li, W.; Zhao, Y.; Xu, G.; Yan, C.; Lin, H.; Lang, L. Hypertension modifies the short-term effects of temperature on morbidity of hemorrhagic stroke. *Sci. Total Environ.* **2017**, *598*, 198–203. [CrossRef]

40. World Health Organization. 2019. Available online: http://www.who.int/mediacentre/factsheets/fs310/en/ (accessed on 1 December 2019).

41. World Heart Federation Stroke. 2019. Available online: http://www.world-heart-federation.org/cardiovascular-health/stroke/ (accessed on 1 December 2019).

42. Yang, W.S.; Wang, X.; Deng, Q.; Fan, W.Y.; Wang, W.Y. An evidence-based appraisal of global association between air pollution and risk of stroke. *Int. J. Cardiol.* **2014**, *175*, 307–313. [CrossRef]

43. Yousufuddin, M.; Bartley, A.C.; Alsawas, M.; Sheely, H.L.; Shultz, J.; Takahashi, P.Y.; Young, N.P.; Murad, M.H. Impact of multiple chronic conditions in patients hospitalized with stroke and transient ischemic attack. *J. Stroke Cerebrovasc. Dis.* **2017**, *26*, 1239–1248. [CrossRef]