Dovepress
Taylor & Francis Group

ORIGINAL RESEARCH

Uncovering Diagnostic Correlations Between Traditional Chinese and Western Medicine Using Instrumental Variable Models in Cardiometabolic Patients: Evidence from 1.2 Million Records

Jingjing Zhou [1,2], Shangcheng Zhou [3], Huijing Chen [1], Xiyin Chen [2], Guanyang Zou [1], Yulin Gao [4], Shangwen Jing [5,6], David Makram Bishai [2], Jinxin Li [7], Ailin Zhong [8], Zhenyuan Liu [2], Ailing Liu [1]

[1]School of Public Health and Management, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, People's Republic of China; [2]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, People's Republic of China; [3]School of Humanities and Management, Zhejiang Chinese Medical University, Hangzhou, Zhejiang, People's Republic of China; [4]School of Nursing, Southern Medical University, Guangzhou, Guangdong, People's Republic of China; [5]Science and Technology Innovation Center, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, People's Republic of China; [6]School of Chinese Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, People's Republic of China; [7]The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangdong Provincial Hospital of Chinese Medicine, Guangzhou, Guangdong, People's Republic of China; [8]Office of International Exchange and Cooperation, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, People's Republic of China

Correspondence: David Makram Bishai, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, G/F, Patrick Manson Building (North Wing), 7 Sassoon Road, Pokfulam, Hong Kong, People's Republic of China, Tel + 852 3917 9916, Email dbishai@hku.hk

Background: Little is known about the relationship between syndromes in Traditional Chinese Medicine (TCM) and chronic diseases coded by Western Medicine (WM). TCM hospitals where both WM and TCM are practiced offer an opportunity to assess this relationship. TCM, based on syndrome differentiation and treatment, may aid in guiding treatment and predicting length of stay and healthcare costs. However, inconsistent coding of TCM syndromes arises due to variations in diagnostic interpretation, subjective assessment, and lack of standardized coding practices. The objective was to assess the correlation between WM diagnoses and TCM syndromes, accounting for diagnostic miscoding in the data.

Methods: We examined discharge data from 1,218,224 records for patients aged 45 and above, diagnosed with cardiometabolic diseases and admitted to TCM hospitals between 2017 and 2022, stays ranging from 24-hours to 90 days. Latent class analysis (LCA) was used to measure the correlation between TCM syndromes and WM. To address potential diagnostic errors, we applied bivariate probit models with instrumental variable (IV).

Results: There were 580,698 (47.67%) records for males, while 989,702 (81.24%) records from Tertiary-A hospitals. The LCA and probit models revealed that TCM syndrome diagnoses had a high ratio of noise to signal. After correcting for diagnostic errors, significant associations were found between WM diagnoses and TCM syndromes. Notably, diabetes mellitus was strongly associated with syndrome/pattern of qi and yin deficiency (coefficient = 2.711); cerebrovascular diseases showed strong associations with syndrome/pattern of qi deficiency with blood stasis (coefficient = 2.433) and syndrome/pattern of wind and phlegm blocking collaterals (coefficient = 3.176). All patterns had strong marginal effects ($P < 0.001$).

Conclusion: This large-scale study quantitatively maps the relationship between TCM and WM diagnoses. It introduces a new statistical approach to understanding the correlation between these two diagnostic systems, offering insights into integrated medicine for secondary prevention.

Keywords: cardiometabolic diseases, western medicine, TCM syndromes, instrumental variable methods, errors in variable

3827

Background

Current research on chronic diseases is mostly based on disease classifications of Western medicine (WM).[1–3] By cross-coding leading patterns of WM with Traditional Chinese Medicine (TCM) syndromes, one can examine patterns of alignment between the TCM practitioners' syndromes and the WM syndromes. These patterns may help in clinical practice because they are the basis of syndrome differentiation and treatment, which can help enabling personalized integrative care, predicting disease severity as measured by length of stay or medical costs and improving cross-disciplinary training.[4]

TCM has a unique theoretical system in the treatment and prevention of diseases. Previous studies showed that TCM constitutions' identification can divide the population into nine types, eg1) Ping-he (balance), 2) Qi-xu (qi deficiency), 3) Yang-xu (yang deficiency), 4) Yin-xu (yin deficiency), 5) Tan-shi (phlegm and dampness), 6) Shi-re (dampness and heat), 7) Xue-yu (blood stasis), 8) Qi-yu (qi depression) and 9) Te-bing (allergic).[5] Constitutions play an important role in improving clinical treatment for related diseases.[5,6] Separate from TCM constitutions are TCM syndromes, which are classified according to etiology, pathogenesis, and clinical manifestations and are the patterns or sums of symptoms of a disease at a certain stage in diseases development process.[7] In TCM hospitals, patients with the same WM diagnosis will have different TCM syndromes because of different etiology, location, nature, and the relationship between pathogenic factors and the body's resistance in the disease.[8,9] TCM doctors' efforts in syndrome differentiation guide treatment.[7] Patients in TCM hospitals in China are treated according to the combination of their WM classified diseases and TCM syndromes using a holistic consideration. The diagnostic efforts of dual WM and TCM diagnoses routinely generates a large quantity of data about each patient. To date, there have been few statistical efforts to find the cross-correlation patterns between the TCM and Western traditions.[8,10,11]

Finding correlations in the distribution of WM diseases with TCM syndromes can provide more evidence for the effective implementation of integrated TCM and WM and provide new ideas for secondary prevention. A study in Hunan province and Hong Kong derived eight TCM pattern differentiation rules for functional dyspepsia patients, which provide new evidence for disease diagnosis and prevention.[12] Although there have been studies of the correlation between TCM syndromes and diseases, most literatures used questionnaires or data from one department in the hospital.[8,13] The healthcare records data in TCM institutions maintain dual WM codes and TCM codes for each patient. However, the TCM syndromes distribution may vary because of spatial differences of topography and climate where the patient is situated that are included in TCM diagnoses.[12] Differences in medical norms and localized medical practice cultures may also affect doctors' habits.[14] A study[15] from Shanghai showed that the TCM syndromes diagnosis were inconsistent because of variations in diagnostic interpretation across different schools of TCM, subjective syndrome assessment, or lack of standard enforcement in coding practices, etc. Therefore, we aimed to use the total TCM records in a city from 2017 to 2022 to find the correlation between WM disease diagnosis and TCM syndrome diagnosis in cardiometabolic diseases, which have the highest burden and similar risk factors.[1,16,17] To reduce the inconsistencies in TCM syndrome classification, we explored the use of Latent Class Analysis (LCA) to reveal the correlation between WM and TCM by treating each diagnosis as a latent variable.[18] We also used bivariate probit models with instrumental variables (IV)[19] to reduce the endogeneity bias and reduce measurement error so as to reveal the correlations. The ultimate goal is to provide evidence that the dual coding offers patients benefits in the form of secondary prevention derived by integrated medicine.

Methods
Data Sources

De-identified hospitalization data were obtained from the Structured Query Language (SQL) database of health-care in a city of South China. Eligible data came from hospitalization records for inpatients who had been admitted to one of 25 different TCM institutions from 2017 to 2022. Inclusion criteria were records for patients aged 45 or above with length of stay for at least 24-hours and no more than 90 days. Patients had to have been

diagnosed with hypertension diseases (HTN), type 2 diabetes mellitus (DM), disorders of lipoprotein metabolism (DLM), coronary heart diseases (CHD), heart failure (CHF), cerebrovascular diseases (CVD). These conditions corresponded to ICD-10 codes (I10-I13, I20-I25, I50, I60-I66, I69, E11, E12, E14, E78).

In order to employ the instrumental variable method, a set of exogenous predictor variables was needed. These instrumental variables included the hospital's observed proportion of each TCM diagnosis by gender. We also used air quality, temperature, Relative Humidity (RH), and day of the week of the admission as covariate variables. Air quality data came from the National urban air quality real-time release platform of China (https://air.cnemc.cn:18007/). Temperature data came from the National Climatic Data Center (NCDC, https://www.ncei.noaa.gov/pub/data/noaa/isd-lite/), part of the National Oceanic and Atmospheric Administration (NOAA). RH came from the following formula.[20]

$$RH = 100 * \left[\frac{e^{\frac{17.625*Dp}{243.04+Dp}}}{e^{\frac{17.625*T}{243.04+T}}} \right]$$

where Dp means the average dew point temperature (degree Celsius), T means the average temperature (degree Celsius).

Data Analysis

We used R software[21] version 4.4.1 to conduct data cleaning, database administration, descriptive analysis, and LCA. We used Stata MP[22] version 18 to conduct the Wilcox chi-squared test, Kruskal–Wallis's chi-squared test, probit analysis, and bivariate probit analysis.

Tidyverse package[23] in R software was used to clean the data and select the records with the diagnosis of cardiometabolic diseases. The flow chart of data cleaning is shown in Figure 1. Then, we standardized the TCM syndromes (Supplementary Table 1) entered by the TCM practitioners in each record according to the terminology released in 1995 (GB1995) and revised in 2021 (GB2021)[24] and by consulting the doctors in TCM hospitals.

We deleted improper records of healthcare costs by reference to a table of basic healthcare costs and remove the records of 69 patients who had each been admitted more than 50 times. As the amount of TCM syndromes missing data was relatively small (177 missing data in 1,218,224 records) and no meaningful way to impute the missing values was available based on the existing records, we used the removing setting in the bivariate probit analysis. Descriptive statistical analysis was used for general information. Numbers and percentages were used to describe categorical data. The *mean ± sd* was used to summarize continuous variables that followed a normal distribution. The median and interquartile range were used for variables that did not follow a normal distribution.

We firstly used the LCA method[25] to measure the correlation between TCM syndromes and WM and divide the records into different classes using the *poLCA* package. We analyzed the latent class for pure WM diagnoses, pure WM diagnoses with 18 TCM syndromes, WM diagnoses and added the TCM syndromes one by one according to the proportion and suggestion from TCM doctors. The overall model fits were reported with Akaike information criterion (AIC), Bayesian information criterion (BIC), Likelihood ratio (G^2), Chi-square goodness of fit (X^2), and P value. We set the maximum number of iterations = 3000, tolerance value = 1e-10, na.rm = FALSE, times of model is estimated (nrep) = 10. P value was determined using the likelihood ratio test, which tests the improvement in fit of a k class model with k–1 class model. We searched for classifications in which adding TCM diagnoses improved the AIC, BIC, and G^2 of the overall classification. We also used an ordinary probit model[26] in Stata to assess the correlation between WM and TCM syndromes.

We were concerned that the TCM syndromes might have suffered from miscoding leading to the well-known errors in the variables (EIV) problem.[27] When independent variables are measured with imprecision, coefficients are biased towards zero because noise dominates signal in the data. A well-known solution to EIV bias is to use a set of instrumental variables Z that would be correlated with the TCM syndromes (used here as independent variables) directly, but not with the dependent variables (ie the WM diagnoses). By predicting error miscoded TCM syndromes from the instruments Z, much of the noise can be removed provided there is a good fit in the instrumental variable step. Because the WM diagnosis and TCM syndromes were all dichotomous, we used a bivariate probit model.[27–29] The bivariate probit model is a joint model for two binary dependent variables that
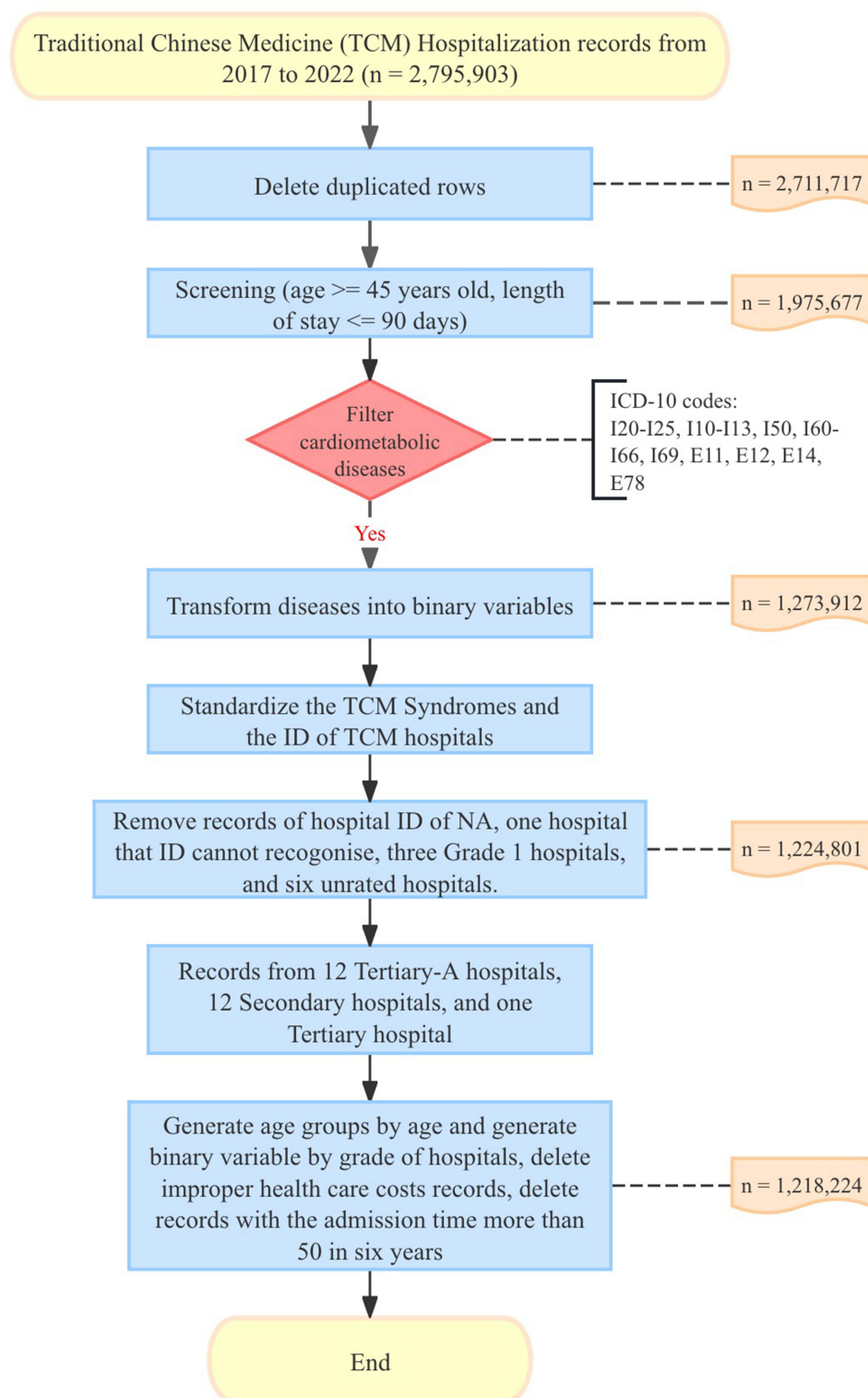
Figure 1 Flow chart of data cleaning.

finds the correlation index from one latent variable to two latent variables,[29] and the formulas for two latent variables are presented as[30]

$$[1a]\ WM_j^* = x_1\beta_1 + TCM_k\beta_2 + \varepsilon_1$$

$$[1b]\ TCM_k^* = x_1\gamma_2 + z_3\gamma_3 + \varepsilon_2$$

Where j is the WM diagnosis from 1 to 6, k is the TCM syndrome diagnosis from 1 to 20, ε_1 and ε_2 are joint normal with means were zero, variances were one, and correlation rho ρ). The bivariate probit model outcome formulas were[30]

$$WM_j = \begin{cases} 1\ if\ WM_j^* > 0 \\ 0\ if\ WM_j^* \leq 0 \end{cases}$$

$$TCM_k = \begin{cases} 1\ if\ TCM_k^* > 0 \\ 0\ if\ TCM_k^* \leq 0 \end{cases}$$

The bivariate probit model is estimated as a joint system using maximum likelihood. Notice that TCM is an independent variable in Equation [1a] and a dependent variable in Equation [1b]. The Z variables that predict TCM in Equation [1b] are excluded in [1a] and serve to reduce the noise in the raw measures of TCM. For instruments, we chose to use the hospital's observed proportion of each TCM diagnosis by gender. We defend the exclusion of these Z variables from [1a] by asserting that TCM hospitals may have a local culture with peer effects that make certain TCM syndromes more or less common.

The marginal effects were calculated for the joint probability by the formula

$$Marginal\ effects = E_{WM_j} / \Delta_{TCM_k}$$

Where E_{WM_j} means the expected value of the dependent variable WM_j, Δ_{TCM_k} means the change in the dependent variable TCM_k.

The covariate variables (x variables) included were age group, the daily mean value of carbon monoxide (CO), the daily mean value of ozone (O_3), daily average temperature (TEMP), dew point temperature (DEWP), admission time from 2017 to 2022, day of week, holiday or not, RH, and PM2.5. The overall fit of the models was estimated with coefficients, marginal effects, Wald $\chi2$ and P value. Then, we compared the correlations between WM and TCM, and chose the WM-TCM patterns. The rho statistic in a bivariate probit model assesses the correlation between error terms in [1a] and [1b]. If rho is statistically no different from zero, then the bivariate model offers no gain in statistical efficacy and the model could just as well have been estimated as separate equations. Essentially rho = 0 is consistent with the null hypothesis of no EIV in the coding of TCM diagnoses. For bivariate probit models, there is no method to test its overidentification. We applied a linear probability two-stage model to the dichotomous diagnostic data in order to perform an overidentification test using estat overid after running the model with *ivreg2*, and the weak instrument test was used to evaluate the explanatory power of the instrumental variables at the first stage.[31] The modeling sequence for statistical analysis was shown in Figure 2.

Results

Basic Information of Hospitalized Records

Among 1,975,677 TCM hospitalized records with patients aged 45 or above and length of stay in hospitals from 24 hours to 90 days, there were 1,273,912 (64.48%) records with diagnosis of cardiometabolic diseases from 2017 and 2022. There were 1,218,224 episodes in total with 580,698 (47.67%) for males and 637,526 (52.33%) for females, while 989,702 (81.24%) were from Tertiary-A hospitals and 228,522 (18.77) were not from Tertiary-A hospitals. The average age of subjects was 69.42 ± 12.106 for all, 68.14 ± 11.957 for males, and 70.59 ± 12.132 for females. Among all the cardiometabolic diseases, patients with HTN had the most records (833,881, 68.45%), followed by DM (403,139, 33.09%), CVD (395,835, 32.49%), CHD (279,845, 22.97%), DLM (256,600, 20.98%), CHF (140,323, 11.52%). Among all the TCM syndromes, the syndrome/pattern of qi stagnation and blood stasis (QSBS, 119,345, 9.74%) had
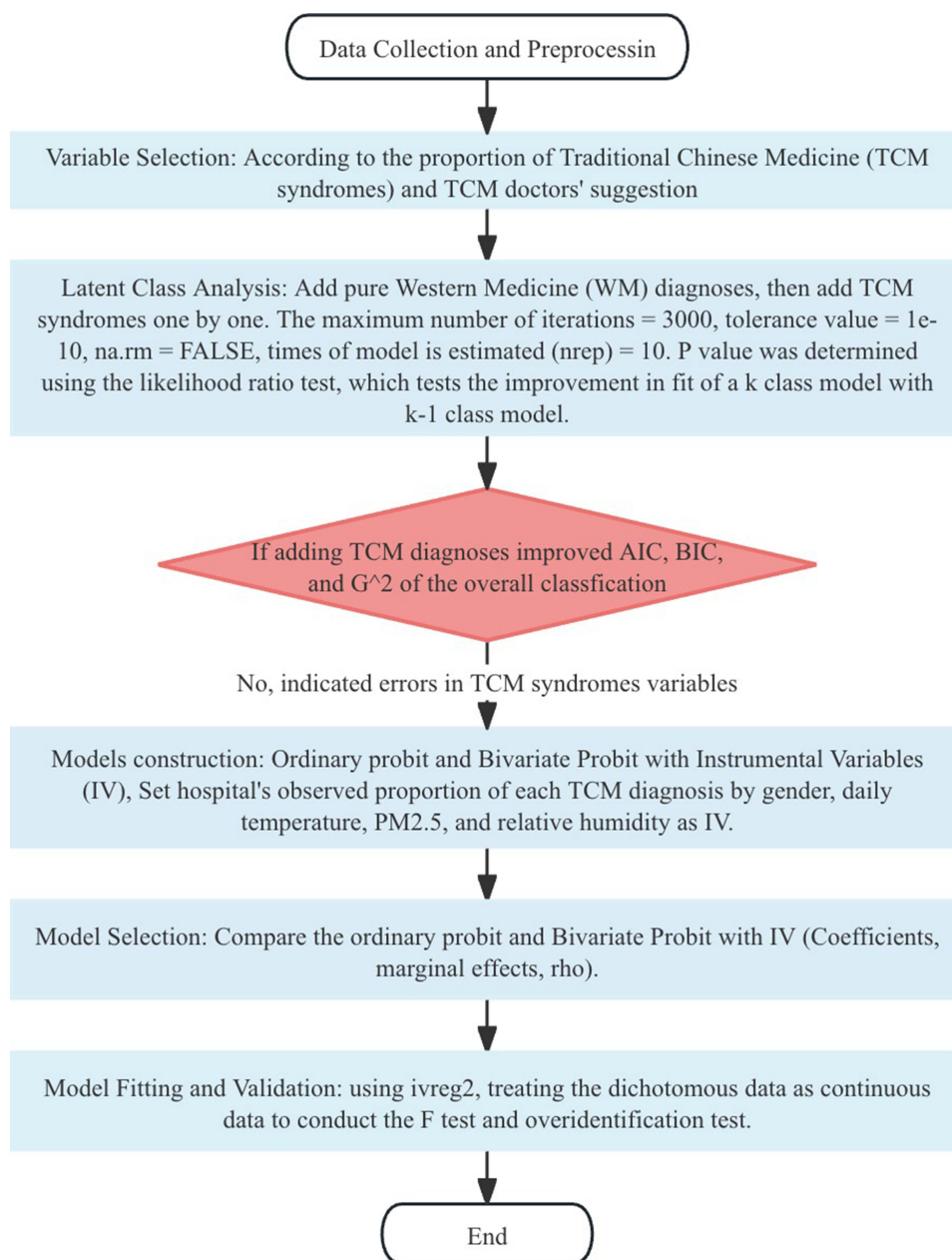
Figure 2 Flow diagram of the modeling sequence for statistical analysis.

the most records, followed by the syndrome/pattern of qi deficiency with blood stasis (QDBS, 105,976, 8.65%), syndrome/pattern of dual deficiency of liver and kidney (DDLK, 104,079, 8.50%). Shown in Table 1.

The Leading Traditional Chinese Medicine Syndromes in Each Cardiometabolic Disease

The leading TCM syndromes in cases with the Western medicine diagnosis of DLM was QSBS (27,405, 2.25%), followed by the syndrome/pattern of qi deficiency with phlegm binding (QDPB, 23,968, 1.97%) and DDLK (21,621, 1.77%). The leading TCM syndromes when the Western medicine diagnosis was HTN was also QSBS (80,255, 6.59%), followed by QDBS (77,833, 6.39%) and DDLK (76,392, 6.27%). The leading TCM syndromes in diabetes mellitus were QDBS (35,591, 2.92%), followed by QSBS (34,770, 2.85%) and DDLK (31,930, 2.62%). The leading TCM syndromes

Table 1 Basic Information of 1,218,224 Hospitalized Records

	Total (n = 1,218,224)	Male (n = 580,698, 47.67%)	Female (n = 637,526, 52.33%)	Tertiary-A Hospitals (n = 989,702, 81.24%)	Other Hospitals (n = 228,522, 18.76%)
Episodic age (x ± s)	69.42 ± 12.106	68.14 ± 11.957	70.59 ± 12.132	68.37 ± 11.858	73.98 ± 12.115
Age groups (n, %)					
45–59 years old	284,587 (23.36)	152,249 (12.50)	132,338 (10.86)	251,387 (20.64)	33,200 (2.73)
45–49 years old	63,218 (5.19)	36,823 (3.02)	26,395 (2.17)	56,675 (4.65)	6,543 (0.54)
50–54 years old	95,049 (7.80)	50,412 (4.14)	44,637 (3.66)	84,068 (6.90)	10,981 (0.90)
55–59 years old	126,320 (10.37)	65,014 (5.34)	61,306 (5.03)	110,644 (9.08)	15,676 (1.29)
60–79 years old	637,073 (52.30)	309,438 (25.40)	327,635 (26.89)	532,513 (43.71)	104,560 (8.58)
60–64 years old	155,723 (12.78)	79,064 (6.49)	76,659 (6.29)	134,521 (11.04)	21,202 (1.74)
65–69 years old	180,461 (14.81)	89,154 (7.32)	91,307 (7.50)	153,956 (12.64)	26,505 (2.18)
70–74 years old	159,672 (13.11)	76,114 (6.25)	83,558 (6.86)	132,718 (10.89)	26,954 (2.21)
75–79 years old	141,217 (11.59)	65,106 (5.34)	76,111 (6.25)	111,318 (9.14)	29,899 (2.45)
80–99 years old	295,736 (24.28)	118,767 (9.75)	176,969 (14.53)	205,372 (16.86)	90,364 (7.42)
80–84 years old	146,056 (11.99)	61,049 (5.01)	85,007 (6.98)	106,341 (8.73)	39,715 (3.26)
85–89 years old	104,014 (8.54)	41,140 (3.38)	62,874 (5.16)	69,998 (5.75)	34,016 (2.79)
90–94 years old	37,951 (3.12)	14,197 (1.17)	23,754 (1.95)	24,269 (1.99)	13,682 (1.12)
95–99 years old	7,715 (0.63)	2,381 (0.20)	5,334 (0.44)	4,764 (0.39)	2,951 (0.24)
100 years old or above (n, %)	828 (0.07)	244 (0.02)	584 (0.05)	430 (0.04)	398 (0.03)
Inpatient of cardiometabolic diseases by year (n, %)					
2017	181,381 (14.89)	84,269 (6.92)	97,112 (7.97)	141,482 (11.61)	39,899 (3.28)
2018	187,825 (15.42)	87,876 (7.21)	99,949 (8.20)	158,688 (13.03)	29,137 (2.39)
2019	213,933 (17.56)	101,296 (8.32)	112,637 (9.25)	183,163 (15.04)	30,770 (2.53)
2020	208,266 (17.10)	99,527 (8.17)	108,739 (8.93)	162,968 (13.38)	45,298 (3.72)
2021	221,886 (18.21)	107,265 (8.81)	114,621 (9.41)	178,207 (14.63)	43,679 (3.59)
2022	204,933 (16.82)	100,465 (8.25)	104,468 (8.58)	165,194 (13.56)	39,739 (3.26)
Inpatient of cardiometabolic diseases (n, %) [a]					
Coronary heart disease	**279,845 (22.97)**	**139,299 (11.43)**	**140,546 (11.54)**	**23,464 (18.34)**	**56,381 (4.63)**
Angina	43,620 (3.58)	19,885 (1.63)	23,735 (1.95)	37,732 (3.10)	5,888 (0.48)
Myocardial infarction	14,576 (1.20)	9,971 (0.82)	4,605 (0.38)	13,554 (1.11))	1,022 (0.08)
Recurrent myocardial infarction	1,738 (0.14)	932 (0.08)	806 (0.07)	1,135 (0.09)	603 (0.05)
Complications of myocardial infarction	404 (0.03)	244 (0.02)	160 (0.01)	276 (0.02)	128 (0.01)
Coronary thrombosis	11,425 (0.94)	5,977 (0.49)	5,448 (0.45)	10,222 (0.84)	1,203 (0.10)
Arteriosclerotic cardiovascular disease	253,497 (20.81)	126,282 (10.37)	127,215 (10.44)	201,425 (16.53)	52,072 (4.27)
Hypertensive diseases	833,881 (68.45)	391,085 (32.10)	442,796 (36.35)	673,724 (55.30)	160,157 (13.14)
Heart failure	140,323 (11.52)	69,696 (5.72)	70,627 (5.80)	115,360 (9.47)	24,963 (2.04)
Diabetes mellitus	403,139 (33.09)	194,091 (15.93)	209,048 (17.16)	332,349 (27.28)	70,790 (5.81)
Disorders of lipoprotein metabolism	256,600 (20.98)	103,761 (8.52)	151,839 (12.46)	217,458 (17.85)	38,142 (3.13)
Cerebrovascular diseases	**395,835 (32.49)**	**212,645 (17.46)**	**183,190 (15.04)**	**296,648 (24.35)**	**99,187 (8.14)**
Hemorrhagic cerebrovascular diseases	33,619 (2.76)	20,704 (1.70)	12,915 (1.06)	28,480 (2.34)	5,139 (0.42)
Ischemic cerebrovascular diseases	271,100 (22.25)	144,991 (11.90)	126,109 (10.35)	209,159 (17.17)	61,941 (5.08)
Sequela of cerebrovascular disease	134,751 (11.06)	74,163 (6.09)	60,588 (4.97)	92,230 (7.57)	42,521 (3.49)
Number of diseases, N (%) 1	514,223 (42.21)	237,030 (19.46)	277,193 (22.75)	425,363 (34.92)	88,860 (7.29)
2	405,528 (33.29)	198,609 (16.30)	206,919 (16.99)	328,117 (26.93)	77,411 (6.35)
3	220,380 (18.09)	108,411 (8.90)	111,969 (9.19)	175,225 (14.38)	45,155 (3.71)
4	68,521 (5.62)	32,262 (2.65)	36,259 (2.98)	53,471 (4.39)	15,050 (1.24)
5	9,312 (0.76)	4,268 (0.35)	5,044 (0.41)	7,309 (0.60)	2,003 (0.16)
6	260 (0.02)	118 (0.01)	142 (0.01)	217 (0.02)	43 (0.00)
Episodes from Tertiary-A hospitals Yes	989,702 (81.24)	485,937 (39.89)	503,765 (41.35)	/	/
No	228,522 (18.76)	94,761 (7.78)	133,761 (10.98)	/	/

(Continued)

Table 1 (Continued).

	Total (n = 1,218,224)	Male (n = 580,698, 47.67%)	Female (n = 637,526, 52.33%)	Tertiary-A Hospitals (n = 989,702, 81.24%)	Other Hospitals (n = 228,522, 18.76%)
Episodic main TCM Syndromes [b]	**719,935 (59.10)**	**332,078 (27.26)**	**387,857 (31.84)**	**575,210 (47.22)**	**144,725 (11.88)**
QSBS	119,173 (9.78)	43,833 (3.60)	75,340 (6.18)	83,161 (6.83)	36,012 (2.96)
QDBS	105,205 (8.64)	50,298 (4.13)	54,907 (4.51)	68,066 (5.59)	37,139 (3.05)
DDLK	103,585 (8.50)	31,061 (2.55)	72,524 (5.95)	86,986 (7.14)	16,599 (1.36)
QDPB	77,456 (6.36)	42,270 (3.47)	35,186 (2.89)	76,735 (6.30)	721 (0.06)
PSOC	46,038 (3.78)	28,167 (2.31)	17,871 (1.47)	43,383 (3.56)	2,655 (0.22)
PSOI	36,043 (2.96)	18,261 (1.50)	17,782 (1.46)	22,036 (1.81)	14,007 (1.15)
DDQB	32,666 (2.68)	8,539 (0.70)	24,127 (1.98)	26,202 (2.15)	6,464 (0.53)
DDSK	29,246 (2.40)	14,490 (1.19)	14,756 (1.21)	24,639 (2.02)	4,607 (0.38)
WPBC	27,263 (2.24)	16,861 (1.38)	10,402 (0.85)	19,382 (1.59)	7,881 (0.65)
QYD	24,498 (2.01)	11,559 (0.95)	12,939 (1.06)	18,566 (1.52)	5,932 (0.49)
TPOL	20,678 (1.70)	13,076 (1.07)	7,602 (0.62)	14,783 (1.21)	5,895 (0.48)
SDDS	19,218 (1.58)	9,742 (0.80)	9,476 (0.78)	19,136 (1.57)	82 (0.01)
KDC	17,692 (1.45)	8,390 (0.69)	9,302 (0.76)	15,605 (1.28)	2,087 (0.17)
WPIVC	16,954 (1.39)	10,603 (0.87)	6,351 (0.52)	16,482 (1.35)	472 (0.04)
DHPD	16,257 (1.33)	9,627 (0.79)	6,630 (0.54)	14,081 (1.16)	2,176 (0.18)
SDDE	15,841 (1.30)	8,087 (0.66)	7,754 (0.64)	14,786 (1.21)	1,055 (0.09)
DDLS	12,122 (1.00)	7,214 (0.59)	4,908 (0.40)	11,181 (0.92)	941 (0.08)

Notes: [a]Without mutually exclusive; [b]Only calculated the syndromes accounting for one percent or above, 177 missing data. Values in bold indicate larger categories of disease.
Abbreviations: TCM, Traditional Chinese Medicine; QSBS, syndrome/pattern of qi stagnation and blood stasis; QDBS, syndrome/pattern of qi deficiency with blood stasis; DDLK, syndrome/pattern of dual deficiency of liver and kidney; QDPB, syndrome/pattern of qi deficiency with phlegm binding; PSOC, syndrome/pattern of phlegm and stasis obstructing the collaterals; PSOI, syndrome/pattern of phlegm and stasis obstructing the internal; DDQB, syndrome/pattern of dual deficiency of qi and blood; DDSK, syndrome/pattern of dual deficiency of spleen and kidney; WPBC, syndrome/pattern of wind and phlegm blocking collaterals; QYD, syndrome/pattern of qi and yin deficiency; TPOL, syndrome/pattern of turbid phlegm obstructing the lung; SDDS, syndrome/pattern of spleen deficiency with dampness obstruction; KDC, syndrome/pattern of kidney deficiency with complex; WPIVC, syndrome/pattern of wind and phlegm stagnating and impeding vessels and collaterals; DHPD, syndrome/pattern of dampness and heat pouring downwards; SDDE, syndrome/pattern of spleen deficiency with dampness encumbrance; DDLS, syndrome/pattern of dual deficiency in the lung and spleen.

in heart failure were QDBS (15,727, 1.29%), followed by QDPB (7,863, 0.65%) and syndrome/pattern of dual deficiency of spleen and kidney (DDSK, 6,609, 0.54%). The leading TCM syndromes in CHD were QDPB (34,043, 2.79%), followed by QDBS (27,949, 2.29%) and QSBS (20,483, 1.68%). The leading TCM syndromes in CVD were QDBS (75,909, 6.23%), followed by syndrome/pattern of phlegm and stasis obstructing the collaterals (PSOC, 34,904, 2.87%) and QDPB (28,681, 2.35%). The proportion of leading TCM syndromes in different cardiometabolic diseases is shown in Table 2.

Correlations Between Western Medicine Diagnosis and TCM Syndromes Diagnosis

The LCA models showed that the six diseases model had the best fit statistics. Adding the TCM syndrome data did nothing to improve the statistical fit of the classification. This is consistent with our concern that the TCM syndrome data had a high ratio of noise to signal. See Supplementary Table 2–Table 21.

The probit models and bivariate probit models with IV showed that almost all absolute values of ordinary probit coefficients were smaller than coefficients from bivariate probit models with IV, except that the coefficient between CHD and syndrome/pattern of heart blood stasis obstruction (HBSO, ordinary coefficient = 1.775, $P < 0.001$; Bivariate coefficient = 1.149, $P < 0.001$), or the coefficient between CHD and the syndrome/pattern of heart yang insufficiency (HYI, ordinary coefficient = 1.577, $P < 0.001$; Bivariate coefficient = 1.048, $P < 0.001$), and CVD with syndrome/pattern of wind and phlegm stagnating and impeding vessels and collaterals (WPIVC, ordinary coefficient = 1.987, $P < 0.001$; Bivariate coefficient = 0.066, $P < 0.001$). This indicates that the bias in the raw data is "towards zero" which is also consistent with regressors being contaminated with noise. The F-tests indicated that the instrumental variables are strongly correlated with TCM syndromes ($F > 10$). The overidentification test showed that there is evidence of overidentification, which may indicate potential issues with instrument validity, but it is hard to be sure because the linear approximation is imperfect. Shown in Supplementary Table 22–Supplementary Table 27.

Table 2 The Leading Traditional Chinese Medicine Syndromes in Cardiometabolic Diseases

Order	Disorders of Lipoprotein Metabolism		Hypertensive Diseases		Diabetes Mellitus		Heart Failure		Coronary Heart Disease		Cerebrovascular Diseases	
	Syndromes	n, %	Syndromes	n, %	Syndromes	n, %	Syndromes	n, %	Syndromes	n, %	Syndromes	n, %
1	QSBS	27,405 (2.25)	QSBS	80,255 (6.59)	QDBS	35,591 (2.92)	QDBS	15,727 (1.29)	QDPB	34,043 (2.79)	QDBS	75,909 (6.23)
2	QDPB	23,968 (1.97)	QDBS	77,833 (6.39)	QSBS	34,770 (2.85)	QDPB	7,863 (0.65)	QDBS	27,949 (2.29)	PSOC	34,904 (2.87)
3	DDLK	21,621 (1.77)	DDLK	76,392 (6.27)	DDLK	31,930 (2.62)	DDSK	6,609 (0.54)	QSBS	20,483 (1.68)	QDPB	28,681 (2.35)
4	QDBS	15,733 (1.29)	QDPB	50,760 (4.16)	QDPB	21,662 (1.78)	QSBS	6,001 (0.49)	DDLK	17,414 (1.43)	WPBC	24,829 (2.04)
5	DDQB	9,556 (0.78)	PSOC	35,011 (2.87)	QYD	15,866 (1.30)	TPOL	5,768 (0.47)	PSOI	9,895 (0.81)	QSBS	17,978 (1.48)
6	PSOI	7,891 (0.65)	PSOI	26,081 (2.14)	PSOC	15,694 (1.29)	PSOI	4,709 (0.39)	PSOC	9,183 (0.75)	DDLK	16,088 (1.32)
7	PSOC	7,064 (0.58)	DDSK	21,288 (1.75)	DDSK	12,376 (1.02)	PSOC	4,681 (0.38)	HBSO	7,525 (0.62)	QPIVC	15,635 (1.28)
8	SDDS	6,146 (0.50)	DDQB	20,273 (1.66)	PSOI	11,258 (0.92)	QYD	4,508 (0.37)	DDSK	6,296 (0.52)	PSOI	15,065 (1.24)
9	WPBC	5,860 (0.48)	WPBC	20,154 (1.65)	DDQB	10,115 (0.83)	DDLK	4,401 (0.36)	TPOL	5,897 (0.48)	DDQB	9,118 (0.75)
10	KDC	5,325 (0.44)	QYD	14,485 (1.19)	WPBC	8,422 (0.69)	HBSO	2,867 (0.24)	QYD	5,724 (0.47)	TPOL	7,075 (0.58)
11	SDDE	4,708 (0.39)	TPOL	14,273 (1.17)	SDDS	6,349 (0.52)	DDQB	2,638 (0.22)	DDQB	5,210 (0.43)	DDSK	6,016 (0.49)
12	QPIVC	4,252 (0.35)	KDC	11,767 (0.97)	SDDE	5,859 (0.48)	HYI	2,496 (0.20)	HYI	3,672 (0.30)	LYH	5,773 (0.47)
13	QYD	4,215 (0.35)	QPIVC	11,719 (0.96)	QPIVC	5,390 (0.44)	YQD	2,369 (0.19)	WPBC	3,166 (0.26)	QYD	5,087 (0.42)
14	SSD	4,110 (0.34)	SDDS	11,491 (0.94)	TPOL	5,190 (0.43)	DEOQDSK	2,284 (0.19)	DHS	3,256 (0.26)	WPAU	5,025 (0.41)
15	DDSK	3,894 (0.32)	DHPD	11,293 (0.93)	DHPD	5,048 (0.41)	DDHL	2,093 (0.17)	BBH	2,737 (0.22)	PDAL	4,299 (0.35)

Abbreviations: QSBS, syndrome/pattern of qi stagnation and blood stasis; QDBS, syndrome/pattern of qi deficiency with blood stasis; DDLK, syndrome/pattern of dual deficiency of liver and kidney; QDPB, syndrome/pattern of qi deficiency with phlegm binding; PSOC, syndrome/pattern of phlegm and stasis obstructing the collaterals; PSOI, syndrome/pattern of phlegm and stasis obstructing the internal; DDQB, syndrome/pattern of dual deficiency of qi and blood; DDSK, syndrome/pattern of dual deficiency of spleen and kidney; WPBC, syndrome/pattern of wind and phlegm blocking collaterals; QYD, syndrome/pattern of qi and yin deficiency; TPOL, syndrome/pattern of turbid phlegm obstructing the lung; SDDS, syndrome/pattern of spleen deficiency with dampness obstruction; KDC, syndrome/pattern of kidney deficiency with complex; DHPD, syndrome/pattern of dampness and heat pouring downwards; SDDE, syndrome/pattern of spleen deficiency with dampness encumbrance; DDLS, syndrome/pattern of dual deficiency in the lung and spleen; HBSO, syndrome/pattern of heart blood stasis obstruction; LYH, syndrome/pattern of liver yang hyperactivity; HYI, syndrome/pattern of heart yang insufficiency; SSD, syndrome/pattern of spleen and stomach deficiency; YQD, syndrome/pattern of yang qi deficiency; DEOQDSK, syndrome/pattern of dampness-turbidity encumbrance and obstruction and qi deficiency in spleen and kidney; DDHL, syndrome/pattern of dual deficiency of the heart and lung; DHS, syndrome/pattern of dampness and heat stagnation; BBH, syndrome/pattern of blood binding with heat; WPAU, syndrome/pattern of wind and phlegm attacking the upper; PDAL, syndrome/pattern of phlegm and dampness accumulating in the lung.

The bivariate probit models with IV showed that the standout patterns linking WM diagnoses and TCM syndromes were a connection between HTN with syndrome/pattern of liver yang hyperactivity (LYH, coefficient = 2.306, $P < 0.001$); CHF with syndrome/pattern of heart yang insufficiency (HYI, coefficient = 1.911, $P < 0.001$), or with syndrome/pattern of qi and yin deficiency (QYD, coefficient = 1.531, $P < 0.001$); DM with QYD (coefficient = 2.711, $P < 0.001$); DLM with syndrome/pattern of dual deficiency of qi and blood (DDQB, coefficient = 2.088, $P < 0.001$), or with syndrome/pattern of spleen deficiency with dampness obstruction (SDDS, coefficient = 1.616, $P < 0.001$), or with syndrome/pattern of spleen deficiency with dampness encumbrance (SDDE, coefficient = 1.575, $P < 0.001$); CVD with syndrome/pattern of wind and phlegm blocking collaterals (WPBC, coefficient = 3.176, $P < 0.001$), or with QDBS (coefficient = 2.433, $P < 0.001$), or with syndrome/pattern of phlegm and stasis obstructing the collaterals (PSOC, coefficient = 2.403, $P < 0.001$), or with syndrome/pattern of phlegm and stasis obstructing the internal (PSOI, coefficient = 2.031, $P < 0.001$). All the patterns above had strong marginal effects (5.105~27.846, $P < 0.001$). See Table 3, Figure 3 and Supplementary Table 28.

Discussion

Using data on TCM hospitalization records in middle-aged and older adults in a city of South China, we found correlation patterns between WM diagnoses and TCM syndromes in cardiometabolic diseases. These findings reflect how changes in TCM syndromes may influence variations in WM diagnoses, offering insights into the dynamic relationship between the two systems.

In our study, the most common TCM syndromes in our records for patients with cardiometabolic diseases were QSBS, QDBS and DDLK. A related study in a hospital of north China showed that the leading TCM syndromes in heart diseases were QDBS, QDPB, PSOI, and HBSO, which is partly consistent with our study.[11] According to TCM theory, the TCM syndromes distribution in space will vary because of spatial differences of topography and climate. Locations also vary in medical norms and medical practice cultures that may also affect doctors' habits in how they diagnose TCM syndromes.[14] According to WM, CHD, CVD, and CHF share common risk factors of HTN, DM, and DLM.[32,33] We found that HTN, DM, and DLM had a large quantity of records, and the distribution showed spatial variation. Because our data came from 25 different hospitals, there may be patterns due to hospital based norms and cultures within how practitioners interpret and apply TCM theory.[14] The LCA models were the first attempt to reveal the correlation between WM and TCM, by treating each diagnosis as a latent variable.[18] The pure WM LCA results showed a good fit. However, when we added TCM syndromes into LCA models, the statistical values of tests of fit such as AICs, BICs, and likelihood ratios became worse. This statistical finding indicates that the TCM syndromes contained substantial noise reflecting statistical errors in the recording of TCM syndrome diagnosis.[27]

The probit models showed that the TCM syndromes were correlated with WM diagnosis, however the ordinary probit coefficients were attenuated and biased towards zero as would be expected if there were a high degree of mis-coding. This may have occurred because of the complexity of diagnosis, different theories and different doctor habit in TCM diagnosis bring challenge to reveal the correlations of TCM and WM. There are some competing theories in TCM,[34] resulting in different school of thought, and different hospitals may have the central doctor or teacher spread an alternative school of thought.

To correct for measurement error, we predicted each patient's assigned TCM syndrome in their different hospitals using instrumental variables (IV) calibrated to diagnostic patterns seen in each hospital. These prediction models produced diagnostic tests (F-Statistics) showing they were statistically justified.[19] Moreover, even though there can also be miscoding of syndromes in WM, it was generally more reliable in our analysis because the dataset was restricted to of ICD-10 codes for CHF, CHD, CVD, DM, HTN and DLM. The bivariate probit with IV showed that different WM diagnoses were correlated with selected specific TCM syndromes. The marginal effects after bivariate probit showed that LYH was statistically significantly associated with HTN, the same finding held for HYI and QYD in CHF, QYD in DM, DDQB, SDDS, and SDDE in DLM, WPBC, QDBS, and PSOI in CVD with the coefficients of bivariate probit were larger than 1.5. The largest effects were seen for WPBC and CVD, QYD and DM, and QDBS and CVD, so we will now discuss the potential significance of these results.

Table 3 Coefficients of Bivariate Probit Models and Marginal Effects

Diagnostic Patterns	Ordinary Probit		Bivariate Probit					Marginal Effects	
	Coefficient [SE]	95% CI	Coefficient [SE]	95% CI	*Wald*	Athrho	Rho	Ey/dx [SE]	95% CI
Hypertension diseases - LYH	0.845** [0.055]	0.737~0.953	2.306** [0.023]	2.261~2.350	52,352.90**	−0.849**	−0.690**	7.312** [0.336]	6.654~7.970
Heart failure - HYI	1.307** [0.319]	0.682~1.932	1.911** [0.038]	1.836~1.987	53,946.37**	−0.258**	−0.252**	5.105** [0.209]	4.696~5.515
Heart failure - QYD	0.294** [0.096]	0.106~0.483	1.531** [0.044]	1.446~1.617	59,386.70**	−0.535**	−0.489**	5.404** [0.313]	4.791~6.018
Diabetes mellitus - QYD	0.839** [0.130]	0.584~1.094	2.711** [0.013]	2.686~2.736	98,650.84**	−1.338**	−0.871**	27.846** [1.531]	24.848~30.845
Disorders of lipoprotein metabolism - DDQB	0.319* [0.147]	0.030~0.607	2.088** [0.016]	2.057~2.120	139,442.91**	−1.002**	−0.763**	13.386** [0.421]	12.560~14.212
Disorders of lipoprotein metabolism - SDDS	0.245* [0.103]	0.044~0.446	1.616** [0.015]	1.586~1.645	106,474.83**	−0.765**	−0.644**	8.341** [0.223]	7.903~8.777
Disorders of lipoprotein metabolism - SDDE	0.200** [0.063]	0.076~0.324	1.575** [0.020]	1.535~1.614	103,123.34**	−0.648**	−0.570**	6.192** [0.181]	5.837~6.547
Cerebrovascular diseases - WPBC	1.922** [0.113]	1.701~2.143	3.176** [0.018]	3.141~3.210	115,829.92**	−0.915**	−0.724**	16.291** [0.801]	14.721~17.860
Cerebrovascular diseases - QDBS	1.094** [0.139]	0.822~1.366	2.433** [0.006]	2.422~2.444	387,090.50**	−1.228**	−0.842**	16.496** [0.371]	15.769~17.223
Cerebrovascular diseases - PSOC	1.244** [0.148]	0.954~1.534	2.403** [0.009]	2.386~2.420	212,996.33**	−0.869**	−0.701**	10.784** [0.183]	10.426~11.142
Cerebrovascular diseases - PSOI	0.262 [0.148]	−0.028~0.553	2.031** [0.011]	2.009~2.052	150,356.05**	−1.171**	−0.825**	14.955** [0.425]	14.124~15.785

Notes: **p < 0.01, *p < 0.05.
Abbreviations: TCM, Traditional Chinese Medicine; IV, instrumental variable; QDBS, syndrome/pattern of qi deficiency with blood stasis; PSOC, syndrome/pattern of phlegm and stasis obstructing the collaterals; PSOI, syndrome/pattern of phlegm and stasis obstructing the internal; DDQB, syndrome/pattern of dual deficiency of qi and blood; WPBC, syndrome/pattern of wind and phlegm blocking collaterals; QYD, syndrome/pattern of qi and yin deficiency; SDDS, syndrome/pattern of spleen deficiency with dampness obstruction; SDDE, syndrome/pattern of spleen deficiency with dampness encumbrance; LYH, syndrome/pattern of liver yang hyperactivity; HYI, syndrome/pattern of heart yang insufficiency.
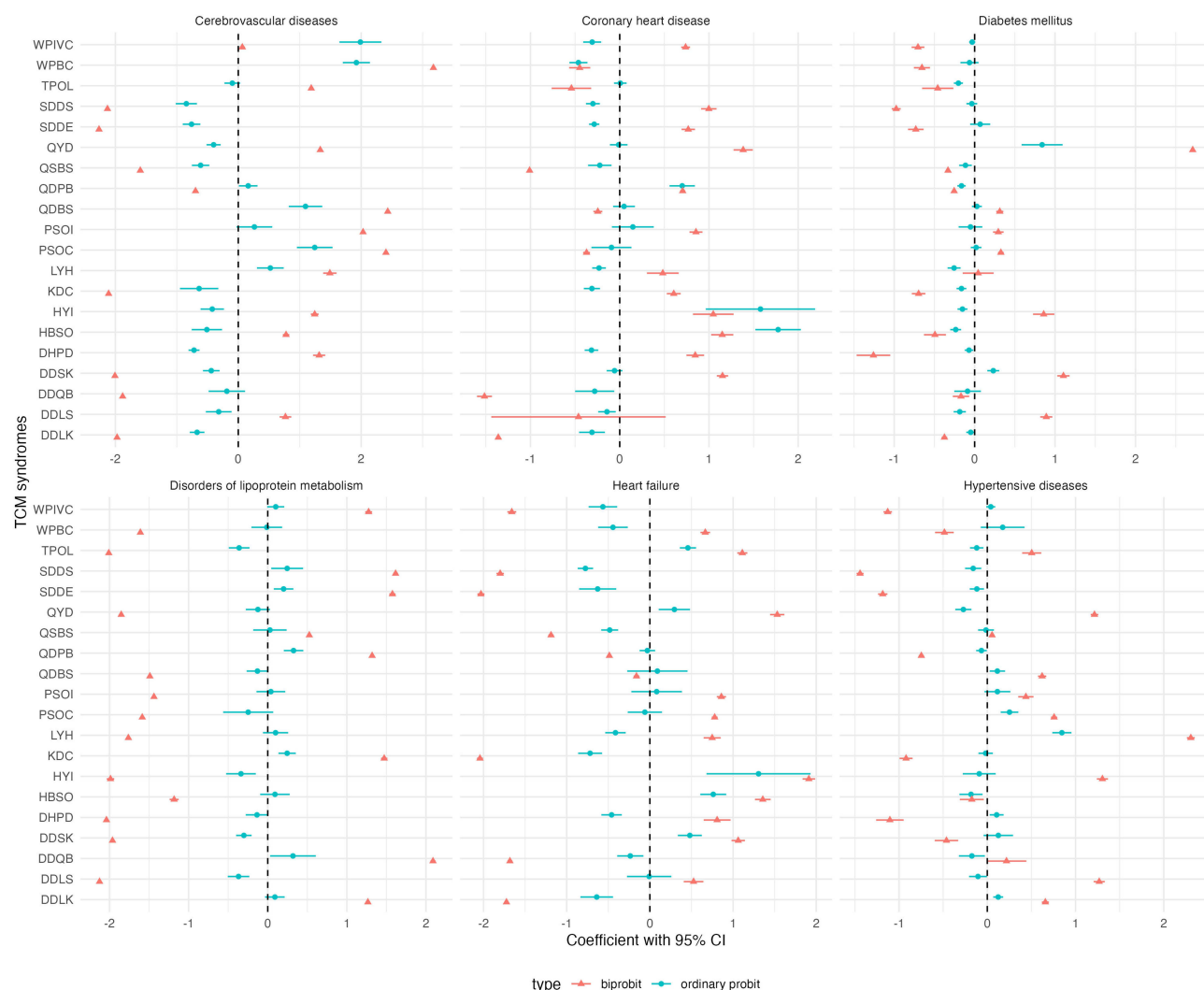
Figure 3 Correlation between Western Medicine and Traditional Chinese Medicine diagnosis.
Abbreviations: WPIVC, syndrome/pattern of wind and phlegm stagnating and impeding vessels and collaterals; WPBC, syndrome/pattern of wind and phlegm blocking collaterals; TPOL, syndrome/pattern of turbid phlegm obstructing the lung; SDDS, syndrome/pattern of spleen deficiency with dampness obstruction; SDDE, syndrome/pattern of spleen deficiency with dampness encumbrance; QYD, syndrome/pattern of qi and yin deficiency; QSBS, syndrome/pattern of qi stagnation and blood stasis; QDPB, syndrome/pattern of qi deficiency with phlegm binding; QDBS, syndrome/pattern of qi deficiency with blood stasis; PSOI, syndrome/pattern of phlegm and stasis obstructing the internal; PSOC, syndrome/pattern of phlegm and stasis obstructing the collaterals; LYH, syndrome/pattern of liver yang hyperactivity; KDC, syndrome/pattern of kidney deficiency with complex; HYI, syndrome/pattern of heart yang insufficiency; HBSO, syndrome/pattern of heart blood stasis obstruction; DHPD, syndrome/pattern of dampness and heat pouring downwards; DDSK, syndrome/pattern of dual deficiency of spleen and kidney; DDQB, syndrome/pattern of dual deficiency of qi and blood; DDLS, syndrome/pattern of dual deficiency in the lung and spleen; DDLK, syndrome/pattern of dual deficiency of liver and kidney.

To understand the most significant findings of a correlation between WPBC, QDBS and CVD, we look to both TCM and WM theories of pathogenesis. According to TCM theory, QDBS is caused by pathogenic factors damaging vital qi or qi deficiency failing to move blood, leading to blood stasis. Clinical findings in QDBS that TCM practitioners look for include localized stabbing pain that does not shift, a pale or dark purple tongue and stasis spots or patches on the edges. The pulse is deep, thin, or rough. Other accompanying symptoms include a dull complexion, fatigue, weakness, shortness of breath, and reluctance to speak.[24] WPBC is understood in TCM theory to be caused by wind pathogens stirring up hidden phlegm or liver wind carrying phlegm, which disrupts the meridians and obstructs the muscles and channels. The TCM symptoms characterizing WPBC include post-stroke hemiplegia, limb spasms or tremors, twitching, facial deviation, drooling from the mouth, slurred speech, etc.[24] CVD may also be accompanied by symptoms such as dizziness and numbness in the limbs.[24] In WM, the clinical manifestations of CVD largely depend on the affected area of the brain and the severity of the damage. Common symptoms include sudden numbness or weakness in the face, arm, or leg,

especially on one side of the body. This weakness or numbness may affect speech, walking, or the ability to grasp objects, which may also lead to vision problems and severe headache: A sudden, intense headache with no apparent cause, which may be accompanied by vomiting, dizziness, or altered consciousness. A previous study showed the similar results of acute cerebral infarction were correlated to the syndrome elements of qi deficiency, phlegm, blood stasis, and wind.[35] However, these studies did not consider errors in variables and just had a small number of samples.

Theory can also shed light on the correlation between QYD and DM. The TCM concept of QYD refers to a type of syndrome caused by various factors that lead to the depletion of both qi and yin, where signs of qi deficiency and yin deficiency coexist, resulting in a lack of body nourishment.[24] In TCM, qi deficiency refers to the condition where the body's qi, or vital energy, is insufficient, leading to an inability to regulate and retain essential body functions, such as body fluids and blood. This can result in symptoms like spontaneous sweating, frequent urination, and fatigue, as the body fails to "hold" its resources properly. On the other hand, yin deficiency generating heat reflects a situation where the depletion of yin, the body's cooling and moistening force, leads to internal heat. When yin is deficient, the body loses its balance between yin and yang, resulting in signs of internal heat such as hot flashes, night sweats, dry mouth, and irritability.[7] A mechanism that could explain the symptoms of DM as understood by TCM practitioners could be hyperglycemic diuresis, which would deplete body fluids, hence leading to dry mouth and thirst as described under the "Xiaoke disease" pattern.[7]

Syndrome differentiation to guide treatment is the core of TCM. In TCM theory, a disease characterized with multiple TCM syndromes can reflect the severity of the case.[7] Previous studies also showed that TCM syndrome classification may change as the disease progresses and the TCM syndromic changes can reflect a natural transitions from occurrence, initial, intermediate, and advanced stages, to recovery, relapse, or deterioration of disease.[9] The better understanding of correlations between WM and TCM syndromes may add values to the clinical practice. For example, improving the personalized integrative care, predicting length of stay or healthcare costs and establishing cross-disciplinary training.

Limitations

Data came from the home page of the hospitalization record, which did not cover comprehensive information. There was unmeasured confounding that may affect the results. For example, the inconsistencies between doctors, the lack of biological indicators related to TCM syndromes. Moreover, the findings are correlational rather than causal, which can only reveal the correlation between TCM and WM diagnosis.

Conclusions

We found the correlations between WM and TCM syndromes diagnosis were attenuated due to coding errors and these correlations became more detectable with the usage of IV. Considering possible correspondence between each tradition's theory of pathogenesis can provide a new perspective for integrative medicine and a way to cross walk between two contrasting medical diagnostic traditions. TCM practitioners describe the severity of diseases by assigning TCM syndromes, which also reflect the etiology and pathogenesis and naturally motivate target measures to correct the disorder. This study provides a basis for adding TCM syndrome modules to clinical electronic medical records, as well as for standardizing coding of dual diagnoses and conducting interdisciplinary training in institutions that practice both diagnoses.

Abbreviations

WM, Western Medicine; TCM, Traditional Chinese Medicine; HTN, hypertension diseases; DM, type 2 diabetes mellitus; DLM, disorders of lipoprotein metabolism; CHD, coronary heart diseases; CHF, heart failure; CVD, cerebro-vascular diseases; AIC, Akaike information criterion; BIC, Bayesian information criterion; QSBS, syndrome/pattern of qi stagnation and blood stasis; QDBS, syndrome/pattern of qi deficiency with blood stasis; DDLK, syndrome/pattern of dual deficiency of liver and kidney; QDPB, syndrome/pattern of qi deficiency with phlegm binding; PSOC, syndrome/pattern of phlegm and stasis obstructing the collaterals; PSOI, syndrome/pattern of phlegm and stasis obstructing the internal; DDQB, syndrome/pattern of dual deficiency of qi and blood; DDSK, syndrome/pattern of dual deficiency of spleen and kidney; WPBC, syndrome/pattern of wind and phlegm blocking collaterals; QYD, syndrome/pattern of qi and yin

deficiency; TPOL, syndrome/pattern of turbid phlegm obstructing the lung; SDDS, syndrome/pattern of spleen deficiency with dampness obstruction; KDC, syndrome/pattern of kidney deficiency with complex; WPIVC, syndrome/pattern of wind and phlegm stagnating and impeding vessels and collaterals; DHPD, syndrome/pattern of dampness and heat pouring downwards; SDDE, syndrome/pattern of spleen deficiency with dampness encumbrance; DDLS, syndrome/pattern of dual deficiency in the lung and spleen; HBSO, syndrome/pattern of heart blood stasis obstruction; LYH, syndrome/pattern of liver yang hyperactivity; HYI, syndrome/pattern of heart yang insufficiency.

Data Sharing Statement

The data that support the findings are restrictions apply and not publicly available. We will provide the details of the data with reasonable request. The code will be available upon publication.

Ethics Approval and Consent to Participate

This study utilizes anonymized medical record data and complies with the principles of the Declaration of Helsinki. Ethical approval for the research was obtained from ethics committee of Southern Medical University (Approval No. of ethics committee: 2024-75), and all procedures followed data privacy protection and ethical guidelines to ensure the confidentiality of participants.

Acknowledgments

Funding

Disclosure

The authors declare that they have no competing interests.

References

1. Chinese Preventive Medicine Association, Branch of Heart Disease Prevention and Control, Chinese Preventive Medicine Association, Chinese Diabetes Society, Branch of Stroke Prevention and Control, Chinese Preventive Medicine Association, Chinese Health Management Society, Branch of Non-communicable Chronic Disease Prevention and Control, Chinese Preventive Medicine Association. Chinese guideline on healthy lifestyle to prevent cardiometabolic diseases. *Chin Circulation J*. 2020;12:141–162.
2. Larsen FB, Pedersen MH, Friis K, Glümer C, Lasgaard M. A latent class analysis of multimorbidity and the relationship to socio-demographic factors and health-related quality of life. a national population-based study of 162,283 Danish adults. *PLoS One*. 2017;12(1):e0169426. doi:10.1371/journal.pone.0169426
3. Zheng Y, Zhou Z, Wu T, et al. Association between composite lifestyle factors and cardiometabolic multimorbidity in Chongqing, China: a cross-sectional exploratory study in people over 45 years and older. *Front Public Health*. 2023;11:1118628. doi:10.3389/fpubh.2023.1118628
4. Lam WC, Lyu A, Bian Z. ICD-11: impact on traditional Chinese medicine and world healthcare systems. *Pharm Med*. 2019;33(5):373–377. doi:10.1007/s40290-019-00295-y
5. Wang Q. Individualized medicine, health medicine, and constitutional theory in Chinese medicine. *Front Med*. 2012;6(1):1–7. doi:10.1007/s11684-012-0173-y
6. Chen KF. *Study on Traditional Chinese Medicine Constitution Factors Related to Health-Related Quality of Life in the General Population of Nine Provinces and Cities [Master]*. Beijing University of Chinese Medicine; 2009. https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C475KOm_zrgu4lQARvep2SAk6at-NE8M3PgrTsq96O6n6cBr3nAhfLqDKYQhjzUomHyyx7awzvApEh0kW7cwArVh&uniplatform=NZKPT.
7. Zhang BL, Wu MH. *Internal Medicine of Traditional Chinese Medicine*. 4th ed. China Traditional Medicine Press; 2017.
8. Lei SY, Chang X, Wang LY, Liu DS, Han XJ. The characteristics of TCM syndromes in 3446 patients suffering CHD complicated with depression based on literature study. *World J Tradit Chin Med*. 2022;35:63–68.
9. Cheng F, Wang X, Song W, et al. Biologic basis of TCM syndromes and the standardization of syndrome classification. *J Tradit Chin Med Sci*. 2014;1(2):92–97. doi:10.1016/j.jtcms.2014.09.005
10. Shuang LI, Shuwei D, Zheyi D, et al. Distribution of traditional Chinese medicine syndromes in diabetic kidney disease chronic kidney disease 1-5: a correlation study. *J Tradit Chin Med*. 2024;44(3):572–580. doi:10.19852/j.cnki.jtcm.20230802.007

11. Huang XL, Li X, Ji HC, et al. A retrospective study on the modern traditional Chinese medicine syndrome characteristics of coronary heart disease based on the clinical diagnosis and treatment data in Tianjin from 2015 to 2021. *J Traditional Chin Med*. 2024;65:1363–1368.

12. Ho L, Xu Y, Zhang NL, et al. Data driven identification and classification of Chinese medicine syndrome types among functional dyspepsia patients: latent tree analysis (abridged secondary publication). *Hong Kong Med J*. 2024;30(4):40–46.

13. Gao H. *Distribution Characteristics of Traditional Chinese Medicine Inquiry Symptoms and Syndrome of 1007 Patients with Coronary Heart Disease at Different Ages [Master]*. Shanghai University of Traditional Chinese Medicine; 2021.

14. Zheng HX. *Basic Theory of Traditional Chinese Medicine*. Beijing: China Traditional Chinese Medicine Press; 2016.

15. Liu GP, Wang YQ, Zhao NQ, et al. Discussion on the information interpretation and diagnostic consistency of the four diagnoses among TCM clinicians. *World Sci Technol*. 2010;12:358–362.

16. GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the global burden of disease study 2021. *Lancet*. 2023;402(10397):203–234. doi:10.1016/S0140-6736(23)01301-6

17. GBD 2021 Risk Factors Collaborators. Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: a systematic analysis for the global burden of disease study 2021. *Lancet*. 2024;403(10440):2162–2203. doi:10.1016/S0140-6736(24)00933-4

18. Skrondal A, Rabe-Hesketh S. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton: Chapman & Hall/CRC; 2004.

19. Hernan MA, Robins JM. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC; 2020.

20. Lawrence MG. The relationship between relative humidity and the dewpoint temperature in moist air: a simple conversion and applications. *Bull Am Meteorol Soc*. 2005;86(2):225–234. doi:10.1175/BAMS-86-2-225

21. R: the R project for statistical computing [Internet]. [cited May 19, 2025]. Available from: https://www.r-project.org/. Accessed June 27, 2025.

22. Stata/MP: advanced multiprocessing capabilities [Internet]. [cited May 19, 2025]. Available from: https://www.stata.com/statamp/. Accessed June 27, 2025.

23. Wickham H. RStudio. tidyverse: easily install and load the "Tidyverse" [Internet]. 2023 [cited May 19 2025]. Available from: https://cran.r-project.org/web/packages/tidyverse/index.html. Accessed June 27, 2025.

24. National medical administration of traditional Chinese medicine (2020) no. 3. notice of the national health commission of the state administration of traditional Chinese medicine on issuing the classification and code of tcm diseases and the terms of tcm clinical diagnosis and treatment _ department documents of the state council _ Chinese government website [Internet]. 2020. Available from: https://www.gov.cn/zhengce/zhengceku/2020-11/24/content_5563703.htm. Accessed June 27, 2025.

25. Maddali MV, Moore AR, Sinha P, et al. Molecular endotypes of idiopathic pulmonary fibrosis: a latent class analysis of two multicenter observational cohorts. *Am J Respir Crit Care Med*. 2024;210(4):455–464. doi:10.1164/rccm.202402-0339OC

26. Khandker SR, Koolwal GB, Samad HA. *Handbook on Impact Evaluation: Quantitative Methods and Practices*. Washington, D.C: World Bank; 2010. https://hdl.handle.net/10986/2693. Accessed June 27, 2025.

27. Griliches Z. Errors in variables and other unobservables. *Econometrica*. 1974;42(6):971–998. doi:10.2307/1914213

28. Christofides LN, Stengos T, Swidinsky R. On the calculation of marginal effects in the bivariate probit model. *Econ Lett*. 1997;54(3):203–208. doi:10.1016/S0165-1765(97)00025-6

29. Seyoum S. Analysis of prevalence of malaria and anemia using bivariate probit model. *Ann Data Sci*. 2018;5(2):301–312. doi:10.1007/s40745-018-0138-3

30. Katchova A. Econometrics academy - bivariate probit and logit models. 2013. Available from: https://sites.google.com/site/econometricsacademy/econometrics-models/bivariate-probit-and-logit-models. Accessed June 27, 2025.

31. Andrews I, Stock JH, Sun L. Weak instruments in instrumental variables regression: theory and practice. *Annu Rev Econ*. 2019;11(1):727–753. doi:10.1146/annurev-economics-080218-025643

32. Han Y, Hu Y, Yu C, et al. Lifestyle, cardiometabolic disease, and multimorbidity in a prospective Chinese study. *Eur Heart J*. 2021;42(34):3374–3384. doi:10.1093/eurheartj/ehab413

33. Li R, Shao J, Hu C, et al. Metabolic risks remain a serious threat to cardiovascular disease: findings from the global burden of disease study 2019. *Intern Emerg Med*. 2024;19(5):1299–1312. doi:10.1007/s11739-024-03605-8

34. Ma Z, Jia C, Guo J, Gu H, Miao Y. Features analysis of five-element theory and its basal effects on construction of visceral manifestation theory. *J Tradit Chin Med*. 2014;34(1):115–121. doi:10.1016/s0254-6272(14)60064-9

35. Wang YY, Hu J, Xu RQ, Lu XL, Wang YW. Analysis on the correlation of traditional Chinese medicine syndromes with collateral circulation, cerebral blood flow perfusion and prognosis of patients with acute cerebral infarction. *J Guangzhou Univ Tradit Chin Med*. 2024;41:1662–1669.