Identification and classification of traditional Chinese medicine syndrome types among senior patients with vascular mild cognitive impairment using latent tree analysis

Chen Fu¹, Nevin Lianwen Zhang², Bao-xin Chen¹, Zhou Rong Chen³, Xiang Lan Jin¹, Rong-juan Guo¹, Zhi-gang Chen¹, Yun-ling Zhang¹

¹. Department of Neurology, Dongfang Hospital, Beijing University of Chinese Medicine, Beijing 100078, China
². Department of Computer Science and Engineering, the Hong Kong University of Science and Technology, Hong Kong, China

ABSTRACT

OBJECTIVE: To treat patients with vascular mild cognitive impairment (VMCI) using traditional Chinese medicine (TCM), it is necessary to classify the patients into TCM syndrome types and to apply different treatments to different types. In this paper, we investigate how to properly carry out the classification for patients with VMCI aged 50 or above using a novel data-driven method known as latent tree analysis (LTA).

METHOD: A cross-sectional survey on VMCI was carried out in several regions in Northern China between February 2008 and February 2012 which resulted in a data set that involves 803 patients and 93 symptoms. LTA was performed on the data to reveal symptom co-occurrence patterns, and the patients were partitioned into clusters in multiple ways based on the patterns. The patient clusters were matched up with syndrome types, and population statistics of the clusters were used to quantify the syndrome types and to establish classification rules.

RESULTS: Eight syndrome types were identified: Qi deficiency, Qi stagnation, Blood deficiency, Blood stasis, Phlegm-dampness, Fire-heat, Yang deficiency, and Yin deficiency. The prevalence and symptom occurrence characteristics of each syndrome type were determined. Quantitative classification rules were established for determining whether a patient belongs to each of the syndrome types.

CONCLUSION: A solution for the TCM syndrome classification problem for patients with VMCI and aged 50 or above was established based on the LTA of unlabeled symptom survey data. The results can be used as a reference in clinic practice to improve the quality of syndrome differentiation and to reduce diagnosis variances across physicians. They can also be used for patient selection in research projects aimed at finding biomarkers for the syndrome types and in randomized control trials aimed at determining the efficacy of TCM treatments of VMCI.

Keywords: vascular mild cognitive impairment; medicine, Chinese traditional; syndrome; latent tree analysis

Citation: Fu C, Zhang NL, Chen BX, Chen ZR, Jin XL, Guo RJ, Chen ZG, Zhang YL. Identification and classification of traditional Chinese medicine syndrome types among senior patients with vascular mild cognitive impairment using latent tree analysis. J Integr Med. 2017 March; Epub ahead of print.
1 Introduction

With the global increase in population and life expectancy, dementia has become a prominent world health problem. It is one of the main diseases that have significant impacts on the health and quality of life among seniors. Dementia can be classified as degenerative, vascular, or other types based on its etiology. Up to 34% of dementia cases show significant vascular pathology. Recently, the construct of “vascular cognitive impairment” (VCI) has been introduced to capture the entire spectrum of cognitive disorders, ranging from mild cognitive impairment to fully developed dementia, that are caused by or associated with vascular factors. VCI has three subtypes: vascular dementia (VaD), Alzheimer’s disease (AD) with a vascular component (mixed dementia), and the pre-dementia stage that is referred to as vascular mild cognitive impairment (VMCI or VaMCI).

There are currently no definitive checklists of diagnostic criteria for VCI and VMCI, and hence their prevalence is difficult to ascertain. However, it is known that the prevalence of VaD doubles every 5.3 years. As the pre-dementia stage, VMCI is believed to be much more common.

Some studies have shown that 46% of patients with VMCI developed dementia after five years. It is therefore important to apply some interventions to prevent and postpone VMCI in order to reduce the occurrence of dementia. The current treatment options for VMCI target at improving symptoms (e.g., memory loss, psychological symptoms such as depression and anxiety), or controlling the various risk factors in order to prevent cognitive function impairment. There are no effective medicines for the management of VMCI in Western medicine (WM).

In traditional Chinese medicine (TCM), cognitive impairment is categorized as “dementia” (Chidai or Daibing). Acupuncture and other TCM treatments have been used to treat VMCI in China for years. Reports on a number of studies indicate that acupuncture techniques may be effective for improving intelligence, stimulating consciousness and enhancing memory, and appear to be beneficial for the management of cognitive and memory functions. Findings from other studies suggest that oral Chinese traditional patent medicine and comprehensive TCM intervention appear to improve the cognitive functions and quality of life for patients with cognitive impairment.

To treat patients with VMCI using TCM, it is necessary to classify the patients into TCM syndrome types and prescribe different treatments for different types. In general, the efficacy of TCM treatments for WM diseases relies heavily on the proper classification of patients into TCM syndrome types. This is an important issue that has so far received little attention. We have developed a data-driven method for solving the classification problem. The method is called latent tree analysis (LTA). It is a generalization of latent class analysis (LCA), a method often used in WM research to identify subtypes of a patient population in the absence of a gold standard. LTA improves LCA by relaxing a key independence assumption made by LCA that is unrealistically strong. More details can be referred in an article published by Li et al.

In this paper, we seek to identify the TCM syndrome types among patients with VMCI and to establish classification rules for them using LTA.

2 Methods

2.1 Data collection

Data used in this study were collected in a project on the TCM prevention and treatment of VMCI supported by the TCM Special Research Projects Program, China Ministry of Science and Technology under Contract No. 200807011. Ethical approval for the project was given by the Ethics Committee on Clinical Research of Dongfang Hospital.

The project involved seven top-graded hospitals in Northern China: the Dongfang Hospital of Beijing University of Chinese Medicine, Peking University People’s Hospital, Wanning Hospital of China Academy of Chinese Medical Sciences, Second Hospital of Tianjin Medical University, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Affiliated Hospital of Changchun University of Chinese Medicine, and Traditional Chinese Medicine Hospital of Hebei Province.

All VMCI patients in the project were recruited between February 2008 and February 2012 from the neurology divisions and wards of the hospitals, and the surrounding communities. Since there are no standard diagnostic criteria for VMCI, the research team set up the following criteria themselves based on the National Institute of Neurological Disorders and Stroke and the Canadian Stroke Network VCI harmonization standards:

1. Age ≥ 50 years.
2. Presence of vascular factors (cerebral hemorrhage, cerebral infarction, leukoaraiosis, hypertension, diabetes, hyperlipidemia, coronary heart disease, etc.).
3. Subjective complaints of cognitive impairment that were corroborated by an informant.
4. Objective impairment in more than one domain of cognitive abilities by neuropsychologic assessments, including mini-mental state examination (MMSE), Clinical Dementia Rating (CDR), and Montreal Cognitive Assessment (MoCA) Beijing version. The thresholds for
these tests were set as follows: MMSE score > 17 points (illiterates), or > 20 points (education years ≤ 6 years), or > 24 points (education years > 6 years); CDR score = 0 or 0.5 points; MoCA score < 25 points (education ≤ 12 years), or < 26 points (education > 12 years).

5) Sustained cognitive decline for more than 6 months based on patient self-report or caregiver report.

6) There is a causal relationship between vascular factors and cognitive impairment.

7) Normal abilities or mild impairment of daily living activities/social activities according to the result of Activities of Daily Living Scale.

8) Not meeting the criteria for the diagnosis of dementia in Diagnostic and Statistical Manual of Mental Disorders.

The items on vascular factors (items 2 and 6) require some clarifications. As mentioned in the introduction, VCI refers to all forms of cognitive deficits caused by or associated with vascular factors, including cerebral hemorrhage, cerebral infarction, leukoaraiosis, hypertension, diabetes, hyperlipidemia, coronary heart disease, and so on.⁷ During data collection, neuroimaging tools such as functional magnetic resonance imaging and CT were used to diagnose cerebral hemorrhage, cerebral infarction and leukoaraiosis. Note that neuroimaging tools are not applicable to hypertension, diabetes, hyperlipidemia and coronary heart disease. Whether there is a causal relationship between vascular factors and cognitive impairment was first assessed independently by two neurology physicians with five or more years of experience. The existence of a causal link was concluded if both physicians agreed it is the case. When they were in disagreement, a third physician was asked to make the final judgment.

The exclusion criteria included the following: patients with mental illness or serious physical illness which could affect neuropsychological examinations; those with Hamilton Depression Rating Scale score ≥ 17 points; and those refused to participate in this study and sign on the informed-consent form.

The data collection teams consisted of neurology physicians and graduate students. The team members were trained in multiple sessions on various aspects of the data collection process and examined on the training materials. Only those who passed the examination were allowed to join the teams. The same training materials were used in all data collection centers.

Data collection was carried out by interviewing all patients using an interviewer-administered questionnaire. The interviewers first decided whether a patient should be included in the study according to the inclusion and exclusion criteria listed above. They then moved on to collect information on TCM signs and symptoms and filled out the questionnaire. Judgements about the TCM symptoms and signs were first made by two interviewers independently, and if they were in disagreement, a third interviewer was asked to make the final judgement. The items in the questionnaire were extracted from both ancient and modern TCM literature on Chidai or Daibing patients. All signs and symptoms mentioned in the literature were initially included. The list was refined through multiple rounds of pilot surveys and expert consultations. Most items were polytomous and they were converted into dichotomous items for this analysis for simplicity, just as in the studies with LCA.¹⁹,²⁰ During preprocessing, a few items were combined. For example, edema on face and edema on limbs are combined into edema. There are 93 items in the final dataset. The items consist of signs and symptoms of interest to TCM. We will sometimes refer to them simply as symptoms for brevity. The data are unlabeled in the sense that the true syndrome types of the patients are unknown.

There was a quality control personnel at each data collection center to ensure the authenticity and completeness of the questionnaires. The project team paid regular visits to the data collection centers and randomly sampled questionnaires for authenticity, accuracy, and completeness checks. Data entry was carried out by two persons, each being responsible for entering half of the questionnaires and for verifying the other half.

2.2 Data analysis

The VMCI data were studied using the LTA method.²¹ The objective of the method is to reveal symptom patterns hidden in the data and to use those patterns to identify patient clusters that correspond to different TCM syndrome types. Statistical characteristics of the patient clusters are then used as quantitative characterizations of the syndrome types, and syndrome classification rules are thereby established.

The analysis was carried out using the Lantern software²¹ with five steps.

2.2.1 Statistical pattern discovery

In the first step, LTA was performed to reveal symptom co-occurrence/mutual-exclusion patterns hidden in the VMCI data. Lantern includes two algorithms for this purpose. In this study, the Extension Adjustment Simplification until Termination algorithm was used.²² The result is a latent tree model with multiple latent variables. Some of the latent variables reveal probabilistic symptom co-occurrence patterns, while others reveal probabilistic mutual-exclusion of symptoms.

2.2.2 Statistical pattern interpretation

The second step was to determine the TCM syndrome connotations of the patterns discovered. In the case of a symptom co-occurrence pattern, the task was to determine what TCM syndrome type(s) can bring about the co-
occurrence of the symptoms in the pattern. The result could be a single syndrome type that can explain all the symptoms in the pattern, or a combination of several syndrome types, each explaining some of the symptoms. If a latent variable represents the probabilistic mutual exclusion of two or more groups of symptoms, each group is a co-occurrence pattern and was interpreted separately. Domain knowledge was used in this step.[23]

2.2.3 Syndrome identification

The statistical pattern interpretation process gave rise to a list of syndrome types. Each syndrome type explains a list of symptoms from various co-occurrence patterns. If all key manifestations of the syndrome type are present on the list, then it was regarded as well supported by the data and was selected as a target for further analysis. Domain knowledge was used in this step also.[23]

2.2.4 Syndrome quantification

For each syndrome type identified at the previous step, cluster analysis was performed on the patient population based on the symptoms that it explains. The symptom variables were not always used as features for the cluster analysis directly. If two or more symptom variables are from the same pattern, they were “combined” into one latent feature. The analysis divided the patients into several clusters. One patient cluster was identified as corresponding to the syndrome type. Population statistics of the patient cluster were used to quantify the syndrome type.

2.2.5 Syndrome classification

Finally, a classification rule was established for each syndrome type. The rule can be used to determine whether a patient belongs to the syndrome type. The main results were obtained at the last two steps and they will be presented in Sections 3.4 and 3.5. The first three steps produced intermediate results that were necessary for the last two steps. They will be presented in Sections 3.1 to 3.3, along with discussions of some subtle practical issues that are not addressed in.[18] Those are important for researchers who wish to apply the method in their own work.

3 Results

A total number of 803 patients were successfully interviewed and their demographic characteristics are summarized in Table 1.

3.1 Statistical pattern discovery

LTA of the VMCI data yielded a model with 32 latent variables. The structure of the model is shown in Figure 3 of Zhang et al.[18] Some of the latent variables reveal probabilistic symptom co-occurrence (i.e., positive correlation) patterns, while others reveal probabilistic mutual exclusion (i.e., negative correlation) of symptoms. The patterns were extracted from the model using the “Model Interpretation function” of Lantern.[18]

The co-occurrence patterns discovered are shown in the second column of Table 2. We see, for example, that the latent variable Y29 reveals the probabilistic co-occurrence of the three symptoms lack of strength, mental fatigue, and loose stool; Y30 reveals the probabilistic co-occurrence of frequent nocturnal urination and dripping urination; Y31 reveals the probabilistic co-occurrence of blurred vision and dry eyes; and so on. Note that there is only one symptom for Y23 in the table. This is because Y23 is directly connected to only one symptom variable in the model.

Each co-occurrence pattern is given as a list of symptoms. Ordering in the list is important. Conceptually, a pattern corresponds to a class of patients.[18] Symptoms at the front of the list are more important than those at the end in determining whether a patient belongs to the cluster. For example, the list for the pattern Y27 consists of short of breath, sighing, hypochondrium distension or pain, tinnitus resembling tide, and stabbing headache. Those symptoms are ordered in descending importance, with short of breath being the most important and stabbing headache being the least important.

The mutual-exclusion patterns discovered are shown in the second column of Table 3. For example, Y02 reveals the probabilistic mutual exclusion of two groups of symptoms: Y02-1 (anorexia, undigested food in stool, etc.) and Y02-2 (darkish tongue). The first group Y02-1 consists of more than one symptom. They are positively correlated and hence tend to co-occur. In other words, Y02-1 is a co-occurrence pattern. Similarly, Y09-1 (slow pulse, moderate pulse) and Y25-1 (insomnia, dreamfulness) are also co-occurrence patterns.

Y14 and Y15 each have three states. They reveal the mutual exclusion of thirst desire no drinks, thirst desire cold drinks, and thirst desire hot drinks. The symptom dry mouth or throat is a higher level symptom. It is subsumed by the forgoing three symptoms and hence removed from further analysis.

3.2 Statistical pattern interpretation

To interpret the co-occurrence of a group of symptoms amounts to identifying the syndrome type(s) that can lead to the co-occurrence of the symptoms. Often there is no syndrome type that can explain the occurrence of all the symptoms in a statistical pattern. Consequently, we have primary interpretations and secondary interpretations. A primary interpretation attempts to explain the co-occurrence of all symptoms in a pattern, and, when that is not possible, tries to explain the leading symptom and as many subsequent symptoms as possible. A secondary interpretation explains other symptoms or gives alternative explanations to some of the symptoms.
The last two columns of Tables 2 and 3 give interpretations of all the statistical patterns discovered by LTA. For example, Y01 reveals the probabilistic co-occurrence of four symptoms: asthenia of defecation, dry stool or constipation, sallow complexion, and clear profuse urination. No single syndrome types can explain all the four symptoms. The primary interpretation is Qi deficiency and it explains the co-occurrence of the first two symptoms. Blood deficiency is a secondary interpretation and it explains the co-occurrence of the second and the third symptoms. In addition, dry stool or constipation can also be caused by Fire-heat, Yin deficiency and Qi stagnation. Clear profuse urination is weakly correlated with the other three symptoms and it is a manifestation of Yang deficiency.

As the second example, consider Y06. It reveals the probabilistic co-occurrence of the two symptoms thick tongue fur and greasy tongue fur. The syndrome type Phlegm-dampness explains both symptoms. It is the primary interpretation of the pattern. There are no secondary interpretations.

As the third example, consider Y25-1. It is about the probabilistic co-occurrence of insomnina and dreamfulness. The two symptoms can be explained by any of the following five syndrome types: Yin deficiency, Fire-heat, Blood deficiency, Qi deficiency, and Phlegm-dampness. They are all primary and alternative interpretations of the pattern. There are no secondary interpretations.

The above examples represent three possible scenarios one might encounter when interpreting a statistical pattern. First, the pattern has only one primary interpretation and no secondary interpretations. The patterns Y06, Y08, Y18, Y23 and Y29, and Y02-2, Y05-2, Y09-1, Y09-2, Y10-1, Y12-1, Y14-2 and Y15-1 fall into this category. Second, the pattern has more than one primary interpretations and no secondary interpretations. The patterns Y13, Y31, Y12-2, Y14-1, Y25-1 and Y25-2 belong to this category. Third, the pattern has both primary interpretations and secondary interpretations. The patterns Y12, Y16, Y17, Y18, Y19, Y20 and Y21 fall into this category.

Table 1  Demographic characteristics of the 803 patients with vascular mild cognitive impairment

<table>
<thead>
<tr>
<th>Item</th>
<th>Overall (N = 803)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean ± standard deviation, year)</td>
<td>68.02 ± 7.72</td>
</tr>
<tr>
<td>Gender (n, %)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>388 (48.30)</td>
</tr>
<tr>
<td>Female</td>
<td>415 (51.70)</td>
</tr>
<tr>
<td>Education (n, %)</td>
<td></td>
</tr>
<tr>
<td>Illiterates</td>
<td>90 (11.30)</td>
</tr>
<tr>
<td>≤ 6 points</td>
<td>208 (25.90)</td>
</tr>
<tr>
<td>6–12 points</td>
<td>364 (45.30)</td>
</tr>
<tr>
<td>&gt; 12 points</td>
<td>140 (17.40)</td>
</tr>
<tr>
<td>Marital status (n, %)</td>
<td></td>
</tr>
<tr>
<td>Separated</td>
<td>1 (0.10)</td>
</tr>
<tr>
<td>Married</td>
<td>730 (91.00)</td>
</tr>
<tr>
<td>Divorced</td>
<td>1 (0.10)</td>
</tr>
<tr>
<td>Widowed</td>
<td>70 (8.80)</td>
</tr>
<tr>
<td>Body mass index (mean ± standard deviation, m²/kg)</td>
<td>24.88 ± 0.11</td>
</tr>
<tr>
<td>Montreal cognitive assessment (mean ± standard deviation)</td>
<td>20.57 ± 3.80</td>
</tr>
<tr>
<td>Hypertension (n, %)</td>
<td>558 (69.50)</td>
</tr>
<tr>
<td>Hypertension duration (mean ± standard deviation, years)</td>
<td>13.14 ± 0.48</td>
</tr>
<tr>
<td>Diabetes (n, %)</td>
<td>223 (27.80)</td>
</tr>
<tr>
<td>Diabetes duration (mean ± standard deviation, years)</td>
<td>8.12 ± 0.43</td>
</tr>
<tr>
<td>Dyslipidemia (n, %)</td>
<td>196 (24.40)</td>
</tr>
<tr>
<td>Dyslipidemia duration (mean ± standard deviation, years)</td>
<td>7.19 ± 0.53</td>
</tr>
<tr>
<td>Coronary heart disease (n, %)</td>
<td>283 (35.20)</td>
</tr>
<tr>
<td>Coronary heart disease duration (mean ± standard deviation, years)</td>
<td>10.67 ± 0.56</td>
</tr>
<tr>
<td>Cerebrovascular disease (n, %)</td>
<td>282 (35.10)</td>
</tr>
<tr>
<td>Cerebrovascular disease duration (mean ± standard deviation, years)</td>
<td>5.31 ± 0.34</td>
</tr>
<tr>
<td>Peripheral vascular disease (n, %)</td>
<td>18 (2.30)</td>
</tr>
<tr>
<td>Peripheral vascular disease duration (mean ± standard deviation, years)</td>
<td>5.76 ± 1.70</td>
</tr>
<tr>
<td>Latent Variables</td>
<td>Symptoms and signs</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Y01</td>
<td>Asthenia of defecation, dry stool or constipation, sallow complexion, clear profuse urination</td>
</tr>
<tr>
<td>Y03</td>
<td>Pale lips, pale complexion, dizzy headache, deep-red tongue, thin tongue, loose stool following dry feces, hemihidrosis</td>
</tr>
<tr>
<td>Y04</td>
<td>Tidal fever or night sweat, spontaneous sweating, dry tongue, fast pulse, fissured tongue</td>
</tr>
<tr>
<td>Y06</td>
<td>Thick tongue fur, greasy tongue fur</td>
</tr>
<tr>
<td>Y07</td>
<td>Vexing heat in chest, edema, fetid mouth odor, muscular twitching, swift digestion rapid hungering</td>
</tr>
<tr>
<td>Y08</td>
<td>Sunken pulse, feeble pulse</td>
</tr>
<tr>
<td>Y09</td>
<td>Varicose sublingual-veins, astringent pulse, tense pulse</td>
</tr>
<tr>
<td>Y10</td>
<td>Bitter taste in mouth, urinary incontinence</td>
</tr>
<tr>
<td>Y11</td>
<td>Acid swallow or epigastric upset, abdominal distension, trembling of limbs</td>
</tr>
<tr>
<td>Y12</td>
<td>Aphtha on month or tongue, throbbing headache</td>
</tr>
<tr>
<td>Y13</td>
<td>Sore waist or knees, expectoration, tinnitus resembling cicada</td>
</tr>
<tr>
<td>Y14</td>
<td>Fat tongue, tongue with ecchymosis, tooth-marked tongue</td>
</tr>
<tr>
<td>Y15</td>
<td>Dizziness, head feels as if swathed, distending headache, nausea or vomiting</td>
</tr>
<tr>
<td>Y16</td>
<td>Dim complexion, blackish lower eyelid, scaly skin</td>
</tr>
<tr>
<td>Y17</td>
<td>Purple or darkish lips</td>
</tr>
<tr>
<td>Y18</td>
<td>Lassitude of limbs or body, fear of cold or cold limbs, numbness</td>
</tr>
<tr>
<td>Y19</td>
<td>Fat oppression, palpitation</td>
</tr>
<tr>
<td>Y20</td>
<td>Short of breath, sighing, hypochondrium distension or pain, tinnitus resembling tide, stabbing headache</td>
</tr>
<tr>
<td>Y21</td>
<td>Lack of strength, mental fatigue, loose stool</td>
</tr>
<tr>
<td>Y22</td>
<td>Frequent nocturnal urination, dripping urination</td>
</tr>
<tr>
<td>Y23</td>
<td>Blurred vision, dry eyes</td>
</tr>
</tbody>
</table>

The numbers following a syndrome type indicate the symptoms in the pattern that the syndrome type explains. For example, Y01 reveals the probabilistic co-occurrence of three symptoms. Qi deficiency explains the first two symptoms, while Blood deficiency explains symptoms number 2 and 3. Y04 reveals the probabilistic co-occurrence of three other symptoms. Yin deficiency explains all the symptoms, and hence it is not followed by any numbers.
interpretations. The patterns $Y_{01}$, $Y_{04}$, $Y_{03}$, $Y_{07}$, $Y_{11}$, $Y_{16}$, $Y_{17}$, $Y_{19}$, $Y_{20}$, $Y_{21}$, $Y_{22}$, $Y_{24}$, $Y_{26}$, $Y_{27}$, $Y_{30}$, $Y_{02}$-1 and $Y_{10}$-2 are in this category.

The interpretation of $Y_{14}$-1 (thirst desire no drinks) requires some discussions. According to TCM theory,[23] the symptom can be due to Blood stasis. In the VMCI data set, however, it is found (in the syndrome quantification step) to be negatively correlated with other manifestations of Blood stasis such as purples or darkish lips, dim complexion, and blackish eyelids. We therefore conclude that, for the population under study, thirst desire no drinks is caused by other factors, namely Yin deficiency and Phlegm-dampness, rather than Blood stasis. Blood stasis is hence not included in Table 3 as an interpretation for the symptom. Similarly, the statistical pattern $Y_{09}$-1 (slow pulse, moderate pulse) can be explained by either Qi deficiency or Phlegm-dampness in theory. However, it is negatively correlated with other manifestations of Phlegm-dampness such as thick tongue fur and greasy tongue fur, and hence Phlegm-dampness is not listed as a possible explanation for the pattern. The pattern $Y_{10}$-1 (pink tongue, hollow headache) can be explained by Qi deficiency, Yang deficiency, or Yin deficiency. However, it is negatively correlated with other manifestations of Qi deficiency and Yang deficiency, and hence only Yin deficiency is listed as its primary interpretation.

In addition, several statistical patterns are not interpreted. $Y_{05}$-1 (white tongue fur) is not interpreted because it is considered normal for the population under study, namely seniors aged 50 or above and with VMCI. $Y_{15}$-2 is not interpreted because it is subsumed by $Y_{14}$-1 and $Y_{14}$-2.

### 3.3 Syndrome identification

Based on the statistical patterns and their interpretations in Tables 2 and 3, we have identified eight syndrome types that are well supported by the data: Qi deficiency, Qi stagnation, Blood deficiency, Blood stasis, Phlegm-dampness, Fire-heat, Yang deficiency and Yin deficiency. Take Yang deficiency as an example. According to

### Table 3 Probing mutual-exclusion patterns and their traditional Chinese medicine interpretations

<table>
<thead>
<tr>
<th>Latent Variables</th>
<th>Symptoms and signs</th>
<th>Primary syndrome elements</th>
<th>Secondary syndrome elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{02}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{02}$-1</td>
<td>Anorexia, undigested food in stool, bland taste in mouth, vomiting of saliva, diarrhea before dawn</td>
<td>Yang deficiency</td>
<td>Qi deficiency (1,3)</td>
</tr>
<tr>
<td>$Y_{02}$-2</td>
<td>Darkish tongue</td>
<td></td>
<td>Blood stasis</td>
</tr>
<tr>
<td>$Y_{05}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{05}$-1</td>
<td>White tongue fur</td>
<td></td>
<td>Fire-heat</td>
</tr>
<tr>
<td>$Y_{05}$-2</td>
<td>Yellow tongue fur</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{09}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{09}$-1</td>
<td>Slow pulse, moderate pulse</td>
<td>Qi deficiency</td>
<td></td>
</tr>
<tr>
<td>$Y_{09}$-2</td>
<td>Taut pulse</td>
<td>Qi stagnation</td>
<td></td>
</tr>
<tr>
<td>$Y_{10}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{10}$-1</td>
<td>Pink tongue, hollow headache</td>
<td>Yin deficiency</td>
<td></td>
</tr>
<tr>
<td>$Y_{10}$-2</td>
<td>Red tongue, little tongue fur</td>
<td>Yin deficiency</td>
<td>Fire-heat (1)</td>
</tr>
<tr>
<td>$Y_{12}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{12}$-1</td>
<td>Slippery pulse</td>
<td></td>
<td>Phlegm-dampness</td>
</tr>
<tr>
<td>$Y_{12}$-2</td>
<td>Thin pulse</td>
<td>Qi deficiency, Blood deficiency</td>
<td></td>
</tr>
<tr>
<td>$Y_{14}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{14}$-1</td>
<td>Thirst desire no drinks</td>
<td>Yin deficiency, Phlegm-dampness</td>
<td></td>
</tr>
<tr>
<td>$Y_{14}$-2</td>
<td>Thirst desire cold drinks</td>
<td>Fire-heat</td>
<td></td>
</tr>
<tr>
<td>$Y_{15}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{15}$-1</td>
<td>Thirst desire hot drinks</td>
<td>Yang deficiency</td>
<td></td>
</tr>
<tr>
<td>$Y_{15}$-2</td>
<td>Not “thirst desire hot drinks”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{25}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{25}$-1</td>
<td>Insomnia, dreamfulness</td>
<td>Yin deficiency, Fire-heat, Blood deficiency, Qi deficiency, Phlegm-dampness</td>
<td></td>
</tr>
<tr>
<td>$Y_{25}$-2</td>
<td>Flushed face</td>
<td>Fire-heat, Yin deficiency</td>
<td></td>
</tr>
</tbody>
</table>
Tables 2 and 3, here are the symptom groups that it explains:

- Y24 (lassitude of limbs or body, fear of cold or cold limbs), Y26 (chest oppression, palpitation), Y30 (frequent nocturnal urination), Y02-1 (anorexia, undigested food in stool, bland taste in mouth, vomiting of saliva, diarrhea before dawn), and Y15-1 (thirst desire hot drinks); Y01 (clear profuse urination), Y03 (pale complexion), Y04 (spontaneous sweating), Y07 (muscular twitching), Y19 (sore waist or knees), Y22 (dim complexion, blackish lower eyelid).

There is a semicolon in the list. The symptom groups before the semicolon are from statistical patterns for which the syndrome type is a primary interpretation, whereas the symptom groups after the semicolon are from patterns for which the syndrome type is a secondary interpretation.

Those symptom groups cover different aspects of the impacts of Yang deficiency, including its impacts on kidney (Y19), on the digestive system (Y02-1), on urination and drinking (Y30, Y15-1), on muscle (Y07), and on the striae and interstice (Y04), as well as its manifestations on the body and limbs (Y24), in the chest (Y26), and on the face (Y03, Y22). Since major aspects of the impacts of Yang deficiency are covered, we conclude that there is sufficient information in the data to quantify Yang deficiency.

The above list is quite long and the symptoms are not in any particular order. In the next section, we will order the symptoms according to their importance in characterizing Yang deficiency. In addition, many of the symptoms will be pruned using the criterion of cumulative information coverage. The result will be a clean and quantitative characterization of Yang deficiency. Although it is not part of the final result, making the list is a necessary intermediate step.

The lists of symptom groups that support each of the other seven syndrome types are listed below.

1. Blood deficiency: Y03 (pale lips, pale complexion, dizzy headache, thin tongue, hemihydrosis), Y31 (blurred vision, dry eyes), Y12-2 (thin pulse), and Y25-1 (insomnia, dreamfulness); Y01 (dry stool or constipation, sallow complexion), Y07 (muscular twitching), Y17 (trembling of limbs), Y21 (dizziness), Y24 (numbness), and Y26 (palpitation).

2. Blood stasis: Y13 (varicose sublingual-veins, astringent pulse, tense pulse), Y22 (dim complexion, blackish lower eyelid, scaly skin), Y23 (purples or darkish lips), and Y02-2 (darkish tongue); Y20 (tongue with ecchymosis), Y21 (dizziness), Y24 (numbness), Y26 (palpitation), and Y27 (stabbing headache).

3. Qi deficiency: Y01 (asthenia of defecation, dry stool or constipation), Y08 (sunken pulse, feeble pulse), Y20 (fat tongue, tooth-marked tongue), Y26 (chest oppression, palpitation), Y29 (lack of strength, mental fatigue, loose stool), Y09-1 (slow pulse, moderate pulse), Y12-2 (thin pulse), and Y25-1 (insomnia, dreamfulness); Y02 (anorexia, bland taste in mouth), Y03 (pale complexion, dizzy headache, loose stool following dry feces), Y04 (spontaneous sweating), Y16 (urinary incontinence), Y19 (sore waist or knees), Y21 (dizziness), Y24 (lassitude of body and limbs), Y27 (short of breath), and Y30 (dripping urination).

4. Qi stagnation: Y13 (varicose sublingual-veins, astringent pulse, tense pulse), Y27 (short of breath, sighing, hypochondrium distention or pain), and Y09-2 (taut pulse); Y01 (dry stool or constipation), Y17 (abdominal distension), and Y26 (chest oppression).

5. Fire-heat: Y16 (bitter taste in mouth, urinary incontinence), Y17 (acid swallow or epigastric upset, trembling of limbs), Y18 (aphtha on month or tongue, throbbing headache), Y21 (dizziness, distending headache, nausea or vomiting), Y05-2 (yellow tongue fur), Y14-2 (thirst desire cold drinks), and Y25-1 (insomnia, dreamfulness); Y01 (dry stool or constipation), Y04 (spontaneous sweating, dry tongue, fast pulse, fissured tongue), Y07 (fetid mouth odor, swift digestion rapid hungering), Y11 (agitation or short temper), Y27 (tinnitus resembling tide), Y10-2 (red tongue), and Y25-2 (flushed face).

6. Phlegm-dampness: Y06 (thick tongue fur, greasy tongue fur), Y11 (sticky or greasy feel in mouth), Y20 (fat tongue, tooth-marked tongue), Y21 (dizziness, head feels as if swathed, distending headache, nausea or vomiting), Y12-1 (slippery pulse), Y14-1 (thirst desire no drinks), and Y25-1 (insomnia, dreamfulness); Y03 (dizzy headache), Y16 (urinary incontinence), and Y19 (expectoration).

7. Yin deficiency: Y04 (tidal fever or night sweat, spontaneous sweating, dry tongue, fast pulse, fissured tongue), Y07 (vexing heat in chest, edema, fetid mouth odor, muscular twitching, swift digestion rapid hungering), Y19 (sore waist or knees, expectoration, tinnitus resembling cicada), Y31 (blurred vision, dry eyes), Y10-1 (pink tongue, hollow headache), Y10-2 (red tongue, little tongue fur), Y14-1 (thirst desire no drinks), and Y25 (insomnia, dreamfulness, flushed face); Y01 (dry stool or constipation), Y03 (deep-red tongue, thin tongue), Y17 (trembling of limbs), Y21 (dizziness), Y22 (dim complexion, blackish lower eyelid), and Y26 (palpitation).

3.4 Syndrome quantification

In the previous section, eight syndrome types were identified as well supported by data. A list of symptom groups was associated with each syndrome type. In the next step, we use those symptom groups to identify patient clusters that correspond to the syndrome types, and use the population statistics of the patient clusters to quantify the syndrome types. This is done using the joint clustering function of Lantern.
3.4.1 Quantification of Yang deficiency

Joint clustering for Yang deficiency is performed using the model shown in Figure 1. The symptoms are from the symptom groups associated with Yang deficiency in the previous section. If a group consists of only one symptom variable, then it is directly connected to the joint clustering variable Z at the top. If a group consists of multiple symptom variables, then they are “combined” into a latent feature and the latent feature is connected to Z.

![Figure 1](https://example.com/figure1.png)

Joint clustering partitions the patients in the VMCI data set into two clusters. Population statistics of the two clusters are as follows:

1. Yang deficiency (0.38): sore waist or knees (0.77), lassitude of limbs or body (0.69), frequent nocturnal urination (0.68), blackish lower eyelid (0.27), fear of cold or cold limbs (0.44), palpitation (0.45), chest oppression (0.51), thirst desire hot drinks (0.32), and spontaneous sweating (0.38).

2. Non-Yang deficiency (0.62): sore waist or knees (0.21), lassitude of limbs or body (0.25), frequent nocturnal urination (0.28), blackish lower eyelid (0.02), fear of cold or cold limbs (0.12), palpitation (0.13), chest oppression (0.20), thirst desire hot drinks (0.08), and spontaneous sweating (0.13).

We see that the symptoms sore waist or knees, lassitude of limbs and body, fear of cold or cold limbs, etc. occur with higher probabilities in the first cluster than in the second cluster. Consequently, the first cluster is interpreted as Yang deficiency, whereas the second cluster interpreted as non-Yang deficiency.

According to the results, the prevalence of Yang deficiency among the patients surveyed is 38%. Among the patients with Yang deficiency, the symptom sore waist or knees occurs with probability 0.77, lassitude of limbs and body occurs with probability 0.69, fear of cold or cold limbs occurs with probability 0.44, and so on. Among the patients without Yang deficiency, on the other hand, the corresponding probabilities are 0.21, 0.25, and 0.12 respectively.

Note that, in the quantification given above, the items are ordered in such a way that the symptoms at the front are those whose occurrence probabilities in the two clusters differ the most. Such symptoms are the most important factors to consider when distinguishing between the two clusters. For example, although chest oppression occurs with higher probability than blackish lower eyelids in the Yang deficiency cluster (0.51 vs 0.27), it is not as important as the latter because it also occurs with higher probability in the non-Yang deficiency cluster (0.20 vs 0.02).

Moreover, only the nine most important symptoms are used to characterize Yang deficiency. This is because their cumulative information coverage has reached the cut-off 95% and hence they are sufficient to capture the differences of the two clusters. In contrast, twenty symptoms are listed as relevant to Yang deficiency in the previous section. Many of them are not important.

3.4.2 Quantification of Yin deficiency

Joint clustering for Yin deficiency partitions the patients into three clusters. The population statistics of the clusters are as follows. As before, the symptoms are sorted according to their importance in characterizing the differences among the three clusters.

1. Yin deficiency I (0.38): sore waist or knees (0.75), blurred vision (0.78), dry eyes (0.62), vexing heat in chest (0.08), fetid mouth odor (0.10), tidal fever or night sweat (0.08), insomnia (0.54), tinnitus resembling cicada (0.43), dreamfulness (0.59), and spontaneous sweating (0.18).

2. Yin deficiency II (0.08): sore waist or knees (0.75), blurred vision (0.52), dry eyes (0.37), vexing heat in chest (0.71), fetid mouth odor (0.71), tidal fever or night sweat (0.71), insomnia (0.77), tinnitus resembling cicada (0.43), dreamfulness (0.80), and spontaneous sweating (0.79).

3. Non-Yin deficiency (0.54): sore waist or knees (0.14), blurred vision (0.26), dry eyes (0.16), vexing heat in chest (0.06), fetid mouth odor (0.08), tidal fever or night sweat (0.08), insomnia (0.21), tinnitus resembling cicada (0.10), dreamfulness (0.29), and spontaneous sweating (0.19).

The third cluster is interpreted as non-Yin deficiency, whereas the other two are interpreted as two different subtypes of Yin deficiency. The main differences between the first two clusters are that the symptoms such as vexing heat in chest, fetid mouth odor, and tidal fever or night sweat occur with much higher probabilities in the second cluster. Therefore, the first cluster is interpreted as Yin deficiency, and the second cluster Yin deficiency with internal heat.

3.4.3 Quantification of other syndrome types

The results of joint clustering for other six syndrome types...
types are listed below. For Blood deficiency, Blood stasis, Fire-heat, and Qi stagnation, the patients are partitioned into two clusters. For Qi deficiency and Phlegm-dampness, the patients are partitioned into three clusters, and two of the clusters are combined because, unlike in the case of Yin deficiency, the differences between them are not important for treatment.

1. Blood deficiency (0.32): blurred vision (0.82), dry eyes (0.69), palpitation (0.56), insomnia (0.60), dizziness (0.62), dreamfulness (0.62), numbness (0.53), trembling of limbs (0.20), and dry stool or constipation (0.47).

2. Non-Blood deficiency (0.68): blurred vision (0.32), dry eyes (0.19), palpitation (0.11), insomnia (0.27), dizziness (0.39), dreamfulness (0.36), numbness (0.27), trembling of limbs (0.05), and dry stool or constipation (0.23).

3. Blood stasis (0.30): purple or darkish lips (0.74), dim complexion (0.49), blackish lower eyelid (0.29), numbness (0.58), palpitation (0.42), scaly skin (0.19), and tongue with ecchymosis (0.11).

4. Non-Blood stasis (0.70): purple or darkish lips (0.08), dim complexion (0.10), blackish lower eyelid (0.04), numbness (0.25), palpitation (0.18), scaly skin (0.04), and tongue with ecchymosis (0.02).

5. Qi deficiency (0.44): sore waist or knees (0.70), lack of strength (0.73), latitude of limbs or body (0.66), short of breath (0.62), chest oppression (0.54), palpitation (0.45), insomnia (0.58), urinary incontinence (0.38), mental fatigue (0.46), dreamfulness (0.61), and asthenia of defecation (0.20).

6. Non-Qi deficiency (0.56): sore waist or knees (0.21), lack of strength (0.29), latitude of limbs or body (0.23), short of breath (0.20), chest oppression (0.15), palpitation (0.11), insomnia (0.23), urinary incontinence (0.09), mental fatigue (0.18), dreamfulness (0.31), and asthenia of defecation (0.05).

7. Qi stagnation (0.35): chest oppression (0.80), short of breath (0.77), sighing (0.42), hypochondrium distension or pain (0.20), abdominal distension (0.23), and dry stool or constipation (0.35).

8. Non-Qi stagnation (0.65): chest oppression (0.06), short of breath (0.18), sighing (0.12), hypochondrium distension or pain (0.02), abdominal distension (0.07), and dry stool or constipation (0.29).

9. Fire-heat (0.31): dry stool or constipation (0.55), insomnia (0.62), fetid mouth odor (0.30), agitation or short temper (0.79), trembling of limbs (0.24), acid swallow or epigastric upset (0.31), dreamfulness (0.64), spontaneous sweating (0.39), bitter taste in mouth (0.50), and aphiha on mouth or tongue (0.11).

10. Non-Fire-heat (0.69): dry stool or constipation (0.20), insomnia (0.27), fetid mouth odor (0.06), agitation or short temper (0.48), trembling of limbs (0.03), acid swallow or epigastric upset (0.08), dreamfulness (0.36), spontaneous sweating (0.16), bitter taste in mouth (0.26), and aphiha on mouth or tongue (0.01).

11. Phlegm-dampness (0.58): greasy tongue fur (0.80), slippery pulse (0.60), and sticky or greasy feel in mouth (0.29).

12. Non-Phlegm-dampness (0.42): greasy tongue fur (0.03), slippery pulse (0.27), and sticky or greasy feel in mouth (0.05).

3.4.4 Technical note

We end this section with a subtle technical note concerning joint clustering and we explain the issue using the joint clustering model for Qi deficiency (Figure 2). Two symptom groups Y26 (chest oppression, palpitation) and Y27 (short of breath) are used in the model among others. In the global model (Figure 3 in [18]), the symptom variable short of breath is grouped, under Y27, with sighing and three other symptom variables. In the joint clustering model, sighing and the other variables are left out. Hence, short of breath might need to be re-grouped with other symptom variables from elsewhere so that the dependence among the variables is properly modeled. In Lantern, there is a search procedure that decides whether re-grouping is necessary and how it should be done. In the joint clustering model for Qi deficiency, short of breath is automatically merged with Y26 (chest oppression, palpitation), and a new latent variable W is thereby created.

3.5 Syndrome classification

Classification rules are obtained for the eight syndrome types using the building rule function of Lantern. The rule for a syndrome type tells us how to determine whether an individual patient belongs to the syndrome type.

Here is the classification rule for Yang deficiency: sore waist or knees (3.7), latitude of limbs or body (2.8), frequent nocturnal urination (2.5), blackish lower eyelid (3.8), palpitation (2.5), fear of cold or cold limbs (2.6), chest oppression (2.0), thirst desire hot drinks (2.4), spontaneous sweating (2.0), dim complexion (1.7), undigested food in stool (2.6), muscular twitching (1.4), pale complexion (2.0), and diarrhea before dawn (2.1).

Threshold: 9.1, Accuracy: 0.96.

The rule gives a numerical score for each of the symptoms listed, and a threshold. They are calculated from the symptom occurrence probabilities and syndrome prevalence given in the previous subsection. If the total score of a patient exceeds the threshold, then he is classified into the Yang deficiency class. Otherwise, he is classified into the non-Yang deficiency class. The accuracy of the rule is with respect to classification directly based on the joint clustering model.

Note that the position of a symptom in the classification...
rule is determined not only by its score, but also by how frequently it occurs. For example, the score for the symptom *undigested food in stool* is 2.6, which is the fourth highest. However, the symptom occurs with low probability (< 3%) and is hence applicable to only a small fraction of the patients. Consequently, it is placed toward the end of the rule. On the other hand, the symptom *frequent nocturnal urination* is placed at a much earlier position although its score (2.5) is lower. It occurs with high probability (45%) and hence applies to more patients.

Also note that the last four symptoms in the rule did not appear in the quantification of Yang deficiency given in the previous section. They occur with low probabilities and hence they are not important factors to consider when describing the main differences between patients with Yang deficiency and patients without Yang deficiency. They are included in the rule because, when present, they are important evidence for classifying individual patients.

For Yin deficiency, there are two classification rules, one for Yin deficiency versus non-Yin deficiency, and another for Yin deficiency I versus Yin deficiency II (with internal heat):

1. Yin deficiency vs Non-Yin deficiency: sore waist or knees (4.2), blurred vision (3.0), dry eyes (2.8), tinnitus resembling cicada (2.8), insomnia (2.3), dreamlessness (2.0), expectoration (2.1), blackish lower eyelid (3.2), palpitation (1.9), dizziness (1.6), dry stool or constipation (1.4), vexing heat in chest (1.9), trembling of limbs (2.1), fetid mouth odor (1.7), dim complexion (1.0) and tidal fever or night sweat (1.3). Threshold: 10.6; Accuracy: 0.98.

2. Yin deficiency II vs Yin deficiency I: tidal fever or night sweat (4.9), vexing heat in chest (4.7), fetid mouth odor (4.5), spontaneous sweating (4.1), dry tongue (3.6), edema (3.3), thirst desire no drinks (2.8), fast pulse (3.1), deep-red tongue (3.3), dry stool or constipation (1.9), and swift digestion rapid hungering (3.2). Threshold: 13.9; Accuracy: 0.97.

The classification rules for the other six syndrome types are as follows:

1. Blood deficiency: blurred vision (3.4), dry eyes (3.2), palpitation (3.4), insomnia (2.0), dizziness (1.7), dreamlessness (1.6), numbness (1.6), trembling of limbs (2.4), dry stool or constipation (1.5), thin pulse (1.1), muscular twitching (2.1), sallow complexion (1.6), pale lips (1.9), dizzy headache (1.9), and pale complexion (2.0). Threshold: 10.6; Accuracy: 0.98.

2. Blood stasis: purple or darkish lips (5.2), dim complexion (3.1), blackish lower eyelid (3.1), numbness (2.0), palpitation (1.7), scaly skin (2.5), tongue with ecchymosis (2.7), and darkish tongue (1.0). Threshold: 6.4; Accuracy: 0.98.

3. Qi deficiency: sore waist or knees (3.2), lack of strength (2.7), lassitude of limbs or body (2.7), sore of breath (2.7), chest oppression (2.8), palpitation (2.8), insomnia (2.2), urinary incontinence (2.7), dreamlessness (1.8), mental fatigue (2.0), asthenia of defecation (2.4), sunken pulse (1.4), dizziness (1.3), spontaneous sweating (1.4), dripping urination (1.7), feeble pulse (2.4), thin pulse (1.2), and dizzy headache (2.2). Threshold: 13.0; Accuracy: 0.96.

4. Qi stagnation: chest oppression (5.8), short of breath (3.9), sighing (2.4), hypochondrium distension or pain (3.4), abdominal distension (2.0), and dry stool or constipation (0.4). Threshold: 6.2; Accuracy: 0.97.

5. Fire-heat: dry stool or constipation (2.2), insomnia (2.1), fetid mouth odor (2.6), agitation or short temper

---

**Figure 2** Joint clustering model for Qi deficiency
(2.0), trembling of limbs (3.1), acid swallow or epigastric upset (2.4), dreamfulness (1.6), spontaneous sweating (1.9), bitter taste in mouth (1.6), aphtha on mouth or tongue (4.1), dizziness (1.5), dripping urine (1.6), dry tongue (1.7), thirst desire cold drinks (2.5), and throbbing headache (2.1). Threshold: 9.1; Accuracy: 0.94.

(6) Phlegm-dampness: grey tongue fur (7.1), slippery pulse (2.1), sticky or greasy feel in mouth (2.8), thick tongue fur (1.5), dizzy headache (1.8), tooth-marked tongue (1.0), fat tongue (1.0), and urinary incontinence (0.6). Threshold: 3.7; Accuracy: 0.97.

4 Discussion

TCM syndromes are latent variables that cannot be “measured” directly. They can only be “measured” indirectly through symptoms. Based on symptom cooccurrence patterns discovered in data and domain knowledge, we have identified eight groups of symptoms that can be respectively used to “measure” eight syndrome types among patients with VMCI. The patients are partitioned into clusters based on each group of symptoms, and one of the patient clusters is matched up with the syndrome type. The population statistics of the patient cluster are used to quantify the syndrome type (Section 3.4), and to establish a classification rule (Section 3.5).

4.1 Quality of results

There is no ground truth against which to evaluate our results, which was what motivated our work at the first place. However, there are qualitative understandings in TCM theory as to which symptoms are the most important factors to consider when determining whether a patient belongs to a syndrome type. They are called key factors of syndrome differentiation, or Bian Zheng Yao Dian in Chinese. Our results match those key factors well. They also satisfactorily reflect the characteristics of the VMCI population under study.

4.1.1 Results on Blood deficiency

According to TCM theory, the key factors for determining Blood deficiency are pale lips, pale or sallow complexion, dizziness, palpitation, insomnia, and thin pulse. All those symptoms are present in our classification rule of Blood deficiency (Section 3.5). The symptoms with the highest scores are blurred vision and dry eyes. Those reflect the characteristics of the population under study, which consists of patients with VMCI and aged 50 or above. According to TCM theory, seniors usually suffers from kidney deficiency, which can lead to liver Blood deficiency (liver and kidney are from the same source), and hence poor vision (liver’s opening orifice is at eye).

According to our syndrome quantification results (Section 3.4), blurred vision and dry eyes occur with the highest probabilities (0.82 and 0.69) in the Blood deficiency class. Palpitation and insomnia also occur with high probabilities (0.56 and 0.60). Those suggest that, for the population under study, Blood deficiency is mainly located at the liver and the heart.

4.1.2 Results on Blood stasis

According to TCM theory, the key factors for determining Blood stasis are local unpalpable stabbing pain, and color change on the face, the lips and the tongue. In our classification rule for Blood stasis, concordantly, the symptoms with the highest scores are purple or darkish lips, dim complexion, and blackish lower eyelids. The rule also includes tongue with ecchymosis and darkish tongue.

However, our rule does not match TCM theory well regarding local unpalpable stabbing pain. In the VMCI data set, there is only symptom about stabbing pain, i.e., stabbing headache. Stabbing pains at other locations of the body were not included in the survey due to their lack of presence in the TCM literature on Chidai or Daibing. The symptom stabbing headache occurred with very low frequency (< 3%) in the survey and it is negatively, and weakly, correlated with other Blood stasis symptoms listed above. Its score is –0.4 and is deleted during rule simplification.

4.1.3 Results on Qi deficiency

According to TCM theory, the key factors for determining Qi deficiency are lack of strength, short of breath, and lassitude of limbs and body. In our classification rule for Qi deficiency, those are among the symptoms with the highest scores. In addition, the four symptoms sore waist or knees, urinary incontinence, chest oppression and palpitation also have high scores. Those reflect the characteristics of the population under study. The patients in the population are seniors and hence are likely to suffer from kidney deficiency, which manifests as sore waist or knees and urinary incontinence. The patients in the population all suffer from VMCI. The vascular factors can affect the heart and lead to chest oppression and palpitation.

According to our quantification results, the symptoms chest oppression, palpitation, and sore waist or knees occur with high probabilities in the Qi deficiency class. Those suggest that, for the population under study, Qi deficiency is mainly located at the heart and the kidney.

4.1.4 Results on Qi stagnation

According to TCM theory, the key factors for determining Qi stagnation are gastric distention and fullness, distension pain, and sighing. In our classification rule for Qi stagnation, concordantly, the symptoms with the highest scores are chest oppression, sighing, hypochondrium distension or pain, and abdominal distension.

4.1.5 Results on Fire-heat

According to TCM theory, common manifestations of Fire-heat include flushed face, thirst, dry stool or
constipation, red tongue, yellow tongue fur, and fast pulse. In our classification rule for Fire-heat, dry stool or constipation and thirst desire cold drinks have high scores. The other top score symptoms are aphtha on mouth or tongue, fetid mouth odor, acid swallow or epigastric upset, insomnia, throbbing headache, and agitation or short temper. In TCM, those symptoms are considered manifestations of excess heat.

However, the symptoms flushed face, red tongue, yellow tongue fur, and fast pulse are not present in our classification rule. They correlate weakly with the aforementioned symptoms in the data set.

4.1.6 Results on Phlegm-dampness

According to TCM theory, the key factors for determining Phlegm-dampness are the presence of visible phlegm and invisible phlegm, greasy tongue fur, and slippery pulse. Concordantly, greasy tongue fur and slippery pulse have high scores in our classification rule for Phlegm-dampness. Dizzy headache and dizziness (invisible phlegm) are also present. The important role of dizzy headache and dizziness in the rule is consistent with the TCM claim that the blockage of brain collateral by phlegm is an important pathogenic factor for dementia.

4.1.7 Results on Yin deficiency

According to TCM theory, the manifestations of Yin deficiency mainly fall into two categories: those caused by the deprivation of nourishment of viscera and organs due to the lack of Yin fluid, and those caused by the relative excessiveness of Yang due to the failure by Yin to restrain Yang. In our classification rule for Yin deficiency, the symptoms sore waist or knee, blackish lower eyelids, frequent nocturnal urination, and dim complexion suggest kidney Yang deficiency. In addition, when heart Yang fails to warm and push blood, palpitation and chest oppression occur. Weak spleen Yang leads to malnutrition of muscle and meridians, which causes lassitude of limbs and body.

Note that the three symptoms sore waist or knee, blackish lower eyelids, and dim complexion appear both in the rule for Yang deficiency and the rules for Yin deficiency. Their total score in the former rule exceeds the threshold, but not in the latter. The implication is that the co-occurrence of the three symptoms is sufficient evidence for concluding Yang deficiency in the VMCI population, while in general one can only conclude kidney deficiency. The reason is that, in the population under study, the three symptoms are always accompanied by some coldness symptoms such as fear of cold or cold limbs, and thirst desire hot drink. This is an interesting finding about the population that might be useful for treatment planning.

4.2 Uses of results

Our classification rules are score-based and hence are convenient to use. They can be applied in three scenarios.

First, a major weakness with TCM is the lack of gold standards for syndrome differentiation. In clinic practice, different physicians might draw different conclusions about the syndrome type of the same patient. This affects the efficacy of treatments and undermines patient confidence. The problem can be alleviated substantially if physicians can use the same set of rules as reference. Our classification rules can be used for the purpose. In addition, they can be used as a basis for the development of syndrome differentiation standards for VMCI.

Second, randomized controlled trial (RCT) has been widely used to determine the efficacy of TCM treatments. To conduct RCT on a treatment for a TCM syndrome type, one needs to: (1) identify patients belonging to the syndrome types, and (2) then randomly divide them into treatment and control groups. Our classification rules can be used in the first step for RCT on TCM treatments for VMCI patients.

Third, there has been much interest on TCM syndrome essence study, where the objective is to identify biomedical indicators that can be used as gold standards for TCM syndrome differentiation. To conduct such research, one needs to: (1) divide patients into those who belong to a syndrome type and those who do not, and (2) compare the biomedical measurements of patients in the two classes. TCM syndrome essence study has so far achieved little success. A key reason is that the first step was not done properly. In our work, the patient clusters
are identified based on symptom occurrence patterns hidden in data. If there are indeed biomedical indicators that capture the “essence” of syndrome types, they should correlate well with syndrome occurrence patterns, and hence correlate well with the patient clusters identified by our method. As such, our work might help in finding biomedical indicators with high sensitivity and high specificity for TCM syndrome types.

4.3 Strengths and limitations

In this paper we address the problem of classifying VMCI patients into TCM syndrome types. The results are quantitative. Eight syndrome types are identified. Prevalence and score-based classification rules are obtained for each of the syndrome types.

The results are obtained using the LTA method. Consequently, they enjoy stronger evidence support than solutions one could get using the alternative approaches, namely the expert standardization and the supervised quantification approaches. In expert standardization, syndrome classification solutions are determined by panels of experts. In supervised quantification, experts are asked to determine the syndrome types of individual patients. In both cases, the results depend heavily on who the experts are. The LTA method, on the other hand, relies only on symptom data. The results are relatively more objective.

One limitation of our work is that the data were all collected in Northern China. It would be interesting to conduct the study in other regions of the country, or to include patients from other parts of the country in the study.

Another limitation is that only dichotomous data were used in our study, and hence severity of symptoms was not taken into consideration. This is mainly for the sake of simplicity. While it is possible and would be interesting to use polytomous data, doing so would substantially complicate the methodology, software and the results. The literature on LCA has mostly been focused on dichotomous data for the same reason. Since the results are meant to be guidelines for physicians to use as references, rather than rigid standards for physicians to follow meticulously, it is desirable to keep them as simple as possible.

5 Conclusions

By analyzing the symptom data collected on 803 VMCI patients aged 50 or above using latent tree models, we have reached the conclusion that there are eight TCM syndrome types among the population: Blood deficiency, Blood stasis, Phlegm-dampness, Fire-heat, Qi deficiency, Qi stagnation, Yang deficiency, and Yin deficiency. In addition, two subtypes have been identified for Yin deficiency, namely Yin deficiency with and without internal heat. The prevalence and the symptom occurrence probabilities for the syndrome types and subtypes have been determined, and corresponding classification rules have been established. The results match TCM theory well and enjoy the strongest evidence support currently possible.

6 Fundings

Research on this article was supported by the Hong Kong Research Grants Council under grant No. 16202515 and No. 16212516, Guangzhou HKUST Fok Ying Tung Research Institute, China Ministry of Science and Technology TCM Special Research Projects Program under grants No. 200807011, No. 201007002 and No. 201407001-8, Beijing Science and Technology Program under grant No. Z111107056811040, Beijing New Medical Discipline Development Program under grant No. XK100270569, and project of Beijing University of Chinese Medicine under grant No. 2011-CXTD-23.

7 Competing interests

The authors declare that they have no competing interests, financial or non-financial.

REFERENCES


5 White L, Petrovitch H, Hardman J, Nelson J, Davis


