

## Research Article

# Interrater Reliability of Chinese Medicine Diagnosis in People with Prediabetes

Suzanne J. Grant,<sup>1</sup> Rosa N. Schnyer,<sup>2</sup> Dennis Hsu-Tung Chang,<sup>1</sup>  
Paul Fahey,<sup>3</sup> and Alan Bensoussan<sup>1</sup>

<sup>1</sup> CompleMED, School of Science and Health, University of Western Sydney, Locked Bag 1797, Penrith, NSW 2751, Australia

<sup>2</sup> School of Nursing, University of Texas, 1710 Red River Street, 5.188, Austin, TX 78701, USA

<sup>3</sup> School of Science and Health, University of Western Sydney, Locked Bag 1797, Penrith, NSW 2751, Australia

Correspondence should be addressed to Suzanne J. Grant; [s.grant@uws.edu.au](mailto:s.grant@uws.edu.au)

Received 13 January 2013; Revised 8 April 2013; Accepted 9 April 2013

Academic Editor: Zhaoxiang Bian

Copyright © 2013 Suzanne J. Grant et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Background.** Achieving reproducibility in research design is challenging when patient cohorts under study are inconsistently defined. Traditional Chinese medicine (TCM) diagnosis is one example where inconsistency between practitioners has been found. We hypothesise that the use of a validated instrument may improve consistency. Biochemical biomarkers may also be used to enhance reliability. **Methods.** Twenty-seven participants with prediabetes were assessed by two TCM practitioners using a validated instrument (TEAMSI-TCM). Inter-rater reliability was summarised using percentage agreement and the kappa coefficient. One-way ANOVA and Tukey's post hoc test were used to test links between TCM diagnosis and biomarkers. **Results.** The two practitioners agreed on primary diagnosis of 70% of participants.  $\kappa = 0.56$  ( $P < 0.001$ ). The three predominant TCM diagnostic patterns for people with prediabetes were Yin deficiency, Qi and Yin deficiency and Spleen qi deficiency. The Spleen Qi deficiency with Damp cohort had statistically significant higher fasting glucose, higher insulin, higher insulin resistance, higher HbA1c and lower HDL than those with Qi and Yin deficiency. **Conclusions.** Using the TEAMSI-TCM resulted in moderate interrater reliability between TCM practitioners. This study provides initial evidence of variation in the biomarkers of people with prediabetes according to the different TCM patterns which may suggest a route to further improving interrater reliability.

## 1. Background

INTERRATER reproducibility is regarded as one of the foundations of high quality research design. Achieving this standard is challenging when patient cohorts under study are poorly or inconsistently defined and in the absence of objective laboratory data [1, 2]. The subjective process of clinical diagnosis in traditional Chinese medicine (TCM) is one such example. Studies have demonstrated inconsistency between TCM practitioners in diagnosing the same patients [3–5]. We hypothesise, however, that it is possible to improve diagnostic consistency between practitioners, through the use of a validated instrument designed to reflect clinical practice and systematically guide practitioners. Once a TCM diagnosis is made, it may be possible to identify relationships between biochemical biomarkers and the diagnosis.

This may be used to refine the instrument and improve reliability.

In clinical practice, treatment of the patient depends on the differential diagnosis. Western medicine uses disease-based diagnosis, while TCM emphasizes patient-based diagnosis. In TCM, the patient's symptoms and signs are gathered through inquiry, observation, palpation, and smell. These symptoms and signs are interpreted into a diagnostic syndrome that often guides a patient specific (individualised) treatment [6, 7]. The same disease may have many different syndromes because of differences in symptoms and signs at different stages of the disease [8]. For example, two people presenting with the same (western) medical diagnosis of prediabetes may have a different clinical presentation leading to a different TCM diagnosis. One may be overweight with muscle aches and tiredness, while the other might have

a wiry body, experience sweating on the hands and feet and a dry mouth. In TCM, these would be two different syndromes, Spleen Qi deficiency with Damp and Kidney Yin deficiency, respectively. The Chinese practitioner chooses or develops different herbal treatment formulas according to the patient's particular presentation and syndrome. Clinical research in TCM is made difficult by the use of this complex, individualised treatment [9].

For clinical research to be meaningful and usefully applied, it should reflect clinical practice as closely as possible. In order to best reflect clinical practice, TCM diagnostic principles should be incorporated into any clinical trial of TCM treatment [6, 7]. However, this is challenging, resource intensive and typically does not occur. Most research relies on applying a biomedical diagnosis and a standard herbal or acupuncture formula to all participants, with TCM diagnosis included in relatively few studies to date [10, 11]. Separating treatment from diagnosis may undervalue the efficacy of the TCM treatment being assessed [6, 12]. Methods which have been used to include a TCM diagnosis in clinical trials include the following:

- (1) allowing individualized patient treatment based on the TCM diagnosis [13, 14];
- (2) allocating patients to set treatment groups based on several overarching TCM diagnostic categories [15, 16];
- (3) assigning a fixed treatment with some capacity for additional herbs or acupuncture points based on TCM diagnosis [17].

Whilst these studies are commendable in that they aim to better reflect TCM practice to improve the relevance of the research, demonstrable interrater reliability of TCM diagnoses between practitioners is essential in order to ensure definitive, reproducible outcomes. If the diagnosis is unreliable, the appropriateness of the prescribed treatment may be in doubt.

Achieving a reliable diagnosis requires the use of standardized definitions and data collection methods. Methods of minimising variability in the diagnostic process include sharing the same patient history form so that all practitioners have the same information [18, 19] and using set practitioner questionnaires [4, 7]. Some studies have sought to reflect current clinical practice by allowing practitioners to use their own style of diagnostic assessment [20] or a minimal data collection and diagnosis form [21]. However, the lack of an adequate data collection instrument limits the precision with which a diagnosis can be consistently obtained by different practitioners and repeated.

Poor process and poor definitions are likely to lead to variable results. Schnyer and colleagues have developed the Traditional East Asian Medicine Structured Interview, TCM version (TEAMSI-TCM) to assist the diagnostic process and minimise potential variability between individual practitioners presented with the same patient [12, 22].

Furthermore, it has been hypothesised that the reliability of TCM diagnosis may be enhanced by noting associations between patient clinical biomarkers and TCM diagnostic

syndromes [23]. These relationships have increasingly been investigated in China [24, 25].

Our study aimed to (1) assess the interrater reliability of TCM diagnosis of people with prediabetes using a modification of TEAMSI-TCM and (2) document any relationships between biomarkers and TCM syndromes in people with prediabetes.

## 2. Methods

*2.1. Recruitment and Participants.* Twenty-seven participants diagnosed with prediabetes by blood test were enrolled in the study. Men and women over 18 years of age with prediabetes were recruited in Sydney, Australia. Prediabetes is defined as having a fasting plasma glucose (FPG) level of <7.0 mmol/L and 2 hr plasma glucose load level  $\geq$ 7.8 mmol/L and <11.0 mmol/L. This study was nested within a clinical trial investigating the effectiveness of Chinese herbal medicine in the treatment of impaired glucose tolerance (IGT) and insulin resistance in persons with prediabetes or mild diabetes.

*2.2. Practitioner Assessment.* Two accredited TCM practitioners with 4-year Bachelor degrees in TCM from the University of Western Sydney, Australia, and more than 4-year clinical experience conducted the TCM diagnosis. Both practitioners were accredited with the relevant national professional association. The practitioners were both aware that the patient had been diagnosed with prediabetes. They were not required to provide a treatment plan. The practitioners were blind to each other's diagnosis until data entry was complete.

*2.3. Diagnostic Instrument.* The TEAMSI-TCM was used for data collection, modified to incorporate some syndromes specific to prediabetes. The instrument adopts a common TCM (Eight-Principle) approach to clinical differentiation. This is described more fully in standard texts [26].

The TEAMSI-TCM data collection instrument is designed to guide practitioners to use clinical signs, combine them in a systematic way, and generate a diagnosis. It is designed to be used in conjunction with education and training. It is structured in two parts—a patient questionnaire and a practitioner package [22].

A patient questionnaire was first completed by the participant. This provided self-reported data on the participant's symptoms. Practitioner 1 and Practitioner 2 read a copy of the same completed questionnaire prior to commencing the interview with the participant. Neither practitioner had access to the other's examination notes. To minimise measurement error, the interviews between patient and practitioner were usually conducted within an hour of each other in the same setting. The order in which a practitioner was seen varied according to appointment times and availability of practitioners.

The practitioners' package consisted of a section for recording notes taken during the interview of the participant. Practitioners recorded general symptoms and symptoms

TABLE 1: Frequency of agreement of primary TCM diagnoses.

		Practitioner 1			
		Kidney Yin deficiency	Qi and Yin deficiency	Spleen Qi deficiency	Total
Practitioner 2	Kidney Yin deficiency	6 (22%)	1 (4%)	3 (11%)	10 (37%)
	Qi and Yin deficiency	0	7 (26%)	2 (7%)	9 (33%)
	Spleen Qi deficiency	0	2 (7%)	6 (22%)	8 (30%)
Total		6 (22%)	10 (37%)	11	27 (100%)

related to the main complaint. An evaluation section provided guidance for recording observations relating to tongue, pulse, body, constitution, and complexion.

The modification included a third form specifically designed for use in people with prediabetes. It allowed for a primary diagnostic syndrome or “pattern” and one or more secondary or accompanying patterns to be selected from a list. An option was provided for practitioners to contribute additional diagnoses or comments to modify the proposed diagnostic patterns provided on the form.

**2.4. Diagnostic Criteria.** A search of Chinese medicine journals and common texts identified eight distinct diagnostic patterns presented by people with prediabetes [27]. Nine diagnostic patterns were identified:

- (i) Spleen deficiency with Damp (or Damp-heat),
- (ii) Qi and Yin deficiency,
- (iii) Qi deficiency,
- (iv) Yin deficiency (with Empty heat or Blood stasis),
- (v) Phlegm-Damp Qi and Phlegm stagnation,
- (vi) Liver Qi stagnation (with Blood stasis),
- (vii) Yang deficiency,
- (viii) Blood stasis.

These patterns formed the basis of the diagnostic list of syndromes specific to prediabetes in the practitioner package.

**2.5. Statistical Analysis.** To measure the level of agreement between practitioners on the primary and secondary syndromes selected, we adopted the approach used by Macpherson et al. [7]. We assessed the level of exact agreement on diagnosis and presented these as a percentage [28]. We also used the kappa coefficient as the main statistic to determine interrater reliability. Kappa is a measure of observed agreement between raters corrected for chance. A kappa of zero means the observed agreement is consistent with or less than chance agreement and a kappa of one means there is complete agreement [29]. MacPherson and colleagues recommended presenting both, the percentage of patients with congruent classifications and Cohen’s kappa coefficient with 95% confidence limits.

Possible associations between biomarkers and diagnostic patterns were investigated for those patients where both practitioners agreed on the primary diagnosis. As we sought to establish any links between the TCM diagnosis and the

biomarkers, we preferred to use the diagnosis where reliability was clear. One-way ANOVA and Tukey’s post hoc test were used to test for relationship between the diagnostic TCM patterns and each of the biomarker variables. In response to the modest sample size, we highlight all *P* values of less than 0.10 and interpret these as suggestive of potential relationships. Analyses were performed in SPSS version 18.0 (SPSS Inc., Chicago, IL), and CIs for values were calculated using the Wald approximation for 95% confidence intervals: estimate  $\pm 1.96$  times the standard error.

### 3. Results

Twenty-seven participants with a mean age of 57.3 yrs (range 36–75 yrs) were recruited from June 2007 to December 2009. The mean fasting blood glucose was 6.2 mmol/L, and mean 2 hr OGTT was 10.6 mmol/L. The majority of participants had hypertension ( $n = 16$ ) and/or high cholesterol ( $n = 14$ ).

The two practitioners agreed exactly on 70% of the primary diagnoses for individual participants (see Table 1). Three TCM diagnostic patterns for people with prediabetes featured Yin deficiency, Qi and Yin deficiency, and Spleen Qi deficiency. These three patterns were fairly evenly distributed among the participants. The interrater reliability for the practitioners was found to be kappa = 0.56 ( $P < 0.001$ ), 95% CI (0.25 to 0.81). This is statistically significantly higher than chance and represents a moderate level of agreement.

Secondary diagnostic patterns identified by practitioners were primarily Damp, Damp-heat, Liver Qi stagnation, and Blood stagnation (Table 2). Most participants (89%) received at least one secondary diagnosis by both practitioners. As multiple secondary patterns were identifiable for each participant, no attempt was made to calculate levels of agreement between practitioners. The hierarchical nature of secondary diagnosis within primary diagnoses was not amenable to kappa analysis. Exact agreement on both primary and secondary diagnostic patterns was 41% (Table 3).

An examination of the practitioners’ TEAMSI-TCM notes from patient interviews and evaluations revealed some obvious reasons for different diagnoses. The main reasons appeared to be differences in what was divulged by a patient to one practitioner and not the other, differences in tongue and pulse evaluations, and varying levels of importance or weight according to certain symptoms.

Where there was exact agreement on TCM diagnosis ( $n = 19$ ), associations with biochemical markers were examined (Table 4). Despite the relatively small sample size, there was evidence that the Spleen Qi deficiency with Damp cohort

TABLE 2: Secondary diagnostic syndrome only.

Accompanying syndrome	Practitioner	
	1	2
Damp heat	7	2
Spleen Qi deficiency	3	1
Plus Damp	11	14
Blood stagnation	5	9
Liver Qi stagnation	10	6
Yang xu	1	0
Phlegm	0	1
Yin deficiency	2	0

TABLE 3: Agreed primary and secondary diagnoses.

Pattern	<i>n</i>
Yin deficiency with spleen Qi deficiency	1
Yin deficiency with Damp heat	1
Yin deficiency with blood stagnation	1
Qi and Yin deficiency with Damp and blood stagnation	1
Qi and Yin deficiency with Damp	1
Spleen Qi deficiency with Damp and liver Qi stagnation	4
Spleen Qi deficiency with Damp heat and liver Qi stagnation	1
Spleen Qi deficiency with Damp heat	1

had on average higher fasting glucose ( $P < 0.10$ ), higher insulin ( $P < 0.05$ ), higher insulin resistance ( $P < 0.05$ ) than both those diagnosed with Qi and Yin deficiency and those participants with Yin deficiency alone. Participants with Spleen Qi deficiency and Damp also had a higher HbA1c ( $P < 0.10$ ) than those participants diagnosed with Qi and Yin deficiency. Participants diagnosed with Spleen Qi deficiency and Damp were also distinct from those diagnosed with Yin deficiency, having higher triglycerides ( $P < 0.05$ ), higher BMI ( $P < 0.05$ ), and lower HDL ( $P < 0.10$ ). There was no evidence of differences in mean biomarkers of people with Yin deficiency and Qi and Yin deficiency.

#### 4. Discussion

This study aimed to assess the interrater reliability of traditional Chinese medicine diagnosis of people with prediabetes using a structured assessment instrument and to explore the relationship between TCM patterns and biomarkers of prediabetes. Although modest in size, it significantly contributes to an understanding of TCM diagnostic patterns in people with prediabetes.

The analysis of 54 diagnoses of 27 participants found a moderate level of agreement between the two practitioners on the primary diagnosis. That is, the practitioners agreed on nearly 6 out of 10 diagnoses. The percentage of congruent classifications on the primary diagnosis was slightly higher at 70%.

All but three participants in our study received at least one accompanying or secondary diagnosis. Diagnosis of multiple diagnostic patterns was a feature of previous research on

interrater reliability [7, 30]. The diagnosis of secondary patterns tends to be subject to lower levels of interrater reliability [31]. This was reflected in our study where complete agreement between practitioners on both the primary and the accompanying diagnoses occurred in only 41% of cases.

The level of agreement on primary diagnosis found in our study is consistent with other interrater reliability studies where two practitioners were used [7, 20, 32]. It should be noted that in general when a diagnosis or clinical judgement requires a subjective assessment the interrater reliability between assessors is generally low [19]. This has been shown in studies assessing interrater reliability in diagnosing schizophrenia or assessing suspected stroke [33, 34].

The moderate level of agreement on the primary diagnosis achieved between the two practitioners in this study may be attributed to several factors—the similar backgrounds of the practitioners, the use of a thorough diagnostic instrument, and the limited number of practitioners involved. It is possible that a higher rate of agreement may have been reached if the diagnostic categories were constructed differently or training was conducted to ensure a shared understanding.

Previous studies have found that differences in training affect the way a diagnosis is communicated [19, 21]. Our study limited this confounding factor by selecting practitioners who were trained in the same institution and with a similar length of practice experience. Conversely one study found that utilising TCM practitioners from a similar background and clinical experience made little or no difference to the average agreement of TCM diagnosis [19].

One of the strengths of this study was the use of the diagnostic TEAMSI-TCM tool. The benefits were manifold. The comprehensive patient questionnaire provided both practitioners with the same starting point. This ensured that a substantial amount of information was shared, and thereby limited variations of what was divulged in the patient-practitioner interview. Practitioners were guided to use a thorough assessment protocol rather than rely on pulse and/or tongue alone. A rationale for a diagnosis could clearly be seen through examining patient questionnaires and practitioner notes and evaluations. Our findings are supported by previous studies that have found that the use of more objective tools such as “inquiry” or questionnaire-based diagnoses process improves reliability [4, 35].

One of the most remarkable results of this study was that reliability of the TCM diagnostic framework was not null, given that prediabetes is a biomedical disorder determined by elevated blood glucose levels, without any overt symptoms and signs. Yet the TCM framework seems to provide an internally consistent framework to assess individual patient differences that may prove significant in prognosis or treatment. In TCM terms, these individuals with IGT are diagnostically “out of balance” in subtly different ways. The aim of TCM treatment is to readjust balance and enhance self-healing. Treatment is adjusted at each visit to meet the dynamic nature of the disease and prevent disease, in this case, progression to diabetes. The value of individualised treatment has been established in some studies [16, 36]. O’Brien et al.

TABLE 4: TCM diagnosis for prediabetes and biomarkers.

Marker	TCM diagnosis			One-way ANOVA <i>P</i> value
	Yin deficiency ( <i>n</i> = 6)	Qi and Yin deficiency ( <i>n</i> = 7)	Spleen Qi deficiency with Damp ( <i>n</i> = 6)	
Hypertension <sup>^</sup>	3	5	3 <sup>a</sup>	
Fasting blood glucose mmol/L (SD)	5.8 (0.8)	5.9 (0.7)	7.0 (0.9)*	0.054
2 hr glucose tolerance mmol/L (SD)	9.8 (2.2)	9.3 (1.6)	12.3 (3.4)	0.131
Insulin mmol/L (SD)	12.5 (8.3)	12.1 (4.6)	23.8 (8.6)**	0.027
Insulin resistance (SD)	1.8 (1.1)	1.6 (0.7)	3.2 (1.1)**	0.035
Glycated haemoglobin (HbA1c %) (SD)	6.1 (0.6)	6.0 (0.4)	6.6 (0.3) <sup>†</sup>	0.060
Triglycerides mmol/L (SD)	1.2 (0.6)	1.7 (0.7)	3.2 (2.2) <sup>^^</sup>	0.053
Total cholesterol mmol/L (SD)	4.8 (0.7)	5.1 (1.0)	4.8 (0.9)	0.755
High density lipoprotein mmol/L (SD)	1.8 (0.5)	1.6 (0.5)	1.1 (0.3) <sup>^</sup>	0.074
Body Mass Index (SD)	27.7 (2.5)	28.7 (4.5)	33.7 (3.6) <sup>^^</sup>	0.057

<sup>a</sup>Biochemical data missing for one participant.

<sup>^</sup>Refers to participants who were currently taking medication for hypertension.

\*Difference between spleen Qi deficiency with Damp and the other two diagnostic categories (*P* < 0.10).

\*\*Difference between spleen Qi deficiency with Damp and the other two diagnostic categories (*P* < 0.05).

<sup>†</sup>Difference between spleen Qi deficiency with Damp and Qi and yin deficiency (*P* < 0.10).

<sup>††</sup>Difference between spleen Qi deficiency with Damp and Qi and yin deficiency (*P* < 0.05).

<sup>^</sup>Difference between spleen Qi deficiency with Damp and Yin deficiency (*P* < 0.10).

<sup>^^</sup>Difference between spleen Qi deficiency with Damp and Yin deficiency (*P* < 0.05).

note that reproducibility may be higher in syndromes which are more commonly and better understood by practitioners [32].

No training was given on the diagnostic criteria for the patterns nominated in the TEAMSI-TCM tool. Different understanding and weighting of signs and symptoms of the three syndromes (Spleen Qi deficiency, Yin deficiency, and Qi and Yin deficiency) may have affected interrater reliability. There may have been improved interrater reliability if practitioners were taken through consensus training, or similar, to ensure that practitioners were all “speaking the same language” [35, 37].

Practitioner notes revealed factors which influence clinical judgement and are difficult to control. These are likely to be important barriers to reaching high levels of interrater reliability in TCM diagnosis. It would seem that the best approach is to utilise methods that guide diagnosis such as using a validated instrument. Even if we educate and train practitioners on these instruments and the nominated diagnostic categories, clinical judgement may confound the result to a lesser degree. Where the priority is to reach consensus for a clinical trial, a further strategy, suggested by O’Brien and Birch, may be used. Whereby two practitioners examine each patient and commence treatment only when diagnostic consensus has been reached [18].

Our results suggest that some differences in diagnosis were more likely than others. This implies some diagnostic patterns are more similar to each other than others. As further information accumulates about the relative “distances” between different diagnoses, weights for kappa may be developed to refine quantification of the magnitude of disagreement [38].

The three predominant TCM diagnostic patterns for people with prediabetes found in our study (Yin deficiency, Qi and Yin deficiency, and Spleen Qi deficiency) reflect those that had been found in previous literature. The pathophysiology of the progression from a prediabetic state of impaired glucose tolerance to diabetes can be seen in TCM as a progression from a deficiency of Spleen Qi with an ensuing stagnation of Damp, the generation of heat, consumption of fluids, and the emergence of Yin and Qi deficiency as the dominant pattern. Some patients will progress through this continuum, and different syndromes will dominate, while for others progression may not be so well defined [39–41]. In eight reviews of diagnostic patterns in people with prediabetes, spleen deficiency with Damp, and Qi and Yin deficiency, closely followed by Yin deficiency, were found to be dominant patterns. Blood stasis was a commonly identified secondary diagnostic pattern [24, 25, 39, 42–46].

The use of biomarkers integrated with TCM diagnosis has been proposed as a strategy that may improve reliability [23]. This small but emerging area of research has found relationships between certain diagnostic patterns and certain biomarkers. For example, eosinophils, which play an important role in the pathogenesis of bronchial asthma and allergic rhinitis, have been found to be more elevated in the patients with typical heat (*zheng*) [47]. Xu et al. (1993) found that high levels of platelet aggregation activity corresponded to blood stagnation [48]. Another study found that diagnosing blood stagnation in people with rheumatoid arthritis was possible through using a combination of proteomic and bioinformatics-based classification methods [49].

TCM diagnostic patterns for prediabetes are not well documented; nonetheless some research has been conducted

on linking biomarkers to these patterns [24, 42]. In our research, we found that the Spleen Qi deficiency group had higher fasting blood glucose (FBG), insulin, greater insulin resistance, and higher HbA1c than the Qi and Yin deficiency group. This group also had higher FBG, insulin, BMI, and worse HDL cholesterol than the Yin deficiency group. The biomarker pathology for people with prediabetes diagnosed with Spleen Qi deficiency may therefore constitute a quite distinct group. This biochemical characterisation of Spleen Qi deficiency was similar to previous studies of biomarkers in prediabetes [24, 25, 50].

Chen and colleagues also found that people with prediabetes diagnosed with Qi and Yin deficiency tended to have hypertension and elevated low density lipids (LDL). We found that there was no significant variation in the biomarkers of people with Yin deficiency compared to those with Qi and Yin deficiency. This may have been because these groups were too similar or because the sample was too small.

Further research on the relationships between TCM diagnosis and biomarkers may provide further objective diagnostic criteria for use in clinical trials.

As well as the small sample size and low statistical power, there are some other limitations to the design of this study. We used two practitioners to perform the diagnoses. Using a smaller number of raters was likely to result in higher percent agreement. Patients were diagnosed on one occasion only. This has been raised as a limitation in previous research, stating that patients are not generally seen only once by a practitioner in clinical practice [19, 37]. Typically a TCM diagnosis is refined over a few visits as signs and symptoms become more or less apparent through “inquiry,” “observation,” and “palpation.” There are several well-described limitations to the use of the kappa statistic that can influence the interpretation of results [51–53]. Very low (or high) prevalence results in high levels of expected agreement, and consequently the kappa value is often low despite near perfect agreement. We do not believe this occurred in our study. This study does not consider other diagnostic traditions of Chinese medicine. *Bianzheng lunzhi* or “determining treatment on the basis of discerning the syndrome” has achieved central position as the method for diagnosis in Chinese medicine. But in clinical practice this is one of many diagnostic and therapeutic strategies employed by clinicians [54].

## 5. Conclusion

When using the Eight-Principle Pattern Differentiation model characteristic of TCM, practitioners diagnosing patients with prediabetes commonly selected one of three patterns: Qi and Yin deficiency, Spleen Qi deficiency, or Yin deficiency. There was a moderate level of interrater reliability between practitioners. From the diagnostic data clear diagnostic patterns or TCM categories and subcategories of people with prediabetes are apparent.

Our research contributes to the growing body of knowledge about the reliability of TCM diagnostic techniques and explores the potential use of biomarkers in enhancing this

knowledge. Specifically it expands our understanding of the main TCM patterns present in people with prediabetes.

The interrater reliability of TCM diagnosis merits wider study. This study has demonstrated a viable methodology. Future studies should include larger sample sizes of both practitioners and patients to address differences in training, education and clinical experience, conducting training, and “calibration” exercises. The larger samples sizes and accumulating knowledge of the relationship between diagnoses will inform the development of more refined statistical investigations (such as weights for the kappa statistic). The continued improvement of the face and content validity of the TEAMSI-TCM instrument will result in a robust tool for the conduct of interrater reliability studies and, in the future, clinical trials.

## Acknowledgments

Suzanne Grant would like to thank the Australian Acupuncture and Chinese Medicine Association for contributing financial support to this research and Emma Howe for assisting in this study.

## References

- [1] E. Flossmann, J. N. Redgrave, D. Briley, and P. M. Rothwell, “Reliability of clinical diagnosis of the symptomatic vascular territory in patients with recent transient ischemic attack or minor stroke,” *Stroke*, vol. 39, no. 9, pp. 2457–2460, 2008.
- [2] R. C. Serin, “Diagnosis of psychopathology with and without an interview,” *Journal of Clinical Psychology*, vol. 49, no. 3, pp. 367–372, 1993.
- [3] O. Birkeflet, P. Laake, and N. Völlestad, “Low inter-rater reliability in traditional Chinese medicine for female infertility,” *Acupuncture in Medicine*, vol. 29, no. 1, pp. 51–57, 2011.
- [4] S. Mist, C. Ritenbaugh, and M. Aickin, “Effects of questionnaire-based diagnosis and training on inter-rater reliability among practitioners of traditional Chinese medicine,” *Journal of Alternative and Complementary Medicine*, vol. 15, no. 7, pp. 703–709, 2009.
- [5] C. J. Hogeboom, K. J. Sherman, and D. C. Cherkin, “Variation in diagnosis and treatment of chronic low back pain by traditional Chinese medicine acupuncturists,” *Complementary Therapies in Medicine*, vol. 9, no. 3, pp. 154–166, 2001.
- [6] R. Hammerschlag, “Methodological and ethical issues in clinical trials of acupuncture,” *Journal of Alternative and Complementary Medicine*, vol. 4, no. 2, pp. 159–171, 1998.
- [7] H. MacPherson, L. Thorpe, K. Thomas, and M. Campbell, “Acupuncture for low back pain: traditional diagnosis and treatment of 148 patients in a clinical trial,” *Complementary Therapies in Medicine*, vol. 12, no. 1, pp. 38–44, 2004.
- [8] F. Yu, T. Takahashi, J. Moriya et al., “Traditional Chinese medicine and kampo: a review from the distant past for the future,” *Journal of International Medical Research*, vol. 34, no. 3, pp. 231–239, 2006.
- [9] R. L. Nahin and S. E. Straus, “Research into complementary and alternative medicine: problems and potential,” *British Medical Journal*, vol. 322, no. 7279, pp. 161–164, 2001.
- [10] J. L. Shea, “Applying evidence-based medicine to traditional Chinese medicine: debate and strategy,” *Journal of Alternative*

- and *Complementary Medicine*, vol. 12, no. 3, pp. 255–263, 2006.
- [11] J. L. Tang, S. Y. Zhan, and E. Ernst, “Review of randomised controlled trials of traditional Chinese medicine,” *British Medical Journal*, vol. 318, no. 7203, pp. 160–161, 1999.
  - [12] R. N. Schnyer and J. J. B. Allen, “Bridging the gap in complementary and alternative medicine research: manualization as a means of promoting standardization and flexibility of treatment in clinical trials of acupuncture,” *Journal of Alternative and Complementary Medicine*, vol. 8, no. 5, pp. 623–634, 2002.
  - [13] H. C. Diener, K. Kronfeld, G. Boewing et al., “Efficacy of acupuncture for the prophylaxis of migraine: a multicentre randomised controlled clinical trial,” *The Lancet Neurology*, vol. 5, no. 4, pp. 310–316, 2006.
  - [14] L. A. Kalish, B. Buczynski, P. Connell et al., “Stop Hypertension with the Acupuncture Research Program (SHARP): clinical trial design and screening results,” *Controlled Clinical Trials*, vol. 25, no. 1, pp. 76–103, 2004.
  - [15] S. Joos, C. Schott, H. Zou, V. Daniel, and E. Martin, “Immunomodulatory effects of acupuncture in the treatment of allergic asthma: a randomized controlled study,” *Journal of Alternative and Complementary Medicine*, vol. 6, no. 6, pp. 519–525, 2000.
  - [16] A. Bensoussan, N. J. Talley, M. Hing, R. Menzies, A. Guo, and M. Ngu, “Treatment of irritable bowel syndrome with Chinese herbal medicine: a randomized controlled trial,” *Journal of the American Medical Association*, vol. 280, no. 18, pp. 1585–1589, 1998.
  - [17] B. Brinkhaus, J. Hummelsberger, R. Kohnen et al., “Acupuncture and Chinese herbal medicine in the treatment of patients with seasonal allergic rhinitis: a randomized-controlled clinical trial,” *Allergy*, vol. 59, no. 9, pp. 953–960, 2004.
  - [18] K. A. O’Brien and S. Birch, “A review of the reliability of traditional east asian medicine diagnoses,” *Journal of Alternative and Complementary Medicine*, vol. 15, no. 4, pp. 353–366, 2009.
  - [19] G. G. Zhang, W. Lee, B. Bausell, L. Lao, B. Handwerger, and B. Berman, “Variability in the Traditional Chinese Medicine (TCM) diagnoses and herbal prescriptions provided by three TCM practitioners for 40 patients with rheumatoid arthritis,” *Journal of Alternative and Complementary Medicine*, vol. 11, no. 3, pp. 415–421, 2005.
  - [20] D. Kalaoukalan, K. J. Sherman, and D. C. Cherkin, “Acupuncture for chronic low back pain: diagnosis and treatment patterns among acupuncturists evaluating the same patient,” *Southern Medical Journal*, vol. 94, no. 5, pp. 486–492, 2001.
  - [21] R. R. Coeytaux, W. Chen, C. E. Lindemuth, Y. Tan, and A. C. Reilly, “Variability in the diagnosis and point selection for persons with frequent headache by Traditional Chinese Medicine acupuncturists,” *Journal of Alternative and Complementary Medicine*, vol. 12, no. 9, pp. 863–872, 2006.
  - [22] R. N. Schnyer, L. A. Conboy, E. Jacobson et al., “Development of a Chinese medicine assessment measure: an interdisciplinary approach using the Delphi method,” *Journal of Alternative and Complementary Medicine*, vol. 11, no. 6, pp. 1005–1013, 2005.
  - [23] G. G. Zhang, B. Bausell, L. Lao, B. Handwerger, and B. M. Berman, “Assessing the consistency of traditional Chinese medical diagnosis: an integrative approach,” *Alternative Therapies in Health and Medicine*, vol. 9, no. 1, pp. 66–71, 2003.
  - [24] L. Z. Chen and H. Jau YB, “TCM syndrome categorization on IGT,” *Chinese Journal of Basic Medicine in Traditional Chinese Medicine*, vol. 13, no. 3, pp. 223–224, 2007.
  - [25] Y. Lu, “The relationship between the insulin levels of IGT patients and TCM syndrome differentiation,” *Zhejiang Journal of Traditional Chinese Medicine*, vol. 5, p. 220, 2003.
  - [26] G. Maciocia, *The Foundations of Chinese Medicine*, Churchill Livingstone, New York, NY, USA, 1989.
  - [27] S. J. Grant, *Chinese herbs for the treatment of impaired glucose tolerance and insulin resistance [Doctoral dissertation]*, Dissertations and Theses database, 2009.
  - [28] V. W. Steinijans, E. Diletti, B. Bömches, C. Greis, and P. Solleder, “Interobserver agreement: cohen’s kappa coefficient does not necessarily reflect the percentage of patients with congruent classifications,” *International Journal of Clinical Pharmacology and Therapeutics*, vol. 35, no. 3, pp. 93–95, 1997.
  - [29] J. R. Landis and G. G. Koch, “The measurement of observer agreement for categorical data,” *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.
  - [30] K. J. Sherman, D. C. Cherkin, and C. J. Hogeboom, “The diagnosis and treatment of patients with chronic low-back pain by Traditional Chinese Medical acupuncturists,” *Journal of Alternative and Complementary Medicine*, vol. 7, no. 6, pp. 641–650, 2001.
  - [31] G. G. Zhang, B. Bausell, L. Lao, W. L. Lee, B. Handwerger, and B. Berman, “The variability of TCM pattern diagnosis and herbal prescription on rheumatoid arthritis patients,” *Alternative Therapies in Health and Medicine*, vol. 10, no. 1, pp. 58–63, 2004.
  - [32] K. A. O’Brien, E. Abbas, J. Zhang et al., “An investigation into the reliability of Chinese medicine diagnosis according to eight guiding principles and Zang-Fu theory in Australians with hypercholesterolemia,” *Journal of Alternative and Complementary Medicine*, vol. 15, no. 3, pp. 259–266, 2009.
  - [33] P. J. Hand, J. A. Haisma, J. Kwan et al., “Interobserver agreement for the bedside clinical assessment of suspected stroke,” *Stroke*, vol. 37, no. 3, pp. 776–780, 2006.
  - [34] M. Maj, R. Pirozzi, A. M. Formicola, L. Bartoli, and P. Bucci, “Reliability and validity of the DSM-IV diagnostic category of schizoaffective disorder: preliminary data,” *Journal of Affective Disorders*, vol. 57, no. 1–3, pp. 95–98, 2000.
  - [35] G. G. Zhang, B. Singh, W. Lee, B. Handwerger, L. Lao, and B. Berman, “Improvement of agreement in TCM diagnosis among TCM practitioners for persons with the conventional diagnosis of rheumatoid arthritis: effect of training,” *Journal of Alternative and Complementary Medicine*, vol. 14, no. 4, pp. 381–386, 2008.
  - [36] R. Manber, R. N. Schnyer, D. Lyell et al. et al., “Acupuncture for depression during pregnancy: a randomized controlled trial,” *Obstetrics and Gynecology*, vol. 115, no. 3, pp. 511–520, 2010.
  - [37] J. J. Y. Sung, W. K. Leung, J. Y. L. Ching et al., “Agreements among traditional Chinese medicine practitioners in the diagnosis and treatment of irritable bowel syndrome,” *Alimentary Pharmacology and Therapeutics*, vol. 20, no. 10, pp. 1205–1210, 2004.
  - [38] A. Bialocerkowski, N. Klupp, and P. Bragge, “How to read and critically appraise a reliability article,” *International Journal of Therapy and Rehabilitation*, vol. 17, no. 3, pp. 114–112, 2010.
  - [39] S. Wu, “Incapability of Spleen to disperse essence and its relation with IGT,” *China Journal of Traditional Chinese Medicine and Pharmacy*, vol. 19, no. 8, pp. 463–465, 2004.
  - [40] B. Flaws, L. Kuchinski, and R. Casanas, *The Treatment of Diabetes Mellitus with Chinese Medicine*, Blue Poppy Press, 2002.
  - [41] W. Fang, G. Fan, X. Chen, and Y. Kui, *Diabetes and Obesity*, People’s Medical Publishing House, 2008.

- [42] J. Lian, "The pattern discrimination and treatment of 60 cases of increased glucose tolerance," *Zhejiang Journal of Chinese Medicine*, 2004.
- [43] Y. M. Dong, S. E. Zhang, and X. M. Liu, "Function of pancreatic islet beta cells and features of TCM symptoms and syndromes in the non-diabetic first-grade relatives of patients with type 2 diabetes mellitus," *Chinese Journal of Integrated Traditional and Western Medicine*, vol. 25, no. 12, pp. 1089–1091, 2005.
- [44] S. Li, X. Feng, C. Qin, and G. Lou, "The recent studies, development and outlook on IGT and TCM," *Central China Medicine Magazine*, vol. 30, no. 5, pp. 437–438, 2006.
- [45] K. Chai and G. Song, "Impaired glucose tolerance influenced by TCM disease prevention," *Henan Traditional Chinese Medicine*, vol. 28, no. 1, pp. 17–19, 2008.
- [46] X. Song, "Current status on research on Chinese medicine for impaired glucose tolerance," *Gansu Journal of TCM*, vol. 20, no. 5, pp. 88–91, 2007 (Chinese).
- [47] E. Yu, "Association between disease patterns in Chinese medicine and Serum ECP in Allergic Rhinitis," *Chinese Medical Journal*, vol. 8, no. 1, pp. 19–26, 1999.
- [48] X. Xu, J. Z. Liao, and S. R. Wang, "Relation between traditional Chinese medicinal syndrome differentiation and blood platelet function in 310 cases of blood stasis," *Chinese Journal of Integrated Traditional and Western Medicine*, vol. 13, no. 12, pp. 718–707, 1993.
- [49] C. Matsumoto, T. Kojima, K. Ogawa et al., "A proteomic approach for the diagnosis of "Oketsu" (blood stasis), a pathophysiologic concept of Japanese traditional (Kampo) medicine," *Evidence-based Complementary and Alternative Medicine*, vol. 5, no. 4, pp. 463–474, 2008.
- [50] W. B. Fang FG, X. H. Chen, and Y. Kui, *Diabetes and Obesity*, People's Medical Publishing House, 2008.
- [51] P. Brennan and A. Silman, "Statistical methods for assessing observer variability in clinical measures," *British Medical Journal*, vol. 304, no. 6840, pp. 1491–1494, 1992.
- [52] M. W. Watkins and M. Pacheco, "Interobserver agreement in behavioral research: importance and calculation," *Journal of Behavioral Education*, vol. 10, no. 4, pp. 205–212, 2000.
- [53] P. J. Karanicolas, M. Bhandari, S. D. Walter et al., "Interobserver reliability of classification systems to rate the quality of femoral neck fracture reduction," *Journal of Orthopaedic Trauma*, vol. 23, no. 6, pp. 408–412, 2009.
- [54] V. Scheid, *Chinese Medicine in Contemporary China: Plurality and Synthesis*, Duke University Press, Durham, NC, USA, 2002.