Review Article

Evaluating traditional Chinese medicine diagnostic instruments for functional dyspepsia: systematic review on measurement properties

Leonard TF Ho[a], Vincent CH Chung[a,b,*], Charlene HL Wong[b], Irene XY Wu[c], Kun Chan Lan[d], Darong Wu[e], Jerry WF Yeung[f], Nevin L Zhang[g], Ting Hung Leung[a], Justin CY Wu[h]

[a] School of Chinese Medicine, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong
[b] The Jockey Club School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong
[c] Xiangya School of Public Health, Central South University, Hunan, China
[d] The Institute of Medical Informatics, National Cheng Kung University, Taiwan
[e] The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangdong, China
[f] School of Nursing, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Hong Kong
[g] Department of Computer Science and Engineering, School of Engineering, The Hong Kong University of Science and Technology, Hong Kong
[h] Department of Medicine and Therapeutics, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

ABSTRACT

Background: Pattern diagnosis-guided treatments in Traditional Chinese Medicine (TCM) has been recognised by the eleventh revision of the International Classification of Diseases (ICD-11). Accurate pattern diagnosis requires reliable and valid diagnostic instruments that guide the collection of TCM clinical data without bias. This study synthesised the existing TCM diagnostic instruments for functional dyspepsia (FD) and appraised their quality regarding their development process and measurement properties.

Methods: Seven electronic databases were searched for validation studies on TCM diagnostic instruments for FD. Synthesis and appraisal of the included studies were performed following the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) Initiative guidelines adapted for TCM diagnostic instruments. Risk of bias assessment was conducted using the *COSMIN Risk of Bias Checklist.*

Results: Five studies were included, with five unique TCM diagnostic instruments for FD identified. All five diagnostic instruments were of inadequate quality in terms of their development process, implying a shortcoming in their relevance, comprehensibility, and comprehensiveness. Only the criterion validity of *Stomach Qi Deficiency Pattern Assessment Scale* was of sufficient quality and had no risk of bias in its validation.

Conclusion: The quality of TCM diagnostic instruments for FD warrants urgent improvements. None of them was considered reliable or valid for guiding TCM pattern diagnosis. To support the evidence base of the standardization of TCM patterns in ICD-11, TCM diagnostic instruments should be developed and validated rigorously under the COSMIN guidelines. Amendments should be made on the guidelines to accommodate the features and uniqueness of TCM diagnostic process.

1. Introduction

Pattern diagnosis-guided treatment is an essential component in the practice of Traditional Chinese Medicine (TCM). It refers to the individualization of prescriptions according to patients' TCM pat-

tern diagnosed from an overall analysis of complex clinical information gathered via "Four Diagnostic Methods".[1] To facilitate the incorporation of TCM into primary healthcare systems, the World Health Organization (WHO) has been working on the standardization of TCM terminologies and classifications since 1981.[1–3] The WHO program of *International Classification of Traditional Medicine* (ICTM) aims to standardize disease names, patterns, symptoms, signs, indications for treatment, and interventions within the TCM system.[3] A consensus was reached in 2005 after two informal con-

* Corresponding author at: School of Public Health and Primary Care, Prince of Wales Hospital, Room 413, Shatin, New Territories, Hong Kong.
E-mail address: vchung@cuhk.edu.hk (V.C. Chung).

sultations and one formal meeting involving international Traditional Medicine experts from eight Member States.[1] Results of this standardization were published as the *WHO International Standard Terminologies on Traditional Medicine in the Western Pacific Region,*[1] and this have been included in the eleventh revision of the International Classification of Diseases (ICD-11) as a dedicated supplementary chapter.[4, 5] This is considered as the first attempt to incorporate of Traditional Medicine into the ICD-based structure, with an aim of facilitating interprofessional communication and collaboration between traditional and conventional clinicians.[6, 7] standardizing TCM patterns for conventional diagnoses is expected to improve the inter-rater agreement among TCM practitioners in pattern diagnosis and encourage the incorporation of TCM pattern diagnosis into clinical trials.[8-11] Research on the standardization of pattern diagnosis process resembles research on diagnostic evaluation without gold standard.[12, 13] In this process, reliable and valid measurement instruments are vital to collecting essential patient data, allowing repeatable diagnosis of TCM patterns.[14–16] Such TCM diagnostic instruments have been developed and validated for various conditions, including mental stress,[17] lateral elbow pain,[18] ulcerative colitis,[19] and coronary heart disease.[20]

These instruments are developed to facilitate accurate TCM pattern diagnosis, such that treatments can be better personalised for patients. For instance, electroacupuncture and herbal medications are found to be effective for managing symptoms of functional dyspepsia (FD),[21-23] yet the current evidence base is not established under the principle of pattern diagnosis-guided treatment. Although it is expected that personalised treatments guided by pattern diagnosis are superior to standardised prescriptions in terms of effectiveness,[24] it is unclear which TCM diagnostic instruments are of sufficient reliability and validity for FD.[25]

This systematic review aimed to identify a reliable and valid instrument for facilitating TCM pattern diagnosis of FD. We synthesised the existing TCM diagnostic instruments for FD and appraised their development process and measurement properties using evidence-based guidelines published by the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) Initiative.[25, 26]

2. Methods

2.1. Eligibility criteria

We expected to include both published and unpublished studies in Chinese or English, without date limitation, on the validation of TCM diagnostic instruments for FD. We focused on TCM diagnostic instruments which were designed to help collect clinical data via "Four Diagnostic Methods" (inspection, listening and smelling, inquiry, and palpation).[27] Studies on TCM diagnostic instruments for functional gastrointestinal disorders were also included, as long as dyspeptic patients were sampled. All included studies must attach or properly cite at least one accessible full text of the instrument *per se*. Moreover, they must report the results of at least one measurement property validated. When duplicate studies were identified, only the most recent versions were selected.

2.2. Search strategy

Seven electronic databases were searched from their inception to October 16 2019,. Five of them were Chinese: Wanfang Data, China Academic Journals Full-text Database, China Academic Journal Network Publishing Database, SinoMed, and Index to Taiwan Periodical Literature System. Two were English: MEDLINE and EMBASE. Validated search filters with high sensitivity for measurement instruments were applied on MEDLINE and EMBASE.[28, 29] Details of the search strategies are illustrated in Appendix 1.

We also conducted manual searches for potentially eligible studies from reference lists of included studies.

2.3. Study selection

References from the seven electronic databases were imported into EndNote X9 (Clarivate Analytics, Philadelphia, United States). After deduplication, titles and abstracts were screened, and full text of the potentially eligible citations was retrieved for further assessment and confirmation. Two independent reviewers (Ho and Chung) were responsible for study selection. Disagreements and discrepancies were resolved by consensus between reviewers or by arbitration by a third reviewer (Wong).

2.4. Data extraction

Two independent reviewers (Ho and Chung) extracted information on the characteristics of included TCM diagnostic instruments, including their year of development, origin and source language, construct(s), target population, purpose(s), mode of administration, recall period, number of items, and response options. We also extracted relevant information on the characteristics of the study populations, including their sample size, average age, gender, disease(s) in concern, average disease duration, disease severity, source of participants, geographical location of instrument administration, and response rate.

2.5. Data synthesis

Due to heterogeneity across included studies, the development process and measurement properties of each included TCM diagnostic instrument were summarised narratively.[25]

2.5.1. Appraisal of instrument development process
A well-developed TCM diagnostic instrument should have its items be relevant for the construct of interest within a specific population and context of use (relevance to the TCM patterns), be clearly understood by that specific population (comprehensibility among patients), and be covering all key aspects of the construct (comprehensiveness of aspects related to the TCM patterns).[26]

The quality of the development process of each included instrument was evaluated by two independent reviewers (Ho and Chung) via a three-step procedure [Appendix 2] guided by a 37-item questionnaire customised for the evaluation of TCM diagnostic instrument development [Appendix 3]. Designed based on the COSMIN items for evaluating the development quality of patient-reported outcome instruments,[26] our newly adapted questionnaire includes key items regarding (i) cognitive interview study or pilot test with patients for diagnostic instrument comprehensibility; and (ii) cognitive interview study with TCM experts for diagnostic instrument comprehensiveness. Rating discrepancies were resolved by consensus through discussions or by arbitration under a third reviewer (Wong).

2.5.2. Appraisal of measurement properties
The degree of reliability and validity of a TCM diagnostic instrument is reflected by the quality of measurement properties.[25] Reliability is defined as the degree to which measurement is free from measurement error. This encompasses the measurement properties of internal consistency, test-retest reliability, and measurement error.[25] Validity is defined as the degree to which a measurement instrument measures the constructs it purports to measure. This consists of the measurement properties of content validity, construct validity, structural validity, cross-cultural validity, and criterion validity.[25]
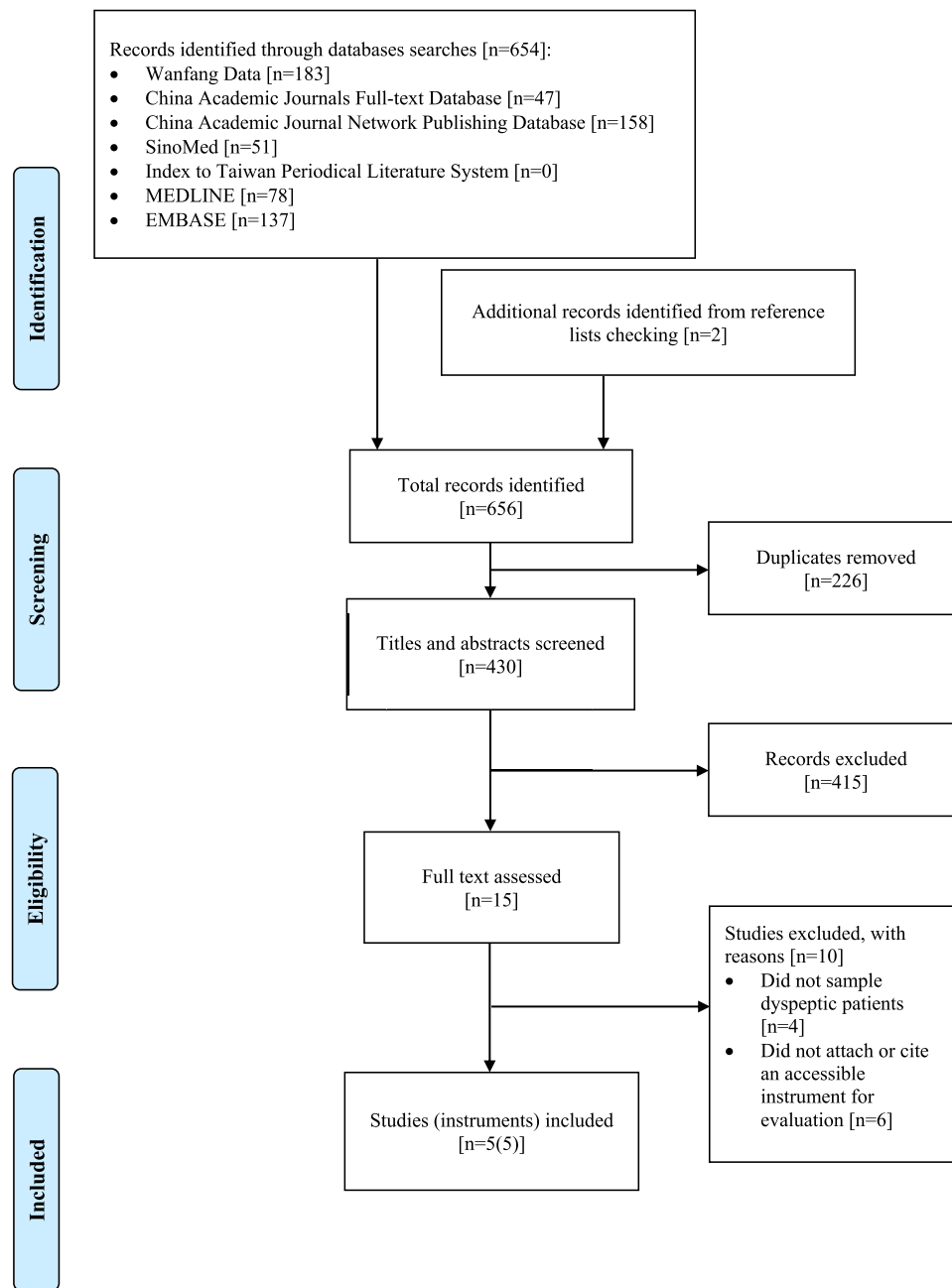
Identification

Records identified through databases searches [n=654]:
- Wanfang Data [n=183]
- China Academic Journals Full-text Database [n=47]
- China Academic Journal Network Publishing Database [n=158]
- SinoMed [n=51]
- Index to Taiwan Periodical Literature System [n=0]
- MEDLINE [n=78]
- EMBASE [n=137]

Additional records identified from reference lists checking [n=2]

Screening

Total records identified [n=656]

Duplicates removed [n=226]

Titles and abstracts screened [n=430]

Records excluded [n=415]

Eligibility

Full text assessed [n=15]

Studies excluded, with reasons [n=10]
- Did not sample dyspeptic patients [n=4]
- Did not attach or cite an accessible instrument for evaluation [n=6]

Included

Studies (instruments) included [n=5(5)]

Fig. 1. Flow of literature search.

The quality of the measurement properties of each included TCM diagnostic instrument was rated against relevant COSMIN criteria.[25, 26] Each measurement property was judged to be either of sufficient quality or insufficient quality.[25, 26] For instance, the test-retest reliability of an instrument would be of sufficient quality if its intraclass correlation coefficient or Cohen's kappa coefficient was shown to be ≥ 0.70 in the validation study.[25, 26] Assessments were performed independently by two reviewers (Ho and Chung) in duplicate. Rating discrepancies were resolved by consensus through discussions or by arbitration under a third reviewer (Wong).

2.6. Risk of bias assessment

Risk of bias assessment was performed by two reviewers (Ho and Chung) independently following the procedures recommended by the COSMIN Initiative.[25, 26] The risk of bias of each included validation study was rated against criteria listed in the *COSMIN Risk of Bias Checklist* on a four-point scale: no risk of bias, serious risk of bias, very serious risk of bias, or extremely serious risk of bias.[30] Rating discrepancies were resolved by consensus through discussions or by arbitration under a third reviewer (Wong).

3. Results

3.1. Study selection

The search yielded 654 citations [Fig. 1]. Two additional references were identified through manual searches on reference lists. After deduplication, 430 studies proceeded to the screening process. After title and abstract screening, 415 studies were excluded and only fifteen studies were qualified for full text assessment.

Table 1
Characteristics of included Traditional Chinese Medicine diagnostic instruments.

Instrument	Year of development	Origin (source language)	Construct(s)	Target population	Purpose(s)	Mode of administration	Recall period	Number of items	Response options
Traditional Chinese Medicine Functional Gastrointestinal Disorders Scale[31]	2010	China (Mandarin)	TCM signs and symptoms of FGIDs	Patients with FGIDs	Diagnose TCM patterns	Self-administered	2 weeks	44	Never[1]; Not that much[2]; Moderate[3]; Frequently[4]; All the time[5]
Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders[32]	2011	China (Mandarin)	TCM signs and symptoms of FGIDs	Patients with FGIDs	Diagnose TCM patterns	Self-administered	2 weeks	54	Five-point Likert scale: 1–5 (1=Never, 5=All the time)
Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders (Amended)[35]	2015	Hong Kong (Cantonese)	TCM signs and symptoms of FD	Patients with FD	Diagnose TCM patterns	Phone-based interview administered by trained research assistants	2 weeks	24	Five-point Likert scale: 1–5 (1=Never, 5=All the time)
Traditional Chinese Medicine Functional Dyspepsia Scale[33]	2015	China (Mandarin)	TCM signs and symptoms of FD	Patients with FD	Diagnose TCM patterns	Interview-based with physical assessment administered by TCM practitioners	2 weeks	33 (24 answered by patients; 9 answered by practitioners)	Never[1]; Not that much[2]; Moderate[3]; Frequently[4]; All the time[5]
Stomach Qi Deficiency Pattern Assessment Scale[34]	2018	South Korea (Korean)	TCM signs and symptoms of SQD	Patients with dyspepsia	Diagnose TCM patterns	Interview-based with physical assessment administered by TCM practitioners	Not reported	12 (9 by patients; 3 by TCM practitioners)	Never[1]; Not that much[2]; Moderate[3]; Frequently[4]; All the time[5]

TCM: Traditional Chinese Medicine; FGIDs: Functional gastrointestinal diseases; FD: Functional dyspepsia; SQD: Stomach qi deficiency

Eventually, five validation studies completely satisfied the inclusion criteria and were included in this review.

3.2. Characteristics of instruments

Five TCM diagnostic instruments for FD were evaluated in these five validation studies published between 2010 and 2018 [Table 1]. Three of the five diagnostic instruments were developed in China and having Mandarin as their source language.[31-33] The remaining two were developed in South Korea and Hong Kong and having Korean and Cantonese, respectively, as their source language.[34, 35] Three of the five diagnostic instruments were developed for diagnosing TCM patterns among patients with FD or dyspepsia.[33-35] The other two were made for serving a similar purpose among patients with functional gastrointestinal disorders.[31, 32] The mode of administration among the five varied. Two were self-administered questionnaires with about fifty items,[31, 32] two required a face-to-face interview with a physical assessment performed by TCM practitioners,[33, 34] and one administered via phone-based interviews by trained personnel.[35]

3.3. Quality of instrument development process

All five included TCM diagnostic instruments performed well in describing clear constructs and their origins, explaining the target population, and clarifying the context of use. However, due to the lack of concept elicitation process involving TCM experts, all five instruments were judged to have an inadequate design quality [Table 2]. Only the development of *Traditional Chinese Medicine Functional Gastrointestinal Disorders Scale* included cognitive interview studies with a representative sample of patients and TCM experts.[31] Apart from the described limitations, comprehensibility and comprehensiveness of the included instruments were not satisfactory either. As the lowest rating across the 37 items decides the overall rating,[26] all included TCM diagnostic instruments were

graded as "inadequate" in terms of the quality of their development process.

3.4. Quality of measurement properties

3.4.1. Traditional chinese medicine functional gastrointestinal disorders scale

Traditional Chinese Medicine Functional Gastrointestinal Disorders Scale (FGIDS) was validated among 100 healthy subjects and 198 patients with FD or irritable bowel syndrome (IBS) [Table 3].[31] The validation study took place in hospitals and specialist outpatient clinics in Guangzhou, Shenzhen, and Maoming, China. Its structural validity was assessed by examining whether the proposed nine-factor model could be jusitified by the data obtained using the diagnostic instrument [Table 4]. The hypothesised factors proposed four possible TCM patterns for FD (stagnation of liver qi, spleen-stomach qi deficiency, liver qi invading the stomach, and stomach dampness-heat); and five possible TCM patterns for IBS (liver qi depression, liver qi invading the spleen, spleen-stomach weakness, cold-heat complex, and large intestinal dryness accumulation).[31] Only one TCM pattern for FD and four TCM patterns for IBS could be revealed from the data, with an unsatisfactory comparative fit index (0.87) and root mean square error of approximation (0.064) [Table 5]. The structural validity of FGIDS was therefore considered to be of insufficient quality.

3.4.2. Traditional chinese medicine syndrome differentiation instrument for functional gastrointestinal disorders

Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders (FGDDI) was validated among 500 subjects, including healthy participants and patients with functional gastrointestinal disorders.[32] The setting was in specialist outpatient clinics in Guangzhou, China [Table 3]. Its structural validity was assessed by investigating whether the proposed four-factor model could be justified by the data obtained using the

Table 2

Quality of development process among included traditional Chinese medicine diagnostic instruments.

	FGIDS[31]	FGDDI[32]	Amended-FGDDI[35]	CMFDS[33]	SSQD[34]
Diagnostic instrument design					
Clear construct (item 1 in Appendix 3)	Very good	Very good	Very good	Very good	Very good
Clear origin of construct (item 2 in Appendix 3)	Very good	Very good	Very good	Very good	Very good
Clear target population for which the instrument was developed (item 3 in Appendix 3)	Very good	Very good	Very good	Very good	Very good
Clear context of use(item 4 in Appendix 3)	Very good	Very good	Very good	Very good	Very good
Instrument development involved a sample from TCM expert community (item 5 in Appendix 3)	Inadequate	Inadequate	Inadequate	Inadequate	Inadequate
Concept elicitation (item 6–13 in Appendix 3)	*Not applicable*	*Not applicable*	*Not applicable*	*Not applicable*	*Not applicable*
Overall quality of diagnostic instrument design	**Inadequate***	**Inadequate***	**Inadequate***	**Inadequate***	**Inadequate***
Cognitive interview study or pilot test (patients)					
Study or test performed among the target population (item 14 in Appendix 3)	Very good	Inadequate	Inadequate	Inadequate	Inadequate
Study or test performed in a sample representing the target population (item 15 in Appendix 3)	Doubtful	*Not applicable*	*Not applicable*	*Not applicable*	*Not applicable*
Comprehensibility assessment (item 16–25 in Appendix 3)	Doubtful#	*Not applicable*	*Not applicable*	*Not applicable*	*Not applicable*
Overall quality of cognitive interview study or pilot test (patients)	**Doubtful***	**Inadequate***	**Inadequate***	**Inadequate***	**Inadequate***
Cognitive interview study (TCM experts)					
Study performed among the TCM expert community (item 26 in Appendix 3)	Very good	Inadequate	Inadequate	Inadequate	Inadequate
Study performed in a sample representing the TCM expert community (item 27 in Appendix 3)	Very good	*Not applicable*	*Not applicable*	*Not applicable*	*Not applicable*
Comprehensiveness assessment (item 28–37 in Appendix 3)	Doubtful#	*Not applicable*	*Not applicable*	*Not applicable*	*Not applicable*
Overall quality of cognitive interview (TCM experts)	**Doubtful***	**Inadequate***	**Inadequate***	**Inadequate***	**Inadequate***
Overall quality of instrument development process	**Inadequate@**	**Inadequate@**	**Inadequate@**	**Inadequate@**	**Inadequate@**

FGIDS: Traditional Chinese Medicine Functional Gastrointestinal Disorders Scale; FGDDI: Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders; Amended-FGDDI: Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders (Amended); CMFDS: Traditional Chinese Medicine Functional Dyspepsia Scale; SSQD: Stomach Qi Deficiency Pattern Assessment Scale; TCM: Traditional Chinese Medicine

* The lowest rating across items 1–13, 14–25, and 26–37 in Appendix 3 was taken as the overall qualities of diagnostic instrument design, cognitive interview study or pilot test (patients), and cognitive interview (TCM experts), respectively.

The lowest ratings across items 16–25 and 28–37 in Appendix 3 were taken as the ratings of comprehensibility assessment and comprehensiveness assessment, respectively.

@ The lowest rating across items 1–37 in Appendix 3 was taken as the rating of the overall quality of instrument development process.

instrument [Table 4]. The four factors corresponded to the four TCM patterns of FD, including liver qi depression, spleen deficiency with dampness encumbrance, spleen-stomach qi deficiency, and qi stagnation.[32] As the root mean square error of approximation of the model was smaller than 0.06, the structural validity of FGDDI was judged to be of sufficient quality [Table 5].

3.4.3. Traditional chinese medicine syndrome differentiation instrument for functional gastrointestinal disorders (Amended)

Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders (Amended) (Amended-FGDDI) was validated among 270 FD patients recruited from specialist outpatient clinics in Hong Kong [Table 3].[35] Its structural validity was evaluated via exploring replicability of the TCM patterns contained in FGDDI using data collected in Hong Kong [Table 4]. The structural validity of Amended-FGDDI was considered to be of insufficient quality as only exploratory factor analysis were conducted [Table 5].

3.4.4. Traditional chinese medicine functional dyspepsia scale

Traditional Chinese Medicine Functional Dyspepsia Scale (CMFDS) was validated among ninety-one FD patients recruited from

specialist outpatient clinics in Guangzhou and Shanghai, China [Table 3].[33] CMFDS proposed a five-factor model, which corresponded to TCM patterns of spleen-stomach qi deficiency, liver qi depression, heat, stagnation of phlegm-dampness, and qi stagnation.[33] Its test-retest reliability and criterion validity were evaluated [Table 4]. Since Cohen's kappa of at least 0.70 was not achieved by all factors, the test-retest reliability of CMFDS was rated to be of insufficient quality [Table 5]. The criterion validity was of indeterminate quality [Table 5], as the factors' correlation with the gold standard (diagnosis made by TCM practitioners) or area under the receiver operator curve results were not reported.

3.4.5. Stomach qi deficiency pattern assessment scale

Stomach Qi Deficiency Pattern Assessment Scale (SSQD) was validated among thirty healthy subjects and thirty patients with dyspepsia in hospitals located in Seoul, South Korea [Table 3].[34] Its internal consistency and criterion validity were evaluated [Table 4]. Due to a lack of sufficient structural validity, its internal consistency was judged as indeterminate [Table 5]. Nevertheless, with an area under the receiver operator curve value of 0.916, its criterion validity was rated as of sufficient quality [Table 5].

Table 3
Characteristics of populations sampled among included validation studies.

Instrument	Sample N	Average age (SD, range)	Gender % female	Disease characteristics			Source of participants	Location	Response rate
				Disease(s)	Average disease duration	Disease severity			
Traditional Chinese Medicine Functional Gastrointestinal Disorders Scale[31]	300 (Healthy: 100 Patients: 200)	15–25 (31.9%) 26–45 (46.6%) 46–60 (13.8%) ≥60 (7.7%)	47%	50% FD 50% IBS	Not mentioned	Not mentioned	Hospital & Specialist outpatient clinic	Guangzhou, Shenzhen, and Maoming, China	99.3%
Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders[32]	500 (Healthy subjects & patients)	15–25 (23.6%) 26–35 (36%) 36–45 (27%) 46–55 (8.4%) 56–70 (5%)	49.5%	FGIDs	Not mentioned	Not mentioned	Specialist outpatient clinic	Guangzhou, China	72.4%
Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders (Amended)[35]	270	50.6 (12.8, 18–80)	78.6%	FD	Not mentioned	Not mentioned	Specialist outpatient clinic	Hong Kong	83.7%
Traditional Chinese Medicine Functional Dyspepsia Scale[33]	91	40.3 (11.95)	46.1%	FD	Not mentioned	Not mentioned	Specialist outpatient clinic	Guangzhou and Shanghai, China	Not mentioned
Stomach Qi Deficiency Pattern Assessment Scale[34]	60 (Healthy: 30 SQD: 30)	Healthy: 37.5 (11.1, 19–74) SQD: 40.47 (10.3, 19–74)	Healthy 63% SQD: 40%	Dyspepsia	Not mentioned	Not mentioned	Hospital	Seoul, South Korea	Not mentioned

FD: Functional dyspepsia; IBS: Irritable bowel syndrome; FGIDs: Functional gastrointestinal diseases; SQD: Stomach qi deficiency

Table 4
Validation of measurement properties among included Traditional Chinese Medicine diagnostic instruments.

	FGIDS[31]	FGDDI[32]	Amended-FGDDI[35]	CMFDS[33]	SSQD[34]
Content validity	Not validated	Not validated	Not validated	Not validated	Not validated
Structural validity	Validated	Validated	Validated	Not validated	Not validated
Internal consistency	Not validated	Not validated	Not validated	Not validated	Validated
Cross-cultural validity	Not validated	Not validated	Not validated	Not validated	Not validated
Measurement invariance	Not validated	Not validated	Not validated	Not validated	Not validated
Test-retest reliability	Not validated	Not validated	Not validated	Validated	Not validated
Measurement error	Not validated	Not validated	Not validated	Not validated	Not validated
Criterion validity	Not validated	Not validated	Not validated	Validated	Validated
Construct validity	Not validated	Not validated	Not validated	Not validated	Not validated
Responsiveness	Not validated	Not validated	Not validated	Not validated	Not validated

FGIDS: Traditional Chinese Medicine Functional Gastrointestinal Disorders Scale; FGDDI: Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders; Amended-FGDDI: Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders (Amended); CMFDS: Traditional Chinese Medicine Functional Dyspepsia Scale; SSQD: Stomach Qi Deficiency Pattern Assessment Scale

3.5. Risk of bias assessment

Results of the risk of bias assessment are illustrated in Table 6. Criterion validity studies of CMFDS and SSDQ were at no risk of bias. The structural validity study of Amended-FGDDI and the internal consistency study of SSDQ were at no risk of bias as well. However, structural validity studies of FGIDS and FGDDI, as well as the test-retest reliability study of CMFDS were at serious risk of bias.

4. Discussion

4.1. Summary of main results

A total of five unique TCM diagnostic instruments were included, of which their validation studies evaluated seven measurement properties conducted among five populations. Among these results, the structural validity of FGDDI and the criterion validity of SSDQ were of sufficient quality. Only the validation study of SSDQ was at no risk of bias. Apart from the inadequate quality of measurement properties, the development process of these instruments failed to demonstrate sufficient quality. Therefore, according to the COSMIN guidelines, none of them could be recommended as a reliable and valid instrument for facilitating TCM pattern diagnosis of FD.

4.2. Implications for practice

The quality of measurement properties is crucial for ensuring reliability and validity of measurement instruments, including TCM diagnostic instruments. Three measurement properties are commonly validated by the developers of TCM diagnostic instruments: (i) content validity – it refers to whether the instrument covers a broad scope of accepted and relevant TCM diagnostic concepts, contains comprehensible diagnostic items, and demonstrates a high consistency between its diagnostic concepts and diagnostic questionnaire items;[19, 36-38] (ii) structural validity – it assesses the strength of statistical relationships between TCM patterns, and the relevant diagnostic questionnaire items;[39-41] and (iii) test-retest reliability – it indicates the consistency of diagnostic results when

Table 5
Summary of measurement property validations.

Instrument	Structural validity Result [Quality of measurement property]*	Internal consistency Result [Quality of measurement property]#	Test-retest reliability Result [Quality of measurement property]@	Criterion validity Result [Quality of measurement property]+

Traditional Chinese Medicine Functional Gastrointestinal Disorders Scale[31]	• Total CFI and RMSEA: 0.87; 0.064
	• CFI and RMSEA of each identified factor (TCM pattern):
	  ○ Stomach dampness-heat (pattern of FD) – 0.52; 0.15
	  ○ Liver qi depression (pattern of IBS) – 0.72; not reported
	  ○ Liver qi invading the spleen (pattern of IBS) – 0.69; 0.17
	  ○ Spleen-stomach weakness (pattern of IBS) – 0.56; 0.16
	  ○ Cold-heat complex (pattern of IBS) – 0.81; 0.20
	• [Insufficient]
		Not validated	*Not validated*	*Not validated*

Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders[32]	• Total CFI and RMSEA: 0.83; 0.052 • [Sufficient]	*Not validated*	*Not validated*	*Not validated*

Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders (Amended)[35]	• KMO: 0.777
	• Bartlett's sphericity test: p<0.001
	• EFA:
	  ○ 8 factors with eigenvalue >1, contributed to 59.4% of total variance
	  ○ Factors were not related to TCM patterns
	• [Insufficient]
		Not validated	*Not validated*	*Not validated*

Traditional Chinese Medicine Functional Dyspepsia Scale[33]	*Not validated*	*Not validated*	• Cohen's kappa of each identified factor (TCM pattern):
			  ○ Spleen-stomach qi deficiency – 0.65
			  ○ Liver qi depression– 0.82
			  ○ Heat – 0.74
			  ○ Stagnation of phlegm-dampness – 0.76
			  ○ Qi stagnation – 0.68
			• [Insufficient]
				• Gold standard: Diagnosis made by Chinese medicine experts
				• Sensitivity and specificity of each factor (TCM pattern):
				  ○ Spleen-stomach qi deficiency – 0.73; 0.95
				  ○ Liver qi depression – 0.88; 0.94
				  ○ Heat – 0.90; 0.90
				  ○ Stagnation of phlegm-dampness – 0.90; 0.89
				  ○ Qi stagnation – 0.76; 0.93
				• [Indeterminate]

Stomach Qi Deficiency Pattern Assessment Scale[34]	*Not validated*	• Cronbach's alpha: 0.87 • [Indeterminate]	*Not validated*	• Gold standard: Diagnosis made by Korean medical doctors • Sensitivity: 93.33% • Specificity: 86.67% • AUC: 0.916 • [Sufficient]

CFI: Comparative fit index; TCM: Traditional Chinese Medicine; FD: Functional dyspepsia; IBS: Irritable bowel syndrome; RMSEA: Root mean square error of approximation; KMO: Kaiser-Meyer-Olkin Test; EFA: Exploratory factor analysis; AUC: Area under the Receiver Operator Curve

* Sufficient: CFI >0.95 OR RMSEA <0.06; indeterminate: not all information for "sufficient" reported; insufficient: criteria for "sufficient" not met

Sufficient: Sufficient structural validity AND Cronbach's alpha(s) ≥0.70 for each unidimensional scale or subscale; indeterminate: criteria for "sufficient structural validity" not met; insufficient: sufficient structural validity AND Cronbach's alpha(s) <0.70 for each unidimensional scale or subscale

@ Sufficient: Cohen's kappa ≥0.70; indeterminate: Cohen's kappa not reported; insufficient: Cohen's kappa <0.70

+ Sufficient: Correlation with gold standard ≥0.70 OR AUC ≥0.70; indeterminate: not all information for "sufficient" reported; insufficient: correlation with gold standard <0.70 OR AUC <0.70

Table 6

. Risk of bias among measurement property validation studies.

Instrument	Structural validity	Internal consistency	Test-retest reliability	Criterion validity
Traditional Chinese Medicine Functional Gastrointestinal Disorders Scale[31]	**Serious risk of bias**	*Not validated*	*Not validated*	*Not validated*
Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders[32]	Serious risk of bias	*Not validated*	*Not validated*	*Not validated*
Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorders (Amended)[35]	**No risk of bias**	*Not validated*	*Not validated*	*Not validated*
Traditional Chinese Medicine Functional Dyspepsia Scale[33]	*Not validated*	*Not validated*	**Serious risk of bias**	No risk of bias
Stomach Qi Deficiency Pattern Assessment Scale[34]	*Not validated*	No risk of bias	*Not validated*	No risk of bias

the instrument is administered twice on the same patient over a while, assuming that the condition would stay stable during the period.[38, 42-44] Yet, despite the significance of these measurement properties to pattern diagnosis, they were not rigorously validated among the existing TCM diagnostic instruments for FD.

4.3. Implications for research

Future researchers should also conduct critical appraisals on the relevance, comprehensibility, and comprehensiveness of TCM diagnostic instruments with the amended COSMIN items proposed by this study, of which the operation procedure is reported in Appendix 2. Assessment criteria for measurement properties in the COSMIN guidelines should also be amended to accommodate the appraisal of TCM diagnostic instruments. Criterion validity demonstrates the degree of which assessment results from a TCM diagnostic instrument is an adequate reflection of result from the "gold standard" test.[25] Nevertheless, considering the absence of a gold standard test in TCM pattern diagnosis,[45-48] as well as the poor inter-rater agreement among TCM practitioners in pattern diagnosis,[49] it is doubtful whether diagnoses made by TCM practitioners should serve as the gold standard for criterion validity assessments of TCM diagnostic instruments. Latent class analysis is commonly performed nowadays to explore TCM patterns, with examples from studies conducted among patients with systemic lupus erythematosus,[50] cardiovascular disease,[51] depression,[52] persistent insomnia,[53] and vascular mild cognitive impairment.[48] Since the construct of TCM patterns revealed by latent class analysis may not be a linear combination of relevant TCM signs and symptoms,[35, 48, 54] factor analysis may not be suitable for evaluating the structural validity of those diagnostic instruments. As TCM pattern is a summary of pathological changes in a patient at a specific time point and is constantly changing,[1] such dynamic feature may not favor standard evaluation of test-retest reliability which requires patients to be stable for at least two weeks.[25, 43] To understand the dynamic feature of TCM pattern, prognostic tests may be conducted to study the possible changes of pattern diagnosis along the course of the disease. For example, empirical evidence illustrated that the progression from impaired glucose tolerance to diabetes mellitus resembles the progression from spleen deficiency with dampness encumbrance to spleen-stomach dampness-heat, then eventually to yin and qi deficiency.[55]

Recently, many TCM diagnostic instruments have been developed to enable an systematic and reproducible measurement of TCM diagnostic patterns.[17–20, 38, 40–42, 44, 56–58] Unlike the disease specific instruments included in this review, some of those instruments measure specific TCM diagnostic construct directly without considering the disease entity *per se*. The *Traditional Chinese Medical Diagnostic Descriptor* was developed under this approach, with a goal of measuring TCM patterns regardless of the disease defined under Chinese or conventional medicine.[57] Another example

is the *Blood Stasis Questionnaire*, which measures the presence of blood stasis generically.[58] Comparative diagnostic research should be conducted to compare how different measurement approaches would influence prescription decision, and subsequently, patient outcomes. Last but not least, multisite cross-sectional studies involving patients from countries could be conducted to explore potential geographic variations on pattern diagnosis between populations.[35]

4.4. Strengths and limitations

One of the strengths of this study is that searches were performed on seven Chinese and English electronic databases to ensure the comprehensiveness of the literature search. Second, the sensitivity of the literature search was strengthened by the use of validated search filters for measurement instruments. Third, this systematic review was formulated under the COSMIN guidelines with novel amendments made to accommodate the assessment of the development process of TCM diagnostic instruments.

One of the main limitations is that criteria for appraising measurement properties proposed by COSMIN guidelines may not be entirely applicable to TCM diagnostic instruments, as the COSMIN guidelines were mainly developed for the evaluation of patient-reported outcome instruments. That said, systematic reviews on measurement properties of instruments other than patient-reported outcome have been conducted using the COSMIN guidelines.[59-61] Further amendments on the guidelines should be made for appraising TCM diagnostic instruments.

4.5. Conclusions

This systematic review revealed that the quality of existing TCM diagnostic instruments for FD is unsatisfactory. We were not able to identify any reliable and valid instruments for TCM pattern diagnosis of FD. Future researchers are recommended to develop and evaluate TCM diagnostic instruments following the guidance of the COSMIN Initiative, with relevant amendments to accommodate the features and uniqueness of TCM diagnostic process. Such research initiative will establish the evidence base of TCM pattern diagnosis, supporting further standardization of diagnostic practice under the auspice of ICD-11.

Author contributions

Conflict of interest

Vincent CH Chung is an Editorial Board Member of Integrative Medicine Research. Other authors declare no conflicts of interest.

Funding

This research received no grant from any funding agency.

Ethical statement

This research did not require an ethical approval as it does not involve any human or animal experiment.

Data availability

There is no available data as this work used available literature.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.imr.2020.100713.

References

1. World Health Organization Regional Office for the Western Pacific. *WHO International Standard Terminologies on Traditional Medicine in the Western Pacific Region*. Manila: World Health Organization Regional Office for the Western Pacific; 2007 [cited 2020 May 18]Available from *https://apps.who.int/iris/handle/10665/206952* .

2. World Health Organization. Declaration of Alma-Ata. Geneva: World Health Organization; 1978 [cited 2020 May 18]. Available from: https://www.who.int/publications/almaata_declaration_en.pdf?ua=1.

3. Morris W, Gomes S, Allen M. International classification of traditional medicine. *Glob Adv Health Med*. 2012;1(4):38–41.

4. Lindmeier C. *WHO Releases New International Classification of Diseases (ICD 11)*. Geneva: World Health Organization; 2018 [cited 2020 May 18]. Available from *https://www.who.int/news-room/detail/18-06-2018-who-releases-new-international-classification-of-diseases-(icd-111* .

5. World Health Organization. *ICD-11 For Mortality and Morbidity Statistics (Version: 09/2020)*. Geneva: World Health Organization; 2020 [cited 2020 Sep 18]. Available from *https://icd.who.int/browse11/l-m/en* .

6. . *World Health Organization. International Classification of Traditional Medicine: ICTM*. Geneva: World Health Organization; 2006 [cited 2020 May 18]. Available from *https://apps.who.int/classifications/apps/icd/meetings/2006meeting/WHOFIC2006%20-%20C408%20-%20ICTM%20proposal.pdf* .

7. Xu W, Zhu L, Zu L, Yan S, Zhang H, Dou D. Formulation and consideration of world health organization international classification of traditional medicine. *J Tradit Chin Med*. 2020;40(1):157–161.

8. Schnyer RN, Citkovitz C. Inter-rater reliability in traditional chinese medicine: challenging paradigmatic assumptions. *J Altern Complement Med*. 2019;25(11):1067–1073.

9. Eigenschink M, Dearing L, Dablander TE, Maier J, Sitte HH. A critical examination of the main premises of traditional Chinese medicine. *Wien Klin Wochenschr*. 2020;132(9):260–273.

10. Zhang X, Tian R, Zhao C, Birch S, Lee JA, Alraek T, et al. The use of pattern differentiation in WHO-registered traditional Chinese medicine trials – A systematic review. *Eur J Integr Med*. 2019;30.

11. Lam WC, Lyu A, Bian Z. ICD-11: impact on traditional chinese medicine and world healthcare systems. *Pharmaceut Med*. 2019;33(5):373–377.

12. World Health Organization. Regional Office for the Western Pacific. *The Regional Strategy for Traditional Medicine in the Western Pacific (2011-2020)*. Manila: WHO Regional Office for the Western Pacific; 2012.

13. World Health Organization. *Regional Office for the Western Pacific. The Regional Strategy for Traditional Medicine in the Western Pacific 2002*. Manila: WHO Regional Office for the Western Pacific; 2002.

14. Chung VCH, Ho RST, Wu X, Wu JCY. Incorporating traditional Chinese medicine syndrome differentiation in randomized trials: Methodological issues. *Eur J Integr Med*. 2016;8(6):898–904.

15. Birch S, Alraek T. Traditional East Asian medicine: how to understand and approach diagnostic findings and patterns in a modern scientific framework? *Chin J Integr Med*. 2014;20(5):336–340.

16. Jiang M, Lu C, Zhang C, Yang J, Tan Y, Lu A, et al. Syndrome differentiation in modern research of traditional Chinese medicine. *J Ethnopharmacol*. 2012;140(3):634–642.

17. Zheng S, Kim C, Meier P, Sibbritt D, Zaslawski C. Development of a novel questionnaire for the traditional Chinese medicine pattern diagnosis of stress. *J Acupunct Meridian Stud*. 2017;10(4):276–285.

18. Gadau M, Zhang SP, Yeung WF, Bian ZX, Lu AP. TCM pattern questionnaire for lateral elbow pain: development of an instrument via a delphi process. *Evid Based Complement Alternat Med*. 2016;2016.

19. Chen XL, Wen Y, Wu ZC, Zhang BP, Hou ZK, Xiao JL, et al. Development of a traditional chinese medicine syndrome-specific scale for ulcerative colitis: the large intestine dampness-heat syndrome questionnaire. *Evid Based Complement Alternat Med*. 2018 2018.

20. Fang G, Zhang LL, Ren Q, Zhou XW, Wang B, Zhou X, et al. Development of a diagnostic questionnaire for damp phlegm pattern and blood stasis pattern in coronary heart disease patients (CHD-DPBSPQ). *Evid Based Complement Alternat Med*. 2019;2019.

21. Ho RST, Chung VCH, Wong CHL, Wu JCY, Wong SYS, Wu IXY. Acupuncture and related therapies used as add-on or alternative to prokinetics for functional dyspepsia: overview of systematic reviews and network meta-analysis. *Sci Rep*. 2017;7(1):10320.

22. Chu MHK, Wu IXY, Ho RST, Wong CHL, Zhang AL, Zhang Y, et al. Chinese herbal medicine for functional dyspepsia: systematic review of systematic reviews. *Therap Adv Gastroenterol*. 2018;11.

23. Chung VC, Wong CH, Wu IX, Ching JY, Cheung WK, Yip BH, et al. Electroacupuncture plus on-demand gastrocaine for refractory functional dyspepsia: Pragmatic randomized trial. *J Gastroenterol Hepatol*. 2019;34(12):2077–2085.

24. Bensoussan A, Talley NJ, Hing M, Menzies R, Guo A, Ngu M. Treatment of irritable bowel syndrome with Chinese herbal medicine: a randomized controlled trial. *JAMA*. 1998;280(18):1585–1589.

25. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–1157.

26. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a delphi study. *Qual Life Res*. 2018;27(5):1159–1170.

27. World Health Organization. *Regional Office for the Western Pacific. WHO International Standard Terminologies on Traditional Medicine in the Western Pacific Region: Manila: WHO Regional Office for the Western Pacific*; 2007.

28. Alberta University Canada. *Translations for Other Databases - MEDLINE*. Amsterdam: COSMIN Initiative; 2020 [cited 2020 July 18]. Available from *https://www.cosmin.nl/tools/pubmed-search-filters/* .

29. Jansma E. *Translations for Other Databases - EMBASE*. Amsterdam: COSMIN Initiative; 2020 [cited 2020 July 18]. Available from *https://www.cosmin.nl/tools/pubmed-search-filters/* .

30. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171–1179.

31. Zhuang K. *Development and Evaluation of Traditional Chinese Medicine Functional Gastrointestinal Disorders Scale (Chinese) [Thesis]*. Guangzhou, China: Guangzhou University of Chinese Medicine;; 2010.

32. Lin C. *Analysis of Traditional Chinese Medicine Syndrome Differentiation Instrument for Functional Gastrointestinal Disorder Based on Syndrome Differentiation Using Structural Equation Model and Item Response Theory Methods* (Chinese) *[Dissertation]*. Guangzhou, China: Guangzhou University of Chinese Medicine; 2011.

33. Cao Y. *Analysis of Traditional Chinese Medicine Functional Dyspepsia Scale Based on Syndrome Differentiation Using Multidimensional Item Response Theory Method (Chinese) [Dissertation]*. Guangzhou, China: Guangzhou University of Chinese Medicine; 2015.

34. Lee JH, Park JW, Ko SJ, Kim JS. Development and validation of a new pattern identification scale for stomach Qi deficiency. *Eur J Integr Med*. 2018;17:56–63.

35. Shuldiner SR, Chung VCH, Wu X, Ching J, Ho RST, Cheong PK, et al. Methodological challenges in mapping chinese medicine syndrome with conventional diagnosis: Implications for multi-centre trials in integrative medicine. *Eur J Integr Med*. 2015;7(4):358–364.

36. Schnyer RN, Conboy LA, Jacobson E, McKnight P, Goddard T, Moscatelli F, et al. Development of a Chinese medicine assessment measure: an interdisciplinary approach using the delphi method. *J Altern Complement Med*. 2005;11(6):1005–1013.

37. Wong W, Lam CLK, Su Y-C, Lin SJ-S, Ziea ET-C, Wong VT, et al. Measuring body constitution: validation of the body constitution questionnaire (BCQ) in Hong Kong. *Complement Ther Med*. 2014;22(4):670–682.

38. Wong YC, Pang MC. Initial validation of the Yin-Yang assessment questionnaire for persons with diabetes mellitus. *World J Diabetes*. 2015;6(11):1198–1206.

39. Shen M, Cui Y, Hu M, Xu L. Quantifying traditional Chinese medicine patterns using modern test theory: an example of functional constipation. *BMC Complement Altern Med*. 2017;17(1):44.

40. Wu D, Huang C, Mo X, Liu J, Cai J, Liu C, et al. Development and initial validation of a Traditional Chinese Medicine symptom-specific outcome measure: a Zheng-related atopic dermatitis symptom questionnaire (ZRADSQ). *Health Qual Life Outcomes*. 2013;11:212.

41. Fu TC, Lin YC, Chang CM, Chou WL, Yuan PH, Liu MH, et al. Validation of a new simple scale to measure symptoms in heart failure from traditional Chinese medicine view: a cross-sectional questionnaire study. *BMC Complement Altern Med*. 2016;16(1):342.

42. Park YJ, Cho SW, Lee BH, Park YB. Development and validation of the Yin deficiency scale. *J Altern Complement Med*. 2013;19(1):50–56.

43. Huang Z, Hou Z, Liu X, Liu F, Wu Y. Quantifying liver stagnation spleen deficiency pattern for diarrhea predominate irritable bowel syndromes using multidimensional analysis methods. *Evid Based Complement Alternat Med*. 2018;2018.

44. Lee S, Park J, Lee H, Kim K. Development and validation of Yin-deficiency questionnaire. *Am J Chin Med*. 2007;35(1):11–20.
45. Wang XN, Zhou V, Liu Q, Gao Y, Zhou X-H. Evaluation of the accuracy of diagnostic scales for a syndrome in Chinese medicine in the absence of a gold standard. *Chin Med*. 2016;11(1):35.
46. Yeung WF, Chung KF, Ng KY, Yu YM, Zhang SP, Ng BF, et al. Prescription of chinese herbal medicine in pattern-based traditional Chinese medicine treatment for depression: a systematic review. *Evid Based Complement Alternat Med*. 2015;2015.
47. Li SM, Xu H, Chen KJ. The diagnostic criteria of blood-stasis syndrome: considerations for standardization of pattern identification. *Chin J Integr Med*. 2014;20(7):483–489.
48. Zhang NL, Fu C, Liu TF, Chen B-x, Poon KM, Chen PX, et al. A data-driven method for syndrome type identification and classification in traditional Chinese medicine. *J Integr Med*. 2017;15(2):110–123.
49. Jacobson E, Conboy L, Tsering D, Shields M, McKnigh P, Wayne PM, et al. Experimental studies of inter-rater agreement in traditional Chinese medicine: a systematic review. *J Altern Complement Med*. 2019;25(11):1085–1096.
50. Wu WH, Liu JY, Chang HH. Latent class model based diagnostic system utilizing traditional Chinese medicine for patients with systemic lupus erythematosus. *Expert Syst Appl*. 2011;38(1):281–287.
51. Xu Z, Zhang NL, Wang Y, Liu G, Xu J, Liu T, et al. Statistical validation of traditional Chinese medicine syndrome postulates in the context of patients with cardiovascular disease. *J Altern Complement Med*. 2013;19(10):799–804.
52. Zhao Y, Zhang NL, Wang T, Wang Q. Discovering symptom co-occurrence patterns from 604 cases of depressive patient data using latent tree models. *J Altern Complement Med*. 2014;20(4):265–271.
53. Yeung WF, Chung KF, Zhang NLW, Zhang SP, Yung KP, Chen PX, et al. Identification of Chinese medicine syndromes in persistent insomnia associated with major depressive disorder: a latent tree analysis. *Chin Med*. 2016;11(1):4.
54. Zhang NL, Yuan S, Chen T, Wang Y. Latent tree models and diagnosis in traditional Chinese medicine. *Artif Intell Med*. 2008;42(3):229–245.
55. Grant SJ, Schnyer RN, Chang DH, Fahey P, Bensoussan A. Interrater reliability of chinese medicine diagnosis in people with prediabetes. *Evid Based Complement Alternat Med*. 2013;2013.
56. Ryu H, Lee H, Kim H, Kim J. Reliability and validity of a cold-heat pattern questionnaire for traditional Chinese medicine. *J Altern Complement Med*. 2010;16(6):663–667.
57. Popplewell M, Reizes J, Zaslawski C. A novel approach to describing traditional chinese medical patterns: the "traditional Chinese medical diagnostic descriptor". *J Altern Complement Med*. 2019;25(11):1121–1129.
58. Park YJ, Yang DH, Lee JM, Park YB. Development of a valid and reliable blood stasis questionnaire and its relationship to heart rate variability. *Complement Ther Med*. 2013;21(6):633–640.
59. Boet S, Etherington N, Larrigan S, Yin L, Khan H, Sullivan K, et al. Measuring the teamwork performance of teams in crisis situations: a systematic review of assessment tools and their measurement properties. *BMJ Qual Saf*. 2019;28(4):327.
60. Bellagamba D, Vionnet L, Margot-Cattin I, Vaucher P. Standardized on-road tests assessing fitness-to-drive in people with cognitive impairments: a systematic review. *PLoS ONE*. 2020;15(5).
61. Rahmatpour P, Sharif Nia H, Peyrovi H. Evaluation of psychometric properties of scales measuring student academic satisfaction: a systematic review. *J Educ Health Promot*. 2019;8:256.