

Development of a Chinese Medicine Assessment Measure: An Interdisciplinary Approach Using the Delphi Method

ROSA N. SCHNYER, L.Ac.,^{1,2} LISA A. CONBOY,^{1,2} ERIC JACOBSON,³ PATRICK McKNIGHT,⁴
THOMAS GODDARD,⁵ FRANCESCA MOSCATELLI,⁵ ANNA T.R. LEGEDZA,^{2,6}
CATHERINE KERR,² TED J. KAPTCHUK,² and PETER M. WAYNE¹

ABSTRACT

Background: The diagnostic framework and clinical reasoning process in Chinese medicine emphasizes the contextual and qualitative nature of a patient's illness. Chinese medicine assessment data may help interpret clinical outcomes.

Objectives: As part of a study aimed at assessing the validity and improving the inter-rater reliability of the Chinese diagnostic process, a structured assessment instrument was developed for use in clinical trials of acupuncture and other Chinese medical therapies.

Study design: To foster collaboration and maximize resources and information, an interdisciplinary advisory team was assembled. Under the guidance of two group process facilitators, and in order to establish whether the assessment instrument was consistent with accepted Chinese medicine diagnostic categories (face validity) and included the full range of each concept's meaning (content validity), a panel of Traditional Chinese Medicine (TCM) expert clinicians was convened and their responses were organized using the Delphi process, an iterative, anonymous, idea-generating and consensus-building process. An aggregate rating measure was obtained by taking the mean of mean ratings for each question across all 10 experts.

Results: Over three rounds, the overall rating increased from 7.4 (SD = 1.3) in Round 1 to 9.1 (SD = 0.5) in Round 3. The level of agreement among clinicians was measured by a decrease in SD.

Conclusions: The final instrument TEAMSI-TCM (Traditional East Asian Medicine Structured Interview, TCM version) uses the pattern differentiation model characteristic of TCM. This modular, dynamic version was specifically designed to assess women, with a focus on gynecologic conditions; with modifications it can be adapted for use with other populations and conditions. TEAMSI-TCM is a *prescriptive* instrument that guides clinicians to use the proper indicators, combine them in a systematic manner, and generate conclusions. In conjunction with treatment manualization and training it may serve to increase inter-rater reliability and inter-trial reproducibility in Chinese medicine clinical trials. Testing of the validity and reliability of this instrument currently is underway.

INTRODUCTION

At the heart Chinese Medicine (CM)* is a unique diagnostic framework and clinical reasoning process that

emphasizes the contextual and qualitative nature of a patient's illness. This process includes the assessment and synthesis of a large number of variables (including signs and symptoms), into a dynamic portrait of the patient's state of

¹New England School of Acupuncture, Watertown, MA.

²Osher Institute, Harvard Medical School, Boston, MA.

³Department of Social Medicine, Harvard Medical School, Boston, MA.

⁴University of Arizona, Tucson, AZ.

⁵Integral Healthcare Solutions, Falls Church, VA.

⁶Beth Israel Deaconess Medical Center, Boston, MA.

*The term CM (Chinese Medicine) is used to differentiate it from TCM (Traditional Chinese Medicine), which is characterized by the use of Pattern differentiation. CM includes many other styles of acupuncture that do not use this model.

health, which serves as the framework for developing and implementing specific treatment strategies.^{1,2}

In most clinical research on CM therapies many elements related to the experimental design (e.g., patient screening and eligibility, outcome measures, etc.) are defined only in biomedical terms. Despite its central role, CM diagnosis has not been well-integrated into such research, and few formal attempts have been made to evaluate its validity and reliability.³⁻⁷ The integration of a standardized CM diagnostic process in clinical trials has the potential of providing a unique, ecologically valid perspective for designing trials, assessing clinical outcomes, formulating evidence-based clinical recommendations, and developing future studies.⁸⁻¹⁰ The development of a valid and reliable assessment and outcome instrument is a fundamental step in this direction.

A small number of previous innovative studies have evaluated the reliability of the overall CM diagnostic process,³⁻⁶ as well as some of its components¹¹⁻¹³ and provide an important background to the present work. These studies suggest that the inter-rater reliability in the CM diagnostic process is low, except for one study,¹² which found 80% agreement between two raters. However, conclusions from these studies are limited by small sample sizes, analysis of the results on the basis of agreement alone, or lack of information about the validity of the instruments used to collect diagnostic data.

The operationalization of the CM diagnostic process presents many methodologic challenges, most arising from the difficulty of translating a qualitative process into a quantitative algorithm. As part of a study aimed at improving the validity and reliability of the CM diagnostic process, a structured assessment instrument was developed for use in clinical trials, the Traditional East Asian Medicine Structured Interview (TEAMSI-TCM; TCM version).[†] The hypothesis of the larger trial is that a structured assessment instrument will increase the inter-rater reliability of Traditional Chinese Medicine (TCM) diagnosis when used in clinical trials. This version of TEAMSI-TCM was specifically designed as a diagnostic tool for dysmenorrhea and related women's health conditions; however, with some modifications it can be used with other conditions and patient populations.

This article discusses the rationale, process, and experience gained in the development of TEAMSI. The cornerstone of this approach consisted of assembling an interdisciplinary team that included TCM experts, instrument development specialists, experienced acupuncture clinical researchers, a medical anthropologist, a sociologist, and a biostatistician. The goal was to build a common research platform by fostering cross-scientific-cultural communication. In addition, the Delphi Process was used to collect and

synthesize the input of 10 experienced TCM practitioners and quantify the incremental improvement in the instrument's validity during its stages of development. The following text provides an overview of the final instrument.

METHODS

The interdisciplinary advisory team

The exponential growth of CAM research in the United States in the last decade highlights the need for stronger collaboration and interdisciplinary dialogues between traditional medical practitioners and clinical research experts. Interdisciplinary dialogue has the potential to foster synergy and creativity, and maximize resources and information.¹⁴ Based on previous experience with a TCM-based assessment and outcome instrument,^{9,15,16} the authors included the following representatives in the interdisciplinary advisory team: TCM experts (RS, TK), a measurement methodologist (PM), a medical anthropologist (EJ), a sociologist (LC), clinical researchers (PW, RS, TK, LC, CK, EJ), a biostatistician (AL), and two Delphi experts (TG, FM). An initial 3-day (25-hr) meeting was held to listen to and learn from each other's perspectives, identify relevant issues about instrument development, and chart an overall strategy for the project. A second 2-day (19-h) meeting took place 6 months later to evaluate progress and integrate the results of the Delphi process into a final draft of the instrument. The advisory team maintained regular communication via email and telephone between these meetings.

A number of discipline-based issues were regularly raised by members of the advisory team. The expert responses to these issues guided the development of the structure and content of the instrument. This interdisciplinary process advanced long-term goals shared by the team: building communication between the CM and biomedical scientific cultures by developing a shared scientific language; building a common CM-biomedicine research platform for the quantitative and qualitative analysis of acupuncture clinical outcomes, and developing an instrument that could be used for CM and cross-cultural epidemiologic research.

The Delphi Process: Consensus-building among a panel of TCM experts to evaluate face-and content-validity

To establish whether the instrument was consistent with accepted TCM diagnostic categories (face validity) and included the full range of each concept's meaning (content validity), a panel of expert TCM clinicians was convened and their responses were organized using the Delphi process¹⁷⁻²¹: an idea-generating and consensus-building process that minimizes the effects of differences in status and interaction style. Researchers at the Rand Corporation

[†]The name Traditional East Asian Medicine Structured Assessment Interview was selected, and the suffix TCM was added to indicate that although various TEAM systems may share some common features, they each have distinct methods and practices.

developed the Delphi process to gather consensus from experts on topics about which there is great uncertainty.²² It has been used to establish diagnostic criteria and improve patient care in biomedicine^{23–26} and CAM.^{27–29} Four characteristics set the Delphi process apart from other research methods: anonymity, iteration, controlled feedback, and statistical aggregation of group response. This process allows free discussion of views without the influence of personal status; the alteration of personal views during the process without loss of face; the involvement of all groups with an interest in the subject; the synthesis of many opinions into a group response; and completion of all these tasks within a short period of time. The process is particularly suited for contexts where the problem at hand is complex and there is a need to build group consensus on the final decisions. It provides a way for experts to shed light on a topic while avoiding deferral of one to another, overcomes cost and scheduling constraints resulting from geographic dispersion of participants, and provides the benefits of combined open-ended answers and numeric ratings.

The TCM expert panel

Ten expert acupuncturists participated in the Delphi process. The advisory team selected panelists using a convenience sample that reflected a cross-section of TCM practitioners in the United States. The panel included six female and four male practitioners trained in Asia (4) or the United States (6); each with 10 or more years of clinical experience. Most were faculty at acupuncture schools or taught continuing education courses, and all were specialists in either internal medicine or TCM gynecology. A letter of invitation was sent to all potential participants followed by an informational meeting either in person or by telephone. Once participants consented, a detailed description of the project was sent to them outlining their participation.

The Delphi experts explained the process and answered the panelists' questions during an orientation conference call. Each member of the panel was sent an e-mail that contained a description of the entire process and instructions on responding to future e-mails. During a period of approximately 5 months, panel members were asked to complete three different surveys and were given 2 weeks to complete each. Responses to surveys 1 and 2 informed the development of a revised draft of the TEAMS-TCM. Survey 3 consisted of the revised draft along with a summary of the panelists' comments from the second survey and a further set of questions to give participants the opportunity to provide additional input. The administration of questionnaires over a series of three rounds made it possible for the panelists to change their opinions between one round and the next. The advisory team analyzed the results between rounds; their feedback guided the design of each subsequent round. After each round, the Delphi moderators provided an aggregate measure of the group ratings on each particular issue.

Table 1 outlines the main changes made to the working draft between various iterations.

Diagnosis in CM

TCM diagnosis is a process of assessing the *relationship* among many different variables and integrating all relevant information into a description of the patient's illness experience (symptoms), the clinical presentation (signs), and the way in which the illness affects his or her life and behavior.¹ The process of weaving together signs and symptoms into a dynamic and contextual etiologic–pathophysiologic clinical picture is termed *pattern differentiation* (*biàn zhèng*).[‡] A working diagnosis consists of a set of primary patterns and secondary patterns and is not limited to the main complaint. TEAMS-TCM was based on a previously developed instrument³⁰ (PDI) that follows a commonly used TCM format. A review of the literature^{2,31,32} confirmed that all commonly used elements of the assessment process were included in this working draft, which was further revised at the initial advisory team meeting.

RESULTS

Round 1

The purpose of the first round was to obtain an initial response from the TCM experts as to the content (item selection) and format (best sequence to organize items). Experts were asked to offer open-ended comments and rate (on a scale from 1 to 10) the quality, scope, and ease of use of the forms around particular items, assumptions, and concepts contained in the working draft, with the higher numbers indicating better ratings. An aggregate measure was obtained by taking the mean rating for each question across all 10 experts, then calculating the total mean of all those means. As seen in Table 2, the overall rating in Round 1 was 7.4 (standard deviation [SD] = 1.3) with a range of 5.1 to 9.2. The relatively low mean score of 7.4 indicated that the experts thought the form could be improved. The higher SD of 1.3 (compared to later rounds) suggested greater divergence in the experts' ratings in Round 1 than in later rounds.

There was a relatively even split among the consultants about whether TEAMS-TCM should be simplified or made more detailed. Most qualitative and open-ended items in the instrument generated opposition and confusion and eventually were eliminated. The items recording the findings of tongue

[‡]*Pattern* is the translation of the term *zhèng*, which denotes the dynamic nature of disease categories in TCM.^{1,35} The first stage of pattern identification is the differentiation of clinical signs into *yin* and *yang* (Eight Principles); depending on the results obtained, other pattern differentiation models then are applied.³⁵

TABLE 1. CHANGES MADE TO THE INSTRUMENT THROUGH THE DELPHI PROCESS

| | <i>PDI (previously developed instrument)</i> | <i>Working draft (after first advisory team meeting)</i> | <i>Delphi round 1 to round 2</i> | <i>Delphi round 2 to round 3</i> | <i>TEAMS-TCM</i> |
|--|--|---|--|---|--|
| Patient questionnaire: dysmenorrhea | Included only general questions about menstruation as related to TCM. Questions open-ended (e.g., What makes pain better or worse? What is color, amount, consistency of flow?). | Revised TCM textbook and research literature for omissions, discrepancies, and accuracy. Reorganized questions and limited answers to set of multiple choices (e.g., pain better with: heat, rest, activity, pressure). | Based on Delphi panel of experts, refined and expanded instrument content to include questions on other clinically relevant information (e.g., medication use, length of period, define color, consistency). | Further refined and expanded content; reorganized format including sequence of items and layout. | Simplified language to meet third grade reading level. Formatted to include yes, no, and not sure answers. Coded for data entry. |
| Patient questionnaire: general questions | Open-ended questions about symptoms in other areas to include the "Ten Questions" of TCM. Included quality of life issues and life style choices (e.g., diet, exercise). | Revised TCM textbook literature to include most relevant issues in each of the Ten Questions. Converted to closed-ended questions followed by a limited set of multiple choice answers. | Explored expanding length of this section to include all possible clinically relevant questions related to the Ten Questions (e.g., If headache: location, better or worse, quality). Placed open-ended qualitative questions into a separate section. | Reorganized Ten Questions and separated each system (digestion, sleep) into a separate module; kept only the most relevant guiding questions. Created Workbook Modules (see the following). Changed open-ended questions about severity and effect of symptoms to a VAS scale and multiple choice answers. Kept one qualitative question for each module about effect of symptoms on quality of life. | Used textbooks as reference to identify most important guiding questions. Coded instrument for data entry and analysis. |
| Workbook | Practitioner's notes embedded in patient questionnaires. | Developed fold-out for practitioner's notes, matched to questions in patient questionnaires. | Converted fold-out format to less cumbersome notation space for practitioners at base of each page. | Developed workbook to include most important follow-up questions for each medical or organ system and symptom endorsed (i.e., if headaches are endorsed by the patient on the corresponding module, follow up questions about headache quality or location are further developed in the workbook). | Matched every question to the corresponding module in the practitioner workbook. |

TABLE 1. CHANGES MADE TO THE INSTRUMENT THROUGH THE DELPHI PROCESS (CONTINUED)

| | <i>PDI (previously developed instrument)</i> | <i>Working draft (after first advisory team meeting)</i> | <i>Delphi round 1 to round 2</i> | <i>Delphi round 2 to round 3</i> | <i>TEAMS-TCM</i> |
|---|---|---|---|---|--|
| Signs (pulse, tongue, observations) positions, tongue | Open-ended endorsement of general categories (e.g., pulse endorsement (e.g., body or coat). | Defined each category and provided multiple choice items for tongue and pulse, pulse: wiry, slippery. | Developed a detailed description for each category of discrepancies with multiple choice items for endorsement. | Focused pulse and tongue to include main categories only and resolve inaccuracies and ambiguities. Reorganized items for multiple choice endorsements | Further reorganized record of signs for easy use and data entry. |
| Pattern differentiation | General patterns organized by vacuity and repletion according to eight principles. | Step-by-step differentiation to include organs or channels involved. Provided primary and secondary patterns involved in dysmenorrhea and main combination for multiple choice endorsement. | To capture diagnostic reasoning, separated pattern differentiation by categories (eight principles, vacuity or repletion, qi or blood, <i>yin</i> or <i>yang</i>), identifying the specific clinical presentation related to each pattern. Separated patterns not directly relevant to making a diagnosis of dysmenorrhea. | Based on Delphi expert panel recommendations, simplified and clarified pattern differentiation for both main and secondary patterns in dysmenorrhea and general patterns. | Further reorganized record of pattern differentiation. |

The "Ten Questions" of TCM (*shí wèn*) constitute one of the Four Evaluations (*sì zhěn*): (inquiry, smelling or hearing, observation, and palpation), and represent the 10 areas on which Chinese doctors focus when conducting a diagnosis. These ten areas are (1) cold and heat, (2) sweating, (3) head and body, (4) stool and urine, (5) diet and appetite, (6) chest and abdomen, (7) eyes and ears, (8) sleep, (9) medical history, and (10) living habits.^{1,35}

and pulse examination, as well as the options for recording the final pattern differentiation, were simplified.

Round 2

In preparation for the second round, suggested changes to the form were made to facilitate use by both practition-

ers and patients and increase agreement among panelists. These consisted of inclusion of previously omitted items, clarification on the wording of questions, deletion of confusing statements, and reformatting of the list of possible findings (pulse and tongue) and TCM diagnostic categories (patterns). All suggested changes were reviewed by the advisory team relative to three criteria: ecologic validity, ap-

TABLE 2. STATISTICAL SUMMARY OF AGREEMENT FOR EACH ROUND OF THE DELPHI PROCESS

| <i>Delphi process</i> | <i>Round 1</i> | <i>Round 2</i> | <i>Round 3</i> |
|---------------------------------------|----------------|----------------|----------------|
| Overall mean rating across 10 experts | 7.4 | 8.5 | 9.1 |
| Standard deviation | 1.3 | 1.0 | 0.5 |
| Minimum | 5.1 | 6.3 | 8.1 |
| Median | 7.7 | 8.5 | 9.2 |
| Maximum | 9.2 | 9.8 | 9.8 |

Minimum rating = 1; maximum possible rating = 10.

appropriate measurement development, and useful data collection.

Although there was still a need to reconcile differences of opinion about detail and clarity with manageable length, after Round 2, the mean overall rating increased 1.1 points to a level of 8.5 (SD = 1.0) and ranged increased from 6.3 to 9.8. The SD declined from 1.3 in the first round to 1.0 in the second round. Further suggestions were made primarily to the recording of tongue and pulse findings and pattern differentiation.

Round 3

Between the second and third rounds, a second advisory team meeting was held to discuss the results of the Delphi process, analyze the changes to the instrument made so far, and decide the best approach on the basis of study objectives. The final draft then was sent to the panelists with a third round of questions. The third survey resulted in a further increase in the overall rating, mean = 9.1 (SD = 0.5).

Progress was assessed by a decrease in the variability of judgments among experts about the instrument and by achieving increasingly higher ratings as the instrument changed from one round to the next. The level of agreement was quantified by the SD with smaller values indicating greater agreement. The SD was reduced in half after Round 2 and the narrow range (8.1, 9.8) fell within the desired target values (8–10). The ratings among experts not only were less variable, but also were consistently higher as the instrument changed.

The Final Product: TEAMSITCM

The instrument as developed in this project focuses specifically on the TCM style of treatment. TCM's body of

knowledge is very broad. No single instrument would be adequate to assess all possible health conditions; yet a single instrument can be optimized for use in a population that shares characteristics that are important from a TCM standpoint. This version was developed specifically to assess women's gynecologic issues. The range of content and logical relationships was narrowed to focus on improving and recording clinical reasoning within that domain.^{33,34}

In order to provide the breadth and depth of content necessary to capture the assessment process of TCM practitioners with a range of different training and experience, a modular instrument was created. This facilitates the instrument's adaptability to clinical trials focused on other health conditions. Each module can be used independently and also can be developed into a main complaint module. Figure 1 provides a quick reference flow chart and a general description of the overall instrument. Figure 2 provides a sample of Module III. Figure 3 a sample of the practitioner's workbook, which goes with it. Once the reliability and validity of TEAMSITCM has been evaluated clinically and experimentally and corrections are made, the instrument is made available to the research community with guidelines for use. Examples of the form can be found at: www.nesa.edu/research/currentresearch/drcr3.html

DISCUSSION

The collaboration of an interdisciplinary team provided multiple perspectives on the development of TEAMSITCM. Only some of the issues involved were articulated and addressed; many others remained unresolved.

TEAMSITCM is not meant to be a *descriptive* instrument that captures the actual clinical reasoning. It is a *pre-*

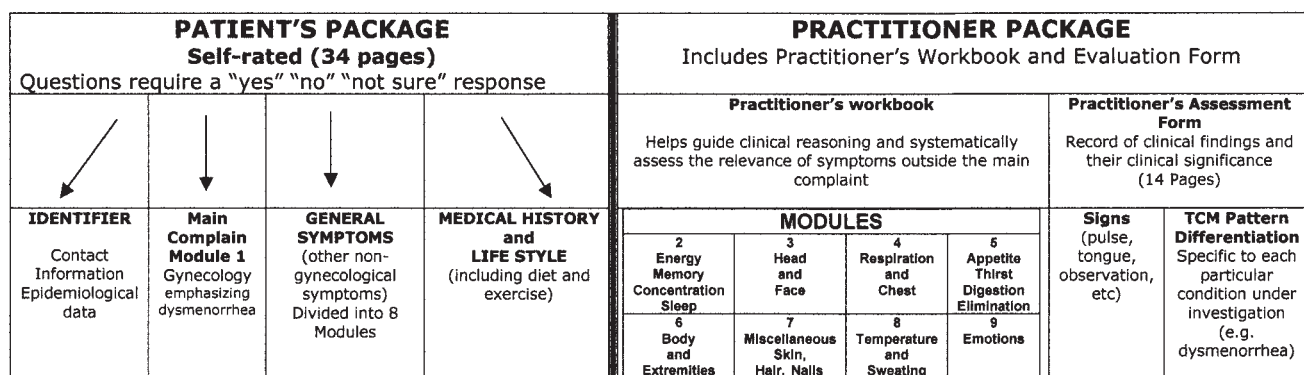


FIG. 1. TEAMSITCM organizational flowchart. The modules constitute the "Ten Questions" of TCM.^{1,2,32,35} The practitioner's workbook provides an interface between the patient and practitioner modules. It includes a "working module" for each module in the patient package. Once the patient endorses a symptom (e.g., headaches), the practitioner refers to the relevant working modules (e.g., Module 3: Head and Face). The module lists the follow-up questions (e.g., onset, location) necessary to differentiate symptoms according to TCM categories (e.g., hyperactive liver *yang*). The symptoms detailed in the modules have been cross-referenced with standard TCM clinical diagnosis texts.^{36–38}

| IN THE PAST YEAR | | | For office use only |
|---|--|--|---------------------|
| Head | | | |
| 1. Do you often have headaches or suffer from migraines ? | <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> Not sure | | GSHEAD1_____ |
| 2. Does your head often feel heavy or full ? | <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> Not sure | | GSHEAD3_____ |
| 3. Do you ever have unusual sensations in your head (for example, it feels hot or cold)? | <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> Not sure | | GSHEAD4_____ |
| 4. Do you ever have any numbness, tingling or itching of your scalp? | <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> Not sure | | GSHEAD5_____ |
| 5. Have you felt dizzy ? | <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> Not sure | | GSHEAD6_____ |
| 6. Have you fainted or lost consciousness ? | <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> Not sure | | GSHEAD7_____ |

FIG. 2. Sample of module III (head and face).

| I. Head | Asked? | For Office use only |
|--|--------------------------|---------------------|
| 1. Headaches or Migraines | | |
| a. Have they been diagnosed by an MD as migraines or other? <input type="checkbox"/> Yes <input type="checkbox"/> No Diagnosis _____ Treatment _____ | <input type="checkbox"/> | GSHEAD1a_____ |
| b. Onset: <input type="checkbox"/> acute/sudden <input type="checkbox"/> chronic/gradual <input type="checkbox"/> recurrent | <input type="checkbox"/> | GSHEAD1b_____ |
| c. Time when worst: <input type="checkbox"/> daytime <input type="checkbox"/> evening <input type="checkbox"/> upon waking <input type="checkbox"/> at night | <input type="checkbox"/> | GSHEAD1c_____ |
| d. Location: <input type="checkbox"/> back of head (occiput) <input type="checkbox"/> top of head (vertex) <input type="checkbox"/> forehead <input type="checkbox"/> sides/temples <input type="checkbox"/> behind the eyes <input type="checkbox"/> whole head | <input type="checkbox"/> | GSHEAD1d_____ |
| e. Quality: <input type="checkbox"/> as if my head were wrapped in cotton <input type="checkbox"/> pulling <input type="checkbox"/> intense <input type="checkbox"/> sharp <input type="checkbox"/> feels like pain is inside head <input type="checkbox"/> throbbing <input type="checkbox"/> dull <input type="checkbox"/> fixed | <input type="checkbox"/> | GSHEAD1e_____ |
| f. Better with: <input type="checkbox"/> resting down <input type="checkbox"/> exercise <input type="checkbox"/> food | <input type="checkbox"/> | GSHEAD1f_____ |
| g. Worse with: <input type="checkbox"/> resting down <input type="checkbox"/> exercise <input type="checkbox"/> food | <input type="checkbox"/> | GSHEAD1g_____ |

FIG. 3. Sample of the practitioner’s workbook for module III (head and face).

scriptive instrument that guides clinicians to use the proper indicators, combine them in a systematic manner, and generate conclusions based on their use and combination. In conjunction with manualization and training⁹ it may serve to increase inter-rater reliability and inter-trial reproducibility in CM research. This hypothesis currently is being tested in a prospective randomized trial that compares inter-rater reliability in TCM diagnosis among acupuncturists using: (1) TEAMSI with training; (2) TEAMSI with no training; and (3) the NESA student clinic intake form with training and no training.

Compared to the TCM assessment questionnaires commonly used in practice, the TEAMSI-TCM leads the practitioner through the process in a more structured manner. Opportunities for open-ended responses were replaced with close-ended response categories, which force a single response. A limitation of this format is that, if multiple an-

swers are correct, a more restricted choice may not necessarily lead to more consistency in responses across raters.

It is not clear that the established statistical definition of reliability as a measure of degree of agreement among raters applies to TCM diagnosis and treatment selection. What exactly constitutes “reliability” in TCM clinical thinking? The assumption, common in biomedicine, that there is one and only one correct diagnostic conclusion for a given individual case may not hold true in TCM. For some patients there may be more than one legitimate TCM diagnosis, and more than one legitimate treatment plan. Some allowance was made for this in TEAMSI-TCM by providing an option for the practitioner to record both primary and secondary diagnoses. Still, for some patients there may be more than one legitimate primary diagnosis. Because of this there is no *a priori* way to determine whether deficits in inter-rater reliability are the result of: (1) errors in the conclusions endorsed by individual raters, or (2) the failure of the analytic process to count multiple legitimate conclusions as instances of inter-rater agreement. This may account for low inter-rater reliability found in the TCM studies mentioned in the preceding.

CONCLUSIONS

The opinion of this TCM expert panel indicated that TEAMSI allows practitioners to accurately record the clinical reasoning. For use in research, they support the adoption of a more rather than less structured instrument. Problems that emerge in actually using TEAMSI will be explored in the testing phase of this project, which is currently underway. Validity testing interviews will assess whether practitioners feel that the form guides the collection of a sufficiently broad range of clinical data, whether it does so in a naturalistic manner, and how their experience using this new instrument compares to their previous training and experi-

ence. Clearly more work is needed to understand the diagnostic process and improve its inter-rater reliability in TCM; the development and evaluation of TEAMSI-TCM constitute an important step.

ACKNOWLEDGMENTS

This project was supported by grant number 5 U19 AT002022 from the National Center for Complementary and Alternative Medicine. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NCCAM, or the National Institutes of Health. The authors thank Kimberly Davis, Dagmar Ehling, Andrew Fitzcharles, Grant Guan Huo, Yan Ping Jin, Robert Kaneko, Cynthia Labruzzo, Peter Valaskatgis, Vivien Zhang, and Zhen Zhen Zhang for their participation in the Delphi Process as part of the expert panel. The authors also thank Julie Buring, Roger Davis, Ellen Connors, Andrea Hrbek, Monica Shields, and Mary Quilty for their assistance with this project. Julie Buring and Richard Hammerschlag gave critical feedback on an earlier draft of this manuscript.

REFERENCES

- Kaptchuk TJ. *The Web That Has No Weaver: Understanding Chinese Medicine*. 2nd ed. Chicago: Contemporary Books (McGraw-Hill), 2000.
- Maciocia G. *The Foundations of Chinese Medicine*. New York: Churchill Livingstone, 1989.
- Hogeboom CJ, Sherman KJ, Cherkin, DC. Variation in diagnosis and treatment of chronic low back pain in Traditional Chinese Medicine acupuncturists. *Complement Ther Med* 2001;9:154–166.
- Kalauokalani D, Sherman KJ, Cherkin DC. Acupuncture for chronic low back pain: Diagnosis and treatment patterns among acupuncturists evaluating the same patient. *South Med J* 2001;94:486–492.
- Zhang GC, Bausell B, Lixing L, Handweger B, Berman MB. Assessing the consistency of traditional Chinese medical diagnosis: An integrative approach. *Altern Ther* 2003;9:66–72.
- Zhang GC, Bausell B, Lixing L, Wen-Lin L, Handweger B, Berman MB. The variability of TCM pattern diagnosis and herbal prescriptions on rheumatoid arthritis patients. *Altern Ther* 2004;10:58–63.
- Zaslowski C. Clinical reasoning in traditional Chinese medicine: Implications for clinical research. *Clin Acupunct Orient Med* 2003;4:94–101.
- Hammerschlag R. Methodological and ethical issues in clinical trials of acupuncture. *J Altern Complement Med* 1998;4:159–171.
- Schnyer RN, Allen JJ. 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. *J Altern Complement Med* 2002; 8:623–634.
- Birch S. Clinical research on acupuncture: Part 2. Controlled clinical trials, an overview of methods. *J Altern Complement Med* 2004;10:481–498.
- Walsh C, Corbin D, Bateman K, Zaslowski C. Feeling the pulse: Trial to assess agreement level among TCM students when identifying basic pulse characteristics. *Eur J Orient Med* 2001;3:25–31.
- King E, Cobbin D, Walsh S, Ryan D. The reliable measurement of radial pulse: Characteristics. *Acupunct Med* 2002;20: 150–159.
- King E, Cobbin D, Ryan D. The reliable measurement of radial pulse: Gender differences in pulse profiles. *Acupunct Med* 2002;20:160–167.
- Morreale SP, Hower CB. The benefits of interdisciplinary collaboration: Down with silos and up with engagement. 2002. Online document at www.aahe.org/interdisciplinary.pdf Accessed May 2005.
- Allen JJB, Schnyer RN, Hitt SK. The efficacy of acupuncture in the treatment of major depression in women. *Psychol Sci* 1998;9:397–401.
- Manber R, Schnyer RN, Allen JJB, Rush J, Blasey C. Acupuncture: A promising treatment for depression during pregnancy. *J Affect Disord* 2004;83:89–95.
- Mangelsdorff AD, Rogers JR, Zucker KW, et al. Vision 21 Delphi panel: Senior Army Medical Service Corps officers' vision of behaviors for success of future Medical Service Corps officers. *Milit Med* 1997;162:450–452.
- White MB, Edwards SA, Russell CS. The essential elements of successful marriage and family therapy: A modified Delphi study. *Amer J Family Ther* 1997;25(3):213–231.
- Pill J. The Delphi method: Substance, context, a critique and an annotated bibliography. *Socio-Econ Planning Sci* 1971;5: 57–71.
- Hiltz SR, Turoff M. *The Network Nation: Human Communication via Computer*. Upper Saddle River, NJ: Addison-Wesley, 1978.
- Delbecq AL, VaneVen AH, Gustafson DH. *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*. Glenview, IL: Scott-Foresman, 1975.
- Linstone HA, Turoff M, eds. *The Delphi Method: Techniques and Applications*. London: Addison-Wesley, 1975.
- Jones J, Hunter D. Qualitative research: Consensus methods for medical and health services research. *BMJ* 1995;311: 376–380.
- Chocholic J, Bouchard SKH, Tan J, Ostrow D. The determination of relevant goals and criteria used to select an automated patient care information system. A Delphi approach. *J Am Med Inform Assoc* 1999;6:219–233.
- Olney R, Lewis RA, Putnam T, Campellone JV. Consensus criteria for the diagnosis of multifocal neuropathy. *Muscle Nerve* 2003;27:117–121.
- Graham B, Reghr G, Wright JG. Delphi method to establish consensus for diagnostic criteria. *J Clin Epidemiol* 2003;56: 1150–1156.
- Barnard S, Lewith GT, Kemp T. Researching complementary therapies: A Delphi study to identify the views of complementary and orthodox practitioners, 1997. *J Altern Complement Med* 1997;3:141–147.
- Richardson J. Developing and evaluating complementary therapy services: Part 1. Establishing service provision through the

- use of evidence and consensus development. *J Altern Complement Med* 2001;7:253–260.
29. Webstaer-Harrison, White A, Rae J. Acupuncture for tennis elbow: An email consensus study to define a standardized treatment in GP's surgery. *Acupunct Med* 2002;20:181–185.
 30. Schnyer RN, Allen JB, Hitt SK, Manber R. Acupuncture in the Treatment of Major Depression: A Manual for Research and Practice. London: Churchill Livingstone, 2001.
 31. Wiseman N, Ellis, A. *Fundamentals of Chinese Medicine*, rev ed. Brookline, MA: Paradigm Publications, 1996.
 32. Deng T. *Practical Diagnosis in TCM*. Edinburgh: Churchill Livingstone, 1999.
 33. Flaws B. *My Sister the Moon: The Diagnosis and Treatment of Menstrual Diseases by Traditional Chinese Medicine*. Boulder, CO: Blue Poppy Press, 1992.
 34. Maciocia G. *Obstetrics and Gynecology in Chinese Medicine*. London: Churchill Livingstone, 1998.
 35. Wiseman N, Ye F. *A Practical Dictionary of Chinese Medicine*. Brookline, MA: Paradigm Publications, 1998.
 36. Maciocia G. *Diagnosis in Chinese Medicine: A Comprehensive Guide*. London: Churchill Livingstone, 2004.
 37. Sionneau P, Gang Lü. *The Treatment of Disease in TCM*, vol 1–7. Boulder, CO: Blue Poppy Press, 1996–2001.
 38. Flaws B, Sionneau P. *The Treatment of Modern Western Medical Diseases with Chinese Medicine*. Boulder, CO: Blue Poppy Press, 2001.

Address reprint requests to:

Rosa N. Schnyer, L.Ac.

Osher Institute at Harvard Medical School

401 Park Drive, 22 A Landmark

Boston, MA 02215

E-mail: Rosa_Schnyer@hms.harvard.edu

Copyright of *Journal of Alternative & Complementary Medicine* is the property of *Mary Ann Liebert, Inc.* and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.