THE JOURNAL OF ALTERNATIVE AND COMPLEMENTARY MEDICINE
Volume 15, Number 10, 2009, pp. 1099–1105

DOI: 10.1089/acm.2009.0020

Original Article

Investigating the Reliability of Japanese *Toyohari* Meridian Therapy Diagnosis

Kylie Ann O'Brien, Ph.D.,[1,2] Estelle Abbas, B.App.Sc.(ChinMed),[1] Paul Movsessian, Dip.Appl.Sc.(Acu),[3] Michael Hook, Dip.Acu., B.A.,[3] Paul A. Komesaroff, Ph.D.,[2] and Stephen Birch, Ph.D.[4,5]

Abstract

Background: *Toyohari* meridian therapy (TMT) is a Japanese system of acupuncture. Acupoint selection follows diagnosis of the primary and secondary patterns of disharmony (*sho*) and disturbances in the *yang* channels. Pulse diagnosis and abdominal palpation diagnosis are the two main diagnostic methods used. Little is known about the reliability of pulse, abdominal, and pattern diagnosis in TMT. This is important since diagnosis of the *sho* determines acupoint treatment. If diagnosis is unreliable, there can be less confidence that the patient will receive optimal treatment.

Objective: The objective of this study is to assess the level of agreement between two TMT practitioners on pulse diagnosis, abdominal diagnosis, and diagnosis of the primary and secondary *sho*.

Methods: An inter-rater reliability study was conducted. Two (2) TMT practitioners separately conducted a TMT examination and completed an assessment form, choosing from a range of possible responses relating to pulse characteristics, abdominal diagnosis, and diagnosis of primary *sho* and secondary *sho*. The κ coefficient was used as a measure of inter-rater reliability of the outcome variables.

Results: Sixty-two (62) Australians (22 males, 40 females) aged 20–65 years participated (mean age 49.2 ± 12.2 years). Level of agreement for pulse diagnosis was 57%, 61%, and 77% for pulse depth, speed, and strength, respectively. For abdominal diagnosis, the level of agreement for involvement of the Lung, Kidney, Spleen, and Liver abdominal regions was 58%, 53%, 35%, and 10%, respectively. The overall level of agreement on primary *sho* diagnosis was 48% and for secondary *sho* diagnosis, 44%.

Conclusions: Overall, there was a reasonable level of agreement on basic pulse characteristics and on abdominal diagnosis for two of the abdominal regions. Level of agreement on primary and secondary *sho* diagnosis suggests room for improvement. Further studies are required in order to gain a greater understanding of the reliability of diagnosis in TMT.

Introduction

Traditional East Asian Medicine (TEAM) refers to a diverse set of therapies originating in East Asia that includes Traditional Chinese Medicine (TCM), one of the more commonly recognized forms of TEAM. Meridian therapy is defined as traditional Japanese medicine based on pulse diagnosis.[1] Based on "*hari* traditions" that are thousands of years old, the term "Meridian therapy" was adopted around the late 1930s to distinguish a form of therapy distinct from modern Chinese acupuncture.[1] Central to Meridian therapy are *yin-yang* theory, five-phase theory, and channel theory. Essentially, illness is understood as the result of a disturbance of *qi* and blood (Japanese *ki* and *ketsu*), identified through detection of states of vacuity (*kyo*) and repletion (*jitsu*) of the channels.[1] The *kyo* and *jitsu* are addressed by applying supplementation *ho* and draining *sha* needling techniques, respectively. *Toyohari* meridian therapy (TMT) was initially practiced mainly by blind practitioners in Japan though in modern times, many sighted practitioners and increasingly,

[1]Victoria University, Melbourne, Australia.
[2]Department of Medicine, Monash University, Alfred Hospital, Melbourne, Victoria, Australia.
[3]Australasian Branch of the Toyohari Association, Sydney, New South Wales, Australia.
[4]Stichting (Foundation) for the Study of Traditional East Asian Medicine (STEAM), Amsterdam, The Netherlands.
[5]Norwegian College of Acupuncture, Oslo, Norway.

Western acupuncturists practice this form of therapy. One of the features of TMT is that unlike Chinese style acupuncture, the needles do not necessarily pierce the skin; instead, *qi* is manipulated at the surface of the skin with the needle tip typically held just above the surface of the skin, or in some cases, lightly touching it. In addition, in TCM acupuncture greater emphasis is placed on individual acupuncture point (acupoint) function, whereas in TMT, the emphasis is on meridian function. There are other differences between TMT and TCM. For example, *yin–yang* theory, five-phase theory, and meridian theory are important in both TCM and TMT. However, in TCM, there are many other theories utilized, in particular *Zang-Fu* (organ) Theory: This is not a feature of TMT. In addition, in TCM the Eight Guiding Principles are important principles that guide diagnosis and treatment, whereas in TMT these are not.

TMT places particular diagnostic emphasis on pulse diagnosis and abdominal (*hara*) diagnosis, but little emphasis on tongue diagnosis, unlike TCM, which places comparatively little emphasis on *hara* diagnosis and greater emphasis on tongue diagnosis. Five-phase theory underpins the comparison of the pulse at the *cun, guan,* and *chi* positions on the right and left radial arteries and detection of pathology in the channels.[2] The correlation between the meridians and pulse positions of the wrist are set out in Table 1, which may differ from those identified under the TCM model in the *chi* positions (that is, no recognition of Kidney *Yang* or Kidney *Yin* pulses in TMT). Abdominal diagnosis involves inspection and light palpation of the abdomen, which is divided into particular regions that relate to the *zang* organs and related channels. Five- phase theory and in particular, Clause 69 of the Nanjing (circa 100 CE), forms the basis for diagnosis of the primary and secondary patterns of disharmony, termed *sho*. The "primary *sho*" represents the root (*ben*) of the disorder or disharmony in the body and relates to the *sheng* cycle (mother–son cycle) in five-phase theory. It is always one of vacuity or *kyo*. The "secondary *sho*," reflective of the controlling cycle or *ke* cycle in five-phase theory, in general may be one of *kyo* or *jitsu* (the exception being when there is a *waho* condition, akin to *qi* stagnation). Finally, the practitioner seeks to ascertain the state of the *yang* channels, which may be either one of *kyo* or *jitsu*. Diagnosis of the primary *sho*, secondary *sho,* and the state of the *yang* channels directly informs treatment, that is, choice of acupoints and needling techniques. Unlike TCM acupuncture, in which acupoint selection may be based on a range of theories and in which there is a greater emphasis on individual acupoint function and perhaps greater latitude to choose acupoints, in TMT the focus is on the meridian and there is a relatively strict protocol in terms of acupoints and needling techniques used. For example, if a lung primary *sho* is diagnosed, the acupoints used to treat are typically *Taiyuan* (Lung 9) and *Taibai* (Spleen 3), which are both "supplemented."

Since optimal treatment relies on correct diagnosis, it is important that diagnosis is accurate. However, relatively little is known about the reliability or repeatability of TMT diagnosis, nor of two of its major diagnostic techniques, pulse and abdominal diagnosis. A few early studies of reliability of diagnosis were conducted in the 1960s in Japan, including test–retest studies of pattern diagnosis based on radial pulse palpation alone,[3,4] test–retest reliability of abdominal palpation,[5] and the inter-rater reliability of pattern diagnosis using the four diagnostic methods.[6] Later studies by Birch found variable levels of reliability for characteristics of pulse diagnosis but good indications of agreement on pattern selection in meridian therapy (Birch S, 1997; Birch S, 1999, unpublished data). The majority of inter-rater reliability studies have been conducted within a TCM model and have assessed pulse diagnosis, tongue diagnosis, other diagnostic variables, and pattern diagnosis in a range of conditions and are summarized elsewhere.[7] Given the many differences between TCM and TMT, results of inter-rater reliability studies in TCM cannot necessarily be extrapolated to TMT.

In order to investigate the reliability of TMT diagnosis, we conducted an inter-rater reliability study between 2 TMT practitioners on key features of TMT diagnosis: *sho* diagnosis, pulse diagnosis, and abdominal diagnosis.

Aim and Objectives

The overall aim of this study was to assess the reliability of diagnosis in TMT between 2 TMT practitioners. Specifically, the objectives were to assess the reproducibility of the following.

1. Pulse depth, speed and strength;
2. Abdominal diagnosis;
3. Primary *sho* diagnosis; and
4. Secondary *sho* diagnosis

Methods

The study was part of a larger study conducted to investigate the potential physiologic correlates in the cardiovascular system associated with a TMT root treatment in an Australian population. It was conducted in the School of Health Sciences Teaching Clinics, Victoria University in Melbourne, Australia. Ethics approval was obtained from the Victoria University Faculty of Health, Engineering, and Science Ethics Committee.

Study participants

Sixty-two (62) participants aged between 20 and 65 years (average 49.2 ± 12.2 years), 22 male and 40 females, were recruited from newspaper advertisements, posters displayed at the Victoria University St Albans campus, and via the intra-university Internet noticeboard.

Inclusion criteria included being 18 years or older and in good health. Exclusion criteria included: diabetes, hypertension, taking antihypertensive medications and other vasoactive medications including hormone replacement therapy. Interested participants taking omega-3 supplements were asked to discontinue these at least 3 weeks prior to study participation.

Toyohari *practitioners*

Two experienced TMT practitioners participated in the study. Both practitioners have qualifications in TCM acupuncture and certificates in *Toyohari East Asian Needle Therapy* (a term often used interchangeably with TMT), have completed advanced training in TMT in Japan, and are approved teachers of TMT with the Toyohari Medical Asso-

Table 1. Relationship Between the Pulse Positions and 12 Channels

	Right wrist		*Left wrist*	
Position/depth	Superficial	Deep	Superficial	Deep
Cun (Sunnko)	Large intestine channel	Lung channel	Small intestine channel	Heart (pericardium) channel
Guan (Kanjo)	Stomach channel	Spleen channel	Gallbladder channel	Liver channel
Chi (Shakuchu)	Triple heater channel	Ming men	Bladder channel	Kidney channel

ciation (Japan). One (1) practitioner has been practicing TMT for 10 years and the other for 12 years.

Toyohari *diagnostic outcome variables*

The TMT examination includes four diagnostic methods: inquiry, auscultation, inspection, and palpation. Inquiry or case history taking involves asking questions regarding the patient's presenting and secondary complaints plus other relevant information. Inspection predominantly involves observation of the face and inner forearm, with the course of the channels on the lower leg also being inspected.[1] The abdominal area is also inspected as part of abdominal diagnosis. Auscultation involves listening to sound quality of the voice, including its tone and strength. Palpation includes pulse diagnosis, abdominal diagnosis, and channel palpation, explained below.

Pulse diagnosis. Three basic characteristics of the pulse overall were assessed: depth, speed, and strength. Pulse depth (location) indicates the depth at which the pulse is felt most strongly on palpation. A floating pulse is one that is felt most strongly with very light palpation and that decreases in strength as more pressure is applied, whereas a sinking pulse is one that is felt most strongly on deep palpation almost to the bone. Practitioners were asked to choose from one of the five following descriptors (from most superficial to deepest): "floating," "little floating," "normal," "little sinking," and "sinking."

Pulse "speed" is generally measured as the number of beats per breath of the patient. A normal speed is considered to be 4–5 beats per breath. A slow pulse is defined as less than or equal to 3 beats per breath and a fast pulse is defined as greater than or equal to 6 beats per breath. Practitioners were asked to choose from one of the five following descriptors of pulse speed: "rapid," "little rapid," "normal," "little slow," and "slow."

Strength of the pulse involves a subjective judgment about the strength of the pulsation of the blood in the blood vessel as detected by the practitioner's fingers. Practitioners were asked to choose from one of the following five descriptors of pulse force: "vacuous," "little vacuous," "normal," "little replete," and "replete."

The word "little" in the descriptors above is equivalent to "slightly."

Abdominal diagnosis. In abdominal diagnosis, the abdominal region from the rib margin to the hypogastric region is inspected for subtle changes in appearance including skin luster, discoloration, spider nevi, musculature, and body fat. The abdomen is also palpated for differences in temperature, firmness, and tenderness. Different organ systems are mapped onto different areas of the abdomen. For example, the Heart region is the region directly below the sternum, extending slightly down toward the navel. A subjective judgment is made about whether the abdomen as a whole and particular areas of the abdomen are *kyo* (indicative of a "vacuity" condition) or *jitsu* (indicative of a replete condition). A hard and full abdomen, for example, would be indicative of a *jitsu* condition. If the area around the umbilicus feels soft and spongy like cotton wool, this would be indicative of a "spleen *kyo*" or spleen vacuity.

Practitioners were asked to palpate the five regions of the abdomen and indicate whether any of the five regions was felt to be abnormal and to indicate whether this area was *kyo* or *jitsu*.

Primary and secondary sho. Practitioners were asked to indicate the primary *sho* following the initial collection of clinical data. The four possible primary *sho*s (patterns) are: lung, liver, spleen, and kidney (primary *sho*).* Practitioners were also asked to indicate the likely secondary *sho* (which may be one of either replete or vacuous).

Data collection. A TMT Assessment Form was developed to systematically record the clinical data from the TMT examination. These data included pulse diagnosis (pulse speed, depth, and strength), abdominal diagnosis, and primary and secondary *sho* diagnoses. Information was recorded as categorical variables, and practitioners were required to choose from a limited range of answers.

Study procedure

The practitioners conducted a TMT examination of each participant separately and sequentially, filling in a TMT assessment form that was placed in an envelope and given to a research assistant on completion. The order in which the practitioners conducted the examination was not controlled. Practitioners were instructed not to discuss the diagnosis until both had completed their assessment forms and handed these to the research assistant. Data entry was completed by the research assistant.

Statistical analysis

The κ coefficient was used as a measure of inter-rater reliability of the outcome variables. Since the variables were nominal variables, the weighted κ rather than the simple κ coefficient was used. The κ coefficient provides a measure of

*Although it is possible to diagnoses a Heart *sho*, this is so rare as to be excluded in most cases.

Table 2. Agreement on Pulse Characteristics

Variable	*No. of possible categories (category names)*	*Agreement (%)*	*Weighted κ coefficient (95% CI)*	*Assessment of outcome*[a]
Pulse depth ($n = 58$)	3 (all sinking, normal, all floating)[b]	57	0.37 (0.16–0.57)	Fair
	5 (sinking, little sinking, normal, little floating, floating)	43	0.35 (0.18–0.53)	Fair
Pulse speed ($n = 57$)	3 (all slow, normal, all rapid)[b]	61	0.40 (0.21–0.59)	Fair
	5 (slow, little slow, normal, little rapid, rapid)	56	[c]	
Pulse strength ($n = 61$)	3 (all vacuous, normal, all replete)[b]	77	0.38 (0.11–0.64)	Fair
	5 (vacuous, little vacuous, normal, little replete, replete)	62	0.37 (0.16–0.57)[d]	Fair

[a]According to criterion of Landis and Koch 1977.
[b]Five categories collapsed into three broader categories.
[c]Unequal number of categories, κ unable to be calculated.
[d]Neither practitioner recorded "replete" category, κ calculated using four categories.
CI, confidence interval.

the level of agreement between 2 or more observers beyond that expected by chance. The interpretation of κ values of Landis and Koch[8] was adopted in this study: When $\kappa < 0.0$, level of agreement is "poor"; κ 0.0–0.20 ("slight"); κ 0.21–0.40 ("fair"); κ 0.41–0.60 ("moderate"); κ 0.61–0.80 ("substantial") and κ 0.81–1.00 ("almost perfect"). Level of agreement on each particular outcome variable was measured only for those cases (participants) in which data were recorded by both practitioners.

Results

Pulse diagnosis

There were missing data for pulse depth, speed, and strength (data missing for $n = 4$, 5, and 1 participants, respectively). For each pulse dimension (depth, speed, strength), the five possible choice categories were collapsed into three categories for analysis. For example, for pulse depth the categories "floating" and "little floating" were combined to form the category "floating" and the "sinking" and "little sinking" categories were combined to form the category "sinking." Results are shown in Table 2.

Abdominal diagnosis

In this study, we were interested to know whether 1 practitioner detected the involvement of a particular region of the abdomen, whether the other practitioner agreed for that region of the abdomen. Level of agreement on involvement was therefore calculated for each of the abdomen regions separately, and the κ coefficient calculated in each instance. Results are shown in Table 3.

In the majority of cases where the practitioners agreed on involvement of a specific abdominal region, the diagnoses were that of *kyo* (all cases for kidney and lung regions, 22/29 cases for the spleen region, and 6/7 for the liver region).

Sho *diagnosis*

The overall level of agreement on primary *sho* diagnosis was 47% (weighted κ was 0.25; 95% confidence interval [CI] [0.04–0.45]). The level of agreement on the secondary *sho* was 44% (weighted κ was 0.29, 95% CI [0.11–0.48]). Practitioners more frequently agreed on a primary *sho* diagnosis of kidney *sho* ($n = 13$ cases) followed by lung *sho* ($n = 10$ cases), spleen *sho* ($n = 4$ cases), and liver *sho* ($n = 2$ cases). Practitioners more frequently agreed on a secondary *sho* diagnosis of spleen *sho* ($n = 12$ cases) followed by none (no sho) ($n = 9$ cases), liver *sho* ($n = 4$ cases), kidney *sho* ($n = 2$ cases), and lung *sho* ($n = 1$ case).

Discussion

This study investigated the reliability of TMT pulse diagnosis, abdominal diagnosis, and primary and secondary *sho* diagnoses between 2 TMT practitioners. Overall, there was a reasonable level of agreement on basic pulse characteristics and on abdominal diagnosis. Level of agreement on primary and secondary *sho* diagnosis suggests room for improvement.

Our results found a reasonable level of agreement between practitioners on pulse characteristics: 57%, 61%, and 77% agreement on depth, speed, and strength, respectively. Using the interpretation of κ coefficients of Landis and Koch,[8] this would be considered a "fair" level of agreement. There have been two studies of inter-rater reliability of pulse diagnosis within TMT. Birch assessed the level of agreement among 5 TMT practitioners who examined a total of 9 subjects (Birch S, 1997, unpublished data). He found that for pulse depth and rate, correlation ranged from no correlation to substantial correlation (mean Spearman rank correlation coefficients 0.43 and 0.38, respectively). For pulse strength, results ranged from no correlation to almost perfect correlation (mean Spearman rank correlation coefficient 0.45). In another study, Birch investigated inter-rater reliability between 2 TMT practitioners examining 43 subjects (Birch S, 1999, unpublished data). Agreement was significant for pulse rate and pulse strength, but not pulse depth. A small number of previous studies have assessed the reliability of pulse diagnosis (Birch S, 1997, Birch S 1999, Craddock DS 1997, O'Brien KA, 2006, unpublished data),[9–11] the majority

Table 3. Agreement on Abdominal Diagnosis

Abdominal region	*Proportion of cases where practitioners agreed that abdominal region involved (%)*[a]	*Proportion of cases where practitioners agreed that the region was not involved*	*Simple κ coefficient (95% CI)*[b]	*Assessment of outcome*[c]
Lung	58%	6.5%	0.08 (−0.15–0.31)	Slight
Spleen	48%	11%	0.07 (−0.19–0.32)	Slight
Liver	11%	45%	0.02 (−0.22– 0.27)	Slight
Kidney	53%	8%	0.07 (−0.16–0.30)	Slight
Heart	0%	95%	NA	NA

[a]Based on $n = 62$ study participants.
[b]κ calculated for agreement on the presence of involvement of an area.
[c]According to criterion of Landis and Koch.[10]
CI, confidence interval; NA, not applicable, κ unable to be calculated: Heart region not diagnosed by either practitioner.

within the field of TCM. These studies have found a variable degree of reliability and are summarized elsewhere.[7]

Our study found that according to Landis and Koch's interpretation of κ coefficients,[8] the level of agreement on the four main abdominal regions (lung, liver, kidney, spleen) was "slight," despite the fact that for the lung and kidney regions, there was a reasonable level of agreement (58% and 53%, respectively). There are no other studies that have assessed inter-rater reliability of abdominal diagnosis. One study by Matsumoto assessed the test–retest reliability of abdominal diagnosis.[5] In three separate studies, 6, 4, and 6 blindfolded observers completed abdominal diagnosis in 8, 12, and 10 subjects, respectively, with each subject being examined twice by each observer in random order. The percentage of repeated findings (test–retest) ranged from 25% to 90% across the 16 observers (average 61%). Level of agreement between the 2 highest scoring observers was found to range from 25% to 70%. No formal statistical analyses were conducted. Matsumoto concluded that abdominal palpation appeared to be a reliable diagnostic procedure.[5]

There have been few studies within Japanese meridian therapy that have investigated the test–retest reliability of diagnosis of pattern diagnosis (Birch S, 1997, Birch S, 1999, unpublished data)[3,4] and the inter-rater reliability of pattern diagnosis (Birch S, 1997, Birch S, 1999, unpublished data).[6] Ogawa's study found that in 3 of the 4 subjects examined, 3 of 5 meridian therapy practitioners agreed on the *sho* (four possible choices).[6] In a study by Birch of 5 meridian therapy practitioners who examined 9 subjects, the extent of agreement between practitioners was significant ($W = 0.29$, $p = 0.03$) [Birch S, 1997, unpublished data]. In a follow-up study that assessed level of agreement between 2 meridian therapy practitioners assessing 43 subjects, again the extent of agreement was significant ($W = 0.18$, $p = 0.01$) (Birch S, 1999, unpublished data). The majority of inter-rater reliability studies of pattern diagnosis have been within TCM, with variable results,[12–21] a summary of which is presented elsewhere.[7] Our study found 48% agreement on primary *sho* and 44% agreement on secondary *sho* between 2 practitioners, a "fair" level of agreement between practitioners according to the definition of Landis and Koch.[8] However, caution should be taken in trying to compare findings across studies since studies vary on what they are measuring agreement on and the method of analyses may differ.

Limitations of the study

With respect to the secondary *sho*, it should be noted that generally the secondary *sho* is confirmed after needling the two acupoints indicated for treating the primary *sho*. In our study, practitioners were asked to indicate the secondary *sho* at the same time as the primary *sho*. It is possible that if practitioners had been allowed to treat the primary *sho*, the secondary *sho* diagnosis and level of agreement may have been different.

Other limitations include those inherent in the use of the κ coefficient as a measure of reliability that is known to vary with prevalence[22] and is influenced by the number of possible response categories.[23] Therefore, as pointed out by Joshua,[24] caution needs to be taken in making comparisons between studies. The fact that both practitioners conducted their examinations in the same room directly after one another minimized time of examination and lighting as potential factors that may have influenced diagnosis and contributed to variability in findings. Both practitioners had similar schooling in TMT, though 1 practitioner is slightly more experienced than the other, a factor that may have contributed to variability.

In some ways, since the range of diagnoses in TMT is limited (for example, typically four primary *sho* choices), this should serve to favor better levels of agreement in comparison to studies that have investigated reliability of TCM pattern diagnosis where diagnostic choice has been allowed to be open-ended (as it has been in some studies).

Finally, we conducted our study on relatively "healthy" persons with no biomedically defined disorder or disease in common. It is possible that if we had conducted the study in a group of participants with a common biomedical disorder or a disorder, disease, or symptom that is well-characterized in terms of the understanding of pathogenesis and diagnosis within the TEAM literature, there might have been a greater level of agreement between practitioners. The results of the study therefore should be interpreted with this in mind. The *sho* describes a relative imbalance in the body and locates it in a particular meridian, and the pulse and abdominal diagnoses are simply means for detecting (often subtle) changes in the body. Even though the study recruited those patients who would be considered relatively "healthy" within a biomedical model, subtle subclinical imbalances may still occur

in the body. This is one of the positive attributes of systems of TEAM in general, that is, to be able to detect subtle changes within the body that could progress to more serious disorders at a later time. Health is defined differently within TEAM systems in comparison to the biomedical model. The notion of internal harmony of subtle forces of *yin* and *yang* and the concept of a subtle kind of energy that we term *ki* in meridian therapy or *qi* in TCM are central to defining health in many systems of TEAM. Those deemed healthy within a biomedical framework of health may not necessarily be perfectly healthy within another framework.

Ways forward

Changes in diagnostic variables such as pulse characteristics and patterns of disharmony are used by TEAM practitioners as evidence of change in a patient. There is an argument, therefore, that TEAM variables such as pulse diagnosis or pattern of disharmony should be incorporated into clinical trial design in addition to biomedical variables. However, this can only be justified if such diagnostic variables are reliable in the first place.

Unlike in TCM, where methods of pulse assessment and teaching can vary widely, in TMT teaching of pulse diagnosis is more standardized. Specialized hands-on methods have been developed to teach pulse diagnosis in TMT. Among these are ways of using palpation of the pulses to observe changes as they occur as a feedback tool. This method, the Kozato method, is a teaching tool that helps train pulse diagnosis and needling skills. TMT also uses pulse diagnosis after each point is needled explicitly for feedback as an assessment of whether the needling was sufficient or not. Together, these make pulse diagnosis more important in TMT practice than, for example, in TCM. Thus, further efforts need to be made to reinforce how agreement can be reached in pulse diagnosis skills. While there were "reasonable" levels of agreement between the 2 practitioners on pulse diagnosis, this was achieved without any specific pre-study training of the practitioners. Two (2) recent studies within the TCM literature have demonstrated substantial increases in level of agreement between practitioners in diagnosis with incorporation of training sessions in which practitioners train in the diagnostic technique, examine patients together, and discuss the diagnoses.[17,22] It is expected that with such additional training, the levels of agreement can rise for pulse diagnosis, abdominal diagnosis, and overall *sho* or pattern selection.

Conclusions

This is the first study to systematically and comprehensively investigate the reliability of diagnosis in TMT. Although results suggest a reasonable level of agreement on pulse characteristics and abdominal diagnosis, the level of agreement on primary and secondary *sho* diagnoses suggests there is room for improvement. More studies need to be completed in order to gain a greater understanding of the reliability of diagnosis in TMT.

Acknowledgments

We gratefully acknowledge the support of Helio Supply Ltd., Sydney, Australia for supply of acupuncture needles used in this study. This study was supported by two research grants from Victoria University: a School of Health Sciences grant and a VU Researcher Development Grant.

Disclosure Statement

No competing financial interests exist.

References

1. Fukushima K. Meridian Therapy. Tokyo: Toyo Hari Medical Association, 1991.
2. Fukushima K. Meridian Therapy: A Hands-On Text on Traditional Japanese *Hari* Based on Pulse Diagnosis. Tokyo: Toyo Hari Medical Association, 1991:107.
3. Debata A. Experimental study on pulse diagnosis of *rokubujoi*. Jpn Acup Mox J 1968;17:9–12.
4. Kurosu Y. Experimental study on the pulse diagnosis of *rokubujoi* 11. Jpn Acup Moxib J 1969;18,3:26–30.
5. Matsumoto T. Experimental study on *fukushin* (abdominal palpation). Jpn Acup Mox J 1968;17:13–16.
6. Ogawa T. To establish new "Chinese medicine": Searching for the contemporary significance of the "meridian controversy" [in Japanese]. Chin Med 1978;1:151–158.
7. O'Brien KA, Birch S. A review of the reliability of Traditional East Asian Medicine diagnoses. J Altern Complement Med 2009;15:353–366.
8. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159–174.
9. Dang T, Zaslawski CJ. A pilot study of the diagnostic reasoning processes used by practitioners of traditional Chinese medicine. The Fourth Australasian Acupuncture and Chinese Medicine conference, Victoria University of Technology, Melbourne, Australia, 1998. In: Zaslawski C. Clinical reasoning in traditional Chinese medicine: Implications for clinical research. Clin Acup Oriental Med 2003;4: 94–101.
10. King E, Cobbin D, Walsh S, Ryan D. The reliable measurement of radial pulse characteristics. Acup Med 2002;20:150–159.
11. Walsh S, Cobbin D, Bateman K, Zaslawski C. Feeling the pulse: Trial to assess agreement level among TCM students when identifying basic pulse characteristics. Eur J Oriental Med 2001;3:25–31.
12. O'Brien KA, Abbas E, Zhang J, et al. An investigation of the reliability of Chinese medicine diagnosis according to Eight Guiding Principles and *Zang-Fu* Theory in hypercholesterolaemic Australians. J Altern Comp Med 2009;15:259–266.
13. Coeytaux RR, Chen W, Lindemuth CE, et al. Variability in the diagnosis and point selection for persons with frequent headache by traditional Chinese acupuncture. J Alt Complement Med 2006;12:863–872.
14. Hogeboom CJ, Sherman KJ, Cherkin DC. Variation in diagnosis and treatment of chronic low back pain by traditional Chinese medicine acupuncturists. Complement Ther Med 2001;23:153–155.
15. Kalauokalani D, Sherman KJ, Cherkin DC. Acupuncture for chronic lower back pain: Diagnosis and treatment patterns among acupuncturists evaluating the same patient. South Med J 2001;94:486–492.
16. Sung JJY, Leung WK, Ching JYL, et al. Agreements among traditional Chinese medicine practitioners in the diagnosis and treatment of irritable bowel syndrome. Aliment Pharmacol Ther 2004;20:1205–1210.
17. MacPherson H, Thorpe L, Thomas K, Campbell M. Acupuncture for lower back pain: Diagnosis and treatment of

148 patients in a clinical trial. Complement Ther Med 2003;12:38–44.
18. Zell B, Hirata J, Marcus A, et al. Diagnosis of symptomatic postmenopausal women by traditional Chinese medicine practitioners. Menopause 2000;7:129–134.
19. Zhang GG, Bausell B, Lao L, et al. The variability of TCM pattern diagnosis and herbal prescription on rheumatoid arthritis patients. Altern Ther Health Med 2004;10:58–63.
20. Zhang G, Lee WL, Bausell B, et al. Variability in the traditional Chinese medicine (TCM) diagnoses and herbal prescriptions provided by three TCM practitioners for 40 patients with rheumatoid arthritis. J Altern Complement Med 2005;11:415–421.
21. Zhang GG, Singh B, Lee WL, et al. Improvement of agreement in TCM diagnosis among TCM practitioners for persons with the conventional diagnosis of rheumatoid arthritis: Effect of training. J Altern Complement Med 2008;14:381–386.
22. Hansen M, Sindrup SH, Christensen PB, et al. Interobserver variation in the evaluation of neurological signs: Observer dependent factors. Acta Neurol Scand 1994;90:145–149.
23. Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall, 1996.
24. Joshua AM, Celermajer DS, Stockler MR. Beauty is in the eye of the examiner: Reaching agreement about physical signs and their value. Intern Med J 2005;35:178–187.

Address correspondence to:
Kylie Ann O'Brien, Ph.D.
Faculty of Health, Engineering and Science
Victoria University
McKechnie Street
St. Albans, Victoria 3021
Australia

E-mail: kylie.obrien@vu.edu.au