

Development of an Instrument to Assess the Quality of Acupuncture: Results from a Delphi Process

Caroline A. Smith, PhD,¹ Christopher J. Zaslowski, PhD,² Zhen Zheng, PhD,³ Deidre Cobbin, PhD,² Suzanne Cochrane, BSocial Work,¹ George B. Lenon, PhD,³ Bertrand Loyeung, BHealth Sc (TCM),² Peter C. Meier, PhD,² Sean Walsh, PhD,² Charlie Changli Xue, PhD,³ Anthony L. Zhang, PhD,³ Xiaoshu Zhu, PhD,¹ and Alan Bensoussan, PhD¹

Abstract

Background: Quality acupuncture influences the outcomes of clinical research, and issues associated with effective administration of acupuncture in randomized controlled trials need to be addressed when appraising studies.

Objective: The study objective was to achieve consensus on domains and items for inclusion in a rating scale to assess quality acupuncture administered in clinical research.

Study design and subjects: An active group of Australian acupuncture researchers initially identified a pool of items assessing quality. The Delphi consensus process was then used to select and reduce the number of items, and an additional expert panel of 42 researchers were invited to participate. Participants initially ranked items along a five-point scale for the first Delphi round, and indicated an agree or disagree response during the second round. For an item to be retained into the second round, an item had to attain greater than 80% agreement that the item described a dimension of quality acupuncture and related study design.

Results: Thirty-two (32) experts agreed to participate in the study. After two rounds of the Delphi process, consensus was reached on 14 domains and 26 items relating to quality acupuncture. Domains, items, and minimum standards related to study design; rationale of the intervention; criteria relating to needling stimulation either manual or electrostimulation; duration and frequency of treatment; and practitioner training.

Conclusions: Items for inclusion in an instrument to assess quality acupuncture in clinical research were identified. Further development of the instrument including relative weighting of items and reliability testing is under way.

Introduction

MOST EVIDENCE PYRAMIDS consider systematic reviews as the best evidence to determine the efficacy, effectiveness and safety of any intervention including acupuncture. Variation in the quality of acupuncture administered can influence the outcomes of a clinical trial, and thus result in variability of the estimates of treatment effects reported in both clinical trials and systematic reviews. An assessment of quality acupuncture relies on full reporting; however, with a complex intervention, reporting of the intervention can be superficial. To improve the quality of reporting of clinical trials, there have been several developments over the last 15 years, beginning with the CONSORT (Consolidated Stan-

dards of Reporting Trials) Statement initially published in 1996.¹ CONSORT was subsequently extended to non-pharmacological treatments (including complementary and alternative medicine and therapies) in 2008,² and revised and published in 2010.³ The STRICTA checklist (Standards for Reporting Interventions in Controlled Trials of Acupuncture), which aims to improve the reporting of acupuncture trials, was first published in 2001.⁴ This checklist was to encourage reporting of intervention details deemed to be important for critical analysis and replication of such trials. In 2010, the revised STRICTA checklist was published as an extension of CONSORT, and includes 6 items and 17 subitems.⁵

With improved reporting of clinical trials, methodological assessment of trials can be more consistently applied, and

¹Centre for Complementary Medicine Research, The University of Western Sydney, Penrith South, Australia.

²UTS College of Traditional Chinese Medicine, The University of Technology Sydney, Australia.

³Health Innovations Research Institute and Discipline of Chinese Medicine, School of Health Sciences, RMIT University, Melbourne, Australia.

this has been aided by the development of rating scales to assess the quality of trials. Critical appraisal tools provide an analytical evaluation of the quality of the study, in particular to the criteria that minimize bias.⁶ Although there are a large number of critical appraisal instruments, over 93 have been identified as of 2002,⁷ many lack rigorous development, and there is little consensus regarding the most appropriate items that should be included in any critical appraisal tool.

Acupuncture practice has evolved over time and in response to different cultural contexts, and today represents a broad range of styles of practice. Acupuncturists commonly use several stimulation techniques including needling, cupping, herbs, *tuina*, and acupressure. Indeed, acupuncture is a complex intervention. Practice styles vary, and the skill and expertise of the practitioner may influence the outcome of treatment. To increase transparency and the reproducibility of acupuncture performed in clinical trials, many acupuncture interventions use standardized or semistandardized treatments, and defined treatment parameters. Reporting of acupuncture parameters has improved in response to the STRICTA checklist,⁸ and is likely to further improve in response to the revised STRICTA checklist.⁵

There are now clear standards for reporting acupuncture parameters, and the time is right to address the related but separate issue of quality acupuncture. Reporting quality may improve the reliability and replicability of a study, but does not address the quality of intervention applied (i.e., validity, standards, or adequacy). An assessment of quality can be based on conceptual and operationalized definitions of what quality means and how to measure it. Measurement may involve the development of standards, generally derived from two sources.⁸ Empirical standards can be derived from actual practice and are used to compare care in one setting with that in another setting. Alternatively, normative standards are derived from sources that set standards of knowledge (for example by standard textbooks, publications, panels of practitioners or research staff in consultations with practitioners⁹).

How to assess quality acupuncture administered in trials or systematic reviews remains problematic. Attempts have been made by researchers to address the need for an instrument to assess quality acupuncture. Preliminary developmental work undertaken by White and Ernst¹⁰ stalled due to the unreliability of the method with inconsistent scores. Quality acupuncture was also considered by Dutch epidemiologists¹¹⁻¹³ undertaking acupuncture systematic reviews. More recently, White and colleagues discussed the need to define an adequate dose of acupuncture.¹⁴ The authors examined 47 systematic reviews (2000-2007), and found only six reviews that considered the question of adequacy or quality acupuncture. Although there is some evidence of an assessment of quality acupuncture, there has been no rigorous development of an instrument involving external views of stakeholders through a consensus method to define the inclusion of items describing quality acupuncture.

In 2008, funding was received by the National Institute of Complementary Medicine (NICM) to establish a network of acupuncture researchers in Australia, and for the network to undertake the development of an instrument to assess quality acupuncture administered in clinical research. This article reports on the initial development of the instrument and the ensuing Delphi process to achieve a consensus on

defining the domains and items of quality acupuncture administered within a clinical trial context. A future publication will present the final instrument and explanatory notes including the relative weighting of items.

Materials and Methods

General methods used in previous research were adapted to guide the instrument development.¹⁵ This process consisted of three objectives: defining the scope and purpose of the instrument, generating items, and achieving a predefined level of agreement for inclusion of items.

The 13 Australian acupuncture researchers formed the NICMAN (National Institute for Complementary Medicine Acupuncture Network) group to commence development of the proposed rating scale. The NICMAN group consisted of three Australian tertiary education institutions, the University of Technology Sydney, the University of Western Sydney, the Royal Melbourne Institute of Technology University, and one research institute, the National Institute for Complementary Medicine (NICM).

The goal of the NICMAN group was to determine the aim of the instrument, generate items for inclusion, provide rationale and evidence for item generation, and propose and invite international participants to contribute to the Delphi consensus process. Two (2) members, CS and CZ, were responsible for establishing criteria for the inclusion and exclusion of items during the Delphi process, and for the analysis and ensuing discussion of the additional comments received from participants, as well as providing feedback to the NICMAN group and Delphi participants.

Overall aim of the instrument

Our project aim was to develop an instrument to assess quality acupuncture administered in experimental (randomized controlled trials [RCTs], crossover studies, clinical controlled designs) rather than descriptive studies (qualitative or cross-section studies such as surveys). The instrument was designed to be used for appraising quality acupuncture and related study design items, and for use in conjunction with the established scales that have been developed to assess the quality of RCTs such as the Cochrane Collaboration tool for assessing risk of bias.¹⁶ The focus was on identifying critical features of acupuncture that were deemed to be necessary in determining quality acupuncture when administered within a clinical trial context. At the same time, the NICMAN group was acutely aware that acupuncture as delivered in a nonresearch clinical setting is a complex intervention involving not only technical needling skill but development of a therapeutic relationship, formulation of a diagnosis, provision of lifestyle advice, and often administering co-interventions such as *gua sha* (scrapping), *tuina* (massage), moxibustion, or electrical stimulation.

Item generation

The initial pool of items was identified during three separate day long meetings. Prior to the first meeting, all NICMAN members were asked to identify and define what criteria were deemed necessary for determining quality acupuncture within the confines of an acupuncture clinical trial. At the first meeting, all members brought to the table

relevant criteria and items they identified as relating to quality acupuncture. Further discussion and clarification occurred at the second and third meetings where literature was used by the NICMAN group to support or refute the inclusion of particular criteria in the instrument. The remaining features were then categorized under the following headings or domains: justification of study design; diagnosis; needling; practitioner training; therapeutic alliance, and treatment of the *spirit* of the acupuncture recipient encompassing their consciousness, emotions, and thought. Following completion of the third meeting, only those domains that were supported by experimental evidence were retained. These domains were developed as a normative standard and examples given. The domains were then finalized into a document and grouped into 20 domains with each item consisting of one or more statements and a total of 40 items. Accompanying the document were a set of explanatory notes that were disseminated to the participants during the Delphi process to clarify specific statements and to assist decision-making.

Selection of participants

A total of 42 international and national experts (including the NICMAN group) were invited by e-mail to participate in the Delphi method. The participants were acupuncture researchers and practitioners involved in the area of research methodology or clinical trials. Participants had a broad range of experience in planning and undertaking acupuncture clinical research, including RCTs, as well as undertaking critical appraisal of RCTs for inclusion into systematic reviews.

Selection of items

The Delphi method is based on a structured process for collecting and distilling knowledge from a group of experts by means of a series of questionnaires interspersed with controlled opinion feedback.¹⁷ An invitation to participate was distributed in an e-mail containing a link to an online survey hosted at www.surveymonkey.com. Notes were attached with an explanation and example of the items for consideration. Two (2) electronic reminders were sent. During the first round, survey participants individually ranked along a five-point scale ranging from strongly disagree to strongly agree their opinion of each item defined as a domain of quality acupuncture. Participants had an opportunity to comment on their responses, or to provide additional remarks.

For an item to be retained into the second round, more than 80% of participants had to agree it described a dimension of quality acupuncture. Following analysis of the first round, a report was circulated to participants detailing the results of the first Delphi round and they were invited to participate in round 2. For the second round, participants were asked to select either an "agree" or "disagree" response rather than rank along a five-point scale as was done in the first round for each item. An agreement of 80% or more was required to achieve consensus for the retention of the item. Two (2) rounds only were conducted to achieve consensus.

On completion of round 2 of the Delphi, all 32 invitees were asked to report their sex and their role as either an

acupuncture practitioner, researcher, educator, administrator, or nonacupuncture researcher. Those who identified as acupuncture practitioners were also quizzed as to the style of acupuncture they practiced, having a choice of Traditional Chinese Medicine, Japanese, Korean, medical, five element, or other. For the questions relating to professional role and style of acupuncture practiced, participants could respond with multiple responses.

Analysis

Data were collated by www.surveymonkey.com, and responses were downloaded both as summaries and detailed comments. The responses to each item were reported as the proportion of participants. CS and CZ met to discuss the qualitative and quantitative answers after round 1, and circulated the results of rounds 1 and 2, with proposed changes to the NICMAN group.

Prior to commencing the study, ethical approval was obtained from the Human Research Ethics Committee at the University of Technology, Sydney (reference number 2009-343A).

Results

Of the 42 experts invited to participate, 32 (76%) agreed, and those who consented to being identified are listed in the acknowledgments. Figure 1 shows the Delphi process. Of the 32 who agreed, 28 completed round 1 and a similar number completed round 2 of the Delphi process. The reasons why four experts did not participate in either round is not known. Due to the anonymity of the research it could not be determined whether the same 28 individuals who completed round one also completed round two. The characteristics of 32 invited Delphi participants are described in Table 1.

First round

Items were retained if the item scored 80% or higher when the "strongly agree" and "agree" categories were combined. Where this was not the case, but 20% or more chose "neither agree or disagree," the participants' comments were used to clarify the wording and rephrase the statement.

Table 2 shows the response percentages and the number of responses for the 40 items. Eight items were deleted (19, 24, 25, 31, 32, 33, 34, 36), and seven items modified slightly in response to comments. The following changes were made for the second Delphi round.

Domain 4: Statement 7. Despite obtaining 100% agreement, statement 7 was rephrased as several participants commented that the issue of the control comparator did not relate to the delivery of the acupuncture intervention. The question was reworded as "the acupuncture intervention is designed to address the research question," deleting any inference to the control intervention. The item descriptor was also changed to "Acupuncture intervention design" rather than "Treatment and Control group design."

Domain 6: Statement 10. An alternative phrase was inserted in statement 10, and the heading was reworded as "Acupuncture points needled are consistent with" rather

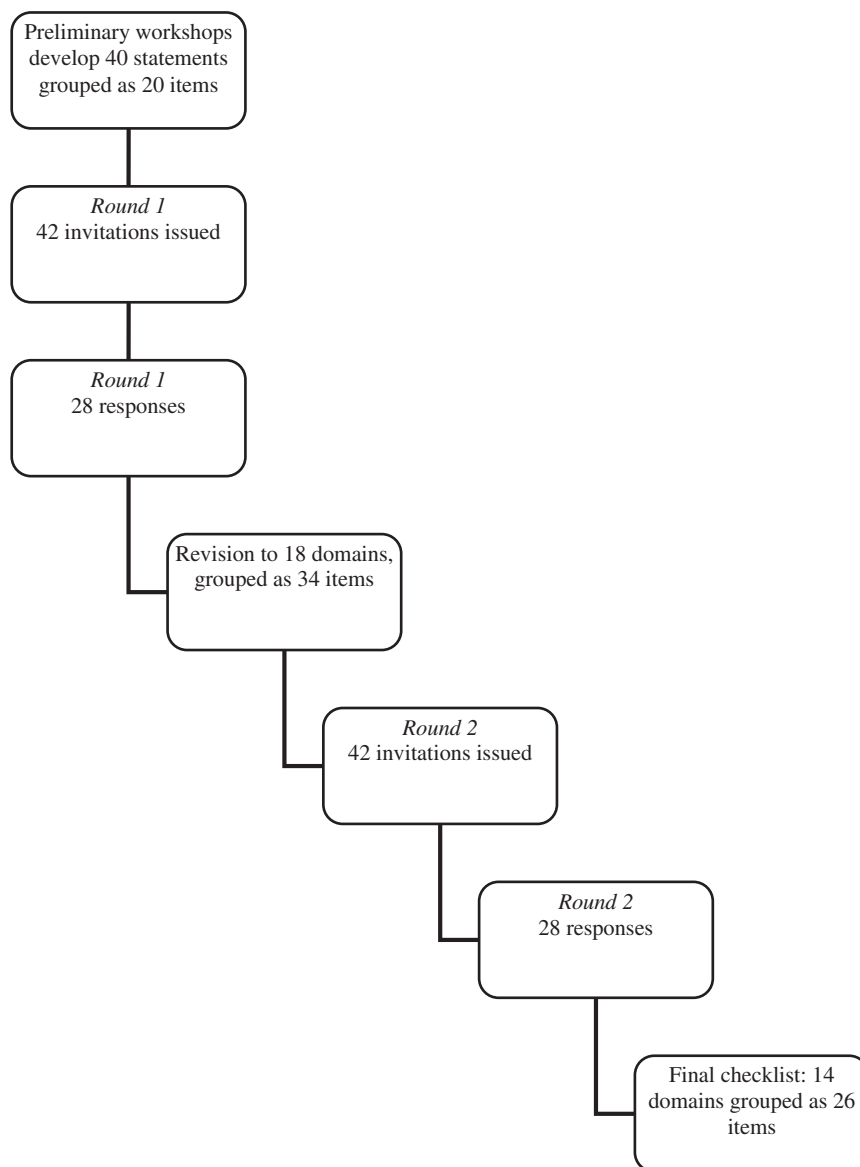


FIG. 1. Flow chart of the Delphi process.

TABLE 1. PROFILE OF PANEL OF EXPERTS INVITED TO PARTICIPATE IN THE DELPHI PROCESS

Characteristic	N	%
Sex		
Female	10	31.2
Male	22	68.8
Profession (multiple responses allowed)		
Acupuncture practitioner	24	
Acupuncture researcher	22	
Acupuncture educator	26	
Administrator	6	
Researcher (nonacupuncturist)	5	
Style of acupuncture practiced (multiple responses allowed)		
TCM	23	
Japanese	5	
Medical	5	
Five element	6	
Other (e.g., <i>Toyo Hari</i> , sports medicine)	3	

TCM, Traditional Chinese Medicine.

than “Acupuncture points are selected according to.” Several participants commented that the development of the acupuncture point prescription for a trial could also be based on acupuncture points that have been used previously in a successful published study. The phrase “Acupuncture points needed are consistent with” was inserted as a statement.

Domain 7: Statement 12. This statement related to the number of sites to be needed proved to be contentious, with 61% agreeing to the retention of the statement, with 18% neither agreeing or disagreeing and another 21% selecting either the disagree or strongly disagree category. The wording was subsequently changed for the second round to include an example of the use of the single acupoint Pericardium 6 (*Nei Guan*), which has been successfully used as the sole point in several previous acupuncture studies for nausea.¹⁸

Domain 8: Statement 13. This statement related to consistent use of a branded needle, gauge and scored 72% for

TABLE 2. RESULTS FROM DELPHI ROUND 1

Domain	Statement no.	Item	Strongly agree/ agree % (no.)	Neither agree or disagree % (no.)	Disagree/ strongly disagree % (no.)
1	1	The research question of the study needs to be clearly described in terms of population	100 (28)	0	0
	2	The research question of the study needs to be clearly described in terms of intervention	100 (28)	0	0
	3	The research question of the study needs to be clearly described in terms of comparator	92.8 (26)	3.6 (1)	3.6 (1)
	4	The research question of the study needs to be clearly described in terms of outcome	96.4 (27)	0	3.6 (1)
2	5	The study design is appropriate for the research question	100 (28)	0	0
3	6	The active intervention is justified by a description of the diagnosis and treatment as per the stated acupuncture paradigm	96.4 (27)	3.6 (1)	0
4	7	The acupuncture and control interventions are appropriately designed to address the research question	100 (28)	0	0
5	8	Justification of the diagnostic process is provided, by evidence linking to the treatment paradigm	96.3 (26)	3.7 (1)	0
	9	Justification of the diagnostic process is provided by evidence linking to the treatment paradigm clinical reasoning	88.5 (23)	11.5 (3)	0
6	10	Acupuncture points are selected according to differential diagnosis	92.6 (25)	7.4 (2)	0
	11	Acupuncture points are selected according to treatment paradigm	96.4 (27)	3.6 (1)	0
7	12	Preference is given to a minimum of six needle sites per treatment. If fewer points are used, justification is presented linking to the treatment paradigm	60.7 (17)	17.9 (5)	21.4 (6)
8	13	Needle brand and gauge is used <i>consistently</i> across all patients	71.5 (20)	14.3 (4)	14.2 (4)
9	14	Point location: Standard acupuncture location texts are used as reference	85.2 (23)	7.4 (2)	7.4 (2)
	15	Point location: Location described in anatomical terms	71.5 (20)	10.7 (3)	17.9 (5)
	16	Point location: An accurate proportional method for locating acupoints used where appropriate	74.0 (20)	18.5 (5)	7.4 (2)
10	17	Appropriate symmetrical or asymmetrical needling sites are <i>relevant</i> to the clinical condition	81.5 (22)	3.7 (1)	7.4 (2)
11	18	Depth of needle insertion expressed in measurement units	75.0 (21)	7.1 (2)	17.9 (5)
12	19	Direction of the needle insertion expressed in degrees and anatomical terms	74.0 (20)	11.1 (3)	14.8 (4)
13	20	Needle retention: Minimum time of 20 minutes unless justification presented reflecting standard scope for different style of acupuncture paradigm	78.5 (22)	10.7 (3)	10.7 (3)
14	21	Number of treatments: If a chronic condition a minimum of six treatments are administered, if fewer treatments are delivered appropriate justification is documented	85.2 (23)	11.1 (3)	3.7 (1)
	22	Number of treatments: If an acute or subacute condition no minimum of treatments are specified, but appropriate justification is to be provided	85.2 (23)	11.1 (3)	3.7 (1)
15	23	Frequency of treatment: Initially a minimum of two treatments per week are administered unless justified by treatment parameters relating to the treatment paradigm	66.6 (18)	18.5 (5)	14.8 (4)

(continued)

TABLE 2. (CONTINUED)

Domain	Statement no.	Item	Strongly agree/ agree % (no.)	Neither agree or disagree % (no.)	Disagree/ strongly disagree % (no.)
16	24	Needling sensation: Assessment of the needling sensation of at least one principal acupuncture point in the study protocol at specified time periods at the start of the treatment, and in the middle of the allocated treatment sessions by 1. verbal report by the subject, or 2. the use of a validated scale	66.6 (18)	11.1 (3)	22.2 (6)
	25	In the absence of <i>de qi</i> , justification is provided of the decision not to obtain <i>de qi</i>	77.7 (21)	7.4 (2)	14.8 (4)
17	26	Needle manipulation must be standardized and/or applied at least once during the treatment session. Manipulation should be expressed in terms of the rotational degrees or depth of lift and thrust (mm) of the needle	59.2 (16)	25.9 (7)	14.8 (4)
	27	Needle manipulation must be standardized and/or applied at least once during the treatment session. Manipulation should be expressed in terms of the duration (seconds) of needle manipulation	67.9 (1)	21.4 (6)	10.7 (3)
	28	Needle manipulation must be standardized and/or applied at least once during the treatment session. Manipulation should be expressed in terms of the number of times the needle was manipulated	75 (21)	17.9 (5)	7.2 (2)
	29	In the absence of needle manipulation justification is provided of the decision not to undertake needle manipulation	88.9 (24)	7.4 (2)	3.7 (1)
18	30	Electro-acupuncture machine should demonstrate approval status and compliance for the country where study is being undertaken	85.1 (23)	11.1 (3)	3.7 (1)
19	31	Electrostimulation parameters. The following aspect associated with the electrostimulation at each individual site of needling should be defined and consistently reported. These include the frequency of pulse expressed in terms of hertz	81.5 (24)	18.5 (5)	0
	32	Electrostimulation parameters. The following aspect associated with the electrostimulation at each individual site of needling should be defined and consistently reported. These include the nature of the pulse expressed in terms of continuous or intermittent	77.8 (21)	22.2 (6)	0
	33	Electrostimulation parameters. The following aspect associated with the electrostimulation at each individual site of needling should be defined and consistently reported. These include identification of the paired needle sites stimulated and the pole	70.4 (19)	25.9 (7)	3.7 (1)
	34	Electrostimulation parameters. The following aspect associated with the electrostimulation at each individual site of needling should be defined and consistently reported. These include the length of pulse width expressed in terms of milliseconds	74 (20)	22.2 (6)	3.8 (1)
	35	The level of patient stimulation response that was sought	76.9 (20)	19.2 (5)	3.8 (1)
	36	If available, the mean and range of electrostimulation for the group expressed in terms of milliamps	74 (20)	22.2 (6)	3.7 (1)

(continued)

TABLE 2. (CONTINUED)

Domain	Statement no.	Item	Strongly agree/agree % (no.)	Neither agree or disagree % (no.)	Disagree/strongly disagree % (no.)
20	37	Level of stimulation response is justified and appropriate	74 (20)	22.2 (6)	3.8 (1)
	38	The acupuncturist administering intervention, is registered with a regulatory authority, or meets at least the minimum WHO standard (WHO 1999)	92.9 (26)	7.1 (2)	0
	39	When differential diagnosis is undertaken, evidence is provided that the acupuncturist has undertaken a full training course as per WHO guideline (WHO 1999)	78.6 (22)	17.9 (5)	3.6 (1)
	40	Evidence is provided of prior clinical training by study personnel relevant to the acupuncture intervention and health condition	92.6 (25)	3.7 (1)	3.7 (1)

WHO, World Health Organization.

the agree/strongly agree category; however, 14% also selected disagree/strongly disagree. Several participants commented that it was not relevant when a pragmatic study was undertaken; therefore, the question was rephrased to specifically note that this question "was less appropriate for pragmatic and effectiveness" acupuncture study designs.

Domain 9: Statements 14–16. These statements sought to operationalize the location of the acupoints by reference to a standard text, modern anatomical terms, and using an accurate proportion method for locating acupoints where necessary. A comment was received that acupuncture can be administered at both traditional acupoint sites, *ashi* points, extra meridian points, or painful sites such as trigger points. Questions were collapsed to the one question to include both acupuncture location methods (i.e., "Standard acupuncture location texts are used as reference or location described in anatomical terms").

Domain 10: Statement 17. This statement required researchers to identify the rationale for needling unilaterally or more commonly bilaterally. One (1) participant suggested that it should be revised to say "justification of a symmetrical or asymmetrical needling for each point." The statement was rephrased to "Symmetrical or asymmetrical needling sites are justified according to the clinical condition."

Domain 11: Statement 18. Rather than state the depth of needling, some participants noted that it was better to reference the needling depth of an acupoint site to a standard text and express as a range rather than specific depth. The statement was rephrased to acknowledge these comments.

Domain 13: Statement 20. This statement sought to establish the appropriate duration for needle retention. While 78% of respondents chose agree/strongly agree for this statement, 11% disagreed or selected neither agree nor disagree category. The sentence was rephrased for clarity for the second round.

Domain 19: Statements 31–37. Several participants commented that most of these statements were related to reporting rather than indicative of the acupuncture being administered. Only two statements were retained for round 2, these being the level of sensory stimulation experienced by the trialist and that the stimulus response was justified.

At completion of the first round, the document had been reduced to 18 domains and 34 items.

Second round

Changes to round 2 included a nonapplicable category (for pragmatic designs). Of the 34 statements, 7 were discarded from the final checklist as they failed to obtain greater than 80% agreement (Table 3). The following item statements were not included in the final checklist: statement 13 (relating to the number of acupuncture points needled), statement 20 (time of needle retention), statement 23 (frequency of acupuncture treatments), statements 24 and 25 (needle manipulation), and statements 29 and 30 (aspects of needle manipulation). The final composition includes 14 domains and 26 items and is presented in Table 4.

Discussion

The aim of this study was to identify domains and items describing quality acupuncture, and to reach a consensus with expert acupuncture researchers determining the inclusion of these items within an instrument to assess quality acupuncture within a clinical research context. Two (2) rounds of the Delphi process achieved a consensus on 28 items. The domains relate to study design, rationale of the intervention, specific criteria relating to needling stimulation either manually or using electrostimulation, duration and frequency of treatment, and practitioner training.

The NICMAN Group identified, and Delphi participants agreed on the need for the inclusions of items relating to study design and rationale. Consideration was given to whether these broader items should be retained, and it was

TABLE 3. RESULTS FROM DELPHI ROUND 2

Domain	Statement no.	Item	Agree %	Not applicable	Disagree %
1	1	The research question of the study needs to be clearly described in terms of population	92.9 (26)		7.1 (2)
	2	The research question of the study needs to be clearly described in terms of intervention	100 (28)		0
	3	The research question of the study needs to be clearly described in terms of comparator	92.9(26)		7.1 (2)
	4	The research question of the study needs to be clearly described in terms of outcome	100 (28)		0
2	5	The study design is appropriate for the research question	100 (28)		0
3	6	The active intervention is justified by a description of the diagnosis and treatment as per the stated acupuncture paradigm	96.3 (26)		3.7 (1)
4	7	The acupuncture intervention is designed to address the research question	96.4 (27)		3.6 (1)
5	8	Justification of the diagnostic process is provided, by evidence linking to the treatment paradigm	92.9 (26)		7.1 (2)
	9	Justification of the diagnostic process is provided by evidence linking to clinical reasoning	96.4 (27)		3.6 (1)
6	10	Acupuncture points needled are consistent with differential diagnosis	92.3 (24)		7.7 (2)
	11	Acupuncture points needled are consistent with treatment paradigm	96.3 (26)		3.7 (1)
	12	Acupuncture points needled are consistent based on literature review or other evidence	88.9 (24)		11.1 (3)
7	13	Preference is given to a minimum of six needle sites per treatment. If fewer points (e.g., PC6) are used, justification is presented linking to the treatment paradigm	71.4 (20)		28.6 (8)
8	14	Needle brand and gauge is used <i>consistently</i> across all patients. (Please note that this question is less appropriate for pragmatic and effectiveness studies, select N/A)	60.7 (17)	25 (7)	14.3 (4)
9	15	Point location: Standard acupuncture location texts are used as reference or location described in anatomical terms	100 (25)		0
	16	Point location: An accurate proportional method for locating acupoints used where appropriate	87.5 (21)		12.5 (3)
10	17	Symmetrical or asymmetrical needling sites are justified according to the clinical condition	96.2 (25)		3.8 (1)
11	18	Depth of needle insertion expressed in millimeters is referenced to a standard text	83.3 (20)		16.7 (4)
	19	Depth of needle insertion expressed in millimeters is expressed as a range	87.5 (21)		12.5 (3)
12	20	Needle retention: Minimum time of 20 minutes unless justification presented reflecting a different acupuncture paradigm	78.6 (22)		21.4 (6)
13	21	Number of treatments: If a chronic condition a minimum of six treatments are administered, if fewer treatments are delivered appropriate justification is documented	88.9 (24)		11.1 (3)
	22	Number of treatments: If an acute or subacute condition, no minimum of treatments are specified, but appropriate justification is to be provided	96.3 (26)		3.7 (1)
14	23	A minimum of two treatments per week are administered unless justified by treatment parameters relating to the treatment paradigm	59.3 (16)		40.7 (11)

(continued)

TABLE 3. (CONTINUED)

Domain	Statement no.	Item	Agree %	Not applicable	Disagree %
15	24	Needle manipulation must be standardized and/or applied at least once during the treatment session. Manipulation should be expressed in terms of the rotational degrees or depth of lift and thrust (mm) of the needle	54.9 (14)	22.9 (6)	25.9 (7)
	25	Needle manipulation must be standardized and/or applied at least once during the treatment session. Manipulation should be expressed in terms of the duration (seconds) of needle manipulation	51.9 (14)	18.5 (5)	29.6 (8)
	26	Needle manipulation must be standardized and/or applied at least once during the treatment session. Manipulation should be expressed in terms of the number of times the needle was manipulated	71.4 (20)	17.9 (5)	10.7 (3)
	27	In the absence of needle manipulation, justification is provided of the decision not to undertake needle manipulation	88.9 (24)	7.4 (2)	3.7 (1)
	16	28	Electro-acupuncture machine should demonstrate approval status and compliance for the country where study is being undertaken. ^a	70.4 (19)	22.2 (6)
17	29	The level of patient stimulation response that was sought (e.g., subsensory threshold, suprasensory threshold, above motor stimulation) ^a	73.1 (19)	19.2 (5)	7.7 (2)
	30	Level of stimulation response is justified and appropriate	65.4 (17)	23.1 (6)	11.5 (3)
18	31	The acupuncturist administering intervention is registered with a regulatory authority, or meets at least the minimum WHO standard (WHO 1999)	100 (28)		0
	32	When differential diagnosis is undertaken, evidence is provided that the acupuncturist has undertaken a full training course as per WHO guideline (WHO 1999)	96.3 (26)		3.7 (1)
	33	Evidence is provided of prior clinical training by study personnel relevant to the acupuncture intervention and health condition	92.0 (23)		8.0 (2)
	34	Evidence is provided of monitoring the administration of acupuncture in the clinical trial setting	79.2 (19)		20.8 (5)

^aNA (not applicable) response applied to those unqualified to answer.
WHO, World Health Organization.

agreed that these items provided a context for assessing quality acupuncture, and would not be otherwise captured in critical appraisal tools. Consensus was easier to achieve for domains and items where standards exist, or there was an awareness of sources informing normative standards such as point location, needling sites, needle depth, and practitioner standards regarding training. For items reflecting the diversity of acupuncture styles (for example, the direction of needling, attainment of *de qi*, the number of acupuncture points needled, time of needle retention, type of needle manipulation, frequency of treatment administration, the individualization of treatment regimen within clinical practice, or the selection of points), lower levels of agreement were attained and those items with inadequate consensus were therefore excluded.

The strengths of this study include the comprehensive systematic approach taken to identifying and defining dimensions of quality acupuncture. The items were developed and ranked following evaluation of evidence from published literature and reviews and comments from experts with different backgrounds. A consensus was achieved from experts in the field, who inevitably bring with them different experience and backgrounds reflecting the diversity of acupuncture styles. All have been involved in clinical research, and the majority had first-hand experience of designing study protocols.

There are some limitations to this work to date. The final composition of items following completion of round 2 was significantly reduced from the initial pool first nominated by the NICMAN group, and may not reflect the complexity of

TABLE 4. COMPOSITION OF FINALIZED ITEMS

<i>Domain</i>	<i>Statement no.</i>	<i>Item</i>
1	1	The research question of the study is clearly described in terms of population
	2	The research question of the study is clearly described in terms of intervention
	3	The research question of the study is clearly described in terms of comparator
	4	The research question of the study is clearly described in terms of outcome
2	5	The study design is appropriate for the research question
3	6	The active intervention is justified by a description of the diagnosis and treatment as per the stated acupuncture paradigm
	7	The acupuncture intervention is designed to address the research question
4	8	Justification of the diagnostic process is provided by evidence linking to the treatment paradigm
	9	Justification of the diagnostic process is provided by evidence linking to clinical reasoning
6	10	Acupuncture points needled consistent with differential diagnosis
	11	Acupuncture points needled consistent with treatment paradigm
	12	Acupuncture points needled consistent with literature review or other evidence
7	13	Needle brand and gauge is used <i>consistently</i> across all participants and sessions. (Please note that this question does not apply for pragmatic and effectiveness studies)
8	14	Point location: Published standard acupuncture location texts are used as reference or location described in anatomical terms
	15	Point location: An accurate proportional method for locating acupoints used where appropriate
9	16	Symmetrical or asymmetrical needling sites are justified according to the clinical condition
10	17	Depth of needle insertion expressed in millimeters as a range and is justified or referenced to a standard text
11	18	Number of treatments: If a chronic condition a minimum of six treatments are administered, if fewer treatments are delivered appropriate justification is documented
	19	Number of treatments: If an acute or subacute condition no minimum of treatments are specified, but appropriate justification is to be provided
12	20	Needle manipulation must be standardized and/or applied at least once during the treatment session. Manipulation should be expressed in terms of the number of times the needle was manipulated and applied
	21	In the absence of needle manipulation, justification is provided of the decision not to undertake needle manipulation
13	22	Electro-acupuncture machine should demonstrate approval status and compliance for the country where study is being undertaken
14	23	The acupuncturist administering intervention is registered with a regulatory authority, or meets at least the minimum WHO standard (WHO 1999)
	24	When a traditional diagnosis is undertaken, evidence is provided that the practitioner has undertaken a full training course as per WHO guideline (WHO 1999)
	25	Evidence is provided of prior clinical training by study personnel relevant to the acupuncture intervention and health condition
	26	Evidence is provided of monitoring the administration of acupuncture in the clinical trial setting

acupuncture in its entirety. Several items were eliminated early on due to a lack of evidence to justify their inclusion (e.g., treatment of the *spirit* of the acupuncture recipient encompassing their consciousness, emotions, and thought, and, recognition of a therapeutic alliance as a valuable aspect of clinical interaction). Additional domains and items were excluded during the Delphi rounds and may reflect the diversity of styles of acupuncture practiced by the expert group, and the differing emphasis given to items such as needling parameters. Comments from participants indicate that the idea of the item being an important component of quality was not disputed but rather the assignment of a normative standard specifying the number of treatments, or the number of acupuncture points was more contentious. Some questioned whether it was appropriate to assign a normative standard separated from the complex factors such as the practitioners' assessment of the individual's response to needle manipulation (e.g., pulses). Our desire to avoid subjectivity has resulted in the exclusion of items that do not reflect how this assessment influences decisions about how often and intensively the needle maybe manipulated.

Although the final set of domains and items achieved by consensus is narrower in scope, the proposed instrument may have a greater application among acupuncturists and researchers as a result of greater external validity achieved by consensus from the diverse expert group. The intention is for the instrument to undergo revisions as new evidence emerges and normative standards are agreed upon. Other limitations could relate to the participation of experts. It is recognized that there may be other styles of acupuncture practiced not represented within the invited group of experts in the Delphi group; however, the views of a diverse group of participants from the NICMAM and Delphi responders were able to be synthesized. Another possible criticism is subjectivity with the retained items. To reduce this, participants with varied backgrounds were invited, and a high threshold of 80% agreement for the retention of items was set. Finally, this instrument has been developed to only evaluate and quantify acupuncture and not other ancillary techniques often used in addition to acupuncture such as moxibustion, cupping, dermal hammers, or bleeding lancets. Consideration should be given in the future to develop appropriate instruments to access these techniques if used in clinical studies.

Some of the items identified by the expert group reflect the activities of other researchers assessing quality acupuncture for systematic reviews. For example, in recent systematic reviews undertaken by Manheimer et al.,¹⁹ and Linde et al.,²⁰ reviewers assessed quality acupuncture by considering the choice of acupuncture points, number of sessions, needling technique, and acupuncturist's experience. Although there are some areas of overlap between treatment parameters defined by the current expert group and these reviewers, there are also divergent areas, and these tended to reflect areas where it was not possible to achieve consensus. This is most likely a result of the diversity of opinion and individualization of a therapeutic approach in relation to specific health complaints.

To contextualize the current work further, it is important to differentiate that STRICTA was developed for *reporting* purposes whereby the current project's aim was to develop an instrument for evaluating quality acupuncture adminis-

tered during a clinical trial. In fact, the ability to assess quality acupuncture is dependent initially on full and transparent reporting, which the STRICTA requires. Retained items have been expressed as a normative standard, irrespective of whether they represent an ideal or more modest level of acupuncture. However, subsequent work will seek to determine a relative weighting of items by consensus and further validate the instrument through pragmatic applications. These standards will in the future help identify a "quality threshold" for primary study inclusion for a systematic review, identify whether heterogeneity in trial results are due to the quality of acupuncture, evaluate the results of a particular study in relation to quality for meta-analysis, and assist interpretation of study results and evaluation of the strength of inferences made.

Further development of this instrument will include establishing the inter- and intrarater reliability of the instrument. The aim is to develop an easy-to-use tool that can be incorporated with use of a methodological critical appraisal tool (for example, the checklist for randomized controlled trial appraisal sheet from the Centre for Evidence Based Medicine,²¹ or AMSTAR for systematic reviews²²). The results of this evaluation and final development of the instrument will be published in the near future.

Conclusions

This is the first systematic approach taken using the Delphi technique to identify domains and items of quality relating to the acupuncture administered in clinical trials. It is intended that the final stage of development will provide an appropriate and robust instrument that will enhance interpretation of the research findings, future trial design, and clinical practice.

Acknowledgments

We wish to thank the following members of the external Delphi group who agreed to be identified and participated in one or both rounds of consultation: Terje Alrek, Stephen Birch, Mark Bovey, John Deare, Andrew Flowers, George Lewith, Jianping Liu, John MacDonald, Eric Manheimer, Ryan Milley, Kylie O'Brien, Charlotte Paterson, Rosa Schner, Karen Sherman, Adrian White, Peter White, Claudia Witt, and Shi Ping Zhang. We wish to thank Charlotte Paterson for her helpful comments on the manuscript. This study was supported by a grant from the National Institute for Complementary Medicine.

Disclosure Statement

No competing financial interests exist.

References

1. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *JAMA* 1996;28:276:637-639.
2. Boutron I, Moher D, Altman DG, et al. Extending the CONSORT statement to randomized trials on non-pharmacological treatment: Explanation and elaboration. *Ann Intern Med* 2008;148:295-309.
3. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: Updated guidelines for reporting

- parallel group randomized trials. *Ann Intern Med* 2010;152:726–732.
4. MacPherson H, White A, Cummings M, et al. Standards for reporting interventions in controlled trials of acupuncture: The STRICTA recommendations. *Complement Ther Med* 2001;9:246–249.
 5. MacPherson H, Altman DG, Hammerschlag R, et al. Revised Standards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA): Extending the CONSORT statement. *Acupunct Med* 2010;28:83–93.
 6. National Health and Medical Research Council. How to Use the Evidence Assessment and Application of Scientific Evidence. Canberra: National Health and Research Council, 2000.
 7. Weste S, King V, Cary TS, et al. Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No 47, Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality, 2002.
 8. Prady SL, Macpherson H. Assessing the utility of the standards for reporting trials of acupuncture (STRICTA): A survey of authors. *J Altern Complement Med* 2007;13:939–943.
 9. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966;44:166–203.
 10. White AR, Ernst E. A trial method for assessing the adequacy of acupuncture treatments. *Altern Ther Health Med* 1998;4:66–71.
 11. Ter Riet G, Kleijnen J, Knipschild P. Acupuncture and chronic pain: A criteria-based meta-analysis. *J Clin Epidemiol* 1990;43:1191–1199.
 12. Ter Riet G, Kleijnen J, Knipschild P. A meta-analysis of studies into the effect of acupuncture on addiction. *Br J Gen Pract* 1990;40:379–382.
 13. Kleijnen J, Ter Riet G, Knipschild P. Acupuncture and asthma: A review of controlled trials. *Thorax* 1991;46:799–802.
 14. White A, Cummings M, Barlas P, et al. Defining an adequate dose of acupuncture using a neurophysiological approach: A narrative review of the literature. *Acupunct Med* 2008; 26:111–120.
 15. Boutron I, Moher D, Tugwell P, et al. A checklist to evaluate a report of a non-pharmacological trial (CLEAR NPT) was developed using consensus. *J Clin Epidemiol* 2005;58: 1233–1240.
 16. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Intervention*. Version 5.01. Online document at: www.cochrane.org/resources/handbook/handbook.pdf Accessed July 26, 2009.
 17. Adler M, Ziglio E, eds. *Gazing into the Oracle: The Delphi Method and Its Application to Social Policy and Public Health*. London: Jessica Kingsley, 1996.
 18. Streitberger K, Ezzo J, Schneider A. Acupuncture for nausea and vomiting: An update of clinical and experimental studies. *Auton Neurosci* 2006;129:107–117.
 19. Manheimer E, Cheng K, Linde K, et al. Acupuncture for peripheral joint osteoarthritis. *Cochrane Database Syst Rev* 2010;1:CD001977.
 20. Linde K, Allais G, Brinkhaus B, et al. Acupuncture for tension-type headache. *Cochrane Database Syst Rev* 2009;1: CD00775587.
 21. Centre for Evidence Based Medicine. Randomised Controlled Trial Appraisal Sheet. Online document at: www.cebm.net/index.aspx?o=1157 Accessed December 23, 2010.
 22. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.

Address correspondence to:

Caroline A. Smith, PhD

Centre for Complementary Medicine Research

The University of Western Sydney

Penrith South 1797

Australia

E-mail: Caroline.smith@uws.edu.au