

Patient-Based Outcome Assessment Instruments in Acupuncture Research

Raheleh Khorsan, M.A.,^{1,2} Alexandra York, M.S.,³ Ian D. Coulter, Ph.D.,⁴⁻⁶
Rachel Wurzman, M.S.,⁷ Joan A.G. Walter, J.D., P.A.,³ and Remy R. Coeytaux, M.D., Ph.D.^{8,9}

Abstract

Background: Outcome assessment can support the therapeutic process by providing a way to track symptoms and functionality over time, providing insights to clinicians and patients, as well as offering a common language to discuss patient behavior/functioning.

Objectives: In this article, we examine the patient-based outcome assessment (PBOA) instruments that have been used to determine outcomes in acupuncture clinical research and highlight measures that are feasible, practical, economical, reliable, valid, and responsive to clinical change. The aims of this review were to assess and identify the commonly available PBOA measures, describe a framework for identifying appropriate sets of measures, and address the challenges associated with these measures and acupuncture. Instruments were evaluated in terms of feasibility, practicality, economy, reliability, validity, and responsiveness to clinical change.

Methods: This study was a systematic review. A total of 582 abstracts were reviewed using PubMed (from inception through April 2009).

Results: A total of 582 citations were identified. After screening of title/abstract, 212 articles were excluded. From the remaining 370 citations, 258 manuscripts identified explicit PBOA; 112 abstracts did not include any PBOA. The five most common PBOA instruments identified were the Visual Analog Scale, Symptom Diary, Numerical Pain Rating Scales, SF-36, and depression scales such as the Beck Depression Inventory.

Conclusions: The way a questionnaire or scale is administered can have an effect on the outcome. Also, developing and validating outcome measures can be costly and difficult. Therefore, reviewing the literature on existing measures before creating or modifying PBOA instruments can significantly reduce the burden of developing a new measure.

Introduction

OUTCOME INSTRUMENTS CAN provide information on the effectiveness of interventions and inform “real life” clinical practice, but selecting suitable outcome instruments can be time consuming and challenging. There are hundreds of outcome measures and instruments available. Instruments can be patient-completed (self-administered) questionnaires, physiologic tests, clinician-completed observation scales, task-specific activities/tests, and impairment tests.

Choosing an outcome instrument is difficult due to numerous factors.^{1,2} There are several important considerations when choosing an outcome measurement. For example, the instrument should “(1) be valid, i.e., must measure what it sets out to test; (2) be reliable and consistent, i.e., it should give reproducible results on different occasions or with comparable groups;³ (3) be responsive to clinical change; (4) be economical and feasible to administer; and (5) make it possible to compare findings to other studies, populations, or standard norms, etc.” (p. 356)⁴ In addition to these

¹Military Medical Research and Integrative Medicine, Samueli Institute, Corona del Mar, CA.

²School of Social Ecology, Department of Planning, Policy and Design, University of California Irvine, Irvine, CA.

³Military Medical Research, Samueli Institute, Alexandria, VA.

⁴UCLA School of Dentistry, University of California Los Angeles, Los Angeles, CA.

⁵RAND Corp., Santa Monica, CA.

⁶Southern California University of Health Sciences, Whittier, CA.

⁷Interdisciplinary Program in Neuroscience, Georgetown University Medical Center, Washington, DC.

⁸Center for Clinical Health Policy Research, and ⁹Department of Community and Family Medicine, Duke University, Durham, NC.

TABLE 1. MOST COMMON PATIENT-BASED OUTCOMES ASSESSMENT INSTRUMENTS USED IN CLINICAL ACUPUNCTURE RESEARCH (TOTAL N = 241)

	Assessing	Validity	Reliability	Clinical responsiveness	Cost	Feasibility	Comparability
Visual Analog Scale (VAS)	General pain intensity (acute or chronic)	✓	✓	✓	Training: none Equipment: none Cost: copying Permission: U	Self-administered, 1 min	✓
Numerical Pain Rating Scale (NRS)	General pain intensity (acute or chronic)	✓	✓	✓	Training: none Equipment: none Cost: copying Permission: U	Self-administered, 1 min	✓
McGill Present Pain Index (MPI or MPQ) (including: the McGill Short Form Pain Questionnaire)	General pain intensity (acute or chronic)	✓	✓	✓	Training: none Equipment: none Cost: copying Permission: U	Self-administered, 5–10 min	✓
Oswestry Pain Disability Index (ODI) (including: revised versions 1.0, 2.0, AAOs/MODEMS, "revised ODI")	Functional disability and back pain (acute or chronic)	✓	✓	✓	Training: none Equipment: none Cost: copying Permission: Needed ^a	Self-administered, 5 min	✓
Roland Morris Disability/Activity Questionnaire (including: revised versions RM-23, RM-18)	Functional disability and back pain (acute and subacute)	✓	✓	✓	Training: none Equipment: none Cost: copying Permission: none ^a	Self-administered, 5 min	✓
SF-36 (including modified versions)	Health-related quality of life	✓	✓	✓	Training: none Equipment: none Cost: varies Permission: Possibly co-copyright holders ^b	Self-administered, 5 min	✓

Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index	Functional disability and arthritis pain (acute or chronic)	✓	✓	✓	Training: none Equipment: none Cost: copying Permission: U	Self-administered, 5–10 min	✓
Beck Depression Inventory (BDI)	Depression	✓	✓	✓	Training: none Equipment: none Cost: copying Permission: U	Self-administered, 5–10 min	✓
Pain Disability Index (PDI)	Functional disability and pain (acute and subacute)	✓	✓	✓	Training: none Equipment: none Cost: copying Permission: U	Self-administered, 5–10 min	✓
Measure Yourself Medical Outcome Profile (MYMOP)	Primary care settings	✓	✓	✓	Training: none Equipment: none Cost: copying Permission: needed: Registration as a user ^c	Structured guidance, 7–10 min	✓
Hospital Anxiety and Depression Scale (HADS)	Hospital general medical outpatients	✓	✓	✓	Training: none Equipment: none Cost: yes Permission: needed ^d	Self-administered, 5–10 min	✓
Nottingham Health Profile (NHP)	Health-related quality of life	✓	✓	✓	Training: none Equipment: none Cost: copying Permission: U	Self-administered, 7–10 min	✓

^aThe instruments (RM-24 & ODI version 2.0) are include in Roland and Fairbank (2000) appendixes (Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine 2000;25:3115–3124). When used in the forms reproduced in the appendixes (Roland and Fairbank 2000), no permission is required from the authors or from Spine.

^bRAND Health, a research division of the RAND corporation, the Medical Outcomes Trust (MOT), Health Assessment Lab (HAL), and QualityMetric Incorporated are co-copyright holders of the Short Form Health Surveys.

^cUser registration is required at www.pms.ac.uk/mymop/index.php?c=contact&s=register.

^dFor purchase of questionnaire contact GL Assessments.

✓, Available; U, unavailable.

considerations, the patient health status must be considered and instruments should measure outcomes specifically targeted (or likely to be affected) by the intervention.

The paradigm and theoretical basis of measures is important when choosing the type of instrument for research or clinical practice. For example, patient-based outcomes assessment (PBOA) instruments (i.e., instruments such as functionality and symptomatology indexes that are reported directly by the patient; also known as patient-reported pencil-and-paper instruments without any interpretation by a clinician or researcher) reflect a paradigm more focused on the individual that includes a greater emphasis on how the individual perceives his or her disease, including whether the person sees himself or herself as ill (in the sick role) or if the disease has resulted in impairment or disability. Some PBOA instruments may reflect this paradigm more than others. The theoretical basis for choosing a metric and the psychometric properties of many common PBOA instruments have been discussed elsewhere for chiropractic research by Khorsan et al.,⁴ and here we apply it to the evaluation of PBOA instruments used in acupuncture research.

The objectives of this review are to (1) assess the available literature and identify the most common PBOA instruments used in acupuncture research; (2) describe a framework for identifying appropriate sets of instruments; and (3) address the challenges associated with these instruments that are relevant to acupuncture.

Methods

This review was conducted in two phases. Initially, a broad systematic search was performed to identify existing instruments for measuring pain, disability, general health status, and well-being found in acupuncture therapy studies.

For the purpose of this study, we define patient outcome instruments as questionnaires, indexes, scales, or other instruments used to measure patients' health status, or change in health status, associated with a given intervention. To be included in the review, an instrument had to be used in at least three separate studies that met our search criteria. Patient outcome instruments were included if they were found cited in abstracts of articles reporting research related to acupuncture.

For each instrument that was included, a second search was conducted to identify articles reporting research into its psychometric properties. This second phase of searching was not restricted to studies of acupuncture. Data on the psychometric properties of identified instruments were subsequently extracted and compared (Table 1).

Inclusion and exclusion criteria

Studies eligible for inclusion in this article included peer-reviewed clinical trials and randomized controlled trials (RCTs) for which a full abstract was available. A sensitive search was originally conducted for key search terms for acupuncture, pain, disability, well-being, and outcome measures. Search terms for acupuncture and outcome were limited to the title or abstract to constrain the magnitude of the review yield to a manageable size. For example, search terms included: Acupuncture, Acupuncture Therapy, Acupuncture Points, Acupuncture Analgesia, patient*, outcome*,

measure*, assessment*, and derivatives of each. A PBOA instrument had to appear in at least three publications to be included in this review.

All abstracts reviewed in this article were in English. Figure 1 outlines our search strategy.

Acupuncture was defined according to the National Institutes of Health (NIH) National Center for Complementary and Alternative Medicine.⁵ This systematic review excluded studies on acupressure without needling, laser acupuncture, transcutaneous electrical nerve stimulation, dry needling, audits, surveys, literature reviews, commentaries, and proceedings. We also excluded case reports. Nearly always, a case report becomes a case report after the fact (after something of interest was noted in a given person). Therefore, it would be unlikely that a research-oriented outcome measure would be administered in a case report.

PubMed was the database searched. The complete search strategy is shown in Figure 1. All articles were screened for PBOA instruments that were reported in the title, abstract, and full article.

Instrument identification and selection

Literature searches were conducted to identify abstracts on outcome instruments cited in studies in which primary or secondary acupuncture care were included. Searches were performed on PubMed from its inception through April 2009.

The search restrictions were human, clinical trial, meta-analysis, practice guideline, RCT, and review for which a full abstract was available. Full-length manuscripts were pulled whenever possible. If an abstract did not include a PBOA instrument and the full-length manuscript was unavailable or the manuscript was in a language other than English, we excluded it from this review. For this review, full-length manuscripts that reported the psychometric properties of the PBOA instruments were obtained.

Results

The search identified 582 articles. After screening of title/abstract, 212 articles were excluded. From the remaining 370 citations, 258 manuscripts identified explicit PBOA, while 112 abstracts did not include any PBOA (Fig. 1). A total of 258 manuscripts were extracted and reviewed (Table 2).

RCTs were the most common design for clinical acupuncture studies included in this review. Musculoskeletal disorders (33.33%), women's health (9.39%), and headache (9.30%) were the most common conditions researched in the included acupuncture studies. Sleep disorders (1.16%) were the least common conditions researched in the included acupuncture studies. The vast majority of studies included PBOA instruments as their primary measure (74.42%).

Assuming that the country of the lead author of a given study was the location of the research, we found that most acupuncture studies were conducted in Europe (55.42%). North America was responsible for about 20% of the research (22.5%), Asia (16.28%), Australia (2.71%) and South America (1.16%).

Of those instruments identified, the visual Analog Scale (VAS) was the most common PBOA measure used. For assessing pain intensity and function, the VAS and the numerical Pain rating Scales (NRS) were the most common scales used. The most commonly used health status and well-

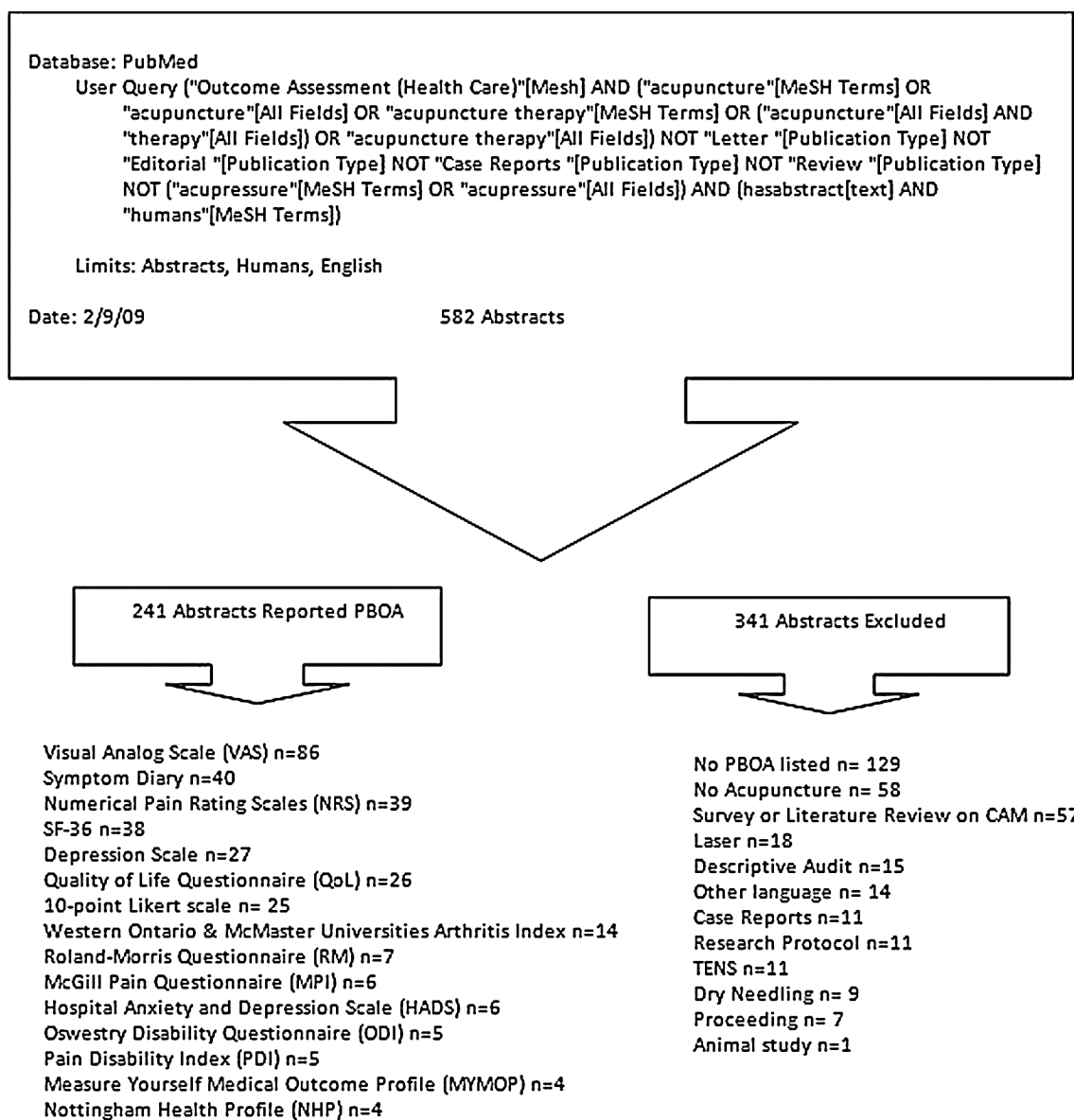


FIG. 1. PBOA, patient-based outcome assessment instruments; TENS, transcutaneous electrical nerve stimulation; CAM, complementary and alternative medicine.

being instrument was the SF-36. Forty studies used a Symptom Diary ($n=40$) to assess patient health and well-being. Table 2 provides a summary and characteristics of the 258 studies included. The most common PBOA instruments used in acupuncture research are reviewed in Figure 1. We also extracted and compared the data on the psychometric properties of identified instruments in Table 1.

The number of study participants in the included studies ranged from 5 ($n=5$) to 14,101 ($n=14,161$) subjects. The mean number of subjects was 251, the median was 56, and the mode was 30. There were 11 studies that included 1000 or more subjects.

The vast majority of studies identified in this review employed a battery of instruments rather than a single instrument. Some articles (about 30%) do not have a specific citation to the PBOA instrument used or incorrectly cite the references. Many studies do not define or explain "modified"

instruments. Most study abstracts included a number of different PBOA instruments.

Discussion

Challenges and psychometric properties associated with PBOA instruments and acupuncture

The goal of this section is not to discuss all the published works relating to the PBOA instruments, but to provide the reader with a rich outline of the important aspects of the PBOA instruments examined in this review. Toward that effort, we will briefly discuss some of the studies that highlight important aspects of a PBOA instrument (i.e., validity) that are significant when considering its use as an outcome measure in acupuncture research or clinical care.

The most common instruments used in acupuncture studies are single-dimensional pain scales pain such as

TABLE 2. STUDY CHARACTERISTICS

	<i>Included studies</i>		<i>Studies with no PBOA Instruments</i>	
	n	%	n	%
Total number of studies	258		112	
Study design				
RCT	182	70.542	39	34.821
CCT	28	10.852	23	20.535
Pilot	25	9.689	9	8.035
CO	8	3.100	0	0
CT	8	3.100	0	0
SR	5	1.937	11	9.821
Mixed design	2	1.162	11	9.821
Unknown design	0	0	19	16.964
General topic				
Musculoskeletal	86	33.333	24	21.428
Women's health	25	9.689	10	8.928
Headache	24	9.302	7	6.25
Other type of conditions	21	8.139	23	20.535
Neurologic condition	20	7.751	7	6.25
Mental health	18	6.976	5	4.464
Gastrointestinal disorders	14	5.426	3	2.678
Addiction	11	4.263	3	2.678
Cardiovascular condition	8	3.100	8	7.142
Cancer and treatment-related symptoms	7	2.713	3	2.678
Pulmonary disorder	5	1.937	2	1.785
Allergy	4	1.550	1	0.892
Urological disorder	4	1.550	5	4.464
Sleep disorder	3	1.162	0	0
Unknown condition	0	0	10	8.928
PBOA Instrument is primary	192	74.418	NA	NA
PBOA Instrument is secondary	21	8.139	NA	NA
PBOA Instrument is valid and reliable	223		NA	NA

PBOA, patient-based outcome assessment; CO, cohort study; CCT, controlled clinical trial; CT, clinical trial; Pilot, pilot study; RCT, randomized control trial; SR, systematic review.

the VAS ($n=86$) and NRS ($n=39$). The VAS is a patient-completed analogue instrument that evaluates pain intensity and function, typically on a 100-mm-long horizontal or vertical line anchored at each end with a statement representing the extremes of the dimension being measured. Unlike the VAS, the NRS is used to measure pain severity using whole numbers. Generally the NRS styles also include a horizontal line marked by whole numbers, typically 0–10.

Both instruments provide pain intensity and function estimates relatively quickly, are highly patient-centered, and have the most value when looking at change within individuals, but are of less value for comparing across a group of individuals at one time point. Also both instruments are quick and simple to administer,⁶ easy to translate into other languages, inexpensive,⁷ and readily available.⁸

The single-dimensional pain intensity and function scales (i.e., the VAS and NRS) have been criticized for their lack of sensitivity, oversimplifying the patient's experience of pain, and their single dimension of pain (e.g., intensity).⁹

For pain-related studies such as headache, musculoskeletal, and cancer- and treatment-related symptoms, the most common scales used in acupuncture clinical research included the VAS, NRS, 10-point Likert scale on subjective experience ($n=24$) and global well-being, and Symptom Diary ($n=40$). However, most studies did not include cita-

tions for the use of the 10-point Likert scale on subjective experience and global well-being and Symptom Diary. Therefore, for these instruments, it is unknown whether there are any validity or reliability issues. The Symptom Diary was also commonly used among acupuncture studies of pulmonary, sleep, and urological disorders. We were unable to determine whether the Symptom Diaries used in acupuncture research were structured, meaning that the research participants recorded particular information related to a specific health event or a particular research question; or unstructured (i.e., journals, used to explore a patient's spontaneous thoughts and feelings in relationship to a particular event).¹⁰ No study in this review cited a standardized and validated Symptom Diary. While there are validated Symptom Diaries available (i.e., the Diagnostic Headache Diary),¹¹ it may be that most researchers are unable to find in the published literature a specific diary that meets their particular research question and therefore each research team devises their own diary.¹²

About 60% of all acupuncture studies in this review used some sort of quality-of-life measure (QoL). The Medical Outcomes Study Short Form-36 (SF-36) ($n=38$) was the most common QoL instrument used across all study designs and conditions. The SF-36, often referred to as the MOS SF-36, and the RAND 36[®]-Item Health Survey 1.0 (distributed by RAND) are identical scales with 36-item general health

status assessment items constructed to survey health status in the 2-year study of patients with chronic conditions.^{13,14} The SF-36 is perhaps one of the most common outcome-assessment instruments in use in contemporary health services research.

There are substantial reliability and validity data for the SF-36.^{13,15-17} Other QoL measures used were the NIH-CPSI, the Nottingham Health Profile, and the EuroQol 5-Dimension form.

In contrast to brief or predictable procedure-related pain, more comprehensive pain assessment requires the determination of other characteristics of the pain, such as location and quality, and its effect on mood and function. Multidimensional pain assessment tools have been developed to capture these aspects of pain.

The most common multidimensional pain scales used by acupuncture pain studies were the McGill pain questionnaire (MPQ), The Oswestry Pain Disability Index (ODI), the Roland Morris Disability/Activity Questionnaire (RM), and the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index.

The MPQ originally attempted to depict pain experience using 78 pain descriptors on three dimensions: (I) sensory (items 1-10), (II) affective (items 11-15), (III) evaluative (items 16); and (IV) an additional miscellaneous descriptor (items 17-20).¹⁸ Later, a short form of the MPQ (SF-MPQ) was derived from commonly used sensory and affective descriptors in the clinical studies.

The ODI, like the MPQ, is a self-report questionnaire designed for assessing the degree of functional limitation in patients with low-back pain seeking consultation in secondary care, while the RM is designed for assessing the degree of functional limitation in patients seeking consultation for low-back pain in primary care.¹⁹ The development, testing, and properties of both measures have been extensively examined and adequately reviewed (Table 1).

The WOMAC is a disease-specific, self-administered questionnaire that evaluates three dimensions: pain (5 questions), stiffness (2 questions), and physical function (17 questions).²⁰ The WOMAC was constructed to evaluate patients' experience of osteoarthritis (OA) of the knee and hip. It was designed in the late 1980s in response to the lack of a multidimensional instrument that could measure clinically important, patient-relevant symptoms of OA in the knee and hip.²¹ The reliability and validity of the WOMAC has been demonstrated in a number of studies.^{22,23}

Disability and depression scales were mainly used in studies on mental health, neurologic conditions, addiction, autoimmune condition, and musculoskeletal disorders. The most common disability and depression scales were the Pain Disability Index (PDI) and Beck Depression Inventory (BDI) ($n = 13$). For studies in hospital settings, the Hospital Anxiety and Depression Scale (HADS) ($n = 7$) was used.

The PDI is a patient-completed, condition-specific functional status questionnaire, based on the Oswestry Low Back Pain Questionnaire. The PDI rates the level of disability on a numerical rating scale (0 = no disability and 10 = maximum disability) assessing seven broad categories of activity including items on recreation, personal care, activities related to home and family, work, frequency and quality of sex life, social activity, and general life-support functions (e.g., eating, sleeping, and breathing).²⁴

The BDI is a multiple-choice self-report inventory that is among the most widely used instruments for measuring the severity of depression and mental health. The instrument was created to measure the intensity, severity, and depth of depression.

The PDI and the BDI have been translated into many languages and are used to evaluate outcomes in a range of populations, settings, and interventions.

Measures for acupuncture in primary care settings and hospitals

The Measure Yourself Medical Outcome Profile (MYMOP), and the HADS^{25,26} were used in several acupuncture clinical studies, including acupuncture for chronic pain management and the elderly (Fig. 1).

MYMOP is an outcome measure originally developed to measure aspects and an effect of symptoms that a patient determines is most important to her or him. It is a sensitive measure of within-person change over time, is applicable to the whole spectrum of illness seen in primary care, is capable of measuring the effects of a wide variety of care, and enables the patient to score the chosen variables.²⁷ However, since the patient requires some structured guidance, especially the first time the MYMOP is completed, this makes it unsuitable for postal administration on the first occasion.

The HADS consists of 14 questions, 7 for anxiety and 7 for depression. Although it was originally designed for hospital general medical outpatients, it has been extensively used for other populations such as in primary care.^{25,26} The psychometric properties for the HADS and MYMOP are extensively evaluated across a range of populations, settings, and interventions.

PBOA instruments in acupuncture versus conventional therapy research

Hull et al., in a prospective cohort study, researched the various methods of assessing clinically meaningful change associated with a course of acupuncture treatments. According to Hull et al., assessing outcomes associated with acupuncture is particularly challenging compared to other therapies. "Acupuncture may result in subtle improvements in a general sense of well-being in ways that may not be measured by standardized instruments, and acupuncture treatments often are customized to meet individual patients' unique clinical needs and therapeutic goals, thereby making it difficult to define *a priori* a single set of clinical outcomes to assess" (p. 247).²⁸ Hull et al. correctly assert that the effects of acupuncture cannot be easily quantified by commonly examined clinical outcomes due to its whole systems construct.

Complementary and alternative therapies are often based on whole medical systems that are built upon complete systems of theory and practice, as is true for acupuncture.²⁹ Most acupuncture research on acupuncture effects and application remains controversial among contemporary biomedical medical researchers and clinicians because the acupuncture research does not generally involve a standardized protocol. According to Hull et al., "Rather, treatments and therapeutic objectives usually are tailored to meet the individual patient's unique clinical presentation, needs, or desires. Such individualized treatment approaches and therapeutic objectives limit the ability of standardized

instruments to assess meaningful clinical change among groups of patients" (p. 248).²⁸ Yet, acupuncture researchers, like all clinical researchers, should assess the appropriateness of their treatment approach defined and measured by clinically significant change and determine patient satisfaction from the intervention.

We found in this review that acupuncture research includes a combination of validated instruments, such as common standardized questionnaires that assess functional status or health-related QoL before and after the administration of a therapeutic intervention for a specific condition. Examples include the VAS, NRS, and SF-36 and disease-specific QoL instrument such as the WOMAC, BDI, and RM used to quantify change over time. However, acupuncture research also included a wide variety of unvalidated instruments like the 10-point Likert scale on subjective experience and global well-being and Symptom Diary. Both types of measures were used to capture particular health and well-being information from the study participant in relationship to a specific health event or an experience based on a particular research question that no single instrument and no combination of validated scales alone could achieve. The large number of unvalidated PBOA instruments in acupuncture research may be associated with the paradigm of whole systems medicine. Therefore, instruments such as MYMOP can be the most useful for assessing clinical change among patients who present for acupuncture treatment with a variety of symptoms, clinical conditions, and therapeutic objectives.²⁸ Further research is needed to determine whether these results apply across other whole medical system therapies compared to conventional therapy research.

Finding more information on PBOA instruments

There are electronic databases established as sources of information on PBOA instruments for clinical and research use. Some can be accessed free of charge, whereas others require membership or fees. Khorsan et al discuss many common electronic databases and translated version of PBOA instruments available for researchers and clinicians.⁴ One such source is the Patient-Reported Outcomes Measurement Information System (PROMIS).³⁰ PROMIS was originally developed as an NIH Roadmap network project intended to develop, validate, and standardize item banks to measure patient-reported outcomes relevant across common medical conditions.³¹ The PROMIS is a publicly available system that can be added to and modified periodically and that allows clinical researchers to access a common repository of items and computerized adaptive tests. The PROMIS is also a network for researchers and clinicians to collaborate on the collection of self-reported data from diverse populations with a variety of chronic diseases, using agreed-upon methods, models, and questionnaires.

Conclusions

Ideally, the selection of potential core outcome measures should be based on "(1) purpose of study or intended use, (2) appropriateness of the measure's content and conceptual model, (3) patient population, (4) reliability, (5) validity, (6) clinical responsiveness, (7) minimization of respondent and administrator burden and feasibility, (8) respondent and administrator acceptability, (9) costs, (10) comparability, and

(11) availability and equivalence of versions for different cultures and languages. Other criteria may also include interpretability, precision of scores, and availability and equivalence of alternate forms and methods of administration (e.g., self-report, interviewer)."⁴ However, like all health measures, the way a questionnaire or scale is administered can have an effect on the outcome. Also, developing and validating outcome measures can be costly and difficult. Therefore, reviewing the literature on existing measures before creating or modifying PBOA instruments can significantly reduce the burden of developing a new measure. Many articles in this review (30%) do not identify the specific outcome measure or incorrectly cite the references.

It is also important to consider the comparability of data collected to previous studies. This requires some knowledge of what measures have been used, which are the most common that are being used in similar populations that would allow comparative data to be used.

Acknowledgments

This work is supported by the U.S. Army Medical Research and Materiel Command under Award No. W81XWH-06-1-0279 through the Samueli Institute. The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision unless so designated by other documentation.

Disclosure Statement

No competing financial interests exist.

References

1. Ware JE Jr, Brook RH, Davies AR, Lohr KN. Choosing measures of health status for individuals in general populations. *Am J Public Health* 1981;71:620-625.
2. Wollersheim H, Hermens R, Hulscher M, et al. Clinical indicators: Development and applications. *Neth J Med* 2007; 65:15-22.
3. Prady SL, Thomas K, Esmonde L, et al. The natural history of back pain after a randomised controlled trial of acupuncture vs usual care: Long term outcomes. *Acupunct Med* 2007;25:121-129.
4. Khorsan R, Coulter ID, Hawk C, Choate CG. Measures in chiropractic research: Choosing patient-based outcome assessments. *J Manip Physiol Ther* 2008;31:355-375.
5. National Center for Complementary and Alternative Medicine (NCCAM). An Introduction to Acupuncture. July 28, 2008. Online document at: <http://nccam.nih.gov/health/acupuncture/> Accessed Aug 26, 2008.
6. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: A comparison of six methods. *Pain* 1986; 27:117-126.
7. McCaffery M, Pasero C. Assessment Tools. *Pain: Clinical Manual*. 2nd ed. St. Louis: Mosby, 1999:58-88.
8. Scott J, Huskisson EC. Graphic representation of pain. *Pain* 1976;2:175-184.
9. Flaherty SA. Pain measurement tools for clinical practice and research. *AANA J* 1996;64:133-140.
10. Ross MM, Rideout EM, Carson MM. The use of the diary as a data collection technique. *West J Nurs Res* 1994;16:414-425.

11. Russell MB, Rasmussen BK, Brennum J, et al. Presentation of a new instrument: The diagnostic headache diary. *Cephalalgia* 1992;12:369–374.
12. Coeytaux RR, Frasier PY, Reid A. Patient-centered outcomes for frequent headaches. *Headache* 2007;47:480–485.
13. Ware JE Jr. The SF-36 health survey. In: Spilker B, ed. *Quality of Life and Pharmacoeconomics*. 2nd ed. Philadelphia: Lippincott-Raven, 1996:337–345.
14. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–483.
15. Ware JE Jr, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *J Clin Epidemiol* 1998;51:903–912.
16. Ware JE Jr, Kosinski M, Bayliss MS, et al. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: Summary of results from the Medical Outcomes Study. *Med Care* 1995;33(suppl):AS264–A279.
17. Ware JE Jr. The status of health assessment 1994. *Annu Rev Public Health* 1995;16:327–354.
18. Melzack R. The McGill Pain Questionnaire: Major properties and scoring methods. *Pain* 1975;1:277–299.
19. Peat G. PPA Recommendations for Low Back Pain-Related Functional Limitation Outcome Measures. London: The Chartered Society of Physiotherapy, July 2004.
20. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): A review of its utility and measurement properties. *Arthritis Rheum* 2001;45:453–461.
21. Bellamy N. Pain assessment in osteoarthritis: Experience with the WOMAC osteoarthritis index. *Semin Arthritis Rheum* 1989;18(suppl 2):14–17.
22. Davis AM, Perruccio AV, Canizares M, et al. Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. *Osteoarthritis Cartilage* 2009;17:829–833.
23. Rothenfluh DA, Reedwisch D, Muller U, et al. Construct validity of a 12-item WOMAC for assessment of femoro-acetabular impingement and osteoarthritis of the hip. *Osteoarthritis Cartilage* 2008;16:1032–1038.
24. Tait RC, Chibnall JT, Krause S. The Pain Disability Index: Psychometric properties. *Pain* 1990;40:171–182.
25. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale: An updated literature review. *J Psychosom Res* 2002;52:69–77.
26. Tang WK, Wong E, Chiu HF, et al. Examining item bias in the anxiety subscale of the Hospital Anxiety and Depression Scale in patients with chronic obstructive pulmonary disease. *Int J Methods Psychiatr Res* 2008;17:104–110.
27. Paterson C. Measuring outcomes in primary care: A patient generated measure, MYMOP, compared with the SF-36 health survey. *BMJ* 1996;312:1016–1020.
28. Hull SK, Page CP, Skinner BD, et al. Exploring outcomes associated with acupuncture. *J Altern Complement Med* 2006;12:247–254.
29. Boon H, Macpherson H, Fleishman S, et al. Evaluating complex healthcare systems: A critique of four approaches. *Evid Based Complement Altern Med* 2007;4:279–285.
30. Fries JF, Bruce B, Cella D. The promise of PROMIS: Using item response theory to improve assessment of patient-reported outcomes. *Clin Exp Rheumatol* 2005;23(suppl 39):S53–S57.
31. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap cooperative group during its first two years. *Med Care* 2007;45(suppl 1):S3–S11.

Address correspondence to:

Raheleh Khorsan, M.A.

Military Medical Research and Integrative Medicine

Samueli Institute

2101 East Coast Hwy

Suite 300

Corona del Mar, CA 92625

E-mail: rkhorsan@siib.org

