Variation in Outcome Measures in Hip and Knee Arthroplasty Clinical Trials: A Proposed Approach to Achieving Consensus

DANIEL L. RIDDLE, PAUL W. STRATFORD, JASVINDER A. SINGH, and C. VIBEKE STRAND

ABSTRACT. OMERACT began work over a decade ago on a consensus effort to identify optimal outcome measures for knee and hip osteoarthritis clinical trials. Recent evidence indicates extensive variation in outcome measures used in clinical trials of knee and hip arthroplasty published since 2000. This heterogeneity leads to confusion, not only for conducting systematic reviews but also for applying evidence to clinical practice. Given the extensive psychometric research conducted in the past 2 decades, the timing seems ideal to design and implement a study to develop consensus on optimal outcome measures for hip and knee arthroplasty trials. We describe a Delphi survey design and an approach for synthesizing the extensive psychometric literature on the outcome measures used in hip and knee arthroplasty trials. Plans for dissemination of the findings are also discussed. This proposed study could have an important influence on the design and reporting of future randomized trials of knee arthroplasty. (J Rheumatol 2009;36:2050–6; doi:10.3899/jrheum090356)

Key Indexing Terms:
ARTHROPLASTY REPLACEMENT KNEE OUTCOME ASSESSMENT

Lower extremity joint replacement surgery is a common and highly effective procedure for many patients with arthritis[1]. In 2002, about 400,000 knee arthroplasty surgeries and about 200,000 hip arthroplasty surgeries were conducted in the USA[2]. Trend data reported by Kurtz, *et al* suggest that by 2010, about 1 million hip and knee arthroplasty surgeries will be conducted with roughly 15% of these being revision surgeries and the remaining 85% primary surgeries[3]. Cost data for joint arthroplasty surgery are also impressive. Ong, *et al* determined the combined hospital and physician procedural charges for Medicare patients receiving joint arthroplasty during the years 1997 to 2003[4]. Mean procedural charges per patient from 1997 to 2003, in 2005 dollars was US$40,000 per primary surgery and $50,000 per revision surgery. If the reimbursement data were extrapolated to the 2006 volume data projections, hospital and physician charges for hip and knee arthroplasty would total about $13 billion in the USA. Given the volume and costs of hip and knee arthroplasty, it is not surprising that an extensive research emphasis has been devoted to the effort[1] with over 160 randomized trials since 2000[5].

Recognizing the need for valid, standardized methods for comparing outcome data, OMERACT used a consensus-based approach in 1997 to identify optimal outcome measures for knee and hip osteoarthritis clinical trials[6]. More than 90% of the participants agreed that pain, physical function, patient global ratings of improvement, and joint imaging procedures should be included in clinical trials of patients with osteoarthritis. Participants were unable to come to consensus regarding specific measures because measures had yet to be identified in the literature as superior for the different outcomes of interest. The consensus of the participants was that in the next 3 to 5 years (2000 to 2003) evidence should be sufficient to identify specific measures for clinical trials.

The past 20 years have seen a tremendous growth in the development and validation of outcome measures for patients with arthritis. Researchers and clinicians now have dozens of measures to choose from when caring for patients with hip or knee arthroplasty. This diversity of measurements, however, comes with a cost. Riddle, *et al* conducted a systematic review of outcome measures used in contemporary clinical trials and found extensive variation in the numbers and types of outcome measures used in hip and knee arthroplasty trials[5]. For example, of the 82 hip replacement trials published since 2000, the Harris Hip Score was used in 43 (52%), but an additional 19 measures were used in the trials. There was extensive variation across trials, not only in the specific measures used but also in the general

From the Department of Physical Therapy, Virginia Commonwealth University, Richmond, Virginia;, USA; School of Rehabilitation Science, McMaster University, Hamilton, Canada; Minneapolis VA Medical Center, Minneapolis, MN; and Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto, CA, USA.

D.L. Riddle PT, PhD, FAPTA, Otto D. Payton Professor, Department of Physical Therapy, Virginia Commonwealth University; P.W. Stratford, MSc, PT, Professor, School of Rehabilitation Science, McMaster University; J.A. Singh, MBBS, MPH; C.V. Strand, MD, Adjunct Clinical Professor, Division of Immunology/Rheumatology, Stanford University School of Medicine.

*Address correspondence to Dr. D.L. Riddle, Department of Physical Therapy, West Hospital Room B-100, PO Box 980224, Virginia Commonwealth University, Richmond, VA 23298-0224,
E-mail: dlriddle@vcu.edu*

construct being measured. Extensive variation also was found in knee trials. For example, 6 different self-report functional status measures were used and none were used in more than half the trials (n = 75).

Heterogeneity in outcome measures across trials potentially leads to several problems. Clinicians who are attempting to integrate findings from multiple trials of an intervention cannot readily interpret findings when different outcome measures are used. Researchers conducting systematic reviews cannot calculate summary measures of effect if measures across trials are different. Finally, some outcome measures have superior measurement properties compared to others, and measures with weaker psychometric properties continue to be used in hip and knee arthroplasty trials[5].

Since the work of OMERACT in 1997 the Osteoarthritis Research Society International (OARSI), has been working to improve clinical trial reporting for patients with osteoarthritis. OARSI has been collaborating with OMERACT in the establishment of criteria for interpreting patient response in OA drug trials[7,8]. The authors established criteria for judging the magnitude of treatment effect but did not identify the specific measures that should be used in OA drug trials. Neither OMERACT nor OARSI have developed a consensus for identifying specific measures for use in randomized controlled trials (RCT) of patients with hip or knee arthroplasty.

Separately, the World Health Organization used a worldwide consensus-based approach in 2001 to develop the International Classification of Functioning Disability and Health (ICF)[9]. We believe that the ICF provides an ideal framework for conceptualizing outcome after knee and hip arthroplasty surgery from a biological, individual, and societal perspective.

We propose to extend and bridge these initiatives through a multistaged approach to establish consensus-based recommendations of specific posthospitalization outcome measures after hospitalization for knee and hip arthroplasty trials. We describe a research design and method for achieving consensus on optimal outcome measures for knee and hip arthroplasty RCT.

METHODS: ICF AS A CONCEPTUAL FRAMEWORK FOR CATEGORIZING OUTCOME MEASURES

The World Health Organization formally adopted the ICF as the standard language for describing health related states and conditions. The ICF is now the internationally agreed upon standard language to describe health.

The framework for the ICF model is illustrated in Figure 1. Each major component within the ICF model will be briefly defined. For a more thorough examination of ICF, several comprehensive descriptions are available. As applied to patients with joint arthroplasty, "health conditions," the component at the top of Figure 1, equate to the arthritis and to any complications arising from treatment such as infection or deep vein thrombosis.

"Body function and structure" refers to the functioning and structural integrity of specific body organs and systems. Patients with hip or knee arthroplasty may, for example, have reduced muscle strength, joint swelling, pain, and psychological distress[10,11]. "Activity" is defined as the completion of a task or action by an individual. Limitations in commonly performed tasks for patients with hip or knee arthroplasty may be walking, bending, sitting, and stairclimbing[12,13].

"Participation" is the term used to describe a patient's involvement in everyday life. When a person's everyday life is disrupted, participation is restricted. For example, if a patient's ability to attend religious services was compromised, that person would have a participation restriction. In addition to the 4 components described above, there are 2 additional components that directly affect body structures, functions, and activity, and participation. These are termed contextual factors. Two broad categories of contextual factors address the interacting influence of personal and environmental factors. "Environmental factors" are external to the person and influence that person's daily life. These include all features including policies, laws, and values in that person's environment. "Personal factors" include gender, race, lifestyle and daily routines. The ICF is receiving worldwide support from a variety of areas in medicine and seems the ideal conceptual model to frame a study of the variation in outcome measures used in hip and knee arthroplasty trials[10,11,14].

Overview of proposed study design. The proposed study will focus on 4 of the 6 main components of the ICF model: body structure and function, activity, participation, and personal factors. For the personal factors component, the study will focus on measurement of patient satisfaction, the most commonly measured personal factor outcome in the hip and knee arthroplasty literature[5]. Health conditions (e.g., surgical complications such as venous thromboembolism or postsurgical infection) will not be examined because these outcomes are most commonly assessed during inpatient care, and the proposed study design is focused on outcomes after hospitalization. The proposed study also will not address outcome measures related to biomechanical issues such as prosthetic loosening. Environmental factors are not included primarily because they are rarely measured in hip and knee arthroplasty trials[5].

Hip and knee replacement outcome measures will be considered separately. In addition, primary replacement will be considered as separate and distinct from revision surgery. Trials will be considered in 3 separate categories, much like those described by Riddle and colleagues[5]. Optimal outcome measures for trials of surgical interventions, trials of nonsurgical physical interventions (i.e., physical therapy), and trials of nonsurgical medical interventions (i.e., medication) will each be identified.

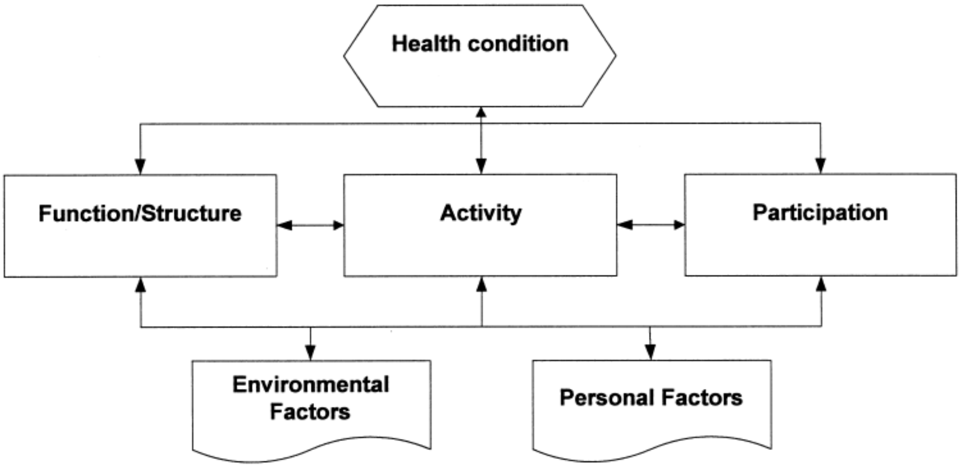


Figure 1. The World Health Organization's ICF model of health and health conditions.

We have designed the study using a multistaged process. For stage 1, literature that has examined the psychometric properties of the various outcome measures identified in the study by Riddle and colleagues[5] will be identified and summarized. All instruments will be categorized into one of the 4 key domains of the ICF. For stage 2, a multistep Delphi survey will be conducted. Experts participating in the Delphi survey will be provided succinct summaries of the psychometric properties of all instruments to guide them during the Delphi survey process. See Figure 2 for a summary of the flow of the study.

The final outcome of the Delphi survey will be a consensus summary that identifies (a) which of the 4 ICF components should be measured for each of the 3 types of RCT conducted on patients with hip arthroplasty; knee arthroplasty, and revision hip or knee arthroplasty, (b) the optimal outcome measures for each ICF component for primary and revision hip and knee arthroplasty. The consensus summary will be presented at an upcoming OMERACT conference to determine if international consensus can be achieved.

Stage 1: Synthesis of psychometric literature. The goal of the literature synthesis is to locate relevant articles reporting the psychometric properties of the outcome measures identified in the systematic review of Riddle, *et al*[5]. We will search the Medline database and limit our search to articles in English. The general search strategy will be to present the name of the instrument, the terms arthroplasty or replacement, the location of hip or knee, and a comprehensive set of measurement terms. We expect to conduct searches for approximately 60 outcome measures. All searches will be conducted by the investigative team.

The following are search examples for the Western Ontario and McMaster Universities Osteoarthritis Index: The WOMAC search ("Western Ontario and McMaster Universities Osteoarthritis Index" OR WOMAC) and (arthroplasty OR replacement) and (hip OR knee) and (change OR valid* OR reliab* OR sensitiv* OR responsive* OR psychometric OR clinimetric) yielded 97 articles; "*" is used in PubMed literature searches.

Data from the searches will be abstracted and critiqued by 6 trained abstractors chosen from graduate students or clinicians with a background in rehabilitation science and who are familiar with the concepts of reliability, validity, and responsiveness. Eligible candidates will complete a training program comprising instruction in the ICF model, sample exercises, and sample critiques with guided feedback. Following the training program, candidates will be evaluated on 6 research papers, 3 of which will be head-to-head comparison studies of competing measures. Each candidate will be required to identify the reliability, validity, and responsiveness coefficients from each of 6 previously selected papers and correctly report these data on 2 structured forms (one for studies of individual measures and one for studies of competing measures), which will include sections that address the cross-sectional and longitudinal validity of each measure. The correct responses for the "gold standard" set of 6 papers will be established by the investigators prior to the start of the study. Only candidate abstractors whose responses are consistent with the gold standard on all 6 papers will be accepted as abstractors for this component of the study.

The results will be compiled for 2 types of articles, those that describe psychometric properties for individual measures and those that compare the psychometric properties of competing measures. The results will be compiled at 2 levels. The first level groups information by ICF category; the second level will summarize the material for competing measures within each ICF component. Summaries will present the psychometric properties of each measure, and when available, results from head-to-head comparison studies of competing measures. Results will be presented in rounds 2 and 3 of the consensus exercise; summaries will present

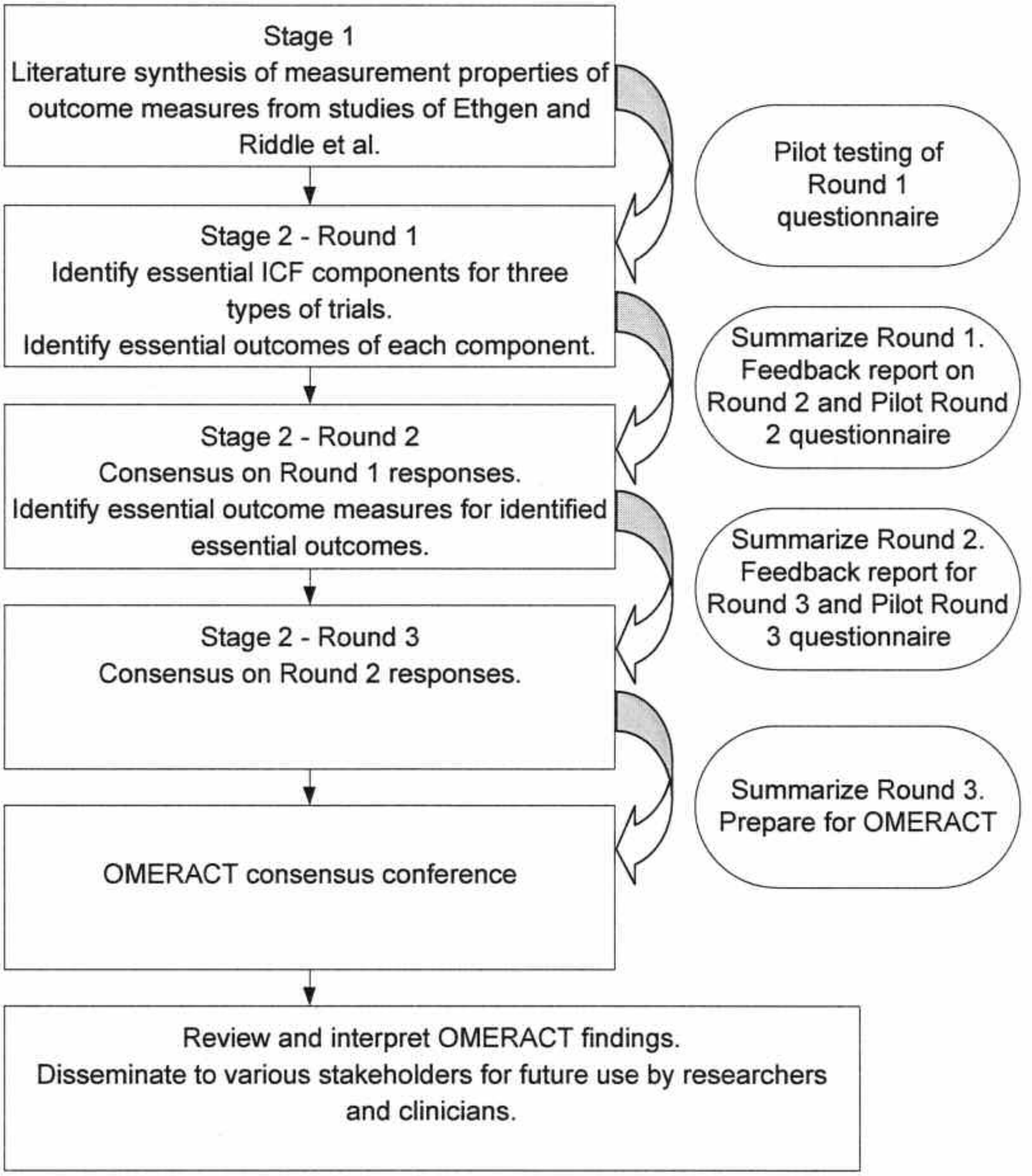

Stage 1
Literature synthesis of measurement properties of outcome measures from studies of Ethgen and Riddle et al.
Pilot testing of Round 1 questionnaire
Stage 2 - Round 1
Identify essential ICF components for three types of trials.
Identify essential outcomes of each component.
Summarize Round 1. Feedback report on Round 2 and Pilot Round 2 questionnaire
Stage 2 - Round 2
Consensus on Round 1 responses.
Identify essential outcome measures for identified essential outcomes.
Summarize Round 2. Feedback report for Round 3 and Pilot Round 3 questionnaire
Stage 2 - Round 3
Consensus on Round 2 responses.
Summarize Round 3. Prepare for OMERACT
OMERACT consensus conference
Review and interpret OMERACT findings.
Disseminate to various stakeholders for future use by researchers and clinicians.

Figure 2. Flow of the proposed study.

"typical" values or information rather than an exhaustive review of the literature concerning a measure. A glossary of terms will accompany the psychometric summaries.

Stage 2: Consensus Delphi approach. The Delphi procedure is a method for achieving consensus of opinion among a panel of experts on a topic when there is lack of agreement or an incomplete state of knowledge[12,13]. The classic Delphi method usually consists of 3 rounds of questions with the format of the first round questions being open-ended[12,13]. However, modified versions of this method abound and they are often categorized by semistructured or structured first round questions[12,15,16]. Our proposed consensus design will apply a modified Delphi method that combines structured and open-ended questions in the first round with a total of 3 rounds to achieve consensus. All communication with Delphi participants will be via email correspondence.

The principal determinant of sample composition is credibility with the target audience[13]. We believe the target audience to be those who conduct clinical trials in the field of hip or knee arthroplasty and consumers of their work. It is with this group in mind that we define our expert panel. Jairath and Weinstein suggest that experts should not only be knowledgeable with the topic area, but must also be impartial to the findings[17]. Delbecq, *et al* have noted that heterogeneous groups, categorized by panel members with different perspectives produce a higher proportion of high quality solutions than homogeneous groups[18].

Our expert panel will consist of a North American representation of orthopedic surgeons, rheumatologists, and physical therapists who have published a minimum of 5 peer reviewed papers relevant to the assessment of patients post hip or knee arthroplasty. At least half the experts will meet the criterion of 5 papers for the hip, the knee, or both. Our rationale for only selecting experts who have published peer-reviewed work is that these individuals have been evaluated by their peers and found to produce credible work. We chose this approach over, for example, select members of professional societies because we did not want potential selection to be politically driven. In addition, we will exclude those who published work describing the development of an outcome measure. Developers of outcome instruments may be particularly biased toward their own measures. Panel members will be identified from the body of literature dealing with hip and knee arthroplasty. The proportion of experts sampled from a specific discipline will be representative of the number of studies authored by members from that discipline, with the maximum representation from any one discipline being 66%. Because orthopedic surgeons are the providers of the intervention, they will represent two-thirds of the members of the panel. Authors will be stratified by discipline and geographical location and purposely sampled within strata. Preference will be given to authors who have contributed most to the literature, as determined by publication counts of researchers identified by the research team and by PubMed searches using key words (replacement OR arthroplasty) AND (hip OR knee).

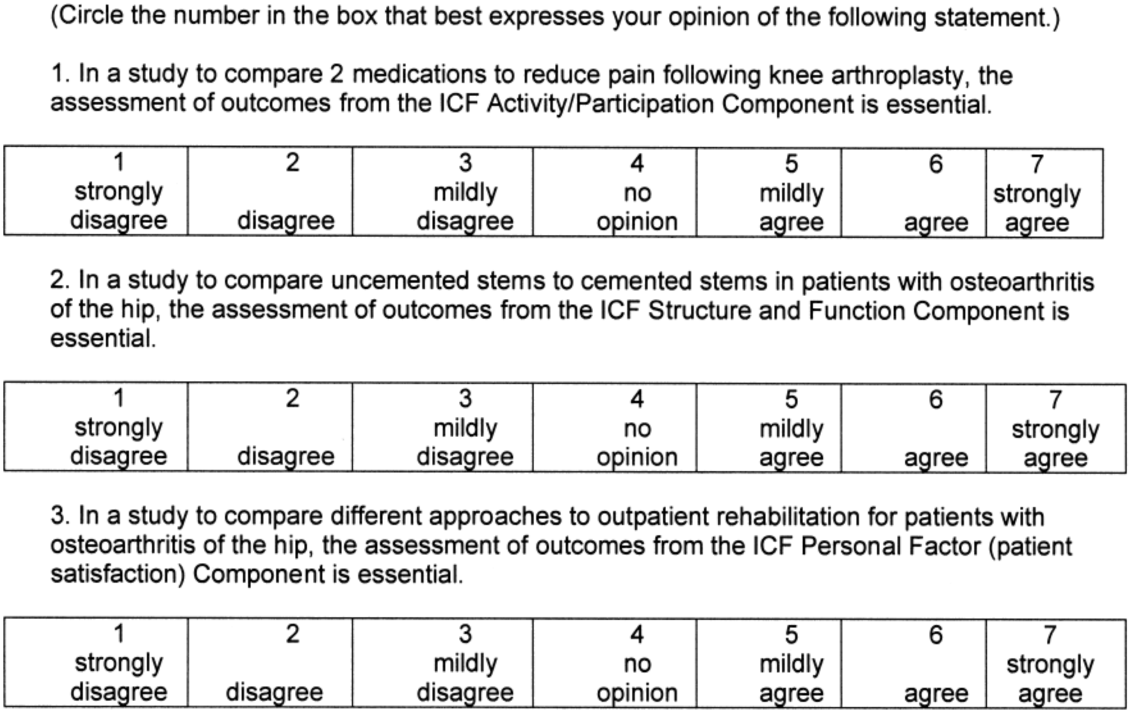
Delphi exercise panel sizes have varied widely (e.g., 10 to over 1000) and appear to be driven by available resources including the pool of likely participants, and the time and cost associated with managing and summarizing data[19]. Given the lack of consensus on a method for estimating the requisite sample size, our goal is to have complete data on 30 panel members. A sample of this size is consistent with other reported Delphi exercises in related areas[20,21]. In addition, sample sizes greater than 30 have seldom been found to improve results[22,23].

The goal of round 1 is to initiate the consensus process regarding "What should be measured?" This includes both ICF components and outcomes (not specific outcome measures) within components. Specifically round 1 questions will address: (a) the ICF components relevant to hip or knee arthroplasty with respect to surgical technique studies, nonsurgical physical interventions, and nonsurgical medical interventions; and (b) the relevant outcomes (not specific outcome measures) within ICF components. The expert panel will be provided with an introduction to the task and a summary of the information obtained from the Stage 1 literature review. Specifically, the information will list outcomes (e.g., pain, range of motion, functional status) under the appropriate ICF components and not the measures (e.g., WOMAC pain subscale, goniometer, 6-minute walk test) used to assess these outcomes. Using a structured question format, experts will be asked to rate on a 7-point Likert scale the extent to which each ICF component is essential to RCT targeting surgical techniques, nonsurgical physical interventions, and nonsurgical medical interventions of patients undergoing hip or knee arthroplasty (see Figure 3). We believe that consensus on ICF component items will be achieved in round 1 because the ICF components examined in this study are commonly assessed in most trials[5]. Hip arthroplasty will be assessed separately followed by knee arthroplasty. After primary arthroplasty is completed, the same approach will be applied to revision surgeries for the hip and then for the knee. Relevant outcomes for each ICF component also will be examined in Round 1.

Round 2 has 2 goals: (a) to achieve consensus on "What relevant outcomes should be measured?" and (b) to begin the consensus process on "How should the outcomes be assessed?" The round 2 information package will contain a summary of round 1 results and a review of the psychometric properties of measures that assess outcomes for which consensus was achieved from round 1. Item summary information will contain a ranking of outcomes within each ICF component. In addition, the median score, interquartile range, and a histogram of responses will be provided for each item. Each expert will be shown his/her round 1 item responses relative to group summary data. With respect to the question "How should the outcomes be assessed?" the psychometric properties of interest will include reliability and validity (cross-sectional and longitudinal). Also, a summary of the results from head-to-head comparison studies of competing measures will be provided. Once again, expert panel members will be asked to respond on the 7-point Likert scale as described above. Also, following each question a space will be for clarifying comments. In addition, expert panel members will also have the opportunity to add measures not identified in the phase 2 literature review.

The goal of round 3 is to continue building consensus for the question "How should the outcomes be assessed?" The round 3 information package will contain a summary of the round 2 results and a review of the psychometric properties of measures for which consensus was not achieved in round 2. Item summary information will contain a ranking of measures within each ICF component. Additionally, the median score, interquartile range, and a histogram of responses will be provided for each item. Each expert will be shown his/her round 2 item responses relative to the group summary data. The round 3 administration will replicate that described in round 2 pertaining to question "How should the outcomes be assessed?"

Reaching consensus and disseminating findings. Although there is no agreed method or standard for defining consensus or convergence of opinion, often a percentage level is applied when considering an item for inclusion[12,13]. Clearly, the choice of the percentage cutoff value for inclusion of an item is arbitrary, and values have varied widely (e.g., 55% to



(Circle the number in the box that best expresses your opinion of the following statement.)

1. In a study to compare 2 medications to reduce pain following knee arthroplasty, the assessment of outcomes from the ICF Activity/Participation Component is essential.

1 strongly disagree	2 disagree	3 mildly disagree	4 no opinion	5 mildly agree	6 agree	7 strongly agree

2. In a study to compare uncemented stems to cemented stems in patients with osteoarthritis of the hip, the assessment of outcomes from the ICF Structure and Function Component is essential.

1 strongly disagree	2 disagree	3 mildly disagree	4 no opinion	5 mildly agree	6 agree	7 strongly agree

3. In a study to compare different approaches to outpatient rehabilitation for patients with osteoarthritis of the hip, the assessment of outcomes from the ICF Personal Factor (patient satisfaction) Component is essential.

1 strongly disagree	2 disagree	3 mildly disagree	4 no opinion	5 mildly agree	6 agree	7 strongly agree

Figure 3. Examples of potential round 1 Delphi questions.

100%)[13,24]. For the proposed study we define consensus as having been met if 70% of the responding expert panel members endorse an item at the "agree" or "strongly agree" level (or in the negative "disagree" or "strongly disagree"). We chose the 70% criterion because this is the criterion generally supported by OMERACT[25,26]. Findings will be presented at a future OMERACT meeting along with thorough summaries of the literature synthesis and voting results obtained during the 3 rounds of the Delphi exercise.

Dissemination to clinical researchers will be accomplished via publication in the peer-reviewed literature. Results will be distributed to and discussed with the Centers for Medicare and Medicaid Services, the key US policy maker for patients with hip and knee arthroplasty, and to The American Academy of Hip and Knee Surgeons, The Hip Society, and The Knee Society. Availability of these standardized outcome measures will be critical for future US National Institutes of Health consensus and state-of-the-science conferences related to joint arthroplasty surgery and will enhance the application and generalizability of data collected through future federally and privately funded research.

REFERENCES

1. NIH Consensus Panel. NIH consensus statement on total knee replacement December 8-10, 2003. J Bone Joint Surg Am 2004;86:1328-35.
2 . Kurtz S, Mowat F, Ong K, Chan N, Lau E, Halpern M. Prevalence of primary and revision total hip and knee arthroplasty in the United States from 1990 through 2002. J Bone Joint Surg Am 2005;87:1487-97.
3. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am 2007;89:780-5.
4. Ong KL, Mowat FS, Chan N, Lau E, Halpern MT, Kurtz SM. Economic burden of revision hip and knee arthroplasty in Medicare enrollees. Clin Orthop Relat Res 2006;446:22-8.
5. Riddle DL, Stratford PW, Bowman DH. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: A systematic review. Arthritis Rheum 2008;59:876-83
6. Bellamy N, Kirwan J, Boers M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799-802.
7. Pham T, van der HD, Altman RD, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis Cartilage 2004;12:389-99.
8. Pham T, van der Heijde DM, Lassere M, et al. Outcome variables for osteoarthritis clinical trials: The OMERACT-OARSI set of responder criteria. J Rheumatol 2003;30:1648-54.
9. World Health Organization. International classification of functioning, disability and health (ICF). Geneva, Switzerland. 2001.
10. Grill E, Huber EO, Stucki G, Herceg M, Fialka-Moser V, Quittan M. Identification of relevant ICF categories by patients in the acute hospital. Disabil Rehabil 2005;27:447-58.
11. Grill E, Stucki G, Boldt C, Joisten S, Swoboda W. Identification of relevant ICF categories by geriatric patients in an early post-acute rehabilitation facility. Disabil Rehabil 2005;27:467-73.
12. Linstone HA, Turoff M. The Delphi method: Techniques and applications. Don Mills, ON: Addison-Wesley Publishing Company; 1975.
13. Powell C. The Delphi technique: myths and realities. J Adv Nurs 2003;41:376-82.

14. Cieza A, Ewert T, Ustun TB, Chatterji S, Kostanjsek N, Stucki G. Development of ICF Core Sets for patients with chronic conditions. J Rehabil Med 2004;44:9-11.
15. Duffield C. The Delphi technique: a comparison of results obtained using two expert panels. Int J Nurs Stud 1993;30:227-37.
16. Bond S, Bond J. A Delphi survey of clinical nursing research priorities. J Adv Nurs 1982;7:565-75.
17. Jairath N, Weinstein J. The Delphi methodology (part one): A useful administrative approach. Can J Nurs Adm 1994;7:29-42.
18. Delbecq AL, Van de Ven AH, Gustafson DH. Group Techniques for Program Planning: A Guide to Nominal and Delphi Processes. Glenview, IL: Scott, Foresman and Company; 1975.
19. Reid N. The Delphi technique: its contributions to the evaluation of professional practice. In: Ellis R, editor. Professional Competence and Quality Assurance in the Caring Professions. London: Chapman and Hall; 1988.
20. Mokkink LB, Terwee CB, Knol DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. BMC Med Res Methodol 2006;6:2.
21. Weigl M, Cieza A, Andersen C, Kollerits B, Amann E, Stucki G. Identification of relevant ICF categories in patients with chronic health conditions: a Delphi exercise. J Rehabil Med 2004;12-21.
22. de Villiers MR, de Villiers PJ, Kent AP. The Delphi technique in health sciences education research. Med Teach 2005;27:639-43.
23. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. Am J Public Health 1984;74:979-83.
24. Williams PL, Webb C. The Delphi technique: a methodological discussion. J Adv Nurs 1994;19:180-6.
25. Gladman DD, Strand V, Mease PJ, Antoni C, Nash P, Kavanaugh A. OMERACT 7 psoriatic arthritis workshop: synopsis. Ann Rheum Dis 2005;64 Suppl 2:ii115-ii116.
26. Kirwan J, Heiberg T, Hewlett S et al. Outcomes from the Patient Perspective Workshop at OMERACT 6. J Rheumatol 2003;30:868-72.