

Patient Perspective Workshop: Moving Towards OMERACT Guidelines for Choosing or Developing Instruments to Measure Patient-Reported Outcomes

JOHN R. KIRWAN, JAMES F. FRIES, SARAH E. HEWLETT, RICHARD H. OSBORNE, STANTON NEWMAN, SABINA CICIRIELLO, MART A. van de LAAR, EMMA DURES, PATRICIA MINNOCK, TURID HEIBERG, TESSA C. SANDERSON, CAROLINE A. FLUREY, AMY L. LEONG, PAMELA MONTIE, and PAM RICHARDS

ABSTRACT. The workshop Choosing or Developing Instruments held at the Outcome Measures in Rheumatology (OMERACT) 10 meeting was designed to help participants think about the underlying methods of instrument development. Conference pre-reading material and 3 brief introductory presentations elaborated the issues, and participants broke into discussion groups before reconvening to share insights, engage in a more general discussion of the issues, and vote on recommendations. Tradeoffs between using current imperfect measures and the long and complex process of developing new instruments were considered, together with the need for rigor in patient-reported outcome (PRO) instrument development. The main considerations for PRO instrument development were listed and a research agenda for action produced. As part of the agenda for action, it is recommended that researchers and patient partners work together to tackle these issues, and that OMERACT bring forward proposals for acceptable instrument development protocols that would meet an enhanced “Truth” statement in the OMERACT Filter. (J Rheumatol 2011;38:1711–15; doi:10.3899/jrheum.110391)

Key Indexing Terms:

PATIENT REPORTED OUTCOMES CONSENSUS DEVELOPMENT CLINICAL TRIALS

From the University of Bristol, Academic Rheumatology Unit, Bristol Royal Infirmary, Bristol, UK; Stanford University School of Medicine, Palo Alto, California, USA; University of the West of England, Academic Rheumatology Unit, Bristol Royal Infirmary, Bristol, UK; Deakin University, Public Health Innovation, Melbourne, Australia; University of London, Centre for Behavioural and Social Sciences in Medicine, London, UK; Centre for Rheumatic Diseases, The University of Melbourne, Department of Medicine, Royal Melbourne Hospital, Melbourne, Australia; Arthritis Centre Twente, University Twente and Medisch Spectrum Twente, Enschede, The Netherlands; Rheumatology and Rehabilitation, Our Lady's Hospice, Dublin, Ireland; Oslo University Hospital, Lovisenberg Diakonal College, Oslo, Norway; Healthy Motivation and Bone and Joint Decade, Santa Barbara, California, USA; and the Consumer Advisory Board, Arthritis Research Centre, Vancouver, BC, Canada.

J.R. Kirwan, MD, University of Bristol, Academic Rheumatology Unit, Bristol Royal Infirmary; J.F. Fries, MD, Stanford University School of Medicine; S. Hewlett, PhD, University of the West of England, Academic Rheumatology Unit, Bristol Royal Infirmary; R.H. Osborne, BSc, PhD, Deakin University, Public Health Innovation; S. Newman, PhD, University of London, Centre for Behavioural and Social Sciences in Medicine; S. Ciciriello, MBBS, Centre for Rheumatic Diseases, The University of Melbourne, Department of Medicine, Royal Melbourne Hospital; M.A. van de Laar, MD, Arthritis Centre Twente, University Twente and Medisch Spectrum Twente; E. Dures, PhD, University of the West of England, Academic Rheumatology Unit, Bristol Royal Infirmary; P. Minnock, RGN, MSc, Rheumatology and Rehabilitation, Our Lady's Hospice; T. Heiberg, PhD, Dean, Oslo University Hospital, Lovisenberg Diakonal College; T.C. Sanderson, PhD, University of the West of England, Academic Rheumatology Unit, Bristol Royal Infirmary; C.A. Flurey, MSc, University of the West of England, Academic Rheumatology Unit, Bristol Royal Infirmary; A.L. Leong, MBA, Healthy Motivation and Bone and Joint Decade, Santa Barbara; P. Montie, Consumer Advisory Board, Arthritis Research Centre; P. Richards, Academic Rheumatology Unit, Bristol Royal Infirmary.

Address correspondence to Dr. Kirwan.
E-mail: John.Kirwan@Bristol.ac.uk

AIM

Previous OMERACT meetings have focused on testing outcome measures once they are available. The Choosing or Developing Instruments Workshop at OMERACT 10 was one of 4 linked “Patient Perspective” workshops with the collective aim of turning attention towards guiding initial development of outcome measures to ensure they include or capture the patient’s perspective. It was designed to help participants think about the underlying methods of instrument development that are common to measuring patient perspective in different diseases and different settings. We sought to derive a set of principles or “quality standards” that will be applicable in general and thence to use these principles to approve future outcome measures.

METHODS

Pre-reading material. Conference pre-reading material¹ reviewed development of the patient perspective on outcome assessment at OMERACT over recent years. It drew participants’ attention to the need to consider the patient-reported outcome (PRO) instrument development process, as well as post-development performance testing. Three brief introductory presentations elaborated these issues, and then participants broke into discussion groups before reconvening to share insights, engage in a more general discussion of the issues, and vote on some recommendations. The main conclusions and recommendations from the workshop were taken to the final OMERACT plenary session for consideration in con-

junction with those from the other patient perspective workshops, and to seek endorsement for the next step forward².

Introductory presentations. In the introductory presentations, J. Fries discussed the innovative approach to instrument development taken by the patient-reported outcome information system (PROMIS) organization³, and the potential for this large item bank to be used efficiently following evaluation by item response theory (IRT) methodology, and particularly in combination with computer adaptive testing⁴. Using the example of physical function, it was shown that floor and ceiling effects were substantially reduced compared to current instruments widely used in rheumatology, such as the Health Assessment Questionnaire. Fries suggested that the time had now come to use the PROMIS system directly in clinical trials and compare its performance directly to current PRO outcome instruments.

S. Hewlett reviewed the development of the Bristol Rheumatoid Arthritis Fatigue (BRAFF) scales, a suite of 3 visual analog scales (VAS) and a 20-item questionnaire that contains 4 robust independent dimensions. The intense iterative process involving patients as both participants/subjects and research partners was described, and the specific benefits of this interaction and changes incorporated into the questionnaire as a result were enumerated⁵. In particular, the development of both detailed question wording and the overall structure of the questionnaire was influenced by a combination of different types of patient engagement, such as participants thinking aloud as they completed the questions, and a patient research partner gauging the overall flow of the questions and how that might affect patients' readiness to complete the whole questionnaire⁶.

R.H. Osborne presented a digest of classical and modern approaches to instrument development. He emphasized the importance and challenge of the first step in questionnaire development: ensuring that both the questionnaire items and the construct being measured reflect, as far as possible, the patients' and clinicians' experiences. Structured approaches to item and construct development were discussed. These include well considered evidence-based item writing rules using grounded nominal group consultation techniques. Reflective clinicians and patients from the "coal face of care" should be consulted to ensure content and face validity⁷. Osborne recommended the use of Rasch analysis and confirmatory factor analysis in a 2-step process for item selection and scale construction. Consultation with other stakeholders, such as commissioners and managers of healthcare, was also highlighted as an important element in establishing what needs to be measured and how. Examples of the processes were drawn from the development of the multidimensional patient-centered questionnaire that measures the impact osteoarthritis (OA) has on individuals (called the OA Quest⁸) and the Multi-attribute Prioritisation Tool for hip and knee joint replacement surgery⁹.

Breakout group presentations. Breakout groups, prior to their

own discussions, first listened to one of 4 mini-presentations used to illustrate some of the methods mentioned in the main presentations. In one such presentation an approach to the development of a robust and accurate patient-reported "uncertainty" outcome measure to be validated and applied in systemic lupus erythematosus (SLE) was reported by S. Newman. Feelings of uncertainty have been reported to hinder the emotional adjustment in general¹⁰ and in SLE patients in particular^{11,12,13,14}. Nevertheless, uncertainty has never been explored empirically in rheumatology because of the lack of both a conceptualization of uncertainty and hence an absence of an adequate and valid tool to measure it¹⁵. Newman and colleagues aim to address both issues in developing a PRO measure that will reflect the patients' perspective of the uncertainty experience.

To broaden the investigation of uncertainty and provide a contrast to SLE, patients with rheumatoid arthritis (RA) were also recruited as they have also reported experiencing uncertainty issues^{16,17}. The systematic approach to development of the instrument involves 3 phases. First, a qualitative investigation has been conducted including a review of the literature across chronic conditions, consultation with rheumatology healthcare professionals (HCP; n = 8) to explore their understanding of uncertainty issues in SLE, and in-depth, semi-structured qualitative interviews with SLE patients (n = 17) and RA patients (n = 15). Transcripts of audiotaped interviews are being coded for uncertainty themes and these will be combined with information from the literature review and HCP opinion to develop a conceptual model. It is intended that item generation for the uncertainty PRO will be conducted on the basis of the conceptual model and wherever possible the words of the patients will be used in the PRO items. Further development will include a postal survey of 400 patients for statistical validation, including thresholds for item response options; item fit statistics; item locations; person separation index and traditional psychometric analyses including standard tests for reliability and validity.

Finally, a cross-sectional cohort study will be conducted using an SLE sample (n = 200) to further explore the construct validity of the new uncertainty PRO in comparison with existing related measures and measures of quality of life and general well-being.

S. Ciciriello reported on the use of concept mapping workshops¹⁸ with patients that incorporate nominal group techniques¹⁹ to inform development of new questionnaires. The critical advantage of this approach is that the perspective of patients is collected in a manner that is not influenced or biased by the researcher. A carefully crafted seeding statement is presented to patients, who work alone, to generate ideas in response to the statement. These ideas are presented, one by one in an egalitarian manner, and recorded with their meaning clarified by the group if necessary. All the ideas are placed on cards and sorted (grouped) individually by patients in any way they see fit. The sort data are entered into a computer, and

multidimensional scaling and cluster analyses then produce a “cluster map,” which is a visual representation of the patients’ ideas and how these ideas are interrelated. The wording that patients use in the workshop is preserved and feeds straight into questionnaire items within a comprehensive set of well defined constructs.

The process of concept mapping was illustrated using the development of the Medication Education Impact Questionnaire (MeiQ)²⁰ and the Methotrexate in Rheumatoid Arthritis Knowledge test (MiRAK)²¹. The MeiQ measures the impact of education about medications in meeting patient needs. It consists of 29 questions across 6 scales: information quality; active communication; coming to terms with diagnosis and treatment; self-management role and responsibility; capacity to self-manage; and self-management support. The MiRAK measures knowledge about methotrexate. The questionnaires have undergone rigorous testing with stakeholders, and have also undergone stringent calibration, validation, and test-retest reliability tests. These have shown that the questionnaires have good content and construct validity, internal consistency, and stability over time. Importantly, the concept mapping resulted in patient-centered questionnaires with well defined short scales, which are highly pertinent to patients.

M.A. van de Laar reported on the Q-sort methodology developed by William Stephenson. The “Q-method” helps to study people’s “subjectivity,” that is, their viewpoint. This contrasts with the traditional factor analysis or “R method,” which involves correlations between variables across a sample of subjects. The name “Q” comes from the form of factor analysis that is used to analyze the data. Q looks for correlations between subjects across a sample of variables²². Q factor analysis reduces the many individual viewpoints of the subjects down to a few “factors,” which represent shared ways of thinking. As an example a study was discussed on fatigue in RA. For this study we questioned: Is it possible to categorize patients according to their fatigue experience? Based on early interview studies as well as on existing measurement instruments for fatigue, a selection of 200 items was used. These 200 items were reduced to 57 relevant statements for RA fatigue. Patients evaluated these fatigue items on cards by sorting them according to a normal distribution. A 4-factor model categorized 2/3 of the patients. These factors can be described as: Little impact of fatigue; Good coping and bad sleep; High distress; Search for balance. This information will be used to create an IRT calibrated item bank to be used in a computer adaptive test instrument that is expected to be finalized at the end of 2011^{23,24}.

E. Dures reported on further studies of fatigue in RA, which is a major problem for patients, but measurement is a challenge. A systematic review by Hewlett, *et al*²⁵ concluded that currently there is no valid tool to measure fatigue in UK patients with RA. Subsequently, 3 years of careful and rigorous development led to the Bristol RA Fatigue group of scales^{26,27}. The BRAF Multi-Dimensional Questionnaire

(BRAF-MDQ) is a 20-item instrument that contains 4 separate subscales (Living with fatigue, and physical, emotional, and cognitive fatigue). The BRAF short scales are 3 single questions on fatigue severity, coping, and effect, validated in visual analog and numerical rating scale versions (BRAF VAS and BRAF NRS). If these PRO are to reflect fatigue, then it is crucial that patients understand them in the way the research intended (and those intentions were based on qualitative data from patients). Qualitative data, focus groups, and “Think Aloud” cognitive interviewing were therefore employed to develop wording for the short scales and to review the draft BRAF-MDQ questions. This exposed some key areas of mismatch between researchers’ intent and patient interpretation, and led to rewording of some stem questions, some response items, and some layout of questions, which strengthened the confidence users can have in these scales as being easily understood by patients. Currently the 2 final stages of evaluating this new PRO are under way. Test-retest reliability raises challenges, because of the inherent fluctuating nature of RA fatigue. Sensitivity to change or responsiveness also raises issues about selecting an appropriate intervention and time-frame within which fatigue should improve.

OUTCOME OF DISCUSSIONS

Breakout discussions in groups of 8–15 were asked to consider their own experiences of working with instrument development, the practical difficulties in ensuring rigorous processes, and the potential usefulness (or otherwise) of OMERACT, specifying in more detail how these aspects may be incorporated into the “Truth” section of the OMERACT Filter²⁸. The verbal reports back to the workshop from the breakout group reporters (together with later written reports), plenary session discussions, and subsequent informal discussions were drawn together, and a summary presented at the final OMERACT 10 plenary session. The main considerations are shown in Table 1 and an action program shown in Table 2. These consist of a mixture of existing information and possible new areas of research.

There was a general appreciation that, while we should strive for rigor and quality, there should be the recognition that exact measures do not exist. This point emerged during discussion about tradeoffs between using current imperfect measures where they exist and the long and complex process of developing new instruments. It was also observed that new measurement approaches such as those proposed by the PROMIS group need to demonstrate that the theoretical benefits can indeed be realized in practice²⁹. Although existing measures might not meet modern standards for questionnaire design³⁰, particularly in the areas of content and construct validity for rheumatological conditions, they may remain useful in the absence of better measures, for example, in the area of sleep quality in RA, where there are good existing generic measures³¹; but a contrary example might be the need for development of a fatigue scale that adequately captures the

Table 1. Main considerations from the discussion group reports.

In relation to instruments

- Patients should know why they are completing an instrument — but this may present challenges in avoiding bias
- Recognize the importance of use within the population in which it was developed
- Trial instruments in appropriate normative populations
- Do not reinvent the wheel, but there are gaps in the literature, and new questionnaires need to be developed on occasions
- Instrument use in clinical trials should be as close as possible to routine clinical practice to enhance the ability to transfer findings to routine practice
- Personalized instruments may help with personalized medicine
- Instruments have to be understood by the responders

In relation to developmental methods

- Nominal group techniques need expert facilitators (for trust, confidentiality, and inclusion)
- Standards for facilitators exist and should be applied
- Patient partner involvement in nominal groups is feasible, important, and valuable
- Patient partner involvement helps with preparations for and running groups; seeding statements; recruitment
- Patients raising issues makes clinicians/researchers think more widely in clinical practice
- Translation of instruments in culture and language requires both forward and backward translation and cultural adaptation
- International collaboration at an early stage; avoid unnecessary duplication
- Reduce responder burden
- Mixed methods approaches are critical (i.e., both qualitative and quantitative approaches). Suggested that international collaborative research begin early in concept, domain analysis, and ultimate item bank generation for instrument development in order to optimize resources, i.e., financial, time, effort, outputs

Table 2. Proposed agenda for action

- Explore how patient knowledge and understanding of measures and their meaning might alter the way questionnaires are completed
- Evaluate and recommend patient partner involvement
- Make available a list of content and methodology experts
- Develop a statement of broad principles for development, in particular with reference to the ‘Truth’ section of the OMERACT Filter
- Compare the techniques through research:
 - What is the best approach and when should each approach be used?
 - What biases are inherent in the available techniques?
 - What is the cost of each approach?
- Compile a list of pros and cons of different methods for the OMERACT web site
- Evaluate ways patient partners contribute to different methods and make specific recommendations
- Compile a compendium of PRO instruments
- Consider providing support for participants engaged in these developments across OMERACT, and advice to help researchers in the field to choose their instruments
- Explore the literature for gaps in instruments currently available to evaluate recognized domains
- Undertake rigorous instrument development to match identified gaps

experience of people with RA^{26,27}. Thus 2 parallel pathways can be followed. First, start using acceptable instruments for the assessment of PRO but take account of their limitations. Second, continue to improve specific as well as generic instruments for the assessment of PRO.

Another issue raised in several discussion groups was the importance of the setting (which disease, which circumstances) and purpose for which a measure was developed and within which it was validated. It is not acceptable to make assumptions that an instrument developed with a specific population can be used generically.

Patients particularly felt that when they complete an instrument they should be informed why they are being asked to do so, and how the results will be used. However, there is the possibility that provision of detailed information may bias data collection and sampling and may therefore compromise the overall mission of the research. This was an unexpected consideration, and it is recommended that researchers and patient part-

ners work together to develop information that is both respectful and enhances the research outcomes for each study on a case by case basis.

Agenda for action and voting. The agenda for action included the potential for OMERACT to have available appropriate references to validated instruments, and contact points for experts in the field of instrument development and testing. There was a feeling that the OMERACT filter might be enhanced if there was advice about the development of instruments incorporated into the “Truth” section of the filter.

At the end of this process participants were asked to vote on 2 primary questions. First, should OMERACT collate a set of principles and procedures to support the OMERACT community to develop high quality questionnaires? Participants voted 93% in favor. Second, would it be helpful to have evidence-based guidelines (quality standards) for the development of instruments to measure patient-reported outcomes? Participants voted 85% in favor. The main points from the dis-

cussion group feedback and the proposed agenda for action were carried forward to a final plenary session, which considered the main issues from all 4 Patient Perspective workshops taken together².

REFERENCES

1. Kirwan JR, Fries JF, Hewlett S, Osborne R. Patient perspective: Choosing or developing instruments. *J Rheumatol* 2011;38:1716-9.
2. Kirwan J, Tugwell PS. Overview of the patient perspective at OMERACT 10 — Conceptualizing methods for developing patient-reported outcomes. *J Rheumatol* 2011;38:1699–701.
3. Patient-Reported Outcomes Measurement System [homepage on the Internet. Accessed March 30, 2011.] Bethesda: National Institutes of Health. Available from <http://www.nihpromis.org/default.aspx>
4. Aletaha D. From the item to the outcome: the promising prospects of PROMIS. *Arthritis Res Ther* 2010;12:104.
5. Nicklin J. The development of scales to measure fatigue in people with rheumatoid arthritis [dissertation]. Bristol; University of the West of England; 2009: 449 p.
6. Nicklin J, Cramp F, Kirwan J, Hewlett S. Using “Think Aloud” to improve questionnaire design [abstract]. *Arthritis Rheum* 2008;58 Suppl:S868.
7. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): An outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. *Patient Educ Couns* 2007;66:192-201.
8. Busija L. Personal and societal burden of osteoarthritis [dissertation]. Melbourne: University of Melbourne; 2010: 481 p.
9. Witt J, Scott A, Osborne RH. Designing choice experiments with many attributes. An application to setting priorities for orthopaedic waiting lists. *Health Economics* 2009;18:681-96.
10. Mast ME. Adult uncertainty in illness: A critical review of the research. *Sch Inq Nurs Pract* 1995;9:3-24; discussion 25-9.
11. Failla S, Kuper B, Nick T, Lee F. Adjustment of women with systemic lupus erythematosus. *Appl Nurs Res* 1996;9:87-96.
12. Wiginton KL. Illness representations: mapping the experience of lupus. *Health Educ Behav* 1999;26:443-53.
13. Mendelson C. Managing a medically and socially complex life: women living with lupus. *Qual Health Res* 2006;16:982-97.
14. Stockl A. Complex syndromes, ambivalent diagnosis, and existential uncertainty: The case of systemic lupus erythematosus (SLE). *Soc Sci Med* 2007;65:1549-59.
15. Politi MC, Han PK, Col NF. Communicating the uncertainty of harms and benefits of medical interventions. *Med Decis Making* 2007;27:681-95.
16. Wiener CL. The burden of rheumatoid arthritis: tolerating the uncertainty. *Soc Sci Med* 1975;9:97-104.
17. Nyman C, Lutzen K. Caring needs of patients with rheumatoid arthritis. *Nurs Sci Q* 1999;12:164-9.
18. Trochim WMK. An introduction to concept mapping for planning and evaluation. *Evaluation Program Planning* 1989;12:1-16.
19. Van de Ven AH, Dalbecq AL. The nominal group as a research instrument for exploratory health studies. *Am J Public Health* 1972;62:337-42.
20. Ciciriello S, Buchbinder R, Wicks IP, Osborne RH. Measuring the impact of patient education on meeting patient needs: Development of the Medication Education Impact Questionnaire (MeiQTM). *Intern Med J* 2010;40 Suppl 3:9.
21. Ciciriello S, Wicks IP, Osborne RH, Buchbinder R. Development and validation of the Methotrexate in Rheumatoid Arthritis Knowledge Test (MiRAK™). *Intern Med J* 2010;40 Suppl 3:9.
22. Brown S. Q-methodology and qualitative research. *Qual Life Res* 1996;6:561-7.
23. Nikolaus S, Bode C, Taal E, van de Laar MA. Four different patterns of fatigue in rheumatoid arthritis patients: results of a Q-sort study. *Rheumatology* 2010;49:2191-9.
24. Lai J-S, Dineen K, Reeve B, Von Roenn J, Shervin D, McGuire M, et al. An item response theory based pain item bank can enhance measurement precision. *J Pain Symptom Manage* 2005;30:278-88.
25. Hewlett S, Hehir M, Kirwan J. Measuring fatigue in rheumatoid arthritis: a systematic review of scales in use. *Arthritis Rheum* 2007;57:429-539.
26. Nicklin J, Cramp F, Kirwan J, Urban M, Hewlett S. Collaboration with patients in the design of patient reported outcome measures: Capturing the experience of fatigue in rheumatoid arthritis. *Arthritis Care Res* 2010;62:1552-8.
27. Nicklin J, Cramp F, Kirwan J, Greenwood R, Urban M, Hewlett S. Measuring fatigue in RA: A cross sectional study to evaluate the Bristol Rheumatoid Arthritis Fatigue Multi-dimensional questionnaire, visual analog and numerical rating scales. *Arthritis Care Res* 2010;62:1559-68.
28. Boers M, Brooks P, Strand V, Tugwell P. The OMERACT Filter for outcome measures in rheumatology. *J Rheumatol* 1998;25:198-9.
29. Boers M. Standing on the promises: First wave validation reports of the Patient-Reported Outcome Measurement Information System. *J Clin Epidemiol* 2010;63:1167-8.
30. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.
31. Wells GA, Li T, Kirwan JR, Peterson J, Aletaha D, Boers M, et al. Assessing quality of sleep in patients with rheumatoid arthritis. *J Rheumatol* 2009;36:2077-86.