

Domains Selection for Patient-Reported Outcomes: Current Activities and Options for Future Methods

PETER S. TUGWELL, INGEMAR F. PETERSSON, MAARTEN BOERS, LAURE GOSSEC, JOHN R. KIRWAN, TAMARA RADER, TESSA C. SANDERSON, MART A. van de LAAR, ERIN UEFFING and JAMES P. WITTER

J Rheumatol 2011;38;1702-1710
<http://www.jrheum.org/content/38/8/1702>

1. Sign up for TOCs and other alerts
<http://www.jrheum.org/alerts>
2. Information on Subscriptions
<http://jrheum.com/faq>
3. Information on permissions/orders of reprints
http://jrheum.com/reprints_permissions

The Journal of Rheumatology is a monthly international serial edited by Earl D. Silverman featuring research articles on clinical subjects from scientists working in rheumatology and related fields.

Domains Selection for Patient-Reported Outcomes: Current Activities and Options for Future Methods

PETER S. TUGWELL, INGEMAR F. PETERSSON, MAARTEN BOERS, LAURE GOSSEC, JOHN R. KIRWAN, TAMARA RADER, TESSA C. SANDERSON, MART A. van de LAAR, ERIN UEFFING, and JAMES P. WITTER

ABSTRACT. *Introduction.* Over the years, Outcome Measures in Rheumatology Clinical Trials (OMERACT) has worked toward consensus on core sets for outcome measurement in specific rheumatologic diseases. OMERACT core sets refer to the minimum number of domains and instruments essential to address the desired outcomes in trials. “Domains” are the attributes of an activity or function. This article discusses the need for an open process for selecting domains, existing frameworks for choosing domains, and the importance of describing the methods for selecting domains.

Methods. We reviewed the domains selection process of 3 OMERACT groups working on patient-reported outcomes (PRO). We categorized these methods in a hierarchy of comprehensiveness and examined the extent to which they address related issues.

Results. There was agreement that a gold standard for domain selection would include 3 important aspects: following a framework, remaining true to the clinical question, and including the clinically relevant outcomes for both benefits and harms.

Discussion. OMERACT participants agreed that a guide for the options for developing domains that meet the OMERACT Filter would be useful. More discussion and explanation is needed to outline outcomes related to the patient perspective that are not covered by the current version of the International Classification of Functioning, Disability and Health (ICF) and to explain the usefulness of the population/intervention/comparison/outcome (PICO) structure in domain selection. Future OMERACT work includes addressing these issues and developing a framework based on the ICF to support comprehensive outcome measurements. (J Rheumatol 2011;38:1702–10; doi:10.3899/jrheum.110389)

Key Indexing Terms:

OUTCOME MEASURES RHEUMATOID ARTHRITIS PATIENT REPORTED OUTCOMES

A pivotal component of Outcome Measures in Rheumatology Clinical Trials (OMERACT) is its focus on patient-reported outcomes (PRO). There are several working groups involved in developing and evaluating various generic methodologic and disease-specific aspects of PRO. As a reflection of their importance at OMERACT 10, 3 plenary sessions were committed to PRO, one devoted to selection of domains, one to selection of instruments, and one to methods for assessing and satisfying the Responsiveness Criteria of the OMERACT Filter. This article summarizes the issues and examples discussed in the Domains Session.

Within OMERACT, the word “domain” has been loosely used as a pointer to relevant areas of outcome in rheumatology studies. Outcome has been defined as how a patient feels, functions, and survives, and OMERACT has frequently pointed to the 5 D’s suggested by Fries: disability, discomfort and pain, drug side effects (toxicity), dollar costs, and death¹. The US Food and Drug Administration (FDA) has recently issued guidance for the development of a PRO instrument as outlined in Figure 1². The processes recommended by OMERACT are entirely consistent with this FDA guidance and indeed provide examples of how these criteria can be

From the Faculty of Medicine, University of Ottawa; Department of Medicine, Ottawa Hospital; Ottawa Hospital Research Institute; and Institute of Population Health, University of Ottawa, Ottawa, Canada; Orthopaedics and Rheumatology, Department of Clinical Sciences, Lund University, Lund, Sweden; Department of Epidemiology and Biostatistics, VU Medical Center, Amsterdam, The Netherlands; Paris Descartes University, Cochin Hospital, Rheumatology B Department, Paris, France; University of Bristol Academic Rheumatology Unit, Bristol Royal Infirmary; and University of the West of England, Bristol, UK; University of Twente, Enschede, The Netherlands; and National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health (NIH), Department of Health and Human Services (DHHS), Bethesda, Maryland, USA.

P. Tugwell is supported in part by the Canadian Institutes of Health Research.

P. Tugwell, MD, MSc, FRCPC, FCHS, Canada Research Chair in Health Equity; Professor of Medicine and Epidemiology, Faculty of Medicine and Principal Scientist, Institute of Population Health, University of Ottawa;

Department of Medicine, Ottawa Hospital; Senior Scientist, Ottawa Hospital Research Institute; I.F. Petersson, MD, PhD, Associate Professor of Rheumatology, Orthopaedics and Rheumatology, Department of Clinical Sciences, Lund University; M. Boers, MD, PhD, Department of Epidemiology and Biostatistics, VU Medical Center; L. Gossec, MD, PhD, Rheumatology B Department, René Descartes University, Cochin Hospital (AP-HP); J.R. Kirwan, MD, Professor of Rheumatic Diseases, Academic Rheumatology Unit, Bristol Royal Infirmary; T. Rader, MLIS, Knowledge Translation Specialist, Cochrane Musculoskeletal Group, University of Ottawa; T.C. Sanderson, PhD, MSc, Postdoctoral Fellow, Faculty of Life Sciences, University of the West of England; M.A. van de Laar, MD, PhD, Arthritis Centre Twente, Department of Psychology and Communication of Health and Risk, University of Twente; E. Ueffing, MHS, Administrator, Campbell and Cochrane Equity Methods Group, University of Ottawa; J.P. Witter, MD, PhD, Program Director, Rheumatic Diseases Clinical Program, Division of Skin and Rheumatic Diseases, NIAMS, NIH, DHHS.

Address correspondence to T. Rader; E-mail: trader@uottawa.ca

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2011. All rights reserved.

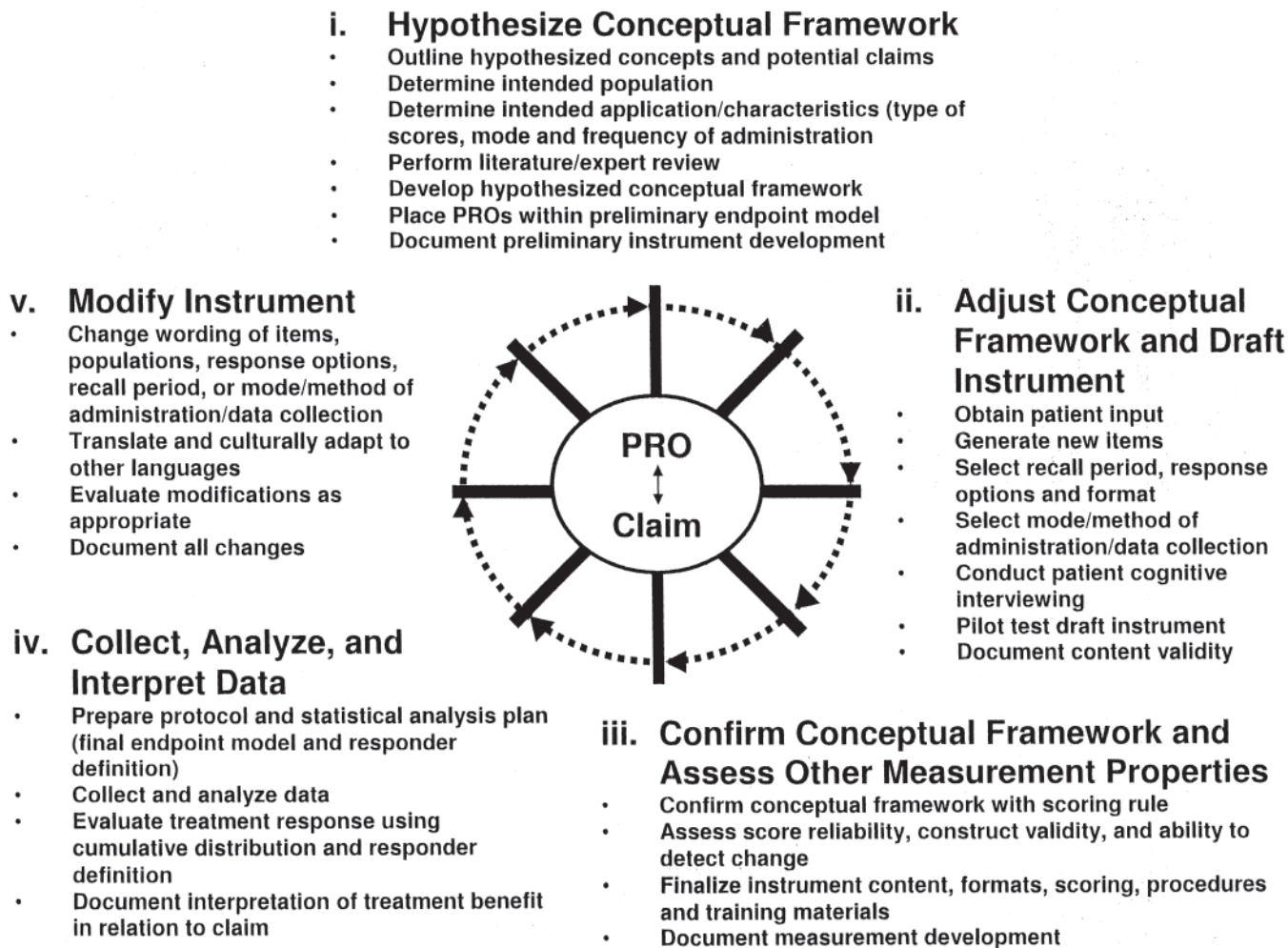


Figure 1. Development of a patient-reported outcome (PRO) instrument: an iterative process. Reprinted from US Department of Health and Human Services².

met. The selection of domains corresponds with Steps (i), (ii), and (iii) in Figure 1. While domains refer to what should be measured, instruments specify how those domains should be measured. The typical process is that first, the domains considered relevant are selected, and then instruments that measure or assess these domains are identified. The main reason to revisit domain selection is increasing insight that the instruments now used, although proven useful, have shortcomings that can be addressed by new methods. Present instruments often address several concepts that cross domains, or have diverse response categories that can result in disordered responses. The first step in instrument development is to again answer the question “What should be measured?”³. This article reviewed the domain selection process of 3 OMERACT groups working on PRO. We categorize these methods in a hierarchy of comprehensiveness, examine the extent to which they address related issues, and discuss the consistency of their process with activities of the FDA. Several key issues were discussed as important considerations in domain selec-

tion. We provide an agenda for future OMERACT meetings and interim work.

Following the FDA approach for domain selection encourages the formulation of a conceptual framework. There are a number of these in use, for example, the International Classification of Functioning, Disability and Health (ICF) and the Patient-Reported Outcomes Measurement Information System (PROMIS).

ICF Framework

The ICF framework (Figure 2)⁴ has replaced the previous spectrum of impairment > disability > handicap, with new domains described from the perspective of the body, the individual, and society in 2 basic categories: (1) Body Functions and Structures (system level); and (2) Activities and Participation (person level and person-environment interaction). In ICF, “functioning” is defined as a “generic term which includes body functions and structures, activities and participation. It indicates the positive aspects of the interac-

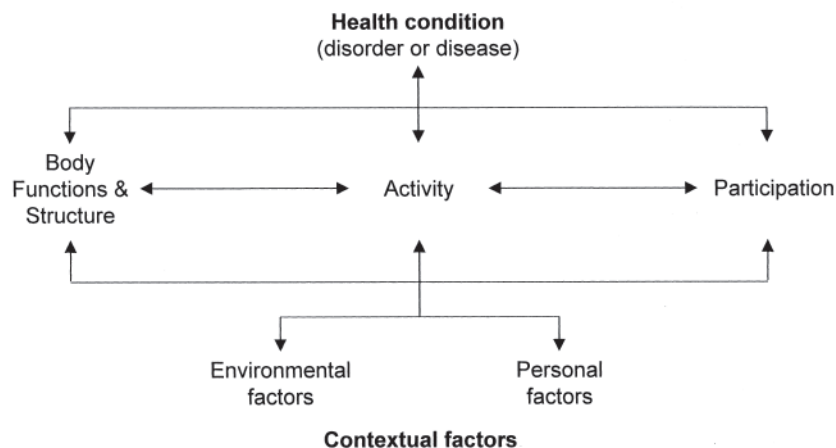


Figure 2. International Classification of Functioning, Disability and Health model of disability. Copyright © World Health Organization (WHO); 2011. All rights reserved.

tion between the individual (with a health condition) and context (personal and environmental factors)^{5,6}. The ICF also offers a hierarchical classification of “domains or categories,” which includes definitions. These domains contained in ICF can be seen as health domains and health-related domains. To improve feasibility of the large and comprehensive ICF, the ICF core sets project was initiated in 2001. The process of the ICF core set addresses the issue of context, specifying the number of categories to be assessed, and focussed on the condition, setting, and preferences of patients and clinicians. The process of disease-specific core sets of ICF categories demonstrates a case of domain identification with a robust process that has achieved acceptance by both rheumatologists and allied health professionals. The ICF core set project attempts explicitly to identify the minimum necessary to adequately represent the experience of most people with the disease of interest. In OMERACT, the ICF Reference Group is tasked to explore and examine the viable and practical interfaces between the OMERACT projects and the ICF to advance outcomes measurement in rheumatology².

PROMIS Framework

The second approach is the PROMIS, another major initiative in domains based upon the World Health Organization’s physical, mental, and social framework. The PROMIS project has expanded this to establish the “detailed articulation of subordinate domains beneath the broad physical, mental, and social headings,” as shown in Figure 3^{7,8}.

PROMIS approach to domain determination. The PROMIS method is based on the concept of domains as “latent traits.” PROMIS provides a strict and well documented procedure using literature reviews, Delphi technique, and empirical data (analyzed both qualitatively and quantitatively) for development of new domains. The process is linked strongly to the interaction between the patient and the physician/health pro-

fessional. The process will be described using PROMIS Domain Hierarchy 2010 and the PROMIS 18 Steps. One aspect of the PRO is that a previously selected domain can sometimes benefit from refinement or further structuring.

CASE STUDIES

During the breakout groups at OMERACT 10, 4 cases illustrating different approaches to domain identification were presented and discussed: (1) Development of the Rheumatoid Arthritis Impact of Disease (RAID) scale; (2) derivation of the Rheumatoid Arthritis Patient Priorities for Pharmacological Interventions (RAPP-PI) outcomes; (3) an exploration of how fatigue emerged as a domain of relevance in RA; and (4) using Q-sort as a method of identifying groups of patients with different domains of interest — the case of fatigue.

Case 1: Rheumatoid Arthritis Impact of Disease (RAID) Scale

The preliminary RAID score is a patient-derived weighted score to assess the impact of rheumatoid arthritis (RA). By impact of disease, we mean outcomes (such as fatigue or well-being) that pertain to domains not captured in the usual PRO of pain, function, and patient global assessment. This score can be used in clinical trials as a new composite index that captures information relevant to patients. The methodology is relevant with a focus on identification of domains of major importance to patients. The methodology is applicable for use in other diseases — both rheumatic and nonrheumatic — and in the development of new instruments.

Initial choice of domains. Ten patients with RA, one from each of 10 European countries, met in Zurich in March 2007. All had definite RA according to the American College of Rheumatology (ACR) criteria⁹, spoke English, and were selected by the principal investigators in each country. They had varying experience in research; 3 were members of the

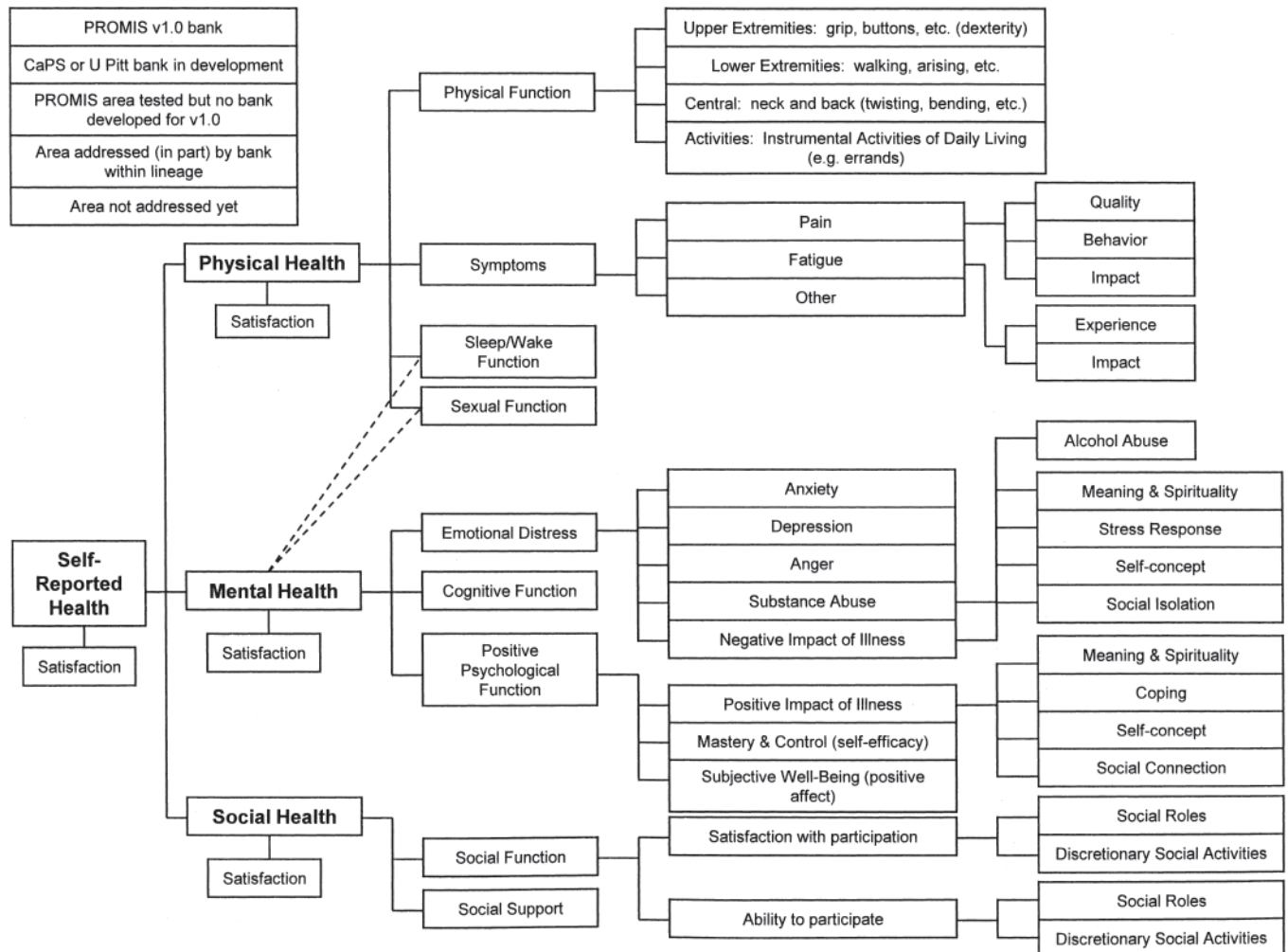


Figure 3. Patient-Reported Outcomes Measurement Information System® (PROMIS®) domain framework⁸. © 2011 PROMIS Health Organization and PROMIS Cooperative Group. Reprinted with permission.

OMERACT patient group. The patients were presented with an extensive literature review on domains of health in RA. During a group discussion and in 3 successive sessions, the participants identified domains of health important for the patient based on their personal experience.

Ranking of domains. For feasibility, the steering committee decided to include in the composite score a maximum of 7 domains. After the first step, the resulting number of domains was too large, so to reduce the number of domains and to obtain better representation a “ranking” strategy was designed. One hundred patients with RA (10 in each country) were contacted by the principal investigator and/or by the patient representative. All had definite RA; there were no other selection criteria. The names of the domains obtained in the previous step were translated into 12 languages with a brief explanation and presented as a list in random order. The participants were asked to rank the domains in order of decreasing importance by giving a number between 1 (most important) and 17 (least important) to the 17 domains. No

other data were collected at this stage (May to June 2007). The 7 highest-ranked domains were retained in the RAID score¹⁰.

Case 2: The RAPP-PI Outcomes

An exploration of how RAPP-PI chose domains. In the RAPP-PI project, a mixed-methods approach was used to develop a patient-generated set of priority treatment outcomes, using in-depth interviews, nominal groups, and a postal survey with RA patients in the UK. The 8 outcomes forming the RAPP-PI were pain, activities of daily living, joint damage, mobility, life enjoyment, independence, fatigue, and valued activities¹¹.

In-depth interviews were conducted with 26 RA patients, sampled purposively for age, sex, medication (anti-tumor necrosis factor or other disease-modifying antirheumatic drugs), disease severity, and work status. Grounded theory guided iterative data collection and analysis. Coding of the data was peer reviewed. A patient research partner collaborated in the research design and analysis.

Nominal group study: patient importance and prioritization of treatment outcomes. Nominal groups¹² were used with the patients (experts with valuable knowledge of living with RA) to rate the importance of and to prioritize the 63 outcomes previously generated in the interview study¹³. Patients with RA⁹ attending outpatient appointments were identified from clinical notes. Participants were purposively sampled for a range of medications, disease duration, disease activity [Disease Activity Score (DAS) patient opinion (general health), visual analog scale (“Considering all the ways your arthritis affects you, how well are you doing?”)]¹⁴, sex, age, and work status. Five nominal groups were held, with 3 rounds. In round 1, patients were asked to individually rate the 63 outcomes using 4 categories of importance: not important/not applicable, important, very important, and most important. Only the most important outcomes were used in round 2. In round 2, after sharing their most important outcomes, debate and discussion followed among these expert groups. A consensus was formed on the most important outcomes to be represented in a core set of measures for drug intervention, and a single group list was created. In round 3, a UK multicenter postal survey enabled the final selection of outcomes for the RAPP-PI. In all, 254 patients were asked to rate the individual importance of the 31 outcomes from the nominal groups and rank the top 6 most important outcomes. A relative importance score was then computed for each outcome. The 8 prioritized outcomes that formed the core of the RAPP-PI were the highest scoring: pain, activities of daily living, avoidance of joint damage, mobility, life enjoyment, independence, fatigue, and valued activities¹³.

Case 3: An Exploration of How Fatigue Emerged as a Domain of Interest in RA

In one of the earliest attempts to quantify outcome assessment, summarized in his paper of 1956, John Lansbury included “Hours after rising before onset of fatigue” as one of the assessment tools¹⁵. He distinguished different types of fatigue, noting that: “Neurotic fatigue is a lack of desire for action and, in our experience, occurs infrequently in rheumatoid patients. Fatigue due to lack of sleep wears off in an hour or so, but fatigue due to rheumatoid arthritis does not”. However, fatigue was not included in the 7 internationally agreed core outcome measures in RA clinical trials¹⁶. Further, although many potential outcome measures were considered during the development of the core set, fatigue was not mentioned. At OMERACT 5, in 2000, the meeting turned its attention to the scores required in the core set measures for them to be considered to have truly changed in response to treatment. There were many technical arguments, but at the end of the day perhaps the most important development was the recognition that taking a patient perspective was required¹⁷. It was the patients at OMERACT 6 who raised awareness of fatigue to health professionals, stimulating new research^{17,18,19,20,21}. Fatigue is an integral part of RA, experienced by almost all

patients at some time and by 40% on most days^{22,23}. It is an important physical and cognitive symptom that is considered overwhelming, uncontrollable, and different from normal tiredness in severity, quality, unpredictability and apparent lack of cause, affecting every aspect of life^{23,24,25,26}; at OMERACT 8 (in 2006), international consensus was reached that fatigue should be measured in all RA studies alongside the core set, using an instrument validated in RA fatigue²⁷. Participation by patients in the process of identifying relevant outcome domains resulted in insights that had been missed by the scientific community working in isolation, and emphasized the need for a robust instrument to measure RA fatigue, including concepts patients consider essential.

Case 4: Q-sort as a Method of Determining Patients with Different Domains of Interest — the Case of Fatigue

Q-methodology may be used to determine patients with different perspectives on outcome domains of interest. As an illustration, a Q-sort study²⁸ aimed to describe different perspectives on fatigue in RA patients was presented. According to Nikolaus, *et al*, patients with RA mention fatigue as one of their most bothersome symptoms²⁸. Three studies on the experience of fatigue in RA^{24,25,26} showed that fatigue is a multidimensional, bothersome symptom with far-reaching consequences. These studies give a first explorative insight into the experience of RA fatigue, but did not address differences between patients in their descriptions of fatigue. The Q-sort method can help examine intraindividual differences in the experience of RA fatigue. The researchers were interested whether fatigue experience differs between patients with RA and whether one patient can have different experiences of fatigue.

Cases in Context

The components of these cases are mapped onto the FDA Guidance Framework in Table 1 to illustrate how OMERACT and the FDA have arrived at similar conclusions and have shown consistency with the iterative process recommended there. As mentioned, some domains related to PRO can benefit from refinement, and this is possible in the FDA’s proposed approach. There is a hierarchy in domains, and PRO can assess different hierarchies. This refinement may include assessment of what patients can report within domains. For example, the concept of difficulty or restriction may be different from patient satisfaction or preferences. These distinctions require increasing interpretation and iteration by patients (Table 2).

DOMAINS — OMERACT AND THE COCHRANE COLLABORATION

The Cochrane Collaboration is a group of more than 28,000 contributors in over 100 countries who review the effects of healthcare interventions²⁹. Each systematic review is a synthesis of all known published controlled clinical trials and

Table 1. Process for domain selection: Similarity between OMERACT group activities and elements of FDA process.

Elements of FDA Recommendation ²	Case 1: RAID ¹⁰ Gossec, et al	Case 2: RAPP-PI ^{10a}	Case 3: Fatigue as a domain of interest in RA	Case 4: Q Sort ¹⁸
Outline hypothesized concepts and potential claims	x	x	x	x
Determine intended population	x	x	x	x
Determine intended application/characteristics (type of scores, mode and frequency of administration)	x	x	x	
Perform literature/expert review	x		x	x
Develop hypothesized conceptual framework			x	x
Place PRO within preliminary endpoint model				
Document preliminary instrument development	x	x		
Obtain patient input	x	x	x	x
Generate new items	x	x	x	
Select recall period, response options and format	x	x	x	
Select mode/method of administration/data collection	x	x	x	
Conduct patient cognitive interviewing	x	x	x	x
Pilot test draft instrument	x		x	
Document content validity	x	x	x	x
Confirm conceptual framework with scoring rule	x			

PRO: patient-reported outcomes; RAPP-PI: Rheumatoid Arthritis Patient Priorities for Pharmaceutical Intervention; RAID: Rheumatoid Arthritis Impact of Disease.

gives the best estimate of health benefits and side effects of a particular therapy; more recently, reviews including nonrandomized and observational studies have been prepared. The Cochrane Musculoskeletal Group produces reliable, up-to-date systematic reviews of interventions for the prevention, treatment, or rehabilitation of musculoskeletal disorders including rheumatology³⁰. These Cochrane Reviews are published in the Cochrane Library. The Cochrane Library has recently changed its approach to focus on the top 7 patient-important outcomes to include both benefit and harm for the first page of each systematic review that contains the abstract, plain-language summary, and a table of the summary of findings. Systematic review authors and editors have identified the strong need to put all PRO in the context of the Clinical Intervention Question, which broadly asks, “Does this intervention do more good than harm?”.

Just as a sentence has a certain written structure, a researchable question has specific parts. Writing the question with a specific structure makes the research process more clear and focused. Terms used to formulate the question can be used directly for the literature search, discussion of the results, planning the intervention, and for our purposes an explicit description of the outcome in the context of the population and intervention. A question formulated with this specific structure is termed the “PICO” question³¹, which as mentioned, includes the following parts: patient/population, intervention, comparison, and outcome (which will include the important domains — the clinically relevant outcomes for benefit and harm).

To make sure the outcome is intended and relevant to the clinical question, there must be agreement and clarity on the PICO. At OMERACT 10, a majority (61%) of participants

indicated through voting that they had not “heard of the PICO structure for the research question for trials.”

As noted, this new format has a summary of findings table that requires the consideration of up to 7 patient-important outcomes including both benefit and harm (see Figure 3). For example, in RA, the default 7 outcomes to be included in Cochrane reviews are:

- ACR 50 response
- Pain
- Health Assessment Questionnaire (HAQ) for function
- Disease Activity Score (DAS)
- Radiographic or appropriate imaging changes
- Short-term serious adverse effects from trials
- Longterm adverse effects or toxicity from observational studies

These 7 outcomes appear arbitrary to some, but they were developed by consensus in a series of editorial meetings with input from the Cochrane Musculoskeletal Consumer Group. It is possible to examine the chosen outcomes and map them to important domains identified by other groups in other ways, perhaps by mapping the outcomes to larger concepts. The 7 outcomes were chosen because they represent domains that reflect patient interests and are useful for clinical decision-making. These are a minimum set of outcomes, and review authors can add more outcomes if they are relevant to the PICO elements of the research question. Editors at the *British Medical Journal* have further expanded the use of the PICO.

This framework should reflect all important outcomes, even when no trials are found during a literature search; when there are no data, this should stimulate the research community to conduct the appropriate research study. Thus, use of this tool ensures that the audience will have the infor-

Table 2. Example of summary of findings table for abatacept for rheumatoid arthritis. © 2011 PROMIS Health Organization and PROMIS Cooperative Group. Reprinted with permission.

Abatacept (2 and 10mg/kg) + DMARD/biologic versus placebo + DMARD/biologic for rheumatoid arthritis						
Patient or population: patients with rheumatoid arthritis						
Settings: Intervention: Abatacept (2 and 10 mg/kg) + DMARD/biologic versus placebo + DMARD/biologic						
Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Control	Abatacept (2 and 10mg/kg) + DMARD/biologic versus placebo + DMARD/biologic				
ACR 50% Improvement Follow-up: 12 months	Medium risk population 168 per 1000	371 per 1000 (291 to 474)	RR 2.21 (1.73 to 2.82)	993 (3 studies)	⊕⊕⊕O moderate ^{1,2,3}	Absolute risk difference= 21% (95% CI, 16% to 27%) NNT=5 (95% CI 4 to 7)
Improvement in physical function (HAQ: greater than 0.3 increase from baseline, 0-3 scale) Follow-up: 12 months	Medium risk population 393 per 1000	637 per 1000 (531 to 766)	RR 1.62 (1.35 to 1.95)	638 (1 study)	⊕⊕⊕O moderate ¹	Absolute risk difference= 24% (95% CI 16% to 32%) NNT=5 (95% CI 4 to 7)
Achievement of low disease activity state (DAS 28 less than 3.2, scale 0-10) Follow-up: 12 months	Medium risk population 98 per 1000	424 per 1000 (278 to 646)	RR 4.33 (2.84 to 6.59)	638 (1 study)	⊕⊕⊕O moderate ¹	Absolute risk difference=33% (95% CI 26% to 39%) NNT=4 (95% CI 3 to 5)
Withdrawals due to adverse events Follow-up: 6 to 12 months	Medium risk population 35 per 1000	45 per 1000 (32 to 63)	OR 1.30 (0.91 to 1.85)	3105 (4 studies)	⊕⊕⊕O moderate ^{1,2,3,4}	
Total serious adverse events Follow-up: 6 to 12 months	Medium risk population 121 per 1000	126 per 1000 (106 to 151)	OR 1.05 (0.86 to 1.29)	3151 (4 studies)	⊕⊕⊕O moderate ^{1,2,3,4}	
Radiographic assessment Genant-modified Sharp erosion score Follow-up: 12 months	The mean radiographic assessment in the control groups was 0.27 units	The mean Radiographic assessment in the intervention groups was 0.27 lower (0.42 to 0.12 lower)		586 (1 study)	⊕⊕⊕O moderate ^{1,5}	
Long-term serious adverse events Follow-up: 2 years	See comment	See comment	Not estimable	950 (2 studies)	⊕⊕⊕O low ⁶	Based on 2 longterm extension studies (LTE) of RCT. Participants on placebo in the RCT switched to abatacept treatment. Number of patients with SAE: Genovese 2005: 103/357; 23.4 SAE/100 patient-years; 70% completed the LTE. Kremer 2006: 149/593; 16.3 SAE/100 patient-years; 90.5% completed the LTE.

*The basis for the assumed risk (e.g. the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk ratio; OR: Odds ratio;

GRADE Working Group grades of evidence

High quality: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Very low quality: We are very uncertain about the estimate.

¹ Kremer 2006: Intention to treat analysis not performed. 9 patients in abatacept group and 5 in placebo group excluded from analysis.

² Weinblatt 2006: 15 people randomized were not treated and not included in analysis

³ Kremer 2003: Risk of attrition bias - less than 80% completion rate in treatment group at 12 months

⁴ Weinblatt 2007: Risk of attrition bias - less than 80% completion rate in the treatment group at 12 months

⁵ Radiographic data obtained for 90% of study participants

⁶ Long-term serious adverse events based on observational data. Two RCTs had a long-term extension (LTE) phase in which people in the placebo group during the RCT switched to abatacept for the LTE.

mation needed for decision-making, including decisions about setting new research priorities. Similarly, uptake of the use of this tool could help identify outcomes that need further development³².

DISCUSSION

In this article, we have given case examples of frameworks used to select domains, placed these frameworks in the context of OMERACT, and described OMERACT activities

around domain selection. OMERACT participants agreed that domain selection needs further discussion and clarification, and that a literature review of domain selection should be undertaken to inform the direction of future activities (Table 3).

The consideration of domain identification and selection could lead to the development of a core set of domains. There is scope also for developing ideas for the initial implementation of "core domains."

Discussion of Possible Domains

1. Death. Measures within this domain would include total and disease-specific survival, time to death, etc.

2. Burden of disease, that includes all-cause, and subsections on disease-specific burden, treatment-associated for example; and mostly or exclusively patient-reported (including quality of life) participation. Ideally, this domain would mostly link with the ICF that focuses on functioning, including activities and participation aspect of patients; however, during informal OMERACT discussions, we identified discordance between ICF and domains of importance to patients. Conceptual sub-domains include: (a) The actual function, activity, participation that is affected, including symptoms and cost (perhaps cost is a separate, main domain, but that would imply it would need a place in every core set); (b) The impact on and/or importance to the patient of these domains (not addressed by the ICF).

3. Disease-specific and treatment-specific process measures necessary to assess specific effects: this would include most measures used currently in trials as endpoints (e.g., forced expiratory volume, tumor response assessed on imaging, RA disease activity, damage, etc.). It would also include adverse events as assessed by laboratory tests, etc.

Efforts of OMERACT have focused on identifying, standardizing, and collecting validated outcome measures, to help interpret results from randomized controlled trials for use in everyday clinical practice. Although the most global question for the patient is “How are you?”, a more in-depth analysis is needed to better grasp the nature of the outcomes of intervention. Also, it is possible that outcomes may change, due to influences on the overall impact of disease, which are subject to other considerations, such as the patient’s self-management or changes in their environment that alter the importance of a particular outcome³³. Further, there may be changes in relevant covariables that may capture the outcome of interest in an unexpected way (for example, an analgesic for knee arthritis may allow increased physical function, but pain levels remain unchanged because the patient increases functional activity to the limit of his or her pain tolerance). In terms of the OMERACT filter (Truth, Discrimination, Feasibility)³⁴, proper domain selection helps meet the requirement for “truth,” a

Table 3. Toward OMERACT 11: A specific research agenda.

Completion of a comprehensive literature review on domain selection
Collaborate and propose key definitions such as domain, impact, function
With regards to the PROMIS approach, analyze content validity in the context of computer adapted testing and other ways of tailoring domain selection to the individual, rather than the disease
Explore the possibility of a ‘core set’ of domains
Develop guidance for domain selection
Consider domains that allow for adequate coverage of benefits and harms in results reporting

word that captures issues of face, content, construct, and criterion validity. This helps the researcher to show that the outcome measures what is intended, and that it is relevant. OMERACT has proposed core sets for outcome measurement in specific diseases. These core sets refer to the minimum number of domains and instruments that are needed to describe outcomes in trials or clinical practice, but it is possible that, as with fatigue in RA, the research community has neglected some relevant domains of outcome.

Future OMERACT will focus on these issues and work towards developing a framework, based on the ICF, to support comprehensive outcome measurements. This includes the analyses of the use of fixed items versus open items, as well as time issues and possible questionnaire overload. At the end of the OMERACT meeting, the vote was to develop further guidance in the coming years and report back. The group plans to define the term “domain,” and discuss domains in a broad context (pain, fatigue, participation) as well as specific domains identified in instruments such as the HAQ, for example: disability, discomfort and pain, drug side effects (toxicity), dollar costs, and death. There is also a strong need to discuss the nature of different perspectives on outcomes. Outcomes are assessed by different measures and this may be due to the nuances of the domain selection. Domains that are understandable and relevant to patients may assess and express the response to treatment differently than domains relevant to researchers. Further, the balancing of benefits and harms takes place in the broad overview of the effects of disease and of its treatment — hence our increasing interest in life impact measures, which will be more informative about this balance, but less informative about the mechanism of action of diseases and their treatments (including the presentation of beneficial and harmful effects). OMERACT participants agreed that a guide for the options for developing domains that meet the OMERACT filter would be useful. More discussion and explanation is needed to outline outcomes related to the patient perspective that are not covered by the current version of the ICF, and to explain the usefulness of the PICO structure in domain selection.

REFERENCES

1. Fries JF. The chronic disease data bank model: a conceptual framework for the computer-based medical record. *Comput Biomed Res* 1992;25:586-601.
2. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. 2009. [Internet. Accessed April 7, 2011.] Available from: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>
3. Boonen A, Stucki G, Maksymowych W, Rat AC, Escorpizo R, Boers M. The OMERACT-ICF Reference Group: integrating the ICF into the OMERACT process: opportunities and challenges. *J Rheumatol* 2009;36:2057-60.

4. World Health Organization. Towards a common language for functioning, disability and health. ICF. Geneva: World Health Organization; 2002. [Internet. Accessed April 7, 2011.] Available from: <http://www.who.int/classifications/icf/training/icf/beginnersguide.pdf>
5. Weigl M, Cieza A, Cantista P, Reinhardt JD, Stucki G. Determinants of disability in chronic musculoskeletal health conditions: a literature review. *Eur J Phys Rehabil Med* 2008;44:67-79.
6. Swinkels RAHM. Measurement instruments for patients with rheumatic disorders: a clinimetric appraisal. Amsterdam: Vrije Universiteit; 2005.
7. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, 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.
8. PROMIS domains. Domain hierarchy framework. PROMIS Health Organization and the PROMIS Cooperative Group; 2008. [Internet. Accessed May 4 2011] Available from: http://www.nihpromis.org/Documents/PROMIS_Full_Framework.pdf
9. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315-24.
10. Gossec L, Dougados M, Rincival N, Balanescu A, Boumpas DT, Canadelo S, et al. Elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: a EULAR initiative. *Ann Rheum Dis* 2009;68:1680-5.
- 10a. Sanderson T, Morris M, Calnan M, Richards P, Hewlett S. An additional core set generated by people with rheumatoid arthritis: RA Patient Priorities for Pharmacological Interventions (RAPP-PI) [abstract]. *Arthritis Rheum* 2009;60 Suppl:1205.
11. Sanderson T, Morris M, Calnan M, Richards P, Hewlett S. Patient perspective of measuring treatment efficacy: the rheumatoid arthritis patient priorities for pharmacologic interventions outcomes. *Arthritis Care Res* 2010;62:647-56.
12. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995;311:376-80.
13. Sanderson T, Morris M, Calnan M, Richards P, Hewlett S. What outcomes from pharmacologic treatments are important to people with rheumatoid arthritis? Creating the basis of a patient core set. *Arthritis Care Res* 2010;62:640-6.
14. van der Heijde DM, van 't Hof M, van Riel PL, van de Putte LB. Development of a disease activity score based on judgment in clinical practice by rheumatologists. *J Rheumatol* 1993;20:579-81.
15. Lansbury J. Numerical method of evaluating the status of rheumatoid arthritis. *Ann Rheum Dis* 1958;17:101-7.
16. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis Rheum* 1993;36:729-40.
17. Wells G, Anderson J, Beaton D, Bellamy N, Boers M, Bombardier C, et al. Minimal clinically important difference module: summary, recommendations, and research agenda. *J Rheumatol* 2001;28:452-4.
18. Hewlett S, Hehir M, Kirwan JR. Measuring fatigue in rheumatoid arthritis: a systematic review of scales in use. *Arthritis Rheum* 2007;57:429-39.
19. Carr A, Hewlett S, Hughes R, Mitchell H, Ryan S, Carr M, et al. Rheumatology outcomes: the patient's perspective. *J Rheumatol* 2003;30:880-3.
20. Ahlmen M, Nordenskiöld U, Archenholtz B, Thyberg I, Ronnqvist R, Linden L, et al. Rheumatology outcomes: the patient's perspective. A multicentre focus group interview study of Swedish rheumatoid arthritis patients. *Rheumatology* 2005;44:105-10.
21. Hewlett S, Carr M, Ryan S, Kirwan J, Richards P, Carr A, et al. Outcomes generated by patients with rheumatoid arthritis: how important are they? *Musculoskeletal Care* 2005;3:131-42.
22. Wolfe F, Hawley DJ, Wilson K. The prevalence and meaning of fatigue in rheumatic disease. *J Rheumatol* 1996;23:1407-17.
23. Milton HV, Hewlett S, Kirwan JR. Fatigue in rheumatoid arthritis [abstract]. *Rheumatology* 2002;41 Suppl 2:73.
24. Hewlett S, Cockshott Z, Byron M, Kitchen K, Tipler S, Pope D, et al. Patients' perceptions of fatigue in rheumatoid arthritis: overwhelming, uncontrollable, ignored. *Arthritis Rheum* 2005;53:697-702.
25. Repping-Wuts H, Uitterhoeve R, van Riel P, van Achterberg T. Fatigue as experienced by patients with rheumatoid arthritis (RA): a qualitative study. *Int J Nurs Stud* 2008;45:995-1002.
26. Tack BB. Fatigue in rheumatoid arthritis. Conditions, strategies, and consequences. *Arthritis Care Res* 1990;3:65-70.
27. Kirwan JR, Minnock P, Adebajo A, Bresnihan B, Choy E, de Wit M, et al. Patient perspective: fatigue as a recommended patient centered outcome measure in rheumatoid arthritis. *J Rheumatol* 2007;34:1174-7.
28. Nikolaus S, Bode C, Taal E, van de Laar MAFJ. New insights into the experience of fatigue among patients with rheumatoid arthritis: a qualitative study. *Ann Rheum Dis* 2010;69:895-7.
29. The Cochrane Collaboration. Working together to provide the best evidence for health care. 2010. [Internet. Accessed April 7, 2011.] Available from: <http://www.cochrane.org/>
30. Maxwell L, Santesso N, Tugwell PS, Wells GA, Judd M, Buchbinder R. Method guidelines for Cochrane Musculoskeletal Group systematic reviews. *J Rheumatol* 2006;33:2304-11.
31. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995;123:A12-A13.
32. Brown P, Brunnhuber K, Chalkidou K, Chalmers I, Clarke M, Fenton M, et al. How to formulate research recommendations. *BMJ* 2006;333:804-6.
33. Sanderson T, Hewlett S, Flurey C, Dures E, Richards P, Kirwan J. The impact triad (incorporating severity, importance and self-management) as a method of enhancing the measurement of the personal life impact of rheumatic diseases. *J Rheumatol* 2011;38:191-4.
34. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998;25:198-9.