Toward a Generalized Framework of Core Measurement Areas in Clinical Trials: A Position Paper for OMERACT 11

Maarten Boers, Leanne Idzerda, John R. Kirwan, Dorcas Beaton, Reuben Escorpizo, Annelies Boonen, Susan Magasi, Ian Sinha, Gerold Stucki, and Peter Tugwell

ABSTRACT. ***Objective.*** The Outcome Measures in Rheumatology (OMERACT) international consensus initiative has successfully developed core sets of outcome measures for trials of many rheumatologic conditions, but its expanding scope called for clarification and updating of its underlying conceptual framework and working process. To develop a core set of what we propose to call outcome measurement instruments, consensus must be reached both on what to measure and how to measure. This article deals with the first part: a framework necessary to ensure comprehensiveness of the domains chosen for measurement. We formulated a conceptual framework of core measurement areas in clinical trials, for discussion at the OMERACT 11 conference.

Methods. We formulated a framework and definitions of key concepts adapted from the literature, and followed an iterative consensus process (small group processes and an Internet-based survey) of those involved including patients, health professionals, and methodologists within and outside rheumatology.

Results. The draft framework comprises 4 core "areas": death, life impact (all aspects of how a patient feels or functions), resource use (monetary and other costs of the health condition and interventions), and pathophysiologic manifestations (disease-specific clinical and psychological signs, biomarkers, and potential surrogate outcome measures necessary to assess specific effects). The survey responses (262 of 2293, response rate 11%) indicated broad agreement with the draft framework and the proposed definitions of key concepts, including understandability and feasibility. A total of 283 comments were processed.

Conclusion. In an iterative process, we have developed a generic framework for outcome measurement and working definitions of key concepts ready for discussion at the OMERACT 11 conference. (First Release March 1 2014; J Rheumatol 2014;41:978–85; doi:10.3899/jrheum.131307)

Key Indexing Terms:
CLINICAL TRIALS
CORE MEASUREMENT
OUTCOME AND PROCESS ASSESSMENT
OMERACT 11

The Outcome Measures in Rheumatology (OMERACT) initiative formulates, for individual health conditions (for example, rheumatoid arthritis, RA), internationally agreed core sets of outcome measures for randomized controlled trials and longterm observational studies[1]. These encompass the outcomes that must always be measured to

From the Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands; Centre for Global Health Research, Institute of Population Health, University of Ottawa, Ottawa; University of Bristol, Academic Rheumatology Unit, Bristol Royal Infirmary, Bristol, UK; Mobility Program Clinical Research Unit, Li Ka Shing Knowledge Institute of St. Michael's Hospital, Toronto, Canada; International Classification of Functioning, Disability, and Health (ICF) Research Branch in cooperation with the World Health Organization (WHO) Collaborating Centre for the Family of International Classifications in Germany; Swiss Paraplegic Research (SPF), Nottwil; Department of Health Sciences and Health Policy, University of Lucerne, Lucerne; and at SPF, Nottwil, Switzerland; Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Center and Caphri Research Institute, Maastricht, The Netherlands; Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA; Respiratory Unit, Alder Hey Children's Hospital, Liverpool, UK; Department of Medicine, University of Ottawa, Ottawa, Canada.

M. Boers, MD, PhD, Department of Epidemiology and Biostatistics, VU University Medical Center; L. Idzerda, MSc, Centre for Global Health Research, Institute of Population Health, University of Ottawa; J.R. Kirwan, MD, FRCP, Professor of Rheumatic Diseases, University of Bristol, Academic Rheumatology Unit, Bristol Royal Infirmary; D. Beaton, BScOT, PhD, Scientist and Director, Mobility Program Clinical Research Unit, Li Ka Shing Knowledge Institute of St. Michael's Hospital; R. Escorpizo, PT, DPT, MSc, Research Scientist; G. Stucki, MD, MS, Director and Chair, ICF Research Branch in cooperation with the WHO Collaborating Centre for the Family of International Classifications in Germany, SPF, and Department of Health Sciences and Health Policy, University of Lucerne; A. Boonen, MD, PhD, Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Center and Caphri Research Institute; S. Magasi, PhD, Assistant Professor, Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University; I. Sinha, PhD, Specialist Registrar in Paediatric Respiratory Medicine, Respiratory Unit, Alder Hey Children's Hospital; P. Tugwell, MD, MSC, FRCPC, Professor, Department of Medicine, University of Ottawa, and Canada Research Chair.

Address correspondence to Dr. Boers, Department of Epidemiology and Biostatistics, VU University Medical Center, PK 6Z 185, PO Box 7057, 1007 MB Amsterdam, The Netherlands. E-mail: eb@vumc.nl

properly assess the harms and benefits of the condition and its treatment. A core set does not replace or define the primary study question, and does not limit the choice of the primary outcome measure. Rather, reporting the results of a core set of outcomes in every trial ensures that a consistent dataset will be available for comparison with other studies, independent of the primary study question and associated outcome measures.

The key to this consensus-based process approach has been to apply the "OMERACT Filter" of "Truth, Discrimination, and Feasibility" to each candidate instrument within each domain of interest (Table 1)[2]. This pragmatic approach was successful and the definition of truth, discrimination, and feasibility added much clarity, but (perhaps because they were a relatively close-knit set of committed researchers in 1 medical subspecialty) the participants shared many unvoiced assumptions about what to include in core set definition. For example, the notion of "comprehensiveness" (content validity, part of Truth) in RA was based on common clinical experience, not questioning enquiry. OMERACT was implicitly using a framework for content validity based on the work of Fries, *et al*[3] and expanded by Kirwan[4], but there was no clear process to determine the comprehensiveness (or other Filter requirements) of the core set as a whole. This common background, while initially beneficial, became problematic as the areas of work expanded. For example, when patient participants were introduced to OMERACT, the comprehensiveness of the RA core set was questioned[5], thus highlighting the need for a broader and more transparent conceptual framework and clarification of the protocols used to select outcome measures.

The first step was to search the literature to find existing frameworks that could be used to define outcome within the OMERACT process. A systematic review[6] identified several existing conceptual models for health (conditions), the most influential being the World Health Organization (WHO) International Classification of Diseases (ICD)[7] and the WHO International Classification of Functioning, Disability, and Health (ICF)[8], formerly International Classification of Impairments, Disabilities, and Handicaps[9]. The ICD lists all known diagnoses and conditions and is grounded in the biomedical model. The ICF is grounded in the integrative and biopsychosocial model and, as the name implies, provides a taxonomy of functioning, disability, and participation. Models that unify these 2 perspectives include the quality of life model developed by Wilson and Cleary[10], and those of Bruce and Fries[11] and Porter[12]. None of the above models are fully applicable to OMERACT: they are mostly aimed at describing or classifying health and function, rather than at measuring outcome as a consequence of an intervention. In addition, none were derived from a documented broad consensus over their underlying philosophy (although the ICF was ratified by WHO member nations), each had been promulgated by an individual or a small group, and any subsequent critique was unstructured or undocumented[6].

Recently the "COMET" (Core Outcome Measures in Effectiveness Trials) initiative emerged, aiming to bring together researchers of all disease areas interested in the development and application of agreed standardized sets of "core outcomes" to be measured in all clinical trials of a specific condition[13]. These aims clearly overlap with those of OMERACT within rheumatology, and there has been a wide, consensus-based cross-fertilization between both groups in producing this position paper for consideration and possible further modification at OMERACT 11.

This article describes the development of a conceptual framework of Core Areas defining what to measure to properly and comprehensively describe the effects of intervention on health conditions. This framework was developed for OMERACT but is likely applicable to health conditions outside rheumatology. Key concepts are introduced here (Table 2); they were discussed at the OMERACT 11 conference[14]; the application of the framework to core set development (i.e., how to measure) is described in other conference articles[15,16,17,18]. As combined, the framework and its application will be termed the "OMERACT Filter 2.0."

Method of Development

The first outline of the framework was developed iteratively based on results of an informal literature review and discussions among experts including the OMERACT and COMET

Table 1. The original OMERACT Filter to determine applicability of a measurement instrument in a setting[2]. From Boers, *et al*. J Rheumatol 1998;25:198–9. Adapted with permission.

Truth	Is the measure truthful, does it measure what is intended? Is the result unbiased and relevant? The word captures issues of face, content, construct and criterion validity. *As gold standards are usually not available, criterion validity is mostly not tested*
Discrimination	Does the measure discriminate between situations of interest? The situations can be states at one time (for classification or prognosis) or states at different times (to measure change). The word captures issues of reliability and sensitivity to change
Feasibility	Can the measure be applied easily, given constraints of time, money, and interpretability? The word captures an essential element in the selection of measures, one that may be decisive in determining a measure's success

Table 2. Key concepts used in the OMERACT Filter 2.0 Framework.

Health	**A state of complete physical, mental and social well-being and not merely the absence of disease or infirmity (WHO 1948)** While this definition is widely accepted, it has a number of disadvantages. It is formulated in absolute terms and may be problematic to measure. Hence, Salomon, *et al* proposed[33]: "Health is comprised of states or conditions of functioning of the human body and mind." It has been suggested that OMERACT forego defining health altogether and consider it to be a concept best described as "The resilience or capacity to cope and maintain and restore one's integrity, equilibrium, and sense of well being" in 3 domains: physical, mental, and social[34]. The current WHO definition was endorsed by more people than the proposed alternative and is thus retained (84% vs 68%)
Health Condition	**A situation of impaired health** While a small minority objected to the term "health condition" replacing the word "disease," this term was deemed best to accommodate conditions of disability where people have impaired function without active disease (e.g., someone with an amputated leg)
Health Intervention	**An activity, performed by, for, with or on behalf of a client(s), whose purpose is to improve individual or population health, to alter or diagnose the course of a health condition, or to improve functioning** The WHO International Classification of Health Interventions (ICHI) is still under development[35]. In particular the word "activity" remains controversial. After feedback we added the word "by" but otherwise included this definition
Core Area	**An aspect of health or a health condition that needs to be measured to appropriately assess the effects of a health intervention. Core Areas are broad concepts consisting of a number of more specific concepts called Domains** We propose the term "Core Area" for the concept of "essential aspect of health." i.e., the top-level categories of measurement. This word has been chosen as a neutral placeholder to avoid the definitional confusion that surrounds alternative phrasing
(Sub) Domain	**A concept to be measured, a further specification of an aspect of health categorized within a Core Area** "Domain" is used here for the concept describing the specification of a core area where measurement will need to take place (and measurement instruments need to be selected). Because of lack of consistency we initially suggested avoiding the word domain altogether and replace it with "dimension." This alternative also proved problematic, and the consensus-based decision was to retain the term "Domain"
Outcome	**Any identified result in a (sub) domain arising from exposure to a casual factor or a health intervention** The initial proposal was: "Any identified result in health or a health condition, including those arising as a consequence of exposure to a causal factor or the handling of a (potential) health problem" "Outcome" and "trial endpoint" have been used interchangeably causing considerable confusion. Survey participants made many (often contradictory) suggestions, indicating the concepts were not clearly defined. The term "endpoint" is further complicated because it has been used to denote both outcomes and point in time. Within OMERACT, we have chosen not to use the word "endpoint," and use the word "outcome" only in the context of the Core Outcome Measurement Set; we do not use the Term "Core Outcome Set" as there is currently no consensus on its technical definition
Measurement Instrument	**A tool to measure a quality or quantity of a variable, in this context a (sub) domain or contextual factor** The tool can be a single question, a questionnaire, a score obtained through physical examination, a laboratory measurement, a score obtained through observation of an image, etc.
Outcome Measurement Instrument	**A measurement instrument chosen to assess Outcome** The result of measurement (recently termed "specific metric"[36]) can be expressed as change, as end result, as cumulative result, or as "time to event" in a (sub) domain. Note that some instruments may measure more than one domain (e.g., indices combining the results of several instruments)
Core Domain Set	**For studies of health interventions, the minimum set of Domains and Subdomains necessary to adequately cover all Core Areas, i.e., fully measure all relevant concepts of a specific health condition within a specified setting** As elaborated below, the Core Domain Set defines the minimum requirements of what needs to be measured; the Core Outcome Measurement Set subsequently defines how to measure these Core Domains, i.e., the specific instruments required. The initial proposal was: "The minimum set of Domains and Subdomains necessary to adequately cover all Core Areas, i.e., fully describe all relevant concepts of a specific health condition in a specified Scope"
Core Outcome Measurement Set	**The minimum set of outcome measurement instruments that must be administered in each intervention study of a certain health condition within a specific setting to adequately cover a corresponding Core Domain Set** The initial proposal was: "The minimum set of outcome measures that must be measured in each study of a certain health condition in a specified (set of) setting(s) to adequately cover a corresponding Core Domain Set"
Setting (scope):	**The set of factors that describes the studies and circumstances to which the core outcome set will apply. This is determined by the study questions and includes the health condition(s), target population, interventions, etc.**
Contextual Factor	**Variable that is not an outcome of the study, but needs to be recognized (and measured) to understand the study results. This includes potential confounders and effect modifiers.** The initial proposal was: "Variables related to the scope or setting of the study (and its core outcome set)" During discussions it became clear that context is critical in defining core outcome measurement sets. This is a very broad area in which various terms have been used to describe the different components. We prefer the broad term "contextual factor" as a placeholder for more specific words. Setting (scope) is usually determined by the study question and includes study population, interventions, etc. The contextual factors are not outcomes of the study, but need to be recognized (and measured) to understand and interpret the study results. Such factors include potential confounders and effect modifiers. In the ICF contextual factors have been described as environmental and personal factors that can influence the impact of the disease process on body functions and body structures, activities and participation

executive groups after the first COMET conference in 2010. A more formalized literature review confirmed the lack of an immediately applicable framework and the need to develop one[6]. Given the lack of a clear alternative, we decided to proceed with the Bruce/Fries/Kirwan work. The authors then developed the next (and subsequent) drafts.

Experts in trial and systematic review methodology were identified (n = 53) and invited to comment on the draft by way of targeted survey questions supplemented with open comments. Eighteen of these experts met at the second COMET conference (July 2011) for a structured discussion of the feedback received by then. The draft version of the framework was formally presented in a plenary session at the conference. Participants were invited to submit comments and suggestions. The framework, now written in the form of a draft paper with definitions of key concepts (Table 2), was further refined. This document was sent along with a reviewer survey to a total of 2293 persons: all participants from the second COMET conference (n = 131), all current and former OMERACT conference participants (n = 678), and participants of the Evidence Based Health listserv (n = 1484)[19]. A total of 262 surveys were returned (161 of these were from OMERACT participants; overall response: 11%). The survey responses indicated broad agreement with the draft framework and the proposed definitions of key concepts, including its understandability and feasibility (Supplementary Table 1 available online at jrheum.org). A total of 283 comments were processed.

The definition of the key concepts and the framework, both supplemented by the comments and explanations as set out in this position paper, were presented at OMERACT 11. The subsequent discussions and eventual conclusions from OMERACT 11 are reported elsewhere[14] except where the definition of key concepts has changed, and the final concepts are reported here for clarity (Table 2).

Proposed Framework and Elaboration (see also Figure 1)

The framework guides core set development in the setting of all trials aimed at assessing benefit or harm in humans, i.e., all trials from phase 2. A large majority of respondents supported use of the framework in systematic reviews and observational studies (each, 84%), but not in audits (51%) or clinical care (61%). Systematic reviews follow naturally from trials, but application to observational studies is outside the scope of the current development. The framework encompasses the effect and the pathophysiological manifestations of health conditions.

Impact of health conditions. Impact of health conditions includes all aspects that describe how a patient feels, functions, or survives, covering 3 areas: Death; Life Impact of Health Conditions; and Resource Use. Life Impact can also be described as "the lived experience of health"[20,21]. Resource use is of paramount importance to society, and can be regarded as a reflection of the societal effect of a health condition, and can also relate to the personal resources of all kinds invested by patients and caregivers in their health.

Pathophysiological manifestations of health conditions. This grouping includes reversible and irreversible (damage) manifestations. Reversible manifestations can be either modifiable risk factors for a health condition (in the setting of prevention trials, e.g., hypertension), or actual manifestations of the health condition (e.g., RA disease activity, glycosylated hemoglobin A levels), as in disease activity. These are a Core Area because clinical trials are done not only to assess effects (benefit and harm) of an intervention, but also to document whether the effect of the intervention specifically targets the pathophysiology of the health condition. In the original Bruce and Fries framework[11] as adapted by Kirwan[4] this was termed "Process" and we adopted this term in initial drafts. However, Donabedian used the word "process" in the context of measurement of quality of care, where the word denoted process of care, and feedback from surveys confirmed the potential for confusion, so we adopted the current term[22].

Elaboration

Death. Where the importance of death in studies of life-threatening conditions is obvious, its inclusion in studies of other conditions resulted in much discussion. All participants agreed death should always be reported even where it is a rare occurrence, making it a core area by default. In life-threatening conditions, death would probably be specified in a core domain set and may even be specified by study developers as the primary outcome measure.

Impact of Health Conditions

The term "Life Impact of Health Conditions" represents what previously was called "burden of disease." This concept is largely covered by health-related quality of life, but also includes life activities, participation, etc. It mostly links with 2 parts of ICF: activities, and participation[23]. Many respondents felt this area was (too) broad and needed further specification in the framework. However, these suggestions seemed to be related to the professional background of the respondent, so while providing some examples we prefer to leave such specification to core set developers who can consider Life Impact in their chosen context.

The time-specific nature of domain specification may be very important in the Area of Impact, but also in the other proposed areas. In Porter's "outcome hierarchy to assess value of health care"[12], time considerations are essential. This is also gaining in importance in chronic diseases, where in some areas (such as rheumatology), agents have been developed that have a much more rapid onset of action than traditional treatments. Core set developers will need to explicitly consider how the health condition interacts with the intervention to decide how the effect is best expressed. In chronic health conditions, measures that capture the experience over time in more detail (e.g., "area under the curve" of effect) may be preferable to single end-of-trial assessments. To measure effect, it is important to define the concepts of "(minimum) clinically important change" (to detect a useful response) and "patient acceptable state" for the outcome measures of choice[24,25].

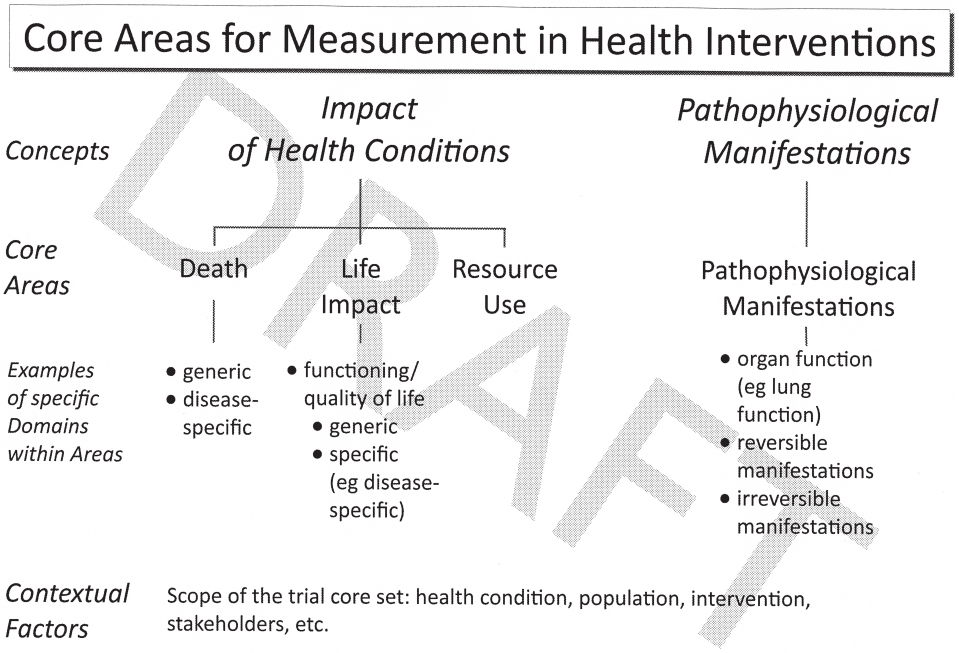

Figure 1. Preliminary framework of core areas for measurement in health interventions to develop trial core sets as proposed to the OMERACT conference. Four areas are defined under 2 broad headings. Core set developers should specify at least 1 domain (and an accompanying instrument) in each area. Measures can be either generic or specified along the dimensions suggested. In addition, developers should define the context of the core set. The end result is a trial core set in a specific setting. Note that important changes were made to the framework during the conference. The final form of the framework, including the development of a core outcome measurement set, is described in the final conference article[37].

Although not included explicitly, the concept of impact includes not only negative but also positive effects, which are more appropriately termed effects on the "amount of health." In this way, measurement of interventions that increase health above the norm can also be placed in the framework; for example, the tradeoffs faced by athletes training for the Olympics, such as the social isolation during training. Currently, the Medical Outcome Study Short Form-36 survey questionnaire is one of the few generic scales to include items on positive health[26].

Resource Use

In economics, resources are defined as inputs required to produce goods and services. These include both tangible factors such as monetary capital and labor, and intangible factors such as opportunity. Both the presence of a health condition and its treatment incur resource use. Most of these are expressed as monetary costs, including costs of the intervention itself, associated costs involved in its application (including costs of treating side effects), and indirect costs associated with productivity loss. The consideration of resource use at the earliest stage of the development of a health intervention has become of paramount importance in recent years, as even the budgets of the richest nations are being threatened by burgeoning healthcare costs. In the developing world, apart from societal and geographical contextual factors discussed below, the applicability of an intervention is strongly driven by its associated resource use. The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system for developing recommendations includes costs as a key component in formulating clinical recommendations[27]. For health economic evaluations, the valuation of health states (utility) is essential. Such valuation is simply another way to measure impact, with death usually valued at zero. While a strong majority favored Resource Use as an important area to assess, many comments were made as to whether this area should always be reported; hence it would be an important discussion point at the OMERACT conference.

Pathophysiological manifestations of health conditions. Despite the recent drive toward measurement of "patient-relevant" outcomes, we argue that in the context of trials, measures of pathophysiological manifestations (PM) should constitute a Core Area alongside Impact. PM is a broad term that includes mostly disease-specific clinical signs, biomarkers[28], and potential surrogate outcome measures[29] necessary to assess specific effects. It can also include psychological manifestations and mostly overlaps with the ICF concepts of body structures and body function. The alignment is not perfect: for instance, we place symptoms such as pain, stiffness, and fatigue under Life Impact, but these are categorized as body functions under

ICF. PM are needed to assess whether the effect of the intervention specifically targets the pathophysiology of the health condition, or alternatively, is nonspecific and unable to change the course of the disease.

PM includes most measures currently being used in trials as primary outcome measures (e.g., forced expiratory volume, tumor response assessed on imaging, most measures of RA disease activity, damage). Many adverse events can be classified as PM (mostly those detected only by biomarkers such as laboratory tests, e.g., liver function abnormalities). PM can sometimes serve as first indication for Impact when this is difficult to measure in the context of a trial (for example, RA joint damage for longterm work disability). OMERACT has developed a framework to help distinguish biomarkers (any PM deemed useful) from surrogate outcome measures (modifiable biomarkers that predict later Impact)[29].

In prevention trials, we perceive risk factors as being surrogates for (usually bad) outcomes that need to be prevented. Note that even in the preventive setting, people at risk of a health condition may experience a real change in their effect of ill health when they become aware of the risk and change their behavior accordingly. We expect core set developers to be most familiar with the PM area. Most PM will be specific to the health condition being studied, and the formulation of what the most important manifestations are will be relatively straightforward. Most discussions will then focus on the choice of instruments.

Adverse events. In most trials the (intended) beneficial effects are studied in much greater detail than the (unintended) harmful effects. The latter are usually only listed and summarized. Integration of benefit and harm into 1 scale would help to determine the total effect of an intervention compared to its alternatives. The development of such a scale needs to overcome several conceptual and practical hurdles[30]. One possibility is the use of utility measurement or quality-adjusted life years: it is assumed that all positive and negative effects are covered under the valuation of a health state. The Cochrane Collaboration Systematic Reviews now require that up to a maximum of 7 of the most important benefits and harms be included in their main summary of findings tables[31], and the GRADE approach to clinical recommendations requires explicit assessment of the tradeoff between benefit and harm[27]. Although adverse events will be measured in one of the core areas and were initially regarded as a domain within these, many respondents felt adverse events should have a recognizable and separate place in the framework, leaving room for further discussion.

Setting and contextual factors. The setting or scope (health condition, target population for the intervention, type of intervention, etc.) for which the core set is being developed will drive most of the deliberations. Contextual factors can be defined as those that are not the primary object of research but that may influence the results or the interpretation of the results. In addition, there are potential confounders and effect modifiers (most of which should be eliminated by randomization), as well as factors that define the generalizability of the study findings. Broadly speaking, contextual factors can be classified as personal (e.g., sociodemographics, comorbidity), societal (e.g., ethnicity, cultural attitudes, traditions) and environmental (e.g., geography). The latter 2 categories are increasingly important as research has become global. Core set developers need to specify the setting, the minimum list of contextual factors to be documented, and also whether such factors are appropriately to be measured at baseline or repeatedly. The specification of contextual factors emerged as an area of contention in both survey and OMERACT 11 participants.

Overlap in classifications. The precise selection of domains within the Core Areas will always depend on the setting of the trial as defined by those developing core outcome measurement sets. In some settings a particular variable will be a primary outcome, in others part of the core outcome measurement set, and in others the same variable may be a contextual factor. For example, adherence to treatment can strongly influence the result of a trial, e.g., whether an intervention reduces disease activity. In some cases adherence can become the target of intervention, and disease activity (or its change) can become the contextual factor.

Discussion

Building on past work, the presented framework for Core Domain Sets unifies the biomedical and quality of life model and remains consistent with other models, particularly the ICD and ICF. The framework is being developed in the setting of trials of health interventions. It closely follows the seminal work by Bruce and Fries in the development of the Health Assessment Questionnaire[11], in turn inspired by Donabedian[22] and others. Its initial application is for randomized trials, but an obvious extension is to observational studies where many if not all the principles underpinning the framework apply.

In the absence of any gold standard, the most important aspect of the framework is its face validity; i.e., it must be acceptable to everyone involved. Therefore, we believe that it is appropriate to follow a consensus process to develop and present this preliminary framework, engaging a wide range of viewpoints including those of patients, caregivers, healthcare providers, researchers, healthcare managers, payers, industry, regulators, and the government. Further, we seek to be explicit and document our stepwise consensus process. Finally, identifying the areas of disagreement can inform what needs to be further developed.

Our survey intended to reach out to a wide audience but many of the people approached (especially from the Evidence Based Health listserv[19]) chose not to respond. It is likely most were not directly engaged in the topic, but the

low overall response rate precludes firm conclusions on the acceptability of the framework in the wider scientific community at this stage. Therefore, the proposals in this position paper, although building on existing models and already subject to extensive discussion, were placed before OMERACT 11 for intense scrutiny and consideration for acceptance as part of the development of OMERACT Filter 2.0. [Note that a standalone article intended for a general (non-rheumatology) audience will appear in the *Journal of Clinical Epidemiology*; it summarizes development of Filter 2.0, described in detail in this part of the OMERACT proceedings[32].]

ACKNOWLEDGMENT

We thank Paula Williamson and Jane Blazeby of the COMET executive, and all members of the OMERACT executive, for valuable advice; and survey participants for their input.

ONLINE SUPPLEMENT

Supplementary data for this article are available online at jrheum.org

REFERENCES

1. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: An international initiative to improve outcome measurement in rheumatology. Trials 2007;8:38.
2. Boers M, Brooks P, Strand V, Tugwell P. The OMERACT Filter for outcome measures in rheumatology. J Rheumatol 1998;25:198-9.
3. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. J Rheumatol 1982;9:789-93.
4. Kirwan JR. A theoretical framework for process, outcome and prognosis in rheumatoid arthritis. J Rheumatol 1992;19:333-6.
5. 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.
6. Idzerda L, Rader T, Tugwell P, Boers M. Can we decide which outcomes should be measured in every clinical trial? A scoping review of the existing conceptual frameworks and processes to develop core outcome sets. J Rheumatol 2014;41:986-93.
7. World Health Organization. International classification of diseases (ICD-10). World Health Organization: 1990. [Internet. Accessed December 17, 2013.] Available from: http://www.who.int/classifications/icd/en
8. World Health Organization. International Classification of Functioning, disability and health. [Internet. Accessed December 17, 2013.] Available from: http://www.who.int/classifications/icf/en/
9. World Health Organization. International Classification of Impairments, Disabilities, and Handicaps. Geneva, 1980. [Internet. Accessed January 9, 2014.] Available from: http://whqlibdoc.who.int/publications/1980/9241541261_eng.pdf
10. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA 1995;273:59-65.
11. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. J Rheumatol 2003;30:167-78.
12. Porter ME. What is value in health care? N Engl J Med 2010;363:2477-81.
13. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials 2012;13:132.
14. Kirwan JR, Boers M, Hewlett S, Beaton D, Bingham CO, Choy E, et al. Updating the OMERACT Filter: Core Areas as a basis for defining core outcome sets. J Rheumatol 2014;41:994-9.
15. Tugwell P, Boers M, D'Agostino M, Beaton D, Boonen A, Bingham CO, et al. Updating the OMERACT Filter: Implications of Filter 2.0 to select outcome instruments through assessment of content, face and construct validity. J Rheumatol 2014;41:1000-4.
16. Wells GA, Beaton D, Tugwell P, Boers M, Kirwan JR, Bingham CO III, et al. Updating the OMERACT Filter: Discrimination and feasibility. J Rheumatol 2014;41:1005-10.
17. Kirwan JR, Bartlett SJ, Beaton D, Boers M, Bosworth A, Brooks PM, et al. Updating the OMERACT Filter: Implications for patient reported outcomes. J Rheumatol 2014;41:1011-15.
18. D'Agostino MA, Boers M, Dougados M, Van Der Heijde D, Iagnocco A, Landewe R, et al. Updating the OMERACT filter: Implications for imaging and soluble biomarkers. J Rheumatol 2014;41:1016-24.
19. JISC. Evidence based health. [Internet e-mail list. Accessed December 17. 2013.] Available from: https://http://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=evidence-based-health
20. Kostanjsek N, Escorpizo R, Boonen A, Walsh NE, Ustun TB, Stucki G. Assessing the impact of musculoskeletal health conditions using the international classification of functioning, disability and health. Disabil Rehabil 2011;33:1281-97.
21. Kostanjsek N, Rubinelli S, Escorpizo R, Cieza A, Kennedy C, Selb M, et al. Assessing the impact of health conditions using the ICF. Disabil Rehabil 2011;33:1475-82.
22. Donabedian A. Evaluating the quality of medical care. Milbank Q 2005;83:691-729.
23. The WHO Family of International Classifications. 2011. [Internet. Accessed December 17, 2013.] Available from: www.who.int/classifications/en/
24. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. Curr Opin Rheumatol 2002; 14:109-14.
25. Tubach F, Ravaud P, Beaton D, Boers M, Bombardier C, Felson DT, et al. Minimal clinically important improvement and patient acceptable symptom state for subjective outcome measures in rheumatic disorders. J Rheumatol 2007;34:1188-93.
26. Essink-Bot ML, Krabbe PF, Bonsel GJ, Aaronson NK. An empirical comparison of four generic health status measures. The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument. Med Care 1997;35:522-37.
27. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924-6.
28. De Gruttola VG, Clax P, DeMets DL, Downing GJ, Ellenberg SS, Friedman L, et al. Considerations in the evaluation of surrogate endpoints in clinical trials: summary of a National Institutes of Health workshop. Control Clin Trials 2001;22:485-502.
29. Lassere MN, Johnson KR, Boers M, Tugwell P, Brooks P, Simon L, et al. Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema. J Rheumatol 2007; 34:607-15.
30. Boers M, Brooks P, Fries JF, Simon LS, Strand V, Tugwell P. A first step to assess harm and benefit in clinical trials in one scale. J Clin Epidemiol 2010;63:627-32.
31. Schünemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P,

Guyatt GH. Chapter 11: Presenting results and 'Summary of findings' tables. In: Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of intervention version 510 (updated March 2011). The Cochrane Collaboration. [Internet. Accessed December 17, 2013.] Available from: http://www.cochrane-handbook.org
32. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol 2014 (E-pub ahead of print).
33. Salomon JA, Mathers CD, Chatterji S, Sadana R, Üstün TB, Murray CJL. Health systems performance assessment. Geneva: World Health Organization; 2003. [Internet. Accessed January 9, 2014.] Available from: http://whqlibdoc.who.int/publications/2003/9241562455.pdf
34. Huber M, Knottnerus JA, Green L, Horst H, Jadad AR, Kromhout D, et al. How should we define health? BMJ 2011;343:d4163.
35. Madden R. ICHI Project Plan Version 2.2, paper no. D020. In: WHO-FIC Network Meeting: 16–22 October 2010, Toronto, Canada. [Internet PDF. Accessed February 11, 2014.] Available from: www.who.int/classifications/network/meeting2010/en
36. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database—update and key issues. N Engl J Med 2011;364:852-60.
37. Boers M, Kirwan JR, Gossec L, Conaghan PG, D'Agostino M, Bingham III C, et al. How to choose core outcome measurement sets for clinical trials: OMERACT 11 approves Filter 2.0. J Rheumatol 2014;41:1025-30.