The
OMERACT Handbook

Maarten Boers, John Richard Kirwan, Peter Tugwell, Dorcas Beaton, Clifton O. Bingham III, Philip G. Conaghan, Maria-Antonietta D'Agostino, Maarten de Wit, Laure Gossec, Lyn March, Lee S. Simon Jaspinder A Singh, Vibeke Strand, George Wells

With thanks to the many OMERACT participants and other colleagues who provided comments on earlier versions of the chapters of this Handbook.


Published by OMERACT

Updated March-31-17
Author Information

The OMERACT Handbook was written and compiled by the OMERACT Executive:

M Boers, MD, PhD, MSc. Departments of Epidemiology & Biostatistics, and Rheumatology; VU University Medical Center, Amsterdam, Netherlands

J R Kirwan, MD, University of Bristol, Academic Rheumatology Unit, Bristol Royal Infirmary, Bristol BS2 8HW, UK

P Tugwell, MD, Department of Medicine, University of Ottawa, Ottawa, Canada

D Beaton, BScOT, PhD, Scientist and Director, Musculoskeletal Health & Outcomes Research, Li Ka Shing Knowledge Institute of St Michael's Hospital; and the Institute for Work & Health, Toronto, ON Canada

C O Bingham III, MD, Division of Rheumatology, Johns Hopkins University, Baltimore, MD, USA

P G Conaghan MB BS, PhD, FRACP, FRCP, Division of Musculoskeletal Disease, University of Leeds, & NIHR Leeds Musculoskeletal Biomedical Research Unit, UK

M A D'Agostino, MD, PhD Versailles-Saint Quentin En Yvelines University, Department of Rheumatology, Ambroise Paré Hospital, APHP, Boulogne-Billancourt, France.

M de Wit, Patient Partner, Department of Medical Humanities, VU Medical Centre, Amsterdam, Netherlands

L Gossec, MD, PhD, Université Pierre et Marie Curie (UPMC) - Paris 6, GRC-UMPC 08 (EEMOIS); AP-HP Pitié Salpêtrière Hospital, Department of Rheumatology; Paris, France.

L March, MD, Department of Rheumatology, Royal North Shore Hospital, Pacific Highway, St Leonards, NSW 2065, Australia

L S Simon, MD, SDG LLC Cambridge, MA, USA

J A Singh, MD, MPH, Birmingham Veterans Affairs Medical Center and University of Alabama at Birmingham, Birmingham, Alabama, USA

V Strand, MD, Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto, CA, USA

G Wells, MSc, PhD Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada
Table of Contents

CHAPTER 1 - HOW OMERACT EMERGED AND DEVELOPED 10

A. A PLETHORA OF TRIAL OUTCOMES 10
B. INADEQUATE METHODOLOGICAL FRAMEWORK 10
C. THE FIRST OMERACT MEETING 10
D. THE FIRST CORE SET 11
E. OMERACT GETS INTO ITS STRIDE 11
F. PATIENT PARTNERS MAKE A DIFFERENCE 11
G. THE OMERACT FILTER AND FILTER 2.0 11
H. SOME CHALLENGES FOR OMERACT 12
I. REFERENCES 14

CHAPTER 2: THE SPIRIT OF OMERACT 15

A. MATERIALS AND METHODS 17
B. RESULTS 21
C. DISCUSSION 26
D. REFERENCES 28

CHAPTER 3: DEVELOPING CORE OUTCOME MEASUREMENT SETS 30

INTRODUCTION 30

A.1. OMERACT MASTER CHECKLIST for Developing Core Outcome Measurement Sets 35

B. ASSEMBLY OF WORKING GROUP AND WORK PLAN 36
1. Forming an OMERACT Working Group 36
2. Stakeholder groups and their contacts identified 36
3. Thorough review of domains and instruments previously used 37
4. Implementation of Delphi and or Focus Groups 37
5. Agreeing on the Core Outcome Measurement Set 38

C. CORE DOMAIN SET SELECTION 40
1. Definition of context: setting (scope) 40
2. Deciding on the inclusion of Resource Use 41
3. Literature review of domains (and instruments), part 1: what has been measured? 41
CHAPTER 4. INSTRUMENT SELECTION FOR CORE OUTCOME MEASUREMENT SETS

Introduction and Background to Instrument Selection

Revision of the OMERACT Filter

Foundation: How do we know if an instrument has passed Filter 2.1?

A. Moving through the Filter 2.1 for Instrument Selection.
   1. Preparing for instrument selection and seeking input from the Technical Advisory Group (TAG).

   1. Question 1: Is it a good match with the target domain in this population?
   2. Question 2: Is it practical to use?
   3. and 4. (1-4): Question 3 and Question 4: Do the numeric scores make sense? & Can it discriminate between groups of interest?
   Searching for the evidence: setting up the literature review and identifying articles.
   4. Synthesis at a measurement property level and filling any gaps (4.B.3 – 4.B.4)

C. Synthesis and filling any gaps
   1. Synthesis of the available body of evidence for each instrument
   2. Reporting of your work to the Technical Advisory Group: TAG
   3. Design studies to fill gaps with review with TAG
   4. Complete the studies, appraise and add to body of evidence (4.C.1)

D. Reporting and seeking endorsement
   1. Prepare Final Report
   2. Prepare for presentation at OMERACT for endorsement of instrument(s)
   3. Results of OMERACT Consensus Vote for endorsing instrument
   4. Set up timeline for update of endorsed instrument(s)
   5. Communication & dissemination plan

Frequently Asked Questions

Conclusion.

Appendices:
   A. Workbook for documenting process of gathering and synthesizing evidence.
   B. The COSMIN-OMERACT Good Methods Checklist adapted for OMERACT Filter 2.1 Instrument Selection needs
   C Adequacy of results review
   D. Supporting background for OMERACT Filter elements
   E. Acknowledgement of international efforts that led to the decisions made in OMERACT Filter 2.1 Instrument Selection Algorithm.
CHAPTER 5 – DEVELOPING A METHODS WORKING GROUP

CHAPTER 6 SPECIAL CONSIDERATIONS FOR IMAGING AND BIOMARKERS

CHAPTER 7 - METHODS FOR REACHING CONSENSUS

A. INTRODUCTION AND THE OMERACT 'PHILOSOPHY' ON CONSENSUS
   1. Efforts to achieve consensus
   2. Structures supporting consensus

B. STEPS TO TAKE IN REACHING CONSENSUS IN OMERACT
   1. Generating ideas
   2. Stakeholder involvement
   3. Initiate and expand your list of domains
   4. Refine your list of domains
   5. Moving on to the instruments

C. GUIDELINES FOR USING CONSENSUS GROUP METHODS

D. GAINING CONSENSUS AT OMERACT MEETINGS
   1. Concepts behind Breakout Group Sessions
   2. Breakout group participants
   3. Preparing and running small group sessions – task and timing
   4. Preparing the report back to the plenary session
   5. Preparing for the report back to the final conference ‘wrap-up’ session
   6. Preparing questions for voting
   7. Voting on core sets
   8. Other types of vote

E. REFERENCES

F. APPENDIX
   1. Check List For Breakout Group Consensus Process
   2. A note on Nominal Group Technique

CHAPTER 8 – PATIENT PARTNERS AND OMERACT

PART 1 – WORKING WITH PATIENT RESEARCH PARTNERS

A. INTRODUCTION

B. PATIENT RESEARCH PARTNERS (PRPS)

C. OVERARCHING PRINCIPLES OF PATIENT INVOLVEMENT IN OMERACT
   1. OMERACT values the experiential knowledge of Patient Research Partners (PRPs)
   2. Engaging patient research partners (PRPs) as integral stakeholders throughout the research process is a fundamental OMERACT principle.

D. RECOMMENDATIONS FOR PUTTING THE OVERARCHING PRINCIPLES INTO PRACTICE
1. Working Group leadership and appropriate representation 82
2. Patient research partners should be identified based on experiential knowledge and language skills, and personal interest 84
3. Patient Research Partners and the Working Group leadership should discuss the goals of the project and mutual expectations. 84
4. Patient Research Partners should be given the opportunity to be involved throughout the research process. 84
5. The working group leadership should provide PRPs with timely and tailored support and information. 85
6. The nature of Patient Research Partner involvement should be reported throughout the OMERACT process 85
7. Involvement of Patient Research Partners should be recognized appropriately including co-chairing, co-presenting and co-authorship if applicable 85

E. METHODS OF SUPPORTING PRP INVOLVEMENT 85
   1. Information 85
   2. Invitations to meetings 86
   3. Support overall 86
   4. Support during meetings and conferences 86
   5. Support between meetings 86
   6. Support from the OMERACT Executive Committee 86
   7. List for Working Group leaders 88

F. DURING THE OMERACT MEETING 89
   1. Patients opening session 89
   2. Patients daily session 89
   3. Patients final session 89

PART 2 – A BRIEF HISTORY OF OMERACT PATIENT RESEARCH PARTNERS 90
   1. Introduction 90
   2. Initial decision to invite PRPs 90
   3. Subsequent role of PRPs 90
   4. Patient contributions to OMERACT meetings and outcome research 90
   5. Five categories of contribution 90
   6. An added perspective 91

G. REFERENCES 92

CHAPTER 9 - THE OMERACT CONFERENCE: PLANNING TO PUBLICATION 94

A. THE OMERACT CONFERENCE 94
   1. Membership 94
   2. Planning the OMERACT conference 95
   3. Obtaining a time slot in the OMERACT conference 95

B. OMERACT PUBLICATIONS 96

C. OMERACT PUBLICATION POLICY 96
   1. Pre-conference materials 96
   2. Conference session reports 97

D. FUNDING 97
   1. Funding support 97
2. Managing the funds 97
3. Availability of papers 98

D. REFERENCES 98

CHAPTER 10 - SUPPORTING PARTICIPANTS WITH EDUCATION AND DEVELOPMENT AT OMERACT (FELLOWS & NEWBIES) 99

A. INTRODUCTION 99

B. OMERACT FELLOWS:
   BACKGROUND 99
   FELLOW ELIGIBILITY CRITERIA & ROLE 99
   A) WG-SPONSORED 100
   B) WG-NOT PRE-SPONSORED 100
   1. Overall process at the conference 101
   2. Fellows opening session 101
   3. Fellows daily session 101

FIRST TIME OMERACT PARTICIPANTS 93
   1. New participants opening session 93
   2. New participants daily session 93

C. SMALL GROUP MODERATORS AND REPORTERS 94

D. WORKING GROUP LEADERS 94

E. REFERENCES 94

The OMERACT Handbook
CHAPTER 1 - HOW OMERACT EMERGED AND DEVELOPED

A. A PLETHORA OF TRIAL OUTCOMES

Between 1983 and 1988 a series of papers demonstrated that Rheumatologists varied considerably in the way they use clinical measures to make judgments about the efficiency of treatment [1]. Clinicians came to markedly different conclusions about individual patient responses to treatment when managing Rheumatoid Arthritis (RA) in routine clinical practice. [2]. Meanwhile, in clinical trials in RA it had been common to use various traditional measures to define the endpoints of the trial. However, the measures chosen were often unique to a study, not comprehensive, insensitive to change, and measured overlapping concepts. [3] Despite conferences, reviews, and editorials [4-9] no consensus emerged on the appropriate endpoints to include in RA clinical trials [10]. In this context, an endpoint is any measure that is used in the evaluation of patients with RA.

B. INADEQUATE METHODOLOGICAL FRAMEWORK

During this period a methodological framework for selecting valid endpoints and indices was proposed [7,10,11]. However, the problems with existing measures were in their validity, their relation with individual patient outcomes, and in their multitude. It became recognized that some measures represent the patho-physiological occurrences that follow from the cause of RA, such as inflammatory activity (whereas outcome measures represent the suffering or loss of health experienced by an individual because of the process of disease [12,13]). Outcome reflects the values of the patient and of society. Emerging from the variety of conferences and discussions, the list of endpoints contained numerous process measures, as well as, mixed process and outcome measures. Furthermore, most of the dimensions of health status were inadequately addressed, and many of the process measures suggested could not serve as proxy for patient outcome such as death or disability. Some recommended endpoints were likely to be invalid because they are redundant (e.g. tender/painful/swollen joints), unreliable (e.g. 50 foot walk time), or insensitive to change (e.g. rheumatoid factor). The lack of standardization meant that each group used its own variant under a common name (e.g. ‘active joint count’). To compound the problem, the multiplicity of outcome measures, assessments, and comparisons in most trials made it extremely difficult to interpret the result of a particular trial.

C. THE FIRST OMERACT MEETING

It was as part of this active questioning of traditional approaches, the recognition of the need for coherence, and an agreed common approach that the first OMERACT conference was convened. (OMERACT was originally an acronym for Outcome Measures in RA Clinical Trials, now it represents the more inclusive scope of ‘Outcome Measures in Rheumatology’.) It was designed to join the more methodologically oriented approaches that had begun separately in the USA through activities within the American College of Rheumatology (ACR) [14] and in Europe [15]. The first OMERACT conference was held in Maastricht in 1992. It was presented under the auspices of the World Health Organization (WHO), the International and European Leagues Against Rheumatism (ILAR and EULAR), and several of the world’s national colleges of rheumatology. This conference brought together 92 rheumatologists, methodologists, drug regulatory officials, and pharmaceutical physicians from all over the world, including many of those involved in formulating the recommendations outlined above [16]. One of the objectives of the conference was to develop consensus on the minimum number of outcome measures to be included in all RA clinical trials. Other objectives included the development of criteria for minimum clinically important improvement in RA patients, and minimum important difference between treatment groups in RA clinical trials; and study of the usefulness of aggregate outcome measures (indices) in the assessment of patients and trials. In plenary sessions and in small groups, various techniques were used to elicit opinions and preferences: direct questioning; rating of sample profiles of patients and trials; and interactive voting before and after discussion.
D. THE FIRST CORE SET

Participants in the first OMERACT meeting found the gathering of intense engagement. The conference hours were long, the discussions sometimes heated, but evidence took precedence over opinion and nominal group consensus [17] took precedence over eminent pronouncement. At times, it could be overwhelming, but by the end the meeting was exhilarating. Agreement was achieved on the outcome domains that later became known as the WHO/ILAR core set [18] (often called the ACR core set as they were subsequently formally approved by an ACR committee [19]). Agreement on a preferred method of measurement for each of the core outcomes could not be achieved, and it was decided to postpone discussions on this issue pending further studies comparing validity of the different methods. An important aspect of this conference was the explicit consideration of methodological issues that emerged, including measurement methodology. There was clear recognition of the need to develop improvement criteria and to explore compound indices of outcome.

E. OMERACT GETS INTO ITS STRIDE

The measures agreed upon were considered preliminary and a proactive program was planned to test not only the validity of these endpoints, but also, the methods for their measurement. This was the start of a continuing process which has resulted in an OMERACT meeting every two years since then. By OMERACT 3, osteoarthritis was under consideration, and subsequently many more disease areas have been incorporated into OMERACT activities (see www.omeract.org). By OMERACT 5 the US Food and Drug Administration had adopted the core set measures for RA and all pharmaceutical companies were including them in their clinical trials. It had appeared most arguments regarding outcome measures in RA had finally been settled. Consideration was now being given to what degree measures would have to be altered for the benefit to be considered ‘clinically important’. It was at this meeting that participants realized that the patients’ voice was missing from the discussions It was resolved to include patients’ perspectives at subsequent conferences, and the consequences of this are briefly reviewed here [20].

F. PATIENT PARTNERS MAKE A DIFFERENCE

Perhaps the biggest impact was to recognize that, despite its success, the RA core set omitted several outcomes of major importance to patients. One of these was fatigue. Since then, a 10-year research programme has resulted in the effective addition of fatigue to the RA core set [21]. Stimulated by the OMERACT process, agreement on core outcome measurement sets has progressed substantially for many other rheumatological conditions. More recently, there are so many conditions being addressed that OMERACT Working Groups have been set up to help carry forward the required development and validation work, and after which, they report back at Special Interest Group meetings during the main OMERACT conference.

G. THE OMERACT FILTER AND FILTER 2.0

Four members of the OMERACT Executive Committee summarised the underlying philosophy of OMERACT by inventing the phraseology of the OMERACT Filter [22]. This encapsulated the concepts of validity in three common and easily remembered words: truth, discrimination and feasibility (Table 1) [22]. But the participation of patients, and the increasingly broad range of rheumatic conditions under consideration, exposed the existence of underlying common assumptions that OMERACT participants had shared when working intensively with one disease (RA) with which they were all familiar. It was recognized that these assumptions, both about the way domains are selected and the way instruments are validated within those domains, need to be made explicit if they are to support core set development more widely, both within Rheumatology, and within medicine generally, where the notion of core outcome measures was beginning to receive greater prominence [23].
In 2010 the OMERACT community set out to develop Filter 2.0 – a clear statement of what OMERACT means by core outcomes and how it agrees on them. At the OMERACT 11 meeting in 2012 these were extensively explored, critically reviewed, and produced in a final form so that the new filter was adopted by a 96% vote at the final plenary session [24]. OMERACT Filter 2.0 incorporates everything in the original OMERACT Filter, but places the original Filter within a broad philosophical approach and provides a careful explanation of how it can be implemented. Summarised in Table 2, the OMERACT Filter 2.0 firmly embraces the role of patient collaborators at all stages of outcome determination. Domains within 4 overall Areas of outcome are selected to represent the minimum number of domains that capture the needs of the condition under consideration – the Core Domain Set. Next, instruments to measure these outcomes are identified (based on evidence and appropriateness of construction). Then they must meet the demands of the original OMERACT Filter. In this way, a minimum Core Outcome Measurement Set can be derived, which will be included in every trial undertaken in the condition, in addition to any other outcomes of primary or secondary interest to the investigators undertaking the trial.

H. SOME CHALLENGES FOR OMERACT

OMERACT has always had a commitment to international collaboration [25]. It is also a ‘bottom up’ organisation in which members select themselves by participating in its activities. In practice the lion’s share of attendees has come from North America and Europe (although Australia has been well represented). There have been few participants from Africa, South America and Asia, although the 20-strong patient group actively sought the inclusion of 3 patients from Indonesia when OMERACT 10 was held in KotaKinabalu. The question was then asked: Would it make a difference if there were wider participation? Different health care systems, different social imperatives and different manifestations of disease might mean that clinicians, researchers and patients from these geographies might have different approaches to the domains and instruments that should be included in all clinical trials. Would the balance of consensus be changed again, like it was when the patient view was incorporated in the RA core set? The only way to find out is to make it happen.

But the methodology of outcome assessment also continues to develop, with greater inspection being paid to the quality of the methods for choosing Core Domain Sets and Core Outcome Measurement Sets. If OMERACT is to keep up to date, it will need to engage with and embrace these developments. Judging by the pace of these developments, and it might not be long before Filter 2.0 becomes Filter 2.1.
### Table 2. – Characteristics of OMERACT Filter 2.0

**Structure**

- There are two concepts to outcome incorporating the impact of health conditions and their pathophysiological manifestations.
- There are four Core Areas of outcome: Death; Life Impact; Resource Use* and Pathophysiological Manifestations. Every clinical trial must include at least one measure under each of these headings.
- Within each Core Area are Domains of interest to particular conditions. Experts and stakeholders should determine at least one Domain to be a core outcome within each Core Area. This is the Core Domain Set. Trial designs are not limited to the Core Domain Set but should include them in all clinical trials in that condition in addition to any other domains that might be relevant to their investigation.
- Within each Core Domain at least one valid outcome measure should be identified. Validity is assured by meeting the requirements of Truth, Discrimination and Feasibility (as described in the original OMERACT Filter).
- The resultant Core Outcome Measurement Set, which includes at least one instrument from each Core Domain, and at least one domain from each Core Area, should be included in the outcomes of all clinical trials in that condition. Trial designers may also incorporate any other outcomes of interest, including a designated primary outcome which is not part of the Core Outcome Measurement Set but is relevant to their investigation.

**Process**

**Identify the Core Domain Sets**

- A literature review of domains and instruments previously used in the condition
- A review of the setting and any contextual factors that need to be taken into account
- Structured enquiry with stakeholders on their views on domains of importance
- Full participation of all stakeholders (including patients) in a consensus process to determine agreement on what to measure – the Core Domain Set

**Identify the Core Outcome Measurement Set**

- Full literature review to identify validated and applicable outcome instruments for each Core Domain
- Validate instruments in the condition of interest if this has not been done
- Develop and validate new instruments for a Domain that does not have an outcome measurement instrument
- Full participation of all stakeholders (including patients) in a consensus process to determine agreement on how to measure – the Core Outcome Measurement Set

*At present strongly recommended only*
I. REFERENCES

20. Gossec L, Kirwan J, de Wit M. Patient perspective in outcome measures developed by OMERACT. JIR This Issue.
25. www.OMERACT.org
CHAPTER 2: THE SPIRIT OF OMERACT

The Spirit of OMERACT: Q Methodology Analysis of Conference Characteristics Valued by Delegates

Caroline A. Flurey, John R. Kirwan, Phillip Hadridge, Pamela Richards, Shawna Grosskleg, and Peter S. Tugwell

**ABSTRACT. Objective.** To identify the major features of OMERACT meetings as valued by frequent participants and to explore whether there are groups of participants with different opinions.

**Methods.** Using Q methodology (a qualitative and quantitative approach to grouping people according to subjective opinion), participants (who attended more than 1 OMERACT conference) sorted 66 statements relating to the “spirit of OMERACT” according to level of agreement across a normal distribution grid. Data were examined using Q factor analysis.

**Results.** Of 226 potential participants, 105 responded (46%). All participants highly ranked the focus on global standardization of methods, outcome measures, data-driven research, methodological discussion, and international collaboration. Four factors describing the “spirit of OMERACT” were identified: “Evidence not eminence” (n = 31) valued the data- and evidence-driven research above personality and status; “Collaboration and collegiality” (n = 19) valued the international and cross-stakeholder collaboration, interaction, and collegiality; “Equal voices, equal votes, common goals” (n = 12) valued equality in discussion and voting, with everyone striving toward the same goal; “principles and product, not process” (n = 8) valued the principles of focusing on outcome measures and the product of guiding clinical trials, but were unsure whether the process is necessary to reach this. The factors did not segregate different stakeholder groups.

**Conclusion.** Delegates value different elements of OMERACT, and thus the “spirit of OMERACT” encompasses evidence-based research, collaboration, and equality, although a small group are unsure whether the process is necessary to achieve the end result. Q methodology may prove useful for conference organizers to identify their delegates’ different needs to tailor conference content. (First Release August 15 2015; J Rheumatol 2015;42:1982–92; doi:10.3899/jrheum.150113)
Outcome Measures in Rheumatology (OMERACT) is an informal international network initiated in 1992 with the aim of improving outcome measurement in rheumatology. Through a biannual conference and working groups that carry out research between conferences, OMERACT has built data-driven consensus for many rheumatologic conditions, including rheumatoid arthritis (RA), ankylosing spondylitis, and osteoarthritis. It has done so by developing widely endorsed “core outcome measurement sets,” each a minimum set of outcome measures covering key domains that must be reported in all randomized controlled trials in a given health condition. Working groups prepare or produce an evidence base to support the identification of domains and instruments to measure those domains in their own areas of interest and expertise (e.g., RA flares) through literature reviews and validation studies. At the conference, plenary presentations alternate with small group sessions (breakout sessions), where delegates have the opportunity to express their views and preferences. The views expressed in the breakout sessions are then reported at a plenary session, where final consensus is formed with the help of interactive voting. Consensus can be used to drive a research agenda, as well as to provide agreement on measures or domains.

The design and format of the OMERACT conference, which has a limited number of delegate places and is held over 5 days at an integrated venue, has the intention of allowing respectful, deliberative dialogue and relationship building among those involved including clinicians, methodologists, regulatory agencies, industry, and patients, who participate as equal partners in the process.

Many delegates repeatedly attend OMERACT conferences, even though this takes a week of their time (including travel) and is an exhausting, intense experience, with working hours lasting from the early morning until late at night. However, it is not known which features or characteristics of the conference particularly appeal to people and encourage them to return. It is important to understand the subjective views of delegates on what contributes to the “spirit of OMERACT” to ensure that future conferences maintain the elements that delegates value the most. This is particularly timely because OMERACT has for the first time set out its underlying philosophy and published the OMERACT Handbook, which details its operational strategies.

Traditional consensus methods such as the nominal group technique and the Delphi technique rely on interactive discussion between participants or repeated interactions until participants are able to reach an agreement. However, this produces an averaging of opinions. The average scores of value ratings attributed by delegates to different aspects of the conference might provide a broad overview, but these methods will not reveal the existence of groups of delegates with different views, nor explore the interactions between different features of the conference program. Q methodology is specifically designed to derive a number of groups (called “factors” in Q methodology), which each represent a different and independent opinion of the issue.

Q methodology combines strengths of qualitative and quantitative approaches to identify factors made up of those who share similar opinions based on their value judgments in prioritizing a large set of descriptors. An appropriate set of descriptors or statements is assembled, each participant sorts them along a continuum of agreement designed to emulate a normal distribution with relatively few extreme values and many central values (Figure 1). The normal distribution is used because it is believed to be the closest distribution to the way that people form opinions.
Participants sort each statement in approximate rank order of the degree to which they agree with that statement in relation to all the other statements. By using a normal distribution grid, the weight assigned to each statement in the analysis increases slightly for each place closer to the extreme ends of the grid, and the positions (scores) of all the statements are included in the analysis for every participant. The statements sorted closer to the outer edges of the distribution have more influence on the factor groupings in relation to the statements closer to the middle of the distribution. Factors are calculated and the results interpreted within the context of the enquiry\textsuperscript{5,6}. Exploring the similarities and differences between the opinions of these groups of participants will provide greater insight into the aspects that different delegates find attractive, and will also allow a comparison of the characteristics of the participants who fall into the different factors. Therefore, we invited all OMERACT delegates who had chosen to return to the conference at least once to take part in an online Q sort survey.

A. MATERIALS AND METHODS

Participants (P-set). Participants gave informed consent by anonymously deciding to participate, and ethics approval was granted by the University of the West of England Research Ethics Committee (Ref: HAS/13/11/151). All OMERACT delegates who had attended at least 2 OMERACT conferences were invited to participate. In all, 332 eligible participants covering all stakeholder groups (clinicians, researchers, fellows, patients, and industry) were identified from previous conference delegate lists and contacted by e-mail to explain the study and invite them to participate online. Two e-mail reminders were also sent, and encouragement to participate continued until participation was closed on the day before a presentation of results from the first 50 respondents was made at an OMERACT conference. Survey reminders were sent to all delegates, because the survey was anonymous. The first delegates attended in 1992, thus an estimate of the potential responder population suggested that 226 would be available to respond, based on UK life expectancy and expected retirements.

Statements (Q set). Several sources were used to collate or produce a wide range of statements regarding the features of OMERACT that might be important to participants. First, data were collected during an internal review of activities conducted 2 years earlier by an independent consultancy (iDENK\textsuperscript{15}). This included interviews, focus groups, and a survey with OMERACT delegates; attendance at executive committee meetings and conference calls; and interactions with working groups. Second, data were collected as part of an investigation into the way patients have been incorporated as participants in OMERACT, including interviews, reports of attendance at patient sessions, analysis of OMERACT documents, and analysis of OMERACT conference proceedings\textsuperscript{16,17}. Finally, to ensure a full range of potential opinions about OMERACT, delegates who had attended only 1 OMERACT conference (either 9 or 10) and chosen not to return were asked to provide their opinions of OMERACT. The statements were refined through discussion with the research team, including a patient research partner. After removing repeated or ambiguous items, 66 statements were included, each worded to follow on from the stem: “To me, the ‘spirit of OMERACT’ is...” (Table 1).
Table 1. By-factor ranking of statements (S1) given in the “spirit of OMERACT” Q study.

<table>
<thead>
<tr>
<th>To me the “spirit of OMERACT” is...</th>
<th>Factor One</th>
<th>Factor Two</th>
<th>Factor Three</th>
<th>Factor Four</th>
<th>Mean Score Across All Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>St1: The credible reputation of the conference</td>
<td>0</td>
<td>0</td>
<td>+1</td>
<td>0</td>
<td>+0.46</td>
</tr>
<tr>
<td>St2: The open and vociferous discussion</td>
<td>+1</td>
<td>+1</td>
<td>+2</td>
<td>0</td>
<td>+1.31</td>
</tr>
<tr>
<td>St3: Everyone’s opinions being treated as equal, regardless of their status</td>
<td>+2</td>
<td>+3</td>
<td>+1</td>
<td>-3</td>
<td>+1.1</td>
</tr>
<tr>
<td>St4: The focus on small group discussions rather than presentations</td>
<td>+2</td>
<td>+7</td>
<td>+3</td>
<td>+2</td>
<td>+2.13</td>
</tr>
<tr>
<td>St5: Patients being invited to the conference</td>
<td>+4</td>
<td>+5</td>
<td>-3</td>
<td>+1</td>
<td>+1.26</td>
</tr>
<tr>
<td>St6: Patients being given the power to drive a research agenda (e.g., fatigue, well-being)</td>
<td>+3</td>
<td>+4</td>
<td>0</td>
<td>-1</td>
<td>+1.24</td>
</tr>
<tr>
<td>St7: Helping therapies to get approved</td>
<td>-1</td>
<td>-5</td>
<td>-2</td>
<td>0</td>
<td>-1.86</td>
</tr>
<tr>
<td>St8: Its innovative nature</td>
<td>+2</td>
<td>+1</td>
<td>+1</td>
<td>0</td>
<td>+1.02</td>
</tr>
<tr>
<td>St9: The opportunity to deal with controversial issues</td>
<td>0</td>
<td>+3</td>
<td>0</td>
<td>+1</td>
<td>+0.79</td>
</tr>
<tr>
<td>St10: Volunteers driving the conference</td>
<td>0</td>
<td>-1</td>
<td>-2</td>
<td>-2</td>
<td>-0.48</td>
</tr>
<tr>
<td>St11: That it is neutral ground for ideas to be discussed</td>
<td>0</td>
<td>+2</td>
<td>0</td>
<td>-4</td>
<td>+0.59</td>
</tr>
<tr>
<td>St12: The commitment to theoretical underpinnings</td>
<td>+2</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>+0.32</td>
</tr>
<tr>
<td>St13: The involvement of a wide variety of stakeholders</td>
<td>+2</td>
<td>+2</td>
<td>+1</td>
<td>+5</td>
<td>+1.6</td>
</tr>
<tr>
<td>St14: The opportunity to hear about progress in areas of work other than my own</td>
<td>0</td>
<td>0</td>
<td>+2</td>
<td>-3</td>
<td>+0.17</td>
</tr>
<tr>
<td>St15: The opportunity to convince my peers that my work is satisfactory</td>
<td>-3</td>
<td>-4</td>
<td>-5</td>
<td>-2</td>
<td>-2.4</td>
</tr>
<tr>
<td>St16: Having to be thick-skinned</td>
<td>-4</td>
<td>-6</td>
<td>-4</td>
<td>-1</td>
<td>-3.28</td>
</tr>
<tr>
<td>St17: The sleepless nights</td>
<td>-5</td>
<td>-6</td>
<td>-6</td>
<td>-4</td>
<td>-3.35</td>
</tr>
<tr>
<td>St18: The research being driven by data/evidence</td>
<td>+6</td>
<td>+2</td>
<td>+7</td>
<td>+2</td>
<td>+3.02</td>
</tr>
<tr>
<td>St19: The opportunity for interactive discussion</td>
<td>+3</td>
<td>+4</td>
<td>+2</td>
<td>+2</td>
<td>+2.46</td>
</tr>
<tr>
<td>St20: The intimacy (small number of delegates)</td>
<td>-1</td>
<td>+5</td>
<td>+3</td>
<td>+4</td>
<td>+0.9</td>
</tr>
<tr>
<td>St21: The chance to get the OMERACT seal of approval</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>-0.37</td>
</tr>
<tr>
<td>St22: The chance to get international recognition for my work</td>
<td>-2</td>
<td>-2</td>
<td>-5</td>
<td>+2</td>
<td>-1.45</td>
</tr>
<tr>
<td>St23: The focus on outcome measures</td>
<td>+7</td>
<td>+6</td>
<td>+5</td>
<td>+7</td>
<td>+3.82</td>
</tr>
<tr>
<td>St24: International collaboration</td>
<td>+5</td>
<td>+3</td>
<td>+4</td>
<td>+6</td>
<td>+2.43</td>
</tr>
<tr>
<td>St25: Senior and junior delegates working together</td>
<td>+3</td>
<td>+1</td>
<td>0</td>
<td>+3</td>
<td>+1.27</td>
</tr>
<tr>
<td>St26: The opportunity to meet “famous” researchers/rheumatologists</td>
<td>-4</td>
<td>-3</td>
<td>-5</td>
<td>-1</td>
<td>-2.48</td>
</tr>
<tr>
<td>St27: The feeling of loyalty</td>
<td>-2</td>
<td>-1</td>
<td>-3</td>
<td>-4</td>
<td>-1.74</td>
</tr>
<tr>
<td>St28: The feeling of belonging</td>
<td>-2</td>
<td>+1</td>
<td>0</td>
<td>+1</td>
<td>-0.32</td>
</tr>
<tr>
<td>St29: Getting work done to tight time scales</td>
<td>-1</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>-0.71</td>
</tr>
<tr>
<td>St30: The organized chaos</td>
<td>-3</td>
<td>+1</td>
<td>-4</td>
<td>+1</td>
<td>+1.5</td>
</tr>
<tr>
<td>St31: The special interest groups</td>
<td>+1</td>
<td>0</td>
<td>0</td>
<td>+4</td>
<td>+0.92</td>
</tr>
<tr>
<td>St32: Having a fellows programme for novice researchers</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>-3</td>
<td>+0.25</td>
</tr>
<tr>
<td>St33: Having a “baddy” system for new patient delegates</td>
<td>-2</td>
<td>-2</td>
<td>-4</td>
<td>-3</td>
<td>-1.3</td>
</tr>
<tr>
<td>St34: The gladiatorial nature (“newbies” have to prove their robustness and worth)</td>
<td>-4</td>
<td>-5</td>
<td>-6</td>
<td>-1</td>
<td>-2.77</td>
</tr>
<tr>
<td>St35: The transparency</td>
<td>+5</td>
<td>-2</td>
<td>0</td>
<td>-5</td>
<td>+0.12</td>
</tr>
<tr>
<td>St36: The focus on goal setting</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>+2</td>
<td>+0.57</td>
</tr>
<tr>
<td>St37: The voting process being at the conference itself (enabling decisions to be made there and then)</td>
<td>+3</td>
<td>+2</td>
<td>+5</td>
<td>-3</td>
<td>+1.56</td>
</tr>
<tr>
<td>St38: The equal voting process (each person is given an equal vote regardless of their experience/interest in the topic)</td>
<td>+2</td>
<td>+1</td>
<td>+3</td>
<td>-6</td>
<td>+0.88</td>
</tr>
<tr>
<td>St39: The focus on striving for consensus</td>
<td>+4</td>
<td>+2</td>
<td>+6</td>
<td>+3</td>
<td>+2.03</td>
</tr>
<tr>
<td>St40: The involvement of a core committed group of people</td>
<td>+1</td>
<td>+3</td>
<td>+2</td>
<td>+4</td>
<td>+1.38</td>
</tr>
<tr>
<td>St41: The emphasis on striving for global standardization and validation of methods</td>
<td>+6</td>
<td>+5</td>
<td>+6</td>
<td>+6</td>
<td>+3.88</td>
</tr>
<tr>
<td>St42: The lively methodological discussion</td>
<td>+4</td>
<td>+4</td>
<td>+3</td>
<td>+4</td>
<td>+2.48</td>
</tr>
<tr>
<td>St43: The opportunity to discuss novel unpublished material</td>
<td>-1</td>
<td>0</td>
<td>-2</td>
<td>-2</td>
<td>-0.38</td>
</tr>
<tr>
<td>St44: Having less visible egos than at other conferences</td>
<td>-3</td>
<td>-4</td>
<td>-4</td>
<td>-7</td>
<td>-2.71</td>
</tr>
<tr>
<td>St45: The beautiful, exotic locations chosen for the conference venue</td>
<td>-5</td>
<td>-3</td>
<td>+1</td>
<td>-3</td>
<td>-2.11</td>
</tr>
<tr>
<td>St46: The remote locations chosen for the conference venue (cut off from civilization)</td>
<td>-5</td>
<td>-2</td>
<td>+3</td>
<td>-2</td>
<td>-1.7</td>
</tr>
<tr>
<td>St47: The final night entertainment</td>
<td>-6</td>
<td>-5</td>
<td>-3</td>
<td>-6</td>
<td>-3.23</td>
</tr>
<tr>
<td>St48: The focus on guiding the conduct of clinical trials</td>
<td>+4</td>
<td>-3</td>
<td>+5</td>
<td>+5</td>
<td>+1.26</td>
</tr>
<tr>
<td>St49: That it focuses and drives the research progress made in between meetings</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1.01</td>
</tr>
<tr>
<td>St50: Reinforcing the rules for adequate clinical trials</td>
<td>+3</td>
<td>-1</td>
<td>+4</td>
<td>+5</td>
<td>+1.31</td>
</tr>
<tr>
<td>St51: The intellectual stimulation</td>
<td>+1</td>
<td>+6</td>
<td>+1</td>
<td>+3</td>
<td>+1.9</td>
</tr>
<tr>
<td>St52: That when consensus is achieved it feels hard-won and deserved</td>
<td>0</td>
<td>0</td>
<td>+4</td>
<td>-4</td>
<td>+0.48</td>
</tr>
<tr>
<td>St53: The exchange of ideas to address shared goals and challenges in different disease areas</td>
<td>+5</td>
<td>+4</td>
<td>+4</td>
<td>+2</td>
<td>+2.15</td>
</tr>
<tr>
<td>St54: The feeling of being part of something unique</td>
<td>-1</td>
<td>+2</td>
<td>+2</td>
<td>-1</td>
<td>-0.54</td>
</tr>
<tr>
<td>St55: The use of the Delphi procedure</td>
<td>+1</td>
<td>-2</td>
<td>+3</td>
<td>-2</td>
<td>+0.3</td>
</tr>
</tbody>
</table>

18 The OMERACT Handbook
Obtaining participants’ preferences. The Q methodology study was completed online using FlashQ, a software package designed for collecting Q methodology data\(^{18}\). It was hosted for the present study on a University of the West of England Website and managed by the first author (CF). A prestudy e-mail was sent to potential participants, advising them when the study would be starting and providing a very brief outline of its intent. This was followed a month later by the invitation e-mail message, which contained a participant information sheet explaining the study and a link to the Website. Participants were advised to allow 45–60 min to complete the online study. Because data were submitted anonymously, reminder e-mails were sent to all participants at intervals of 2 weeks after the initial invitation to thank those who had responded and encourage those who had not yet responded. Participants had the option of asking the research team to stop further reminder e-mails.

The link to the online study presented participants with instructions for completing the Q sort (Table 2). As a first step, participants were presented with the statements one at a time and asked to consider each one in relation to the statement stem “To me, the ‘spirit of OMERACT’ is…” and to sort them into 3 broad categories: agree most; agree least or disagree; neutral. The statements were presented to each participant in a different random order.

When this was complete, the second step asked each participant to review their broad categories and arrange each statement in approximate rank order of the degree to which they agreed with that statement relative to the other statements. Each of the 66 statements was placed in a single box on the Q sort grid (Figure 1) of 66 boxes. The grid pattern allowed for the majority of statements to be agreed or disagreed with mildly or neutrally (for example, there were 10 “0” boxes, and 7 “+1” or “−1” boxes each), but only 1 statement could be placed in the highest agreement box (“+7”) or highest disagreement box (“−7”). Thus each participant’s opinions on the statements were constrained into a quasi-normal distribution of degrees of agreement with the statements (Figure 1)\(^{19}\). The precise shape and limits of agreement/disagreement of this distribution (and the grid) are dependent on the number of statements. A participant could rearrange the position of each statement within the grid as the procedure continued, until satisfied with the distribution.
In the third step, participants were asked to comment on the statements they had placed at the extreme positions at either end of the grid, and were also given the opportunity to provide an open comment about their reasons for how they had sorted the statements overall, or any comments on any particular statements. Finally, additional data were also collected on age, sex, number of OMERACT attendances, and delegate category (e.g., industry, patient).

**Analysis.** The statements are assigned a score for each participant based on which column on the grid they have been placed in (−7 to +7). To provide a broad overview, the mean of the scores given by all the participants to each statement was calculated. For detailed analysis, Q methodology combines qualitative and quantitative methods to produce a rounded interpretation of a single dataset (in contrast to a mixed-methods approach). Q methodology analysis involves factor extraction, rotation, and interpretation. Factor extraction and rotation used the PCQ software package. In Q methodology, participants are treated as variables and are intercorrelated and subjected to by-person factor analysis. The software searches for shared patterns (or sorting configurations) in the data and extracts portions of common variance (factors). For each Q factor to be interpretable, an eigenvalue > 1.0 (indicating factors are unlikely to have grouped participant views by chance), and ≥ 1. Q sort loading significantly upon each factor alone is required. Following extraction, the factors were rotated using orthogonal varimax rotation to ensure each Q sort defined (has a high factor loading in relation to) only 1 of the study factors, so the overall solution maximizes the amount of study variance explained. For ease of interpretation it is standard Q-methodological practice to generate a single exemplary Q sort for each factor by merging (according to a procedure of weighted averaging) the Q sorts of all significantly loading participants on the given factor (termed the factor array). A decision on the final selection of the optimum factor solution was undertaken collectively by the authors, examining the outputs from the different factor analysis solutions (e.g., number of factors, weightings, explained variance, number of participants excluded from factors). Factor interpretation was based
on the factor arrays and the open-ended comments from the factor exemplars (significantly loading participants), which were combined to provide a single gestalt explanation of each factor\textsuperscript{23,24}. Illustrative examples of these open-ended comments are included in the results.

B. RESULTS

Study population. Of the estimated 226 participants available to respond, 105 participated (46%). The participants were 59 researchers (56%), 25 clinicians (24%), 7 patients (7%), 9 from industry (8%), and 5 fellows (5%). A preliminary analysis of the first 50 participants to respond was presented at the OMERACT 12 meeting in 2014. Recruitment ended the day before the presentation, by which time a further 55 respondents had taken part. For these 2 sets of respondents (effectively, early respondents and later respondents who had received a larger degree of encouragement to take part), separate and combined factor analyses provided very similar results, testing multiple factor solutions. Further, the demographic data were very similar in the 2 groups. Therefore we present a combined analysis of all 105 participants. Table 3 shows the study population and demographic data for each factor. Of the 105 people who participated in this study, 60% were male and 56.2% were researchers (mean OMERACT attendance: 4.36 times, SD 2.54). Median time to complete the online study was 25 min (interquartile range 19-35 min).

Overview of consensus statements. The mean score for each statement is shown in Table 1. When ordered from highest to lowest, obvious inflexions suggest that statements scoring

\[ \pm 2.43 \text{ or } \pm 3.23 \] have a very high degree of consensus (see bold type, Table 1). These statements indicate the aspects of OMERACT most valued by participants as a whole: First, the specific focus of OMERACT on striving for global standardization and/or validation of methods [Statement (St) 41: overall mean statement score +3.88]; outcome measures (St23: +3.82); and the research being driven by data and/or evidence (St18: +3.02). As participant (P) 38 commented: “It is and should always be the first letter: ‘OUTCOME’ in OMERACT” (P38, clinician, male). Second, participants value the community aspects of lively methodological and interactive discussion [St42: +2.48; St19: +2.46; “Coming away from the meeting exhausted but stimulated” (P31, clinician, female)]; and inter- national collaboration (St24: +2.43).

The great majority of participants disagreed with the statement “OMERACT is just an elitist clique” (St64: −4.28): “Derogatory words that misrepresent what OMERACT is about: Anyone can join who is willing to shoulder the work.” (P50, researcher, male). There was also consensus disagreement with the ideas that you need to be thick-skinned or have sleepless nights to attend (St16: −3.28; St 17: −3.35). Further, while the final evening entertainment was often discussed fondly, it was not considered important in contributing to the “spirit of OMERACT” (St47: −3.23): “Of course the exotic places and the final night, but you cannot put that [as the item you] like the best!” (P22, clinician, male).

Q methodology overall factor solution. A 4-factor solution was used based on (1) the solutions having a maximal explained variance; (2) the solutions having a maximum number of Q sorts loading significantly on 1 factor; (3) all factors with eigenvalue > 1.00; (4) all factors containing statements distinguishing them from other factors; (5) a minimum number being confounded (i.e., significantly loading on > 1 factor); and (6) the researchers’ judgment. Table 3 provides a summary of characteristics of all factors presented. Table 1 presents the average rating (provided by the factor array) given by each of the 4 factors to each statement. The 4 factors explain 44% of the variance and account for 70 of the 105 participants (67%). Participant loading of \[ \pm 0.40 \] reached significance at \( p < 0.01 \), indicating that each loading participant closely exemplifies the factor they load onto\textsuperscript{23}. The factors will be presented in order of the highest number of loading participants.
Factor One: “Evidence not eminence.” This factor consisted of 31 participants who were predominantly researchers, but included representation from all stakeholder groups; they were predominantly male; and mean OMERACT attendances were representative of the study population (Table 3).

Factor One exemplars highly valued the OMERACT process of data and evidence being used to drive the research agenda [Table 1: St18: score +6]:

“A strong prerequisite to make sound judgments on interventions in clinical trials” (P34, researcher, male;

“it is based on rigorous data assessment and use of the Filter. I think that this places the OMERACT conference as the ‘crucible’ where ideas for outcome measures are tested” (P58, clinician, male);

“The goal is to improve what we do for the benefit of the patients. Evidence is the only thing that has consistently improved how we help patients” (P66, industry, male).

These participants agreed that the OMERACT process is transparent (St35: +5). They valued the exchange of ideas to address shared goals and reported that this sharing of ideas sets OMERACT apart from other conferences:
“This level of commitment and interest in the details can best be summarized by the “geekiness” variable, and is what separates OMERACT attendees from other meetings” (P33, industry, male).

These participants felt strongly that OMERACT is not “just an elitist clique” (St64: −7). Although this statement score was in agreement with Factors Two and Three, in contrast to those factors, the discourse in this factor suggests that this was due to the belief that the data should always win out over personalities or eminence:

“To win over ‘non-specialists’ on the voting process, one has to be clear and concise. OMERACT members are good at letting the presenters know an argument has not been made on a voting point. I don’t vote for something because I trust the presenter. I vote for something if the presenter has educated me enough to believe him/her” (P51, clinician, male);

“I believe all have their ideas put to the public blow torch at OMERACT, and famous investigators have failed to convince, hopefully underpinning that evidence not personalities rule OMERACT” (P90, clinician, male);

“Evidence, not eminence. Eminence can focus efforts towards laudable goals, but it is never the goal itself” (P66, industry, male).

The discourse from this factor indicates that these participants go to OMERACT to further research objectives and not to socialize. They definitely do not use OMERACT as a chance to escape their working life (St60: −6) and they see the exotic and remote locations as an unnecessary distraction to getting the work done (St45: −5; St46: −5):

“I have other (intellectual and urging) engagements to attend to in daily life, so absolutely no reason to escape from these time-consuming obligations” (P25, patient, female);

“We rarely go outside” (P33, industry, male);

“More distraction or deterrents of getting outcomes completed, adds unnecessary complexity” (P68, industry, male).

Factor Two. “Collaboration and collegiality.” This factor consisted of 19 participants who were predominantly researchers, and more female participants were included in this factor than the other 3 factors. They had a slightly higher mean number of OMERACT attendances (5.11) than the study population (4.36; Table 3).

Participants loading onto Factor Two particularly valued the collaboration and collegiality that OMERACT provides. The elements of OMERACT most important to these participants were those that relate to the interaction with other delegates. They valued the intimate aspect of the conference and the limited number of delegates; the discourse suggested that they saw these elements as allowing increased opportunities for collaboration and interaction (St20: +5):

“Having our own area for dining means that we continue debates over meals. All being in the same location (rather than a range of conference hotels spread across a city) lengthens the day at both ends, and also allows for socializing, which further bonds the group” (P03, researcher, female).

The collaborative character of OMERACT valued by this factor contributed to their decision to reject the idea that OMERACT is “just an elitist clique” (St64: −7), in contrast to the reason of evidence over eminence given in Factor One: “I don’t think of us as elitist ... I find it quite collaborative and hard-working” (P52, researcher, female).

These participants were less concerned than those in other factors with helping to get therapies approved (St7: −5). However, the focus on small group discussions was particularly important to them (St04: +7) because these were seen to be more effective for making progress than big plenary sessions alone; and participants indicated that this is what makes OMERACT innovative and distinct from other conferences: “In small groups, communication and reaching consensus and how we get there is easier and more interesting and useful” (P22, clinician, male);
“This is the most importance difference between OMERACT and other conferences, in the dialogue most progress is found” (P55, researcher, male).

A further element valued by these participants was the opportunity to work together with like-minded people on the same intellectual level. They valued the intellectual stimulation and the opportunity to discuss their work with knowledgeable others (St51: +6; St65: +3):

“There is diverse intellectual discussion within individuals from a variety of backgrounds, which is often productive” (P39, researcher, male).

The value these participants placed on the collaborative character of OMERACT may explain why it is more important to them than any other factor that patients are invited to the conference (St05: +5). They also scored higher than the other factors on the statement “the freedom to be a geek with others like myself” (St63: 0). The comments indicated that while not all participants in this factor saw themselves as a “geek,” they did appreciate the collegiate character of OMERACT:

“I do not feel like a geek. It is just passionate people, and even though most of them are famous researchers, most of them act so humbly it is really stimulating” (P22, clinician, male); “I am considered a geek by colleagues ... yet at OMERACT the place is filled with people who want the job of measurement done right” (P52, researcher, female).

The discourse in this factor suggests that OMERACT is inclusive and welcoming, thus these participants strongly believe that one does not need to be “thick-skinned” to attend (St16: –6); nor do they feel OMERACT is gladiatorial in character, with new delegates having to prove themselves (St34: –5):

“This is unnecessary ego-tripping and can be very destructive. We were all ‘newbies’ once” (P03, researcher, female);

“OMERACT is a very level playing field where everyone involved in groups contributes. The leaders have worked hard to create an environment and ethos that everyone has an important voice and is encouraged to contribute” (P89, clinician, female).

Factor Three: “Equal voices, equal votes, common goals”: This factor consisted of 12 participants who were predominantly researchers, with a slightly lower mean OMERACT meeting attendance number (4.17) than the study population (4.36; Table 3).

Factor Three exemplars valued the elements of the OMERACT process that give all delegates an equal voice in the discussions and an equal vote in the final consensus, while striving toward a common goal. Thus they positively rated the voting being equal and being at the conference (St38: +3; St37: +5), and the use of the Delphi procedure (St55: +3). The discourse in this factor suggested that it is important to these participants that everyone has enough information to contribute equally:

“Ensuring the data is there, that everyone understands it, discussions about strengths and weaknesses, voting, back for the next meeting with more data” (P54, clinician, female).

Perhaps owing to the value placed on equality, these participants do not agree that OMERACT is an opportunity to meet “famous” researchers and/or rheumatologists (St26: –5). They do not agree that new delegates have to prove their robustness and worth (St34: –6) nor that OMERACT is “just an elitist clique” (St34: –7). The discourse suggests that these participants feel everyone is encouraged to contribute and that their opinions are valued: “The leaders are enthusiastic and motivating and even though they may be leaders in their field they are approachable and friendly” (P18, researcher, female);

“Not being a rheumatologist, I learned a great deal and was very much encouraged to contribute and explain my own discipline and its value to rheumatology” (P07, researcher, male).

24 The OMERACT Handbook
These participants value the research being evidence- and data-driven (St18: +7), the focus on guiding the conduct of clinical trials (St48: +5), and striving for consensus (St39: +6). These values and goals are more important to this group than the opportunity to convince their peers that their work is satisfactory (St15: −5), or the chance to get international recognition for their work (St22: −5). This attitude may indicate that the participants’ motivation to attend is something other than the desire for individual recognition. “Everyone is there to learn with common goals of helping patients” (P78, industry, male).

Factor 4: “Principles and product, not process”: Although this factor has the lowest eigenvalue (6.56), the number is substantially higher than that due to chance alone (eigenvalue < 1.00). In addition, this factor was included because the 4-factor solution strengthened the other 3 factors compared to the 3-factor solution, and it seemed realistic that a small number of delegates might hold the views expressed within this factor.

Factor Four consisted of 8 participants who were predominantly researchers, but the sex and mean number of times attending OMERACT meetings were representative of the study population (Table 3). Factor Four exemplars valued the overall principles and products of OMERACT, but were unsure as to whether the OMERACT process is necessary or appropriate to achieve this:

“OMERACT is a great idea that needs updating” (P05, researcher, male).

Similarly to the participants in other factors and the overall consensus, these participants value the overall principle of OMERACT, that is, the focus on outcome measures (St23: +7): “OMERACT should focus on this topic” (P16, clinician, male). They also value the products of OMERACT, including guiding the conduct and reinforcing the rules for clinical trials (St48: +5; St50: +5):

“This is a very important feature of OMERACT; we’ll encourage trialists to use the same standardized outcome measures. OMERACT therefore protecting us against selective outcome reporting from industry” (P42, researcher, male).

However, the participants loading onto this factor were more inclined than all other factors to consider OMERACT as “an elitist clique” (St64: +3):

“Too expensive and too exotic” (P16, clinician, male).

These participants do not agree that the OMERACT process is transparent (St35: −5); has fewer visible egos than at other conferences (St44: −7); nor that everyone’s opinions are treated as equal (St03: −3). However, they do value the involvement of a core committed group of people more than the other factors (St40: +4). They did not agree that the voting process is equal or that it should be held at the conference (St37: −3; St38: −6), which the discourse suggests is because they are unsure whether this process is necessary or appropriate to reach the end product:

“This [voting process] is a bit silly since most things are approved by the time the large group votes. Not all attendees know or care about some topics and really are not qualified to vote (not due to position or training, but due to indifference)” (P23, researcher, male);

“I think that much of the work is done prior to the meeting and that those driving the agenda will paraphrase the questions until they get the response they want, or will continue anyway without consensus” (P56, researcher, female).

Despite holding the opinion that the voting process may be unnecessary, these participants still value the collaborative design of the conference, and the involvement of a wide variety of people (St24: +6; St13: +5):

“International ‘buy-in’ is crucial. [It] forces us to seek consensus and inclusion” (P23, researcher, male).
Despite the seemingly more negative views held within this factor, these participants have a mean number of 4.36 times attending OMERACT meetings and therefore it is likely that the perceived positive aspects of OMERACT (e.g., focus on outcome measures and collaboration) outweigh the perceived negative aspects in this factor.

Confounding sorts (sorts loading onto > 1 factor): A participant’s individual sort is considered to be confounded if it significantly loads (≥±0.40, p < 0.01) onto > 1 factor. This study contained 22 confounded sorts: 9 researchers, 5 clinicians, 5 patients, 2 fellows, and 1 person from industry. Half of confounders (5 researchers, 4 patients, 1 fellow and 1 clinician) loaded onto Factor One: “Evidence not eminence” and Factor Two: “Collaboration and collegiality,” and therefore held views that were equally balanced between the 2 factors. Five of the confounders (1 researcher, 2 clinicians, 1 patient, and 1 person from industry) loaded onto Factor One: “Evidence not eminence” and Factor Three: “Equal voices, equal votes, common goals.” The remaining confounders were distributed across different groups. The majority of the patient participants (6/7) loaded either solely onto Factor One, or were confounded across Factor One “Evidence not eminence” and Factor Two “Collaboration and collegiality,” and therefore held views that were equally balanced between these factors. The remaining participant groups were distributed across the different options.

C. DISCUSSION

Almost half of OMERACT delegates who had chosen to attend more than 1 meeting and were likely to be available to participate in this study provided material to review those aspects of OMERACT that embody the “spirit of OMERACT.” Collectively there is a broad consensus that the focus on outcome measures, global standardization of methods, data-driven research, international collaboration, and methodological discussion contribute to creating the “spirit of OMERACT.” Within this, differences of preference emerge that distinguish 4 relatively distinct factors, for which we have coined the following descriptive labels: Evidence not eminence; Collaboration and collegiality; Equal voices, equal votes, common goals; and Principles and product, not process.

While some delegates believe the evidence- and data-driven elements of the conference should be valued above personality (Evidence not eminence), others value the collaborative and interactive elements of the conference and the feeling of collegiality (Collaboration and collegiality). Some valued the focus on equality, with everyone striving toward the same goal (Equal voices, equal votes, common goals) and others value the principles of focusing on outcome measures and the product of guiding clinical trials, but are unsure whether the process is necessary to reach this (Principles and product, not process).

The different participant groups were distributed among the 4 distinct factors, or were confounded between more than 1 factor. This indicates that participant experiences and opinions of OMERACT cannot be categorized according to participant group. If one group had a particularly distinct opinion or experience, this would have been recorded by the method as a separate factor. It is important to note that although patient participants were not greatly represented in the factor solutions, they were present as confounding sorts. Thus, their opinions were not sufficiently distinct to create a separate factor, but in agreement across different areas of opinion (Factor One or confounded across Factor One and Factor Two).

The finding that OMERACT is valued for its focus on outcome measures, data-driven research, collaboration, and equality through consensus agreement supports the description that OMERACT provides for itself on its Website² and in the OMERACT Handbook⁹. The finding in the current study that data-driven research is highly valued by some delegates supports previous research¹⁷ that found one obstacle to accepting patient participation at OMERACT was concern that knowledge brought by patients is experiential. Further, the finding from the current study that collaboration and every delegate being given an equal voice in the discussion are important contributors to the “spirit of OMERACT” supports the finding that many professionals had their perceptions changed by patients:
“Patients were a kind of sparring partner when I entered a relatively new area” 17.

These findings also support an evaluation of a conference focused on the inclusion of minorities into clinical research trials 25, which found that inclusion and collaboration is necessary to make progress.

These are the first data using Q methodology to evaluate a research conference, and it has proved to be a valuable way of measuring delegates’ motivations for attendance. Conference organizers may find this method useful to understand the differences among the delegates, enabling organ-izers to tailor the conference to delegates’ different needs.

A limitation of Q methodology is that participants are given predetermined statements to sort, and thus novel ideas may be overlooked. However, the statements that participants sorted came from a range of sources, including delegates who had attended OMERACT once and chosen not to return and therefore included a wide range of relevant opinions. In addition, any delegate who had ever attended an OMERACT meeting since it began was contacted to take part, ensuring the widest range of opinions possible.

A further limitation is that the opinions of participants who had never returned to OMERACT were not gathered. Although there were 555 individuals who attended only 1

OMERACT meeting, these represented only one-third (32.9%) of the total meetings attended, and thus two-thirds (67.1%) of the participants overall were people who attended more than once. The decision to include delegates who had attended at least 2 OMERACT meetings ensured they had a more rounded experience of the conference and also accounted for the great majority of attendances over the 11 meetings. It should be noted that patients attend the meetings by invitation and thus could only return if they were given the option. However, the majority of patients who were invited to return accepted the invitation; only 1 patient declined to attend a meeting, because of location. The reasons why some delegates did not return cannot be determined in our study, but because the study includes the bulk of attendees, it reflects the features that they value.

It is possible that sorting 66 statements into categories and then across the fixed distribution grid could be tedious for participants and potentially affect their responses. However, the median time to complete the study indicates that this study did not take as long to complete as estimated (24.87 min). Further, participants’ comments indicated that although the task could be challenging (“Sometimes difficult to rank,” P50, researcher, male), it also prompted them to think about their own priorities (“Makes you realise what you most think is really important when limited,” P77, fellow, female). Overall participants seemed to enjoy taking part in the study, saying it was a “fascinating challenge” (P3, researcher, female), an “interesting exercise” (P20, industry, male), and “a few [statements] even made me laugh” (P35, patient, male).

These data were collected online and thus the qualitative methodology may be slightly weaker owing to the lack of face-to-face interaction between researcher and participant. However, participants were asked to provide open-ended comments about their most strongly sorted statements. These comments cohered with the factor groupings and enabled researchers to explain why participants sorted the statements in certain ways, thus enriching the data.

OMERACT delegates value the focus on outcome measures and global standardization of methods that OMERACT provides. There are 4 ways in which delegates’ views of the “spirit of OMERACT” differ, with some placing more importance on evidence-based research, others on collaboration, some on equality, and a small group being satisfied with the principles of OMERACT but unconvinced by the process. These data provide important information about the aspects of OMERACT that are most valued by frequent delegates and support OMERACT’s descriptions of itself. This information may be useful for potential delegates who are considering attending OMERACT for the first time. Thus, it may be beneficial for the OMERACT committee to explicitly recognize these distinct values.
ACKNOWLEDGMENT

The authors thank the OMERACT delegates who participated in this study, the OMERACT Executive for supporting this study, iDENK, and Dr. Maarten de Wit, whose work informed the statements for this study; Tracey Penberthy at the University of the West of England for setting up and hosting the domain for the online study; and FlashQ (www.hackert.biz/flashq/home) for complimentary use of its software.

D. REFERENCES

8. de Wit MP, Abma TA, Koelewijn-van Loon MS, Collins S, Kirwan J. What has been the effect on trial outcome assessments of a decade of patient participation in OMERACT? J Rheumatol 2014;41:177-84.
CHAPTER 3: DEVELOPING CORE OUTCOME MEASUREMENT SETS

INTRODUCTION

As explained in the introductory chapter, lack of standardization of outcome measures limits the usefulness of clinical trial evidence to inform health care decisions. This can be addressed by agreeing on Core Outcome Measurement Sets that describe the minimum necessary to inform clinicians, patients, decision makers and other stakeholders. A Core Outcome Measurement Set ("core set") applies to Randomized Controlled Trials (RCTs) in a specific setting, and is a list of Core Domains with at least one named and applicable measurement instrument for each Domain. Applicability has a specific meaning in regards to OMERACT as explained below. Although investigators may identify and include other domains and instruments of primary or secondary interest to their RCT, the core set should be measured in all clinical trials in that health condition and setting. The core set may also be declared applicable to other study types such as observational studies.

Chapters 3 and 4 comprise a guide to developing first a Core Domain Set and subsequently a Core Outcome Measurement Set, following the requirements of OMERACT Filter 2.0. This is shown in Figures 3.1, 3.2 and 3.3. Filter 2.0 works from a newly developed conceptual framework of Core Areas that contain all potential Core Domains of measurement (Fig 3.1). A list of definitions of concepts used in this Chapter is given in Table 3.1. The steps are also summarized in an OMERACT MASTER Checklist that includes the procedures within OMERACT for convenience (see Appendix to Chapter 3). A more extensive list of concepts and procedures is given by the OMERACT glossary. The background of OMERACT Filter 2.0 and its development process is summarized in an overview paper [1], and more extensively in the proceedings of OMERACT 11.

Once developers have reached the end of the procedures described in these chapters, further work must be done to ensure that each Core Domain is addressed by at least one applicable instrument. This work will be described in Chapter 4 of the OMERACT Handbook which is currently in final stages of development and approval.

The development of an OMERACT core set can be described in 3 steps: 1. Background, preparation and process; 2. Defining the Core Domain Set: What to measure; 3. Defining the Core Outcome Measurement Set: How to measure each domain (including an initial screen, then full documentation of applicability).

There are occasions when it will be necessary to develop new domains: for example, when core set developers decide a new domain is necessary, or an existing candidate domain has not been properly defined or when a group decides to develop a new domain that may be relevant across several health conditions.
Concepts

Impact of Health Conditions

Pathophysiological Manifestations

Core Areas

Death

Life

Impact

Resource Use/
Economical Impact

Pathophysiological Manifestations

Domains

Examples of specific Domains within Areas

- disease
- intervention

- ICF domains: activity and participation
- quality of life
- patient perception of health
- loss of ability to work
- psychosocial impact
- 2ndy impact on family, caregivers
- utility

- societal
- individual
- health care
- direct/indirect (productivity)
- intangible costs

- ICF: body function and structure
- organ function (eg lung function)
- reversible manifestations
- irreversible manifestations
- biomarkers
- surrogate outcomes

Adverse Events
are measured within the core areas, but are labeled separately to allow assessment of benefit and harm.

Choices Influenced by Context

Fig 3.1: Concepts, Core Areas, & Domains for Outcome Measurement in Health Intervention Studies
All important stakeholders are included from the start: patients and their proxies, caregivers, researchers, etc.

update cycle

consensus

agreement on **what** to measure at least one Domain from each Core Area

Fig 3.2: OMERACT Filter 2.0 Developing a Core Domain Set

All important stakeholders are included from the start: patients and their proxies, caregivers, researchers, etc.

update cycle

consensus

agreement on **what** to measure at least one Domain from each Core Area

Fig 3.2: OMERACT Filter 2.0 Developing a Core Domain Set
Figure 3.3: OMERACT Filter 2.0 Developing a Core Outcome Measurement Set
<table>
<thead>
<tr>
<th><strong>Table 3.1: Definitions of key concepts in the development of a Core Outcome Measurement Set.</strong> [1]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HEALTH</strong>&lt;br&gt;A state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.</td>
</tr>
<tr>
<td><strong>HEALTH CONDITION</strong>&lt;br&gt;A situation of impaired health.</td>
</tr>
<tr>
<td><strong>HEALTH INTERVENTION</strong>&lt;br&gt;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.</td>
</tr>
<tr>
<td><strong>CORE AREA</strong>&lt;br&gt;An aspect of health or a health condition that needs to be measured to appropriately assess the effects of a health intervention.</td>
</tr>
<tr>
<td><strong>DOMAIN (AND SUB-DOMAIN)</strong>&lt;br&gt;Component of Core Area: a concept to be measured, a further specification of an aspect of health, categorized within a Core Area.</td>
</tr>
<tr>
<td><strong>OUTCOME</strong>&lt;br&gt;Any identified result in a Domain or Sub-domain arising from exposure to a causal factor or a health intervention.</td>
</tr>
<tr>
<td><strong>MEASUREMENT INSTRUMENT</strong>&lt;br&gt;A tool to measure a quality or quantity of a variable, in this context a Domain or Sub-domain or a 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.</td>
</tr>
<tr>
<td><strong>OUTCOME MEASUREMENT INSTRUMENT</strong>&lt;br&gt;A measurement instrument chosen to assess Outcome.</td>
</tr>
<tr>
<td><strong>CORE DOMAIN SET:</strong>&lt;br&gt;For studies of health interventions, the minimum set of Domains and Sub-domains necessary to adequately cover all Core Areas, i.e. fully measure all relevant concepts of a specific health condition within a specified setting. Describes what to measure.</td>
</tr>
<tr>
<td><strong>CORE OUTCOME MEASUREMENT SET</strong>&lt;br&gt;The minimum set of outcome measurement instruments that must be administered in each intervention study of a certain health condition within a specified setting to adequately cover a corresponding Core Domain Set. Describes how to measure.</td>
</tr>
<tr>
<td><strong>APPLICABILITY (OF A MEASUREMENT INSTRUMENT)</strong>&lt;br&gt;The extent to which a measurement instrument passes the requirements of OMERACT Filter 2.0, i.e. it is documented to be truthful, discriminative, and feasible, and can thus be recommended for inclusion in a Core Outcome Measurement Set.</td>
</tr>
<tr>
<td><strong>SETTING (SCOPE):</strong>&lt;br&gt;The set of factors that describes the studies and circumstances to which the core outcome measurement set will apply. This is determined by the study questions and includes the health condition(s), target population, interventions, etc.</td>
</tr>
<tr>
<td><strong>CONTEXTUAL FACTOR:</strong>&lt;br&gt;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.</td>
</tr>
</tbody>
</table>
# A.1. OMERACT MASTER CHECKLIST for Developing Core Outcome Measurement Sets

## Assembly of working group and work plan

<table>
<thead>
<tr>
<th>#</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>WHITE</th>
<th>RED</th>
<th>AMBER</th>
<th>GREEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.B.1</td>
<td>Forming an OMERACT Working Group</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3.B.2</td>
<td>Stakeholder groups and their contacts identified (see document titled Engaging stakeholders and promoting uptake of OMERACT core outcome sets)</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.B.3</td>
<td>Thorough review of domain and instruments previously used</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.B.4</td>
<td>Implementation of Delphi and or Focus Groups (see document titled Draft Guidelines for using Consensus Group Methods)</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Core Domain Set selection

<table>
<thead>
<tr>
<th>#</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>WHITE</th>
<th>RED</th>
<th>AMBER</th>
<th>GREEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.C.1.1</td>
<td>Definition of context: setting (scope)</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.C.2.1</td>
<td>Deciding on the inclusion of Resource Use</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3.C.3.1</td>
<td>Literature review of domains (and instruments), part 1: what has been measured?</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3.C.4.1</td>
<td>Identification or definition of other domains of interest</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3.C.5.1</td>
<td>Formulation of draft Core Domains - at least 1 per Core Area</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3.C.6.1</td>
<td>Formulation of core contextual factors</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>3.C.7.1</td>
<td>Formulation of core adverse events, if any</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working group vote</td>
<td>Working group agrees they have Draft Core Domain Set prepared</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3.C.8.1</td>
<td>OMERACT consensus on Core Domain Set and timeline for update cycle</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Traffic light scoring.

Throughout the instrument selection process, “traffic light” scoring will be offered.

- **Green** always means “good to go”
- **Amber** always means there is a concern, or caution, or weakness but it is good enough to go forward.
- **Red** always mean stop, do not continue.
B. ASSEMBLY OF WORKING GROUP AND WORK PLAN

1. Forming an OMERACT Working Group

At some stage, usually very near the beginning of the process, core set developers form themselves into an OMERACT Working Group and initiate a ‘Special Interest Group’ activity at OMERACT. (For more details on the OMERACT process overall and how Working Groups fit in see Chapter 7 - Methods For Reaching Consensus) The group formulates the setting (scope) of the core set, which can be modified based on initial discussions at the Special Interest Group activity. Central to this activity is the application of the ‘PICO’ statements used widely in studies to define the patients/population, intervention, comparator/control and outcome. This is discussed more fully in Section 3.C.1.1 below.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.B.1</td>
<td>Forming an OMERACT Working Group</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Stakeholder groups and their contacts identified

In OMERACT, the Working Group always includes patients (for further information see Chapter 8: Patient Partners and OMERACT), clinical researchers, health professionals, and methodologists; and may include policymakers (including regulators, payers), industry, general public, and others. (For further information see Chapter 7 section B Paragraph 2: Stakeholder involvement). The importance of non-mandatory stakeholders should be discussed explicitly, as well as the method of interaction if they are deemed relevant but not part of the Working Group. In terms of representation, we recommend more than 1 representative from each of the stakeholder groups, and a good geographical spread ensuring representation of at least 3 continents.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.B.2</td>
<td>Stakeholder groups and their contacts identified (see document titled Engaging stakeholders and promoting uptake of OMERACT core outcome sets)</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Thorough review of domains and instruments previously used

A systematic review of the literature is the first step in finding information about domains that have been used previously (and instruments that have been used to measure them) in research in the disease group. This review is often led by an OMERACT Fellow attached to a Working Group. (See Chapter 10: Supporting participants with education and development at OMERACT) Groups should not feel restricted to this roster for their final domain set, there are often new domains that need to be identified and operationalized.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3.B.3</td>
<td>Thorough review of domain and instruments previously used</td>
<td>n/a</td>
<td>Red</td>
<td>Stop do not continue</td>
<td>More work needed or a concern, but go</td>
</tr>
</tbody>
</table>

4. Implementation of Delphi and or Focus Groups

The domains that will form the core set derive from the conceptual framework of Core Areas (Figure 3.1). Inclusion of a domain in the Area of Resource Use is strongly recommended, and a decision not to do so requires a discussion as part of the consensus process. Developers initiate stakeholder consultation to determine what each stakeholder group deems essential or desirable to consider for measurement.

During development, developers match the input (from literature review and stakeholder consultation) to specific Domains and, where necessary, further specified Sub-domains in each Core Area relevant to the chosen setting. To ensure face and content validity explicit input from all stakeholders, especially patients, is essential to identify relevant Domains and to expose gaps in what has been previously measured. At the same time, the group is mindful of parsimony and tries to determine the minimum number of Domains necessary to ensure content validity.

Core set developers must consider if there are other factors that are essential to document to support the measurement of the target domain. These may be contextual factors or specific adverse events of interest. (See section C below.) Exploration of contextual factors and adverse events is new to OMERACT. We advise listing these as ‘important’ rather than ‘core’ until further experience has been gained in this process.

Note: groups specifically working on one Domain (e.g. Worker Productivity) will need to explore in which situations their domain should be considered respectively: core; important but not core; or a research agenda topic. Where deemed core, this recommendation needs to be voted on as it probably implies modification of an existing core set. Also, the content of the Domain will most likely need further exploration and specification, replicating the situation found when a core set is being developed.

The validity and acceptability of the agreement on Core Domains is mostly through the quality and documentation of the consensus process and the foundation for that consensus (literature reviews, qualitative studies to identify important domains).

A full discussion for methods of reaching consensus is presented in Chapter 7: Methods for reaching consensus. As part of support to the Working Groups, a member of the OMERACT Executive or Scientific Advisory Committee, supported by a member of the Executive Methods Team, works with the Working Group to develop the research program and help them complete the steps in the process, which are
summarised in a OMERACT MASTER Checklist (see Appendix to Chapter 3) for obtaining first a draft Core Domain Set and then agreement on a Core Domain Set.

The consensus process is a mix of simple surveys, Delphi surveys, small group discussions (outside and also at OMERACT conferences), and plenary discussion plus voting (see Chapter 7: Methods for reaching consensus). A protocol is developed to specify which techniques and which stakeholders are involved, and at which step, and which criteria will be used for decisions. This is reviewed by the Executive Methods Team and either approved or written constructive feedback given. Often a 70% vote in a Delphi survey where participants are forced to prioritize their choices (e.g. by restricting the number of votes for each participant) is considered an appropriate target as a way of selecting potential domains.

Consensus agreement at the OMERACT conference for the Core Domain Set is 70% or more support from those voting at a plenary session.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3.B.4</td>
<td>Implementation of Delphi and or Focus Groups (see document titled Draft Guidelines for using Consensus Group Methods Chapter 9)</td>
<td>n/a</td>
<td>Stop do not continue</td>
<td>More work needed or a concern, but go</td>
<td>Good to go</td>
</tr>
</tbody>
</table>

### 5. Agreeing on the Core Outcome Measurement Set

This section provides an overview of what we will be going through in more detail in Chapter 4.

For clarity, we present the OMERACT Core Outcome Measurement Set selection (selection of instruments for the domains identified above) as the second step of a 2-stage process (core outcome selection followed by core outcome measurement selection). However, work on identifying and agreeing domains often overlaps with that for identifying applicable instruments for those domains. ‘Applicable’ in OMERACT means an instrument passes the original OMERACT Filter of Truth, Discrimination and Feasibility, documented per the guidelines that follow below. [2]

The first step in finding information about instruments that have been used to measure domains of interest is a systematic review of the literature. In practice this is sometimes done at the same time as the search for domains, or starts with that search but is extended in areas that turn out to be of interest. This is often led by an OMERACT Fellow attached to a Working Group (See Chapter 10: Supporting participants with education and development at OMERACT).

The consensus process with OMERACT requires clear reference to published data, requirements for demonstration of the validity of the instrument in the setting in which it is to be used, and a general approach. This is to ensure that an instrument meets the requirements of ‘truth, discrimination and feasibility’ as laid out in the original OMERACT filter and fully encompassed within OMERACT Filter 2.0. The documentation is developed in two steps: an initial to select the most promising instruments for each domain; and then a full documentation for each of those instruments. A protocol is then developed to specify which techniques and which stakeholders are involved, and at which step, and which criteria will be used for decisions. This is reviewed by the Executive Methods Team that either approves or gives written
constructive feedback. If no adequate instrument can be found to measure a Domain, the Working Group may well set out to develop one.

As a support to the Working Groups, a member of the OMERACT Executive or Scientific Advisory Committee (supported by a member of the Executive Methods Team), works with the Working Group. This helps to develop the portfolio and help them complete the steps in the process, which are summarised in a OMERACT MASTER Checklist for obtaining first a draft Core Outcome Measurement Set and then agreement on a Core Outcome Measurement Set.

When a draft Core Outcome Measurement Set is agreed by the Working Group, or they are ready for a final decision in consultation with other OMERACT members, the draft goes to the OMERACT conference for discussion and vote. Often, where one Domain may not yet have an applicable instrument and work is being done to provide evidence of applicability or develop a new instrument, a provisional Core Outcome Measurement Set may be agreed (including a recommendation for the ‘best’ instrument) and subject to further review when a fully applicable instrument becomes available.

Consensus agreement at the OMERACT conference for the Core Outcome Measurement Set is 70% or more support from those voting during the final plenary session. Where an answer option ‘don’t know’ is included in the voting question, the numerator is those voting ‘yes’ or ‘no’; i.e. the ‘don’t know’ votes are not included in the calculation.

To recognize the importance of some domains and/or instruments that are not included in the core set, some developers have viewed them as occupying a series of concentric spheres: the Core Domain Set in the center, surrounded by domains of decreasing importance. One way of viewing this is given in Figure 3.4 and an example from the group working on osteoarthritis (elaborated before Filter 2.0 was made explicit) is given in Figure 3.5.

Figure 3.4 Domains placed in concentric spheres by decreasing importance, and/or the availability of applicable instruments, or for the research agenda

Figure 3.5 ‘Inner’, ‘middle, and ‘outer’ cores in osteoarthritis. [14]
C. CORE DOMAIN SET SELECTION

Within the well-known RCT development structure of Patient, Intervention, Control and Outcome (PICO), the focus of OMERACT is on outcome. Nevertheless, a core set cannot be developed without first defining its setting (scope).

1. Definition of context: setting (scope)

The core set development group will decide what the core set will cover: health condition(s) (disease or disease group); type of interventions being compared (for example same or different class of drugs; drugs vs. surgery); population to which the intervention can be applied; time period of assessment (for example, x-ray changes are required in the core set for rheumatoid arthritis only in studies of 1 year or longer); etc. This is directly equivalent to the ‘PICO’ statements used widely in studies to define the Patients/population, Intervention, Comparator/control and Outcome. The group also needs to decide whether the core set will only apply to randomized trials (or a subset of trials, e.g. effectiveness trials), or also to other types of studies such as longitudinal observational studies. A core set is only useful if the information it provides is adequate to enable treatment decision making. This means that the context in which the decision-making is expected to happen is also part of the setting (scope) of the core set (for example, information for the general public, the patient, the physician, guideline developers, payers, etc.). Note that the word context has a very broad meaning, in contrast to the concept ‘contextual factor’ that is used in a very specific way within OMERACT (see below).

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.C.1.1</td>
<td>Definition of context: setting (scope)</td>
<td></td>
<td>Stop do not continue</td>
<td>More work needed or a concern, but go</td>
<td>Good to go</td>
</tr>
</tbody>
</table>

n/a
2. Deciding on the inclusion of Resource Use

Although there was a large majority of OMERACT participants in favor of including Resource Use as a required Core Area for the definition of OMERACT Filter 2.0 (and hence to be included within the Core Domain Set), there were some reservations felt quite strongly by a minority of participants. Mostly, these were related to the fact that Resource Use instruments had not been widely applied in the context of clinical trials. These participants felt more experience would be helpful. Therefore, Resource Use was only ‘strongly recommended’ for inclusion. The group must decide if there is a special reason why Resource Use should not be included as a Core Area for their core set and if so, the reasons must be clearly documented.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>3.C.2.1</td>
<td>Deciding on the inclusion of Resource Use</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Literature review of domains (and instruments), part 1: what has been measured?

The conduct of a literature review requires careful consideration of the terminology of the search terms and the databases where the search is performed. How detailed this should be has not been determined, but the study should not be limited to RCTs. One good example is the work on low back pain. (See Table 3.2)

<table>
<thead>
<tr>
<th>Table 3.2: List of Life Impact Domains potentially relevant for clinical trials on low back pain research</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Based on a literature search</td>
</tr>
<tr>
<td>Health-Related Quality of Life</td>
</tr>
<tr>
<td>Patient Perception of Illness</td>
</tr>
<tr>
<td>Work/Occupational Function</td>
</tr>
<tr>
<td>Family Life</td>
</tr>
<tr>
<td>Daily</td>
</tr>
<tr>
<td>Psychological</td>
</tr>
<tr>
<td>Social</td>
</tr>
<tr>
<td>Leisure Time</td>
</tr>
<tr>
<td>Sexual Life</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
### 4. Identification or definition of other domains of interest

In each of the Core Areas, at least one Domain of measurement must be defined or selected. The word Domain is closely linked to, or even equates with, the words ‘concept’, ‘attribute’ and ‘construct’ used in the literature (for example ‘domain’ or ‘concept’ used in the Cochrane Handbook). After potential Domains have been collected from the literature, the group needs to decide whether the list is comprehensive, or whether other Domains should be considered. In addition, all potential Domains must be clearly defined (see also 3.C.4.5). It is crucial that a theoretical or conceptual framework should be the base of the identification and understanding of Domains. For example, the International Classification of Functioning, Disability and Health was adopted by the World Health Assembly in 2001 for international use for such a purpose [3]. Given such a framework guiding the conceptualization of domains requires consultation between stakeholders inside and outside of the group, and qualitative research involving focus groups or surveys of relevant stakeholders (such as patients). (See section 3.C.4.5.) In addition, we strongly suggest that developers refer to the website of the Patient Reported Outcomes Measurement Information System (PROMIS) organization: [www.nihpromis.org](http://www.nihpromis.org). PROMIS is a system of highly reliable and precise instruments for patient–reported health status for physical, mental, and social well–being. Many of the PROMIS domains are relevant for consideration, and where this is the case, PROMIS provides an instrument to measure that domain. These can be a short set of fixed items in a paper questionnaire or a Computer Adaptive Test (CAT) (flexible items dependent on preceding response(s)). Although applicability of such instruments in the setting of the core set must be documented through the same process as other potential instruments, these instruments have a head start because they comprise optimized items from legacy tests (such as the SF-36 and the HAQ) and newly developed items that have been validated in several populations.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>3.C.3.1</td>
<td>Literature review of domains (and instruments), part 1: what has been measured?</td>
<td></td>
<td></td>
<td>More work needed or a concern, but go</td>
<td>Good to go</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>3.C.4.1</td>
<td>Identification or definition of other domains of interest</td>
<td></td>
<td></td>
<td>More work needed or a concern, but go</td>
<td>Good to go</td>
</tr>
</tbody>
</table>
It should be noted that the FDA Guidance on the development of Patient-Reported Outcomes [4] expects that Patient reported Outcomes are based upon the patient experience, with saturation of themes in the development of such scales used to measure the selected domains (see also 3.C.4.6). OMERACT will expect developers to show how they have ensured that this consultation has taken place. Some Domains may be further specified into one or more sub-domains in relation to the context of the proposed core set, through discussions with the members of the development group and stakeholders. For the Core Area of Death, simple report of count of deaths is a mandatory reporting requirement in any trial, and further exempt from Filter requirements. However, any instrument assessing a more detailed specification of a Domain within the Area of Death (for example death from a specific cause) would need to pass Filter 2.0 and an instrument in that Domain would need to satisfy all the requirements.

An example of additional domains collected by the Low Back Pain Group were added to the literature review results so as to fully use the Filter 2.0 methodology is shown in Table 3.2 and in the Appendix.

In the end, consensus determines whether the candidate Core Domain Set adequately samples across all Core Areas. (See Chapter 7: Methods for Reaching Consensus)

OMERACT requires a clear definition of each of the selected Domains. For example, for Domains in Life Impact, what is the breadth and depth of the experience in patients? Often qualitative work is done to capture this detail. Alternatively, existing conceptual frameworks could exist and define a concept adequately for the group (i.e., self-efficacy from social cognitive theory and self-management literature). In OMERACT, the work on Fatigue is an example of how breakout sessions paid considerable attention to the meaning of fatigue in persons with arthritis. In a manner, like the creation of a measurement instrument, several avenues can be used to improve the understanding of the Domain. Streiner and Norman [5] discuss methods of item generation which can also be applied to Domain generation as described below. These methods may provide data or steps toward item generation for patient-reported outcome measurement instruments (PROMS), but a full evaluation stage evidence of appropriate methodological rigor will be needed. (See [6] and Chapter 4)

Formal qualitative research is an excellent way of obtaining the experiences of patients, family, and health care providers with the goal to explore the nature and the spectrum of the domain (e.g. fatigue or pain) encountered within the disease. However, rigorous qualitative methods must be used, with the collaboration of a qualitative methodological expert. This is to ensure scientific rigor in study design (theoretical underpinning; patient selection; conduct, recording and transcribing of interviews; data analysis and interpretation). Publication of results is encouraged and having methodological expertise on the team will ensure this is to COREQ standards (equivalent to CONSORT standards for RCTs) (http://www.equator-network.org/reporting-guidelines/coreq). Existing studies found in the literature can be a good source but should also conform to investigative rigor and reporting standards outlined in COREQ. Having an informal discussion with a few patients can be a useful precursor but is not qualitative research.

Conceptual frameworks can be described as available conceptual frameworks that propose a clear understanding of the target concept or authors who have made extensive use of a concept in their work and propose examples of the mechanism of action. For example, if the concept is self-efficacy, the literature supported by Bandura's definition [7] might be used to define self-efficacy as "an individual's belief in his or her capability to produce given attainments". It might also be operationalized in Lorig's work in arthritis [8] as a mediator of care utilization. Finally, the systematic review of available instruments to measure self-efficacy completed by Frei [9] not only provides a suggestion of potential instruments, but also synthesizes the conceptual definitions and their relationship to each other, therefore advancing our understanding of the boundaries (breadth and depth) of the domain that we wish to define.

In regards to clinical observation and research, many of our Domains come from sound clinical observation or published research and can suggest some of the clinical features that might belong to the Domain. Quantitative data or clinical experience can be useful to inform the interview guides to be used in the
qualitative studies. But more in-depth work is needed to ensure the full meaning has been determined and included in the definition.

In their Guidance on the development of Patient-Reported Outcomes [4], the USA Food and Drug Administration (FDA) has suggested a step-wise approach for the conceptual framework in PRO development that closely matches Filter 2.0 (Table 3.3). The Filter 2.0 framework comprises the Core Areas, and steps 1, 2, 4 and 5 are contained in the development of the Core Domain Set. Some OMERACT Working Groups concentrate on developing if a PRO fits within a single Domain rather than a full core set (e.g. Worker Productivity; see also section D. Developing new Domains). Steps 3, 6 and 7 are contained in the section on instruments, below (Chapter 4). The FDA guidance uses somewhat different terminology: potential Domains are called ‘concepts’, further specified as ‘attributes’ (indicators of the disease or condition itself, or of disease activity); the terminology refers to ‘proximal impact’ (functioning, signs and symptoms directly attributable to the disease process) and ‘distal disease impact’ (function – physical and social, and more social roles – wellbeing, work productivity, QOL, satisfaction).

Recently, the FDA has released another guidance document that is highly relevant to OMERACT: Qualification Process for Drug Development Tools. [10] This document is currently under study by the OMERACT Exec. Parallel documents from other government agencies (NHQ, AHRQ, EMA) and from professional associations are reviewed and deviance from our processes explored.

5. Formulation of draft Core Domains - at least 1 per Core Area
For each Core Area at least one Domain must be specified, but no more than necessary. Therefore, core sets will always contain a minimum of 3 Domains (corresponding to 3 Core Areas). We stress that a core set aims to capture the minimum number of Domains necessary to adequately capture what we want to know. Usefulness of a core set strongly depends on parsimony for a small set of core outcome Domains, as these will be required in all relevant studies of the benefits, harms and cost of the interventions of interest. There is some evidence that 7 ± 2 individual items are the maximum the human brain can simultaneously consider. [11] For example, the OMERACT set for RA that led to the development of the ACR 20/50/70 response criteria contains 7 items. Also, the Cochrane Summary of Findings tables allow up to 7 outcomes. Therefore, OMERACT suggests developers should strive for no more than 9 Domains in any core set.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>3.C.5.1</td>
<td>Formulation of draft Core Domains - at least 1 per Core Area</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.3: Steps in hypothesizing the conceptual framework for patient reported outcomes, according to the FDA Guidance.
1 Outline hypothesized concepts and potential claims
2 Determine intended population
3 Determine intended application/characteristics (type of scores, mode and frequency of administration)
4 Perform literature/expert review
5 Develop hypothesized conceptual framework
6 Place PROs within preliminary endpoint model
7 Document preliminary instrument model
The Core Area of Pathophysiological Manifestations requires some special consideration. For a condition without clear pathophysiology (e.g. nonspecific low back pain), developers can recommend that at least one Pathophysiologic Manifestation needs to be measured in all trials, but the specific Domain will be chosen for each (group of) trial(s). Alternatively, Pathophysiologic Manifestation domains can be defined for specific intervention types being studied. For conditions with strongly varying manifestations, such as vasculitis, developers can consider positing a generic Domain of ‘organ function’ or Domains of reversible resp. irreversible manifestations, allowing the conductor of the trial to choose the specific organ or manifestation(s) in the trial design stage.

6. Formulation of Core Contextual Factors

Developers must explore factors that are important to document in order to support the measurement of the target domain. For a contextual factor to be declared ‘core’, a group must have sufficient evidence that it notably modifies the effect of interventions (or its measurement) in most settings. For the initial core set, such evidence will most likely not be available. As the research into contextual factors for core sets is new to OMERACT, we advise naming contextual factors to be ‘important’ rather than ‘core’, and plan further research (research agenda). In addition, we caution that research on the identification or development of applicable instruments of the Core Domains must take priority over the exploration of contextual factors. Note that the use of the term ‘contextual factor’ is strictly defined (see Table 3.1) and differs from ‘context’ as used to define ‘setting’ (see 3.C.1).

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3.C.6.1</td>
<td>Formulation of core contextual factors</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As an example, the *nature* of the work performed impacts on Worker Productivity, and the level of activity impacts on ‘pain on activity’. Another example is the level of ‘self-management’ that modulates the impact of disease. Therefore, when one of these Domains is selected as core, the associated factor is identified as important, and might become a core contextual factor. As consideration of core contextual factors proceeds, it is also possible that no instrument exists for measurement of the contextual factor(s). This was the case of self-management in RA Flares, and this moves to a research agenda. As has been the found on several occasions, such a question may stimulate an entirely new work stream to address newly identified challenges. Again, this process is new to OMERACT so we will need to gain experience.

7. Formulation of Core Adverse Events, if any

Developers must decide if there are any specific adverse events that are important to document in the setting of the core set. As with contextual factors, an adverse event can be declared ‘core’, but only when convincing evidence can be produced that it importantly modifies the interpretation of interventions (or its measurement) in most settings. For the initial core set, such evidence will most likely not be available. We advise naming adverse events to be ‘important’, and plan further research (research agenda). In addition, we caution that research on the identification or development of applicable instruments of the Core Domains must take priority over the exploration of adverse events.
Only adverse events deemed important or core that are not included in Core Domains need to be separately named: adverse events that fall within a chosen Core Domain are already going to be measured. The process of identifying important or core adverse events is new to OMERACT so we will need to gain experience.

8. **OMERACT consensus on Core Domain Set and timeline for update cycle**

When developers are ready to demonstrate the validity of their draft Core Domain Set, having gathered evidence for all the above steps, they seek OMERACT consensus and approval on the Core Domain Set and the timeline to review and if necessary update it. OMERACT suggests a maximum of 10 years duration until the next update.
Chapter 4. Instrument selection for Core Outcome Measurement Sets
Instrument selection: Three pillars, four questions, one answer

Introduction and Background to Instrument Selection

OMERACT (Outcome Measures in Rheumatology) is an international, multi-stakeholder organization aiming to provide an evidence-based decision-making process for agreement on patient-centred and important outcomes for use across clinical trials and observational studies in rheumatology (Boers et al., 1998; Boers et al., 2014a; Boers et al., 2014b; Tugwell et al., 2007; Tugwell et al., 2014).

Since its inception in 1992, OMERACT has worked towards establishing “core domain” sets (i.e. “what to measure” in clinical trials) and “core outcome measurement” sets (i.e. “how to measure”) through evidence-and-consensus-based decision making across key stakeholder groups: Patients and caregivers, Providers, Payers, Product makers, Policy makers, Principal investigators (researchers) the Public, and others (e.g., the Press). OMERACT has always used three pillars of evidence to ensure an instrument is fit for the purpose of use in a core outcome measurement set: Truth, Discrimination and Feasibility (See Figure 1). Instruments that passed with evidence of these pillars were considered having the evidence to support their inclusion in a core outcome measurement set (Boers et al., 1998).

Since its inception in 1992, OMERACT has worked towards establishing “core domain” sets (i.e. “what to measure” in clinical trials) and “core outcome measurement” sets (i.e. “how to measure”) through evidence-and-consensus-based decision making across key stakeholder groups: Patients and caregivers, Providers, Payers, Product makers, Policy makers, Principal investigators (researchers) the Public, and others (e.g., the Press). OMERACT has always used three pillars of evidence to ensure an instrument is fit for the purpose of use in a core outcome measurement set: Truth, Discrimination and Feasibility (See Figure 1). Instruments that passed with evidence of these pillars were considered having the evidence to support their inclusion in a core outcome measurement set (Boers et al., 1998).

A core outcome measurement set does not limit the investigator in choosing the primary outcome from that set, or in fielding other outcomes in their study, but advocates that each study should measure at least the core outcome measurement set. Consistent use of core domain sets and core outcome measurement sets are increasingly recognized for their role in maximizing comparability of findings across trials and facilitating meta-analyses and comparative effectiveness research (Beaton et al., 2015; Boers et al., 2014a; Boers et al., 2014b; Williamson et al., 2012). They are also a way to reduce the risk of selective outcome reporting bias in clinical research. Kirkham has shown that the availability of the RA Core Set has increased consistency in outcome measurement in arthritis where in 2010 70% of trials fielded this core set in their outcomes (Kirkham et al., 2013).
Revision of the OMERACT Filter

In 2014 OMERACT began a deliberate process to refresh the guiding framework for domain and instrument selection (Kirwan et al., 2014; Tugwell et al., 2014). The revisions related to domain selection have been described by Boers et al (Boers et al., 2014b) and in chapter 3 of this handbook highlighting a new framework and the need to have at least one domain represented from each of three core areas (Death, Life Impact and Pathophysiologic Manifestations) and one strongly recommended one (Resource Use). Adverse events and important contextual factors are also decided upon and become part of each core domain set. Once core domain sets are endorsed by OMERACT (see chapter 3 in handbook), working groups move on to identify at least one instrument that passes the Filter requirements for inclusion in a ‘Core Outcome Measurement Set’. In response to the growing number of articles on measurement properties, and the growing number of instruments in the field, the OMERACT Filter 2.1 Instrument Selection process was also revised to help with the finding, appraising and synthesizing available or new evidence of measurement properties to ascertain the one answer that is sought – has the instrument passed the OMERACT Filter 2.1? In January 2017 the Executive of OMERACT endorsed the process described in this chapter for instrument selection and will monitor it with interest as groups begin to use it as a decision making process.

The OMERACT Filter 2.1 instrument selection process has two functions:
1. To define the type of information that is needed in order to ascertain if an instrument has passed the Filter,
2. To suggest a process and provide tools to facilitate moving through filter requirements and to facilitate record keeping and reporting.

Working groups need to document the process they used and work towards a final report and body of evidence. A workbook and specific assistive tools have been developed to support the use of Filter 2.1 for instrument selection and to help track progress and findings (see appendix A for “pull out” workbook). Groups are not obliged to use them so long as they get to the same endpoint with the same standards of evidence. The instrument selection process covers the areas that need to be addressed for outcomes that we think of as coming from a patient’s perspective (pain, quality of life), but it can also be used to guide the appraisal of biomarkers or imaging - two other important areas of activity for OMERACT. We will gradually extend our working examples to include more of these domains. In the meantime, please feel free to ask for help if you are having difficulty with the fit and see Chapter 6 dedicated to unique parts of evaluating imaging and biomarkers as instruments within OMERACT Core Outcome Measurement Sets.

Foundation: How do we know if an instrument has passed Filter 2.1?
The original OMERACT pillars of Truth, Discrimination and Feasibility remain the core or the pillars for instrument selection in OMERACT Filter 2.1 and, as shown in Figure 1, can be broken down into key questions to be answered (four of them) and the measurement performance needed to answer each one (measurement properties).

In order to streamline the process of instrument selection, the Filter 2.1 has suggested a practical re-ordering of the Filter elements (see Figure 2 ). The pillars of Truth, Discrimination and Feasibility are still there, but is now ordered with each step reflecting an increasing investment of time and effort. It also suggests that after the first two steps, a decision can be made to stop considering that instrument before entering the most time-consuming part of the process in the literature review for the last two elements.
This reordered process can be easily converted into a set of steps shown in Figure 3. This includes the four questions that need to be answered affirmatively to ensure the instrument has passed the Filter, and the tracking system of how each was scored. In this diagram you will also see an important decision point between questions 2 and 3 - this is when working groups may decide to let an instrument drop from consideration because it is not feasible or, on closer inspection, it is decided that it is not a good match for their target domain. Only those passing this decision point would go through the process of gathering evidence or making evidence to answer questions three and four.

For scoring at each stage we have used a stoplight set of colours to reflect the working groups appraisal of whether it met the Filter requirements for that attribute. GREEN is used to reflect a high level of confidence that this has been passed and can go forward. AMBER indicating the group sees a need for additional work or has some concern but still considers it good enough to pass that part of the Filter and RED (stop, does not pass). We also added the option of WHITE as an indicator of when no evidence is available (something that is correctable, but must be corrected before the instrument has passed the Filter). WHITE is only offered for questions 3 and 4 which are dependent on available evidence. Again, GREEN and AMBER get a pass, RED and WHITE do not.

Section 4 on the OMERACT Master Checklist is the section that pertains to tracking instrument selection. In this part of the chapter numbered sections refer to sections where the results need to be recorded on the Master Checklist. The questions in the checklist are the same questions as in Figures 2 and 3.
A. Moving through the Filter 2.1 for Instrument Selection.

1. Preparing for instrument selection and seeking input from the Technical Advisory Group (TAG).

The first step in the process for instrument selection is to work as a team to develop a protocol for how you will approach instrument selection. In an appendix to this document, you will see options for documenting the elements of your work in a fillable form (Appendix A, Part A and Part B). Many of these fields at the beginning of Part A are the same information you need to register your protocol with PROSPERO, a free online registry for systematic reviews. Groups may opt to register their protocol in order to document their intent in this field. It sometimes helps with subsequent publication.

When the protocol is complete, Working Groups will meet (in person or virtually) with a group that is “behind the scenes” in instrument selection and that is the Technical Advisory Group (or TAG Team). They are a group of methodologists and statisticians that are part of OMERACT and have volunteered to serve on this advisory group. As well as advising working groups on their progress through the Filter or on the design of studies to fill gaps, they will also help OMERACT to make sure we are using the best, most efficient methods for core set development. Work they are doing now includes testing streamlined versions of the search strategy, and developing a database to help working groups organize their findings not only to
simplify reporting, and publication, but to help create a repository of evidence on measurement properties for future use and updates. You will encounter them along the way, and they are available to help. At this point they should review what you are after, and how you plan to get there. This is your protocol check. When you have done this you can check off 4.A on the master checklist.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>4.A</td>
<td>Review of protocol by technical advisory group.</td>
<td>n/a</td>
<td>Red</td>
<td>Stop do not continue</td>
<td>Good to go</td>
</tr>
</tbody>
</table>

**B. Review of evidence of an instrument’s performance.**

Having sought feedback on the protocol, the groups now move on to gather the information that is need to answer the four questions in the OMERACT instrument selection algorithm in Figure 3.

1. **Question 1: Is it a good match with the target domain in this population?**
   
   The very first step is the assessment of part of the “Truth” pillar addresses whether the instrument appears to be a good match for the target domain and has the right content for the experience of that domain in the intended target population and study situation. Careful consideration is given to the domain of the instrument and the global aim of the instrument, but also to the breadth and depth of the item content. Surveys or checklists for working groups themselves are available, and groups are encouraged to get multiple inputs – particularly from respondents about the adequacy of the content. This covers what is sometimes called face and content validity. We also encourage groups to examine some data of their own or from some publications to look at the distribution of responses, patterns of missing items or floor and ceiling effects – all indicators of potential problems of the fit of the item content with the population of interest.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>4.B.1</td>
<td>Is it a match with the target domain? Conceptual alignment with Domain, face validity, content validity (OMERACT Filter: Truth)</td>
<td>n/a</td>
<td>Red</td>
<td>Stop do not continue</td>
<td>Good to go</td>
</tr>
</tbody>
</table>

**Traffic light scoring.**

Throughout the instrument selection process, “traffic light” scoring will be offered.

- **Green** always means “good to go”
- **Amber** always means there is a concern, or caution, or weakness but it is good enough to go forward.
- **Red** always means stop, do not continue.
2. Question 2: Is it practical to use?

The second step is the assessment of Feasibility, those very practical considerations about cost, burden, and access to the instrument in the necessary language(s) and mode of administration etc. that lead one to decide that it is practical or impractical to use a given instrument. Input is needed from both the users of the instrument to comment on administration, researcher burden and cost issues, and from respondents to comment on burden and suitability of format and administration.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>4.B.2</td>
<td>Is it practical to use?: The patient burden and cost of use (OMERACT Filter: Feasibility)</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Decision point: Does the working group all agree that this instrument has passed these first two questions (GREEN or AMBER = pass)?

We are now at a decision point in the OMERACT instrument selection process and this is unique to OMERACT. If an instrument is not a good match or is not feasible to use in the intended setting, it can be set aside by the working group. Ongoing attention should focus on only those instruments that have passed these two questions with a GREEN or AMBER rating. Many groups have found that a quick check of these first two steps eliminated several instruments that are covering the wrong content for the intended application, or are considered too long, expensive and/or complex to use. It is best to set them aside and continue only with those that have content/concept match and are feasible to use in the intended application.
3. and 4. (1-4): Question 3 and Question 4: Do the numeric scores make sense? & Can it discriminate between groups of interest?

Question 3 and Question 4 are grouped together here because the process is combined to address both of them.

Working groups having approved an instrument’s domain match and feasibility now move into the gathering of evidence of its ability to perform as an indicator of the target domain in the intended context to answer the last two questions in Figure 3. As described above this includes Question 3: Do the numeric scores make sense? or construct validity (Pillar of Truth, Checklist item 4.B.3) and the four measurement properties that together address Question 4: Can it discriminate between groups of interest?

1. Stability in situations of no change (test-retest reliability) (Pillar of Discrimination, 4.B.4.1),

2. Detecting instrument score change in situations of real change (longitudinal construct validity or responsiveness) (Discrimination, 4.B.4.2),

3. Discriminating between groups with the type of change anticipated in a clinical trial setting (Discrimination, 4.B.4.3)

4. Defining established thresholds of meaning (Discrimination, 4.B.4.4).

See the table below for definitions for these terms as ratified by the International Society of Quality of Life Research (ISOQOL)(Reeve et al., 2013). A more theoretical description of these measurement properties and examples of the types of evidence that could be considered for each of them can be found in Appendix D.

Most of this evidence will come from existing evidence in the published literature. The following paragraphs will describe how to find, appraise and synthesize the relevant literature.

<table>
<thead>
<tr>
<th>Filter 2.1 Pillar</th>
<th>Questions related to that pillar, corresponding measurement properties, and its definition (link to checklist item)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth (a)</td>
<td>Is it a match with the target domain? (4.B.1) &lt;br&gt;Content validity – the extent to which the instrument includes the most relevant and important aspects of the domain in the context of the intended measurement situation. &lt;br&gt;Face validity - degree to which the instrument as a whole appears to be a match with the target domain.</td>
</tr>
<tr>
<td>Feasibility</td>
<td>Is it feasible/practical to use? (Checklist items 4.B.2) &lt;br&gt;Burden (respondent and researcher), time, effort, translations, and cost of using this instrument in the intended setting (context of use).</td>
</tr>
<tr>
<td>Truth (b)</td>
<td>Do the numeric scores make sense? (4.B.3)</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Construct validity – the degree to which the scores on the instrument relate to other measures (patient-report or clinical indicators) in a manner that is consistent with theoretically derived, a priori hypotheses concerning the domains that are being measured.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discrimination</th>
<th>Can it discriminate between groups in the setting of interest (clinical trial setting)? (4.B.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test-retest reliability – A measure of the reproducibility of the instrument, that is the ability to provide consistent scores over time in a stable population. (4.B.4.1)</td>
</tr>
<tr>
<td></td>
<td>Responsiveness - The extent to which an instrument can detected changes in the domain of interest over time, when they have occurred. (4.B.4.2)</td>
</tr>
<tr>
<td></td>
<td>Discrimination in Clinical Trials – the degree to which the instruments are sensitive to the related change between the arms of a similar trial (ie, a comparative effectiveness study or a placebo controlled trial, or active comparison arm). (4.B.4.3)</td>
</tr>
<tr>
<td></td>
<td>Thresholds of meaning for proportional summaries (4.B.4.4) – The degree to which one can assign an easily understood meaning to the scores from an instrument. Rates of achievement are compared between arms in clinical trials. This includes thresholds like a minimum important improvement, or a patient acceptable symptom state.</td>
</tr>
</tbody>
</table>

**Searching for the evidence: setting up the literature review and identifying articles.**

Search strategies to find the evidence should be comprehensive, but focused. For this purpose, we suggest a combination of three factors be combined in the search:

1. *Population*. Working groups need to decide on the breadth of population they would like to consider (patient type, attributes, acuity, multiple diseases?).
2. *Instrument names, acronyms or of short forms* for the instrument should be set up to capture each time this instrument has been mentioned in the literature (either in title or abstract).
3. *Measurement properties*. We have provided search terms adapted from the COSMIN search strategies (Terwee et al., 2009) and offer them for use in several databases (see workbook). The search will include terms beyond those five measurement properties we described above, but at this point we suggest using this broader search in case additional measurement properties are hidden in that article. An article on factor analysis for example might also include some construct validation.

The search strategy for several of the currently used database is providing this for working groups in the appended workbook. This can be cut and pasted into the literature search. In this strategy groups will find that when each of *Population, Instrument and Measurement properties* are defined and search strings created and tested, they are connected by Boolean AND’s to produce a much more focused intersection of these large search strings. The source of the evidence will be in the intersection.

Testing the search terms is highly recommended. Working groups usually can find five articles that they know of on validity or reliability for this instrument. Work with the library scientist to test and see if these are captured. If not the library scientist can modify the search if needed to capture these key articles.
Quick screening of the articles can be conducted on titles and abstracts to make sure they are primary studies on measurement properties of the target instrument. Systematic or narrative reviews on the measurement properties of the instrument can be retained and the reference list checked to make sure all the relevant primary articles they included have been captured in your search. Screening questions are provided in the workbook.

At this phase articles that have passed screening are pulled for a fuller review to verify they are primary studies about measurement performance of your instrument in a relevant population, and second to determine the measurement properties that will be addressed in that article if it is relevant. Most groups will be working with about 15-20 studies in the end, but each paper will usually provide evidence on multiple measurement properties.

Introducing the Summary of Measurement Properties Table:

Now is the time to introduce the Summary of Measurement Properties table that will become the one page summary of all your work! We will refer to this frequently as we move through the rest of the chapter and the instrument selection process. At this point the articles that were found in the literature review are placed in the rows, and the X’s placed to identify which measurement properties were addressed in that article. A sum of the X’s in the columns will identify the total number of articles available that could be giving us evidence for that measurement property. If there is a zero or one, we will already know that more information will be required to meet our aim of at least two studies that used good methods.

(Later in the process, as the quality of the methods used in the study are checked (Good Methods Check), colour is added to each of the boxes to indicate if there reviewers found this piece of evidence to be gathered with good methods (GREEN OR AMBER) or not (RED). Empty boxes reflect WHITE, absence of information on that property from that study. X’s are replaced with + or – to indicate that the findings of the study demonstrated adequate or better performance of the instrument (+), equivocal performance (+/-) or poor performance (less than adequate) (-). The evidence across those studies is reviewed and synthesized, and a RAGW rating is given to each measurement property (remembering match and feasibility were done and passed before the literature review and are shown her for completeness). The working group then decides what kind of endorsement they would like to present for a vote).
<table>
<thead>
<tr>
<th>Instrument: ABC</th>
<th>Date completed: Jan 15, 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author/year</strong></td>
<td><strong>Truth</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Construct validity</strong></td>
</tr>
<tr>
<td>Tugwell 1991</td>
<td>+</td>
</tr>
<tr>
<td>Shea, 2004</td>
<td></td>
</tr>
<tr>
<td>Simon, 2004</td>
<td>+</td>
</tr>
<tr>
<td>Marsh, 2005</td>
<td></td>
</tr>
<tr>
<td>De Witt 2006</td>
<td></td>
</tr>
<tr>
<td>Bingham 2006</td>
<td></td>
</tr>
<tr>
<td>Singh 2007</td>
<td></td>
</tr>
<tr>
<td>Gossec 2009</td>
<td></td>
</tr>
<tr>
<td>Strand 2010</td>
<td>+</td>
</tr>
<tr>
<td>Simon 2010</td>
<td></td>
</tr>
<tr>
<td>Mada 2015</td>
<td></td>
</tr>
<tr>
<td><strong>Total available studies for each property</strong></td>
<td>2</td>
</tr>
<tr>
<td><strong>Total studies available for synthesis</strong></td>
<td>2</td>
</tr>
<tr>
<td><strong>Rating (RAGW) [put on Master Checklist]</strong></td>
<td>GREEN From working group</td>
</tr>
<tr>
<td><strong>Overall rating for instrument across properties</strong> [Options: Endorsed, Endorsed with caution, Not endorsed]</td>
<td>Working group is recommending provisional (amber) endorsement of this instrument. Specific research agenda is set to complete to move this up to full endorsement.</td>
</tr>
</tbody>
</table>

**Figure 4** Completed Summary of Measurement Properties table for an instrument (fictitious data as example only).

**Making sure the articles are of good enough quality.**

By counting the X’s on the Summary of Measurement Properties table for each measurement property, it is clear what the pool of potential evidence is for each, i.e. you can see the total available studies for each property. However, some studies may have flaws in their methods that make them at risk for misestimating the true value for the measurement property. Studies like these should be excluded from the review. This is the same as a risk of bias assessment in other types of systematic reviews. There are many tools available to critically appraise the methods used in measurement studies. The TAG team is reviewing them, but few have a focus on this risk of bias. We have chosen a modification of the COSMIN system (with the
collaboration of its developers) focusing on the COSMIN Version 2.0 (2017) checklist. In the 4 point methodological rating system, some COSMIN Version 2.0 items offer an “INADEQUATE” rating to those items which the COSMIN group felt would indicate a methodological flaw that would warrant exclusion from evidence synthesis due to a risk of bias. Only a subset of COSMIN Version 2.12 items offer this rating and OMERACT has focused on this subset.

We assembled those items offering an INADEQUATE rating into a checklist and reworded each to be an affirmative statement. An affirmation of these would suggest avoidance of risk of bias and therefore at least ADEQUATE quality of methods. Reviewers assess each study and give a rating of whether the article did this (YES) or did not report/do it in their study (NO). Based on the array of YES and NO responses and knowing that a NO would normally reflect an inadequate rating and a study that should be excluded, the reviewer makes a summary appraisal of whether, given the results of the Good Methods Check, this piece of evidence is trustworthy enough to be included. We call this the COSMIN-OMERACT Good Methods Check. The following table shows one example for test-retest reliability.

<table>
<thead>
<tr>
<th>COSMIN-OMERACT Good Methods Check for Test-retest reliability. In this system (as is the case in COSMIN v2.12) a “No” or “Red” rating would indicate a serious methodological flaw that would suggest this piece of evidence should not be considered. In the COSMIN-OMERACT Good Methods Check, the reviewer then makes an overall decision about inclusion or exclusion of this evidence.</th>
<th>Notes: (please keep notes about your ratings, and your final decision).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes, good methods</td>
<td>No, not done well</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Were patients stable in the interim period on the construct to be measured?</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Was the time interval appropriate?</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Were the statistical methods appropriate (choose one from below)?</strong></td>
<td></td>
</tr>
<tr>
<td>• A. For continuous scores: Was an intraclass correlation coefficient (ICC), Pearson correlation or Spearman correlation calculated?</td>
<td></td>
</tr>
<tr>
<td>• B. For dichotomous (yes/no) ordinal or nominal scores (named but not ordered categories: red hair/brown hair/blond hair): Was kappa calculated?</td>
<td></td>
</tr>
<tr>
<td><strong>Otherwise good methods? (Free of any other important flaws in design or methods).</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Considering the information available, would you recommend this study as evidence to be considered for this measurement property?</strong> (enter this in Summary of Measurement Properties)</td>
<td></td>
</tr>
<tr>
<td>Yes, good methods used – use this evidence</td>
<td></td>
</tr>
<tr>
<td>Some cautions, but this will be used as evidence</td>
<td></td>
</tr>
</tbody>
</table>
No, there are some problems do not use this evidence.

Notes on this piece of evidence:

It is recommended that two independent reviewers complete the Good Methods Check and then check for consensus. All ratings and the final their consensus vote should be kept for the records. The overall consensus will be entered into the Summary of Measurement Properties Table (above) using the colours GREEN [for good methods], AMBER [some caution but consensus this evidence should go forward] or RED [for problematic methods and an indication that this study will not be used in synthesis]. Look back at the Summary of Measurement Properties table and see that the cells are coloured in for the example studies. An Excel worksheet can be used for tracking these results at an individual and a consensus level, as well as the final decision about this piece of evidence (see workbook for example).

Evaluating the performance of the instrument – so was it good enough?

Studies that have passed the Good Methods Check are now reviewed to extract information on the study and the results of the measurement property tests. Descriptive information on the study, and how it was conducted should be gathered, as well as the results for that property. Remember to gather this information in a way that it is useful for readers and for your future reference when you update! OMERACT TAG will be working with groups to standardize a minimum requirement for reporting. Key features touched on in the Good Methods Check should be included as well as the statistical values and their interpretation. For some properties, for example studies of reliability, there are standards for reporting that could be followed (http://www.equator-network.org/wp-content/uploads/2012/12/GRRAS-checklist-for-reporting-of-studies-of-reliability-and-agreement.pdf).

Defining the standards to indicate when a study is demonstrating that an instrument has good enough reliability and validity is highly variable in the literature. The OMERACT TAG undertook a review of these standards and has developed from them a provisional set of standards for at least adequate performance for each property. Adequacy is only examined in the evidence that has been determined to have “good methods”. The provisional standards are included in the workbook in the appendix. If you are interested in the full review of the standards we found, please let the TAG team know. Standards will be ratified at a future OMERACT and their “provisional” nature removed.

Working groups use a + sign to indicate that that piece of evidence exceeds the provisional standard for that property, a – sign when it does not meet that standard, and a +/- for inconsistent findings – for example in testing construct validity several comparisons could be made. These +/- are added to the Summary of Finding table in the respective slot.

4. Synthesis at a measurement property level and filling any gaps (4.8.3 – 4.8.4)

All studies successfully avoiding risk of bias in their design have now had their findings extracted and compared to the standards. The working group must now consider the synthesis of their information. OMERACT is using the best evidence synthesis approach blending Quality, Quantity, Consistency of findings and Adequate or better Performance. This decision is guided by the work of others in best evidence synthesis groups (Forum, 2011; Schellingerhout et al., 2012; Schmitt et al., 2015; Slavin, 1995). Best
evidence synthesis looks for consistent evidence of good performance across multiple good quality studies of that property.

The literature gathered for each measurement property will be assigned a rating of GREEN (good evidence supporting this property, passes this element of the Filter), AMBER (some caution, but good enough) or RED (stop, evidence against this property or only poor quality evidence) score. If there is no adequate quality evidence available on that property it can be assigned a WHITE rating, and await the creation of that evidence and future update of the rating.

Figure 4 can be used as a guide for assigning the measurement property syntheses, and the summary table is the current suggested summary table for tracking the studies that have conducted measurement property evaluations (what was studied, Good Methods Check, and synthesis of findings). Groups should also compile a descriptive table of the actual results from these studies including a description of the study and study setting itself.

**Figure 4. Synthesis stage for each measurement property (OMERACT Master Checklist items 4.8.3 – 4.8.4.4).** For each property working groups gather all the evidence that they believe should be included in a synthesis. This table can then help them move to a rating of the evidence for each of the five measurement properties required in the filter (construct validity, test-retest reliability, longitudinal construct validity, sensitivity to differences found between treatment arms, thresholds of meaning). Rating is recorded on the checklist (boxes 4.8.3 – 4.8.5).

<table>
<thead>
<tr>
<th>Quality Of studies on measurement properties</th>
<th>Criteria for final rating</th>
<th>Final rating for this measurement property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality Of studies on measurement properties</td>
<td>Quantity of good quality studies</td>
<td>Consistency across studies</td>
</tr>
<tr>
<td>Good methods used</td>
<td>+</td>
<td>At least 2 pieces of evidence</td>
</tr>
<tr>
<td>Good methods</td>
<td>+</td>
<td>At least 2</td>
</tr>
<tr>
<td>Good methods</td>
<td>+</td>
<td>1 study only</td>
</tr>
<tr>
<td>Studies with fatal flaws</td>
<td>...</td>
<td>Not considered</td>
</tr>
<tr>
<td>No evidence</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>All other situations (Final rating not RED or GREEN or WHITE)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summaries for each of the measurement properties are recorded on the OMERACT Master Checklist and on the Summary of Measurement Properties table (Rating RAGW) for each measurement property. White is reserved for situations where no evidence is available.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>4.B.3</td>
<td>Do the numeric scores make sense? Assessing construct validity (OMERACT Filter Pillar: Truth)</td>
<td></td>
<td>n/a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>4.B.4.1</td>
<td>Can it discriminate between groups of interest?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>4.B.4.2</td>
<td>Can it discriminate between groups of interest? 2) Detecting change in situations of change (Sometimes called longitudinal construct validity or responsiveness)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>4.B.4.3</td>
<td>Can it discriminate between groups of interest? 3) Sensitivity to change in the context of an RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>4.B.4.4</td>
<td>Can it discriminate between groups of interest? 4) Thresholds of meaning for individuals defined? Minimum important difference, Patient acceptable state</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Decision point: Working group agrees that there are no gaps in the literature? (GREEN or AMBER = pass)?
C. Synthesis and filling any gaps

1. Synthesis of the available body of evidence for each instrument

Working groups now have a slate of evidence that becomes a profile of the evidence. A slate of AMBER or GREEN ratings would suggest passing the OMERACT Filter 2.1 provisionally or unconditionally (solid green). A group decision needs to be made about how that instrument rates overall. Is it ready for an OMERACT vote? And RED ratings would suggest a poorly performing instrument that will not be considered for endorsement into a core set.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>4.C.1</td>
<td>Synthesis of the available body of evidence for each instrument</td>
<td>n/a</td>
<td>Red</td>
<td>Amber</td>
<td>Green</td>
</tr>
</tbody>
</table>

2. Reporting of your work to the Technical Advisory Group: TAG

By this time, you have amassed a large amount of information that might be impossible to review and teach to the OMERACT constituency for their informed vote. We have inserted a new process to help with this.

The Technical Advisory Group (TAG Team) needs to now review your report, summaries, PRISMA table, Good Methods Checklists and all your ratings. They are available along the way for any type of support, but must be consulted at this point. Their job is to move through the technological parts of your review. They are doing the methodological review of your work and will reduce what you need to present at the OMERACT meeting. OMERACT attenders will know that when you are presenting, this other group has already vetted the details on their behalf.

Your group will still need to present the evidence, and to provide summaries for your final recommendation and also engage the attenders in your results. The TAG Team has just checked the process and offered some comments/suggestions for your conclusions.

Through a review of your work, and a discussion with the working group, the TAG looks at four things:

1. Evidence of your processes – your search terms, results, following through to your final report.
2. Your ratings and your justification of the ratings at each stage.
3. Review any ongoing gaps that need to be filled and your plan to fill that gap. They will be the peer review for any new work that you have had to do to fill gaps.
4. They will discuss your final voting recommendation with you and look at its consistency with OMERACT Filter 2.1 guidance.

Before this is finalized however, groups can consider fixing the evidence gaps/flaws in the table.
3. Design studies to fill gaps with review with TAG
What if there are gaps (WHITE) or only methodologically flawed evidence (RED) in the evidence? If no other candidate instruments with better properties are available, new high quality studies can be designed and performed (technical advisory group can help with design ideas) to fill gaps created by a lack of useable evidence (either absent, or only flawed evidence available). Groups are encouraged to use the “good methods” checklist in the design of the study and to review the design ideas with the TAG team.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>4.C.2</td>
<td>Report to the technical advisory group (TAG) about status of evidence and gaps</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Complete the studies, appraise and add to body of evidence (4.C.1)
These new studies will be assessed for their own quality and then, if merited, included as a piece of evidence along with previous published evidence into the assessment of that property. Given that many may not have been published, a peer review of the new work looking at the “Good methods” attributes will be done by members of the Technical Advisory Group or their designates. OMERACT welcomes this, and would encourage publication of these additional studies.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>4.C.3</td>
<td>Design studies to fill gaps with review with TAG</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>4.C.4</td>
<td>Complete the studies, appraise and add to body of evidence (4.C.1)</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Decision point: Working group agrees that instrument has evidence for passing the Filter 2.1? [GREEN or AMBER = pass]?

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Working Group Vote</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Working group agrees that instrument has evidence for passing the Filter 2.1?</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D. Reporting and seeking endorsement

1. Prepare Final Report

A final submitted report must be made that includes all the forms and reports submitted to the TAG in an iterative process, integrating any feedback until the TAG and the group feel it is the final report.

It is recommended that the final report consist of three things.

1. One page Summary of Measurement Properties table synthesizing literature reviewed, scoring, profile for the instrument of synthesized findings for each property, and the working groups’ decision.
2. 2-4 page summary of the status of the instrument being considered (executive summary).
3. Detailed report of findings – could be following the workbook [Appendix A] including PRISMA flow chart.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>4.D.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prepare final report</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Prepare for presentation at OMERACT for endorsement of instrument(s)

Once the working group has assembled their slate of evidence through a literature review or the creation of new evidence where needed for each of the required measurement properties, and has made an overall assessment that the instrument is good enough to have passed the Filter requirements, they can request a plenary spot at the OMERACT meeting seeking an overall endorsement. The plenary must try to effectively communicate the performance of the instrument across all three pillars of the OMERACT Filter 2.1: Truth, Discrimination and Feasibility to the attenders of the biannual OMERACT meeting through presentation and breakouts to engage people in the evidence and the Working Group’s assessment of it. This must be done in a forum where all OMERACT attenders are present and the members have the results.
of the evidence synthesis described to them (i.e., in a plenary or workshop). Summary of Measurement Properties tables and the OMERACT Master Checklist should be completed within the context of a full written report of the work.

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>White</th>
<th>Red</th>
<th>Amber</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>4.D.2</td>
<td>Prepare for presentation at OMERACT for endorsement of instrument(s)</td>
<td>n/a</td>
<td>Red</td>
<td>Amber</td>
<td>Green</td>
</tr>
</tbody>
</table>

3. Results of OMERACT Consensus Vote for endorsing instrument
The consensus of OMERACT is then sought for the group’s decision (4.D.3, and bottom boxes on Figure 3). The working group makes a recommendation: for example, that this instrument should be judged as a RED – that it should not be endorsed. The attenders then vote if they agree with this recommendation. We recommend the following: I agree, I am not sure but I will go with the will of the group, I disagree, I am strongly opposed to this recommendation.

A 70% agreement or willingness to acquiesce among voting attenders at the session in the absence of 15% or more in strong disagreement suggests consensus with the Working Group’s recommendation.

<table>
<thead>
<tr>
<th>Endorsement</th>
<th>Provisional Endorsement</th>
<th>Not endorsed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full slate of evidence supporting this instrument. Asking OMERACT to endorse it for core set inclusion. Unconditional endorsement based on currently available evidence.</td>
<td>Working group is recommending that we work with this instrument, get additional information to fill gaps, or improve our understanding, address concerns. Asking OMERACT to endorse it for provisional inclusion in core set to get additional information. (summarized in summary table) To be revisited at a subsequent OMERACT meeting.</td>
<td>No available evidence, large gaps in evidence or flawed instrument should not be considered. This instrument should not be considered for core set inclusion at this time.</td>
</tr>
</tbody>
</table>
4. Set up timeline for update of endorsed instrument(s)
Working groups set up a schedule for updating evidence for endorsed instruments, and a communication and dissemination plan for within and outside of the OMERACT community.

5. Communication & dissemination plan
The final step in the OMERACT Filter 2.1 Instrument Selection process is to create a communication and dissemination plan so that other stakeholders hear of the findings and core outcome measurement set. Planning for this as part of the process can help ensure time is allotted for making sure stakeholders have access to the results. Consult KT expertise to think through creative ways to deliver your message effectively through publications, workshops, webinars, websites or other dissemination activities.
Frequently Asked Questions

How does this relate to the development of new instruments?

The OMERACT Filter 2.1 Instrument Selection Algorithm is a process for finding and synthesizing findings of measurement properties to aid in instrument selection. It is not a description of instrument development. Many other considerations go into instrument selection including item / attribute choices, framing of questions, and decisions on scoring and scaling. OMERACT Filter 2.1 Instrument Selection Algorithm would be used after this was established. Many OMERACT working groups have chosen to develop a new instrument because no instrument was available, or the existing instruments could not meet OMERACT Filter 2.0 requirements and received RED or WHITE ratings. Instrument development is a long process. Along the way, evidence of the measurement properties of the final instrument (final set of items, final scoring system) is usually created and could well be part of the gathered evidence for instrument selection phase of the OMERACT Filter.

All new instruments, once developed, will need to go through the OMERACT Filter Instrument Selection Algorithm process and verify that they have enough evidence gathered together to satisfy the working group and the OMERACT attenders that they have met the described above to get to a synthesis statement of GREEN or AMBER. Evidence can come from published work or work that is yet to be published but can be described in enough detail to allow appraisal. The Technical Advisory Group would help the working group with the integration of unpublished work into OMERACT Filter 2.1 Instrument Selection evidence.

In sum, the same requirements need to be met for existing and new instruments developed for OMERACT. All need to be appraised for Truth, Discrimination, and Feasibility, but for new instruments, evidence does not need to be already published.

Conclusion.

The OMERACT Filter remains an important guidepost to the selection of instruments for core measurement sets in clinical trials. Truth, Discrimination and Feasibility continue to be the pillars for decision making about the ability of an instrument to be endorsed for inclusion in a core outcome measurement set for clinical trials. In OMERACT Filter 2.1 we have updated the process to help OMERACT Working Groups deal with a larger volume of information on measurement properties (some good and some not), advances in measurement sciences, and the need to synthesize findings across multiple studies. OMERACT Filter 2.1 builds on the experiences of many other international groups while retaining a clear link to the core OMERACT principles of evidence based decision making, collaboration amongst key stakeholders, and consensus (Flurey et al., 2015).
Appendices:

The OMERACT Handbook group created workbooks with very detailed search strategies and checklists to help working groups move through the steps outlined above. The workgroup facilitates gathering enough information to allow groups to register their review on public platforms for reviews such as PROSPERO (https://www.crd.york.ac.uk/PROSPERO/). Groups are encouraged to do so to facilitate transparency and the publication of the results in the future.

We hope that the accompanying workbooks and appendices help with tracking the steps and organizing information for your own use in publications, and in presentations back to OMERACT.

List of Appendices

A. Workbook for documenting process of gathering and synthesizing evidence.
   Critical appraisal tools for each measurement property

B. The COSMIN-OMERACT Good Methods Checklist adapted for OMERACT Filter 2.1 Instrument Selection needs.

C Adequacy of results review

D. Supporting background for OMERACT Filter elements
   Chapter 4 theory and definitions.

E. Acknowledgement of international efforts that led to the decisions made in OMERACT Filter 2.1 Instrument Selection Algorithm.
Chapter 5 – Developing a Methods Working Group

(DB); GW and JS

This chapter will describe the steps necessary to develop a ‘methods’ activity with OMERACT.

Chapter 6 Special Considerations for Imaging and Biomarkers

MADA, Phil C, & W. Maksymovych, Will Taylor

This chapter will describe the steps necessary to develop an Imaging and Biomarkers activity with OMERACT.
CHAPTER 7 - METHODS FOR REACHING CONSENSUS

A. INTRODUCTION AND THE OMERACT ‘PHILOSOPHY’ ON CONSENSUS

The absolute essence of OMERACT is that formulating a Core Outcome Measurement Set is agreed through consensus. Early activities often begin with brainstorming and the encouragement of contributions from everyone involved. An iterative process of ranking and then re-ranking the relative importance of different ideas and suggestions often generates some intellectual conflict as differing views are debated. While the final result may not be everyone’s preferred choice, the aim is to reach an agreement that all participants can accept as a working arrangement. An important aspect is to refer to evidence wherever it exists, or to generate evidence that might inform the decisions of the participants. Consensus at one step feeds into the next until there is international consensus on a core set of domains and outcome measures to use in all future rheumatologic clinical trials for a particular condition or disease.

1. Efforts to achieve consensus

Efforts to achieve a consensus on developing a core set evolved over time. Initially, a Special Interest Group [SIG] may convene (for 90 min or so) at one or more OMERACT meetings. Continuing support from OMERACT requires progress towards and completion of a successful ‘workshop’ (usually at an OMERACT meeting for a couple of hours. Here all the data supporting a proposed core set are laid out for discussion amongst the OMERACT participants; who consider whether they agree with the core set developers’ proposal. An indicative vote taken at the end of the Workshop may indicate strong support, while weak support may identify issues that can be resolved before the end of the conference. A vote of 70% or greater taken at a final plenary session of the OMERACT conference confirms an OMERACT consensus for the recommended Core Outcome Measurement Set.

2. Structures supporting consensus

There are several unique features of OMERACT that facilitate reaching consensus including the limited number of participants (usually less than 220), close physical integration at a single facility, and sharing meals and breaks together. The meeting programme deliberately includes unscheduled time to allow participants to discuss and debate issues that arise over coffee or in the corridor. Venues are generally in more isolated areas rather than city centers, so that participants stay more focused on the meeting, but there is usually a nearby attraction or natural beauty that makes for a good leisure experience where people can continue to discuss things freely on the road to consensus. Voting with key pads and electronic reporting has been a feature of OMERACT since this technology was first introduced and allows a very rapid and anonymous assessment of the current opinions of participants. The open leadership style and ‘rules of engagement’ in SIGs and Workshops encourage contributions from all participants. Because the OMERACT Filter demands a high level of proof or evidence to support statements or proposals, once this evidence is available disagreement usually becomes less. The interaction between multiple stakeholders with both regional and international representation makes for greater generalizability and validity, although some feel that at times this may slow some of the early fact-finding and consensus reaching steps. Many Working Groups set up pre-meeting surveys and distribute pre-meeting reading, which builds greater awareness and understanding of the issues and allows identification of potential discordance. Finally, as only a brief proportion of the meeting time is given over to didactic content (‘lectures’), there is a major concentration on small group meetings and interactive workshops.
B. STEPS TO TAKE IN REACHING CONSENSUS IN OMERACT

1. Generating ideas

In addition to a systematic literature review of published domains and outcome measures, it is recommended that an OMERACT Working Group conducts a range of fact-finding activities to generate an exhaustive list of potential domains. These activities could include patient and health care professional focus groups and nominal group technique workshops (see Appendix). These activities could be either face-to-face at OMERACT at other meetings, or via international networks combining local patient and health care professional input.

The advent of various on-line survey tools such as Survey Monkey (www.surveymonkey.com), has enabled the acquisition of input from a much wider representation globally. It also offers a cheap and reliable iterative process through which different stakeholders and individuals can contribute even when they cannot attend the meeting, and thus enabling work to proceed between meetings using a range of modified Delphi processes.

The value of the Delphi process is that it allows many different people to contribute as well as it overcomes the logistic and economic challenges of holding face-to-face meetings. The process usually begins when an expert panel (for example, the core set developers) brainstorms ideas and suggestions for the potential setting of their core set and the possible domains of interest. The over-arching principle of the Delphi process is to allow a series of rounds of information sharing in which opinions are collected from each participant without them knowing the opinions of the other participants. This puts everyone on an equal footing and prevents domination by forceful characters. The results are analysed then reported back to participants as the subject of the next round of the Delphi, which provides an opportunity for individuals to reconsider and possibly revise their judgements based on the collective view as it then stands.

2. Stakeholder involvement

A foundational principle of OMERACT is the bringing together of multiple international stakeholders in collaborative research. For the purposes of OMERACT, stakeholders may be defined as individuals or groups who are responsible for or potentially affected by an outcome measure and the healthcare decisions that would result from its implementation. The principle of stakeholder engagement implies a bi-directional relationship between researchers and stakeholders that will inform the development, prioritization, and/or use of the research project. Engagement requires the identification of the appropriate individuals, giving them a voice, and involving them in decision-making and the overall research process.

The first step is to gather the multi-disciplinary network into an OMERACT Working Group that has international representation (representatives from at least 3 continents), at least one patient partner, a Fellow and at least 5 participants with content-specific expertise to brainstorm and generate a wide list of possibilities for outcome domains as applicable to Filter 2.0 Core Areas and intended context and population of intended use. The group will then contact each other by email, talk by Skype, and/or meet at other international meetings as appropriate.

The inclusion of multiple perspectives from the inception phases of research, as well as, throughout the process is critical to maximize the impact of research and facilitate research dissemination. [1] Relevant stakeholders vary between projects, and may even change within a given project over time. Reasons for multiple stakeholder involvement include: increasing the number of ideas, perspectives and depth of questions considered; inclusion of all sectors affected; establishing credibility and ensuring relevance and
meaning to different groups; enhancing quality, increasing the face validity of final proposals; identifying concerns, barriers, and controversies that would not have otherwise been considered; increasing transparency; increasing uptake and dissemination of the outcome measure or research product, and fostering relationships for future research efforts.

For OMERACT projects, examples of potential stakeholders include:

a) Patients and health care consumers;
b) Researchers (and within this, different areas of methodological or content expertise such as outcome measure development, biostatistics, psychometrics, qualitative studies, comparative effectiveness, and clinical trial design);
c) Clinicians (e.g., physicians, surgeons, nurses, psychologists, physical and occupational therapists) and health care providers (meta-level such as institutions and health care systems);
d) Research funders (government, foundations, benefactors);
e) Government regulatory authorities
f) Health care policy groups;
g) Pharmaceutical companies or device manufacturers;
h) Clinical trial management companies (e.g., CROs);
i) Family members and caregivers;
j) Foundations and patient/health care advocacy groups; and
k) Payers (government systems, insurance/managed care benefits managers).

Different nomenclatures for classifying stakeholders have been developed by other groups including the “6 P’s” or “7 P’s”, classifying groups as: patients and the public, providers/practitioners, purchasers, payers, policy makers, product makers/private sector, and principal investigators/professors, and the press. [2] In selecting stakeholders, it is important to avoid tokenism (including a stakeholder just to “check a box”), and make efforts to ensure that stakeholders’ voices are truly integrated in discussions and decision-making.

As Working Groups are established, a brainstorming or thought exercise is recommended to identify potential stakeholders. An early statement from a nascent Working Group regarding the types of stakeholders considered, the rationale for their inclusion or exclusion, the proportional representation considered ideal, etc. would be helpful in guiding the assembly of the larger Working Group, identifying potential gaps, and moving toward meaningful consensus. Discussions with other OMERACT Working Groups, mentors, and other organizations involved in consensus-based exercises, and review of the evolving literature regarding stakeholder engagement may also be helpful in identifying relevant groups.

Asking a question such as, “Who has the necessary expertise and content knowledge for our project?” can help identify an initial list. Asking “In what settings and populations are these outcomes intended to be used?” may provide additional stakeholders groups to consider. For instance, for an outcome intended only for use in a clinical trial, the groups would include those who perform clinical trials, methodologists, pharmaceutical companies, clinicians, regulatory groups, and patients. If a Working Group hoped their work would be adopted in longitudinal observational studies, it would be important to consider a wider range of stakeholders: comparative effectiveness researchers; more broad-based representation of direct health care providers; and health care systems representatives.

Considerations for involving stakeholders:

- Which are the key constituencies that should be represented in the working group considering relevant expertise and intended use of the outcome measure?
- Is this a key stakeholder whose input is essential to the success of the project, or a secondary stakeholder whose participation is desirable but not obligatory?
- At which stage of the research project will a particular stakeholder or group be involved?
- Where will the stakeholder/group participate (e.g. teleconferences, face to face meetings at OMERACT, other interactions)?
- How will participation take place (e.g. surveys, part of Delphi exercise, qualitative work, review of documents, preparation of dissemination to key audiences, etc.)?
Early consultation with those who might have a very different view of outcome measurement in the particular area may be beneficial to better understand their concerns. This could also help identify potential barriers to implementation and feasibility to identify “blind spots” and unrecognized areas of controversy, and identify opportunities for expansion of the research agenda.

Consider and define the specific role anticipated for a stakeholder group or individual, when their contribution would be most useful, and how they will participate. Over the lifespan of a given research project, the particular individuals and groups as well as their roles may vary considerably.

The stages for potential stakeholder engagement in developing a core outcome set include:

a) Defining and refining the study question,
b) Identifying unmet needs and gaps,
c) Determining the appropriate population and settings of intended use,
d) Identifying other stakeholder groups and when to engage them

e) Developing the overall research plan,
f) Participating in the conduct of the research such as the collection of data
g) Serving as part of a Delphi Exercise or other consensus activities
h) Reviewing results
i) Serving as moderators and reporters at OMERACT meetings
j) Developing plans for, and participating in dissemination to respective constituencies
k) Implementing measures into specific settings to help with further refinement
l) Helping with fundraising for research activities

For example, in the development of a PRO measure, the inclusion of patients would be quite critical during a phase of domain identification and prioritization, debriefing questions, and understanding impact. During a phase of psychometric analysis of the performance of items there may not be as much patient involvement; however, in the phase of evaluating the results, patients would be critically needed to ensure that the results were in line with their experience in terms of establishing content and face validity. A situation could arise in which an aspect of disease was deemed to be highly relevant by patients, yet as existing measures were tested, there was not an adequate instrument to address this. Patient engagement would be critical in addressing this discrepancy and in developing a research agenda to further evaluate this particular domain or sub-domain. As different research projects develop, evolve, and progress over time, there may be a need to engage new or different stakeholders.

An important aspect of stakeholder engagement is providing at the beginning some estimation of the level of commitment of time required as well as an understanding of the specific role and expectations for participation. Communication with and between stakeholders is a key aspect of meaningful engagement and participation. It is important for a Working Group to consider how and when communication will occur over the course of the project with the various group members and stakeholders. There are a number of methods for engaging stakeholders for participation in a particular project which would include phone calls/teleconferences (ideally with international access numbers, e.g. FreeConferenceCall.com), emails, tele-meetings (e.g. GoToMeeting, WebEx), shared document resources (e.g. Dropbox, Sharepoint, Basecamp), websites with public and secured areas, chat rooms/discussion boards, in person meetings coupled with other meetings (ACR, EULAR, OARSI, etc.), and biennial OMERACT meetings. If attendance at a specific meeting is required, the group may need to consider how to obtain funding to cover travel for a stakeholder who would not normally attend, or to provide some opportunity for remote participation.

Adequately engaging stakeholders may require additional education concerning a project, its goals, the methodologies, and vocabulary and terminologies. This might include relevant background and reading, one-on-one pre-briefings and debriefings, and additional discussions to ensure that the stakeholders understand the work and their roles. The OMERACT Glossary, originally developed for patients, is a useful resource to those new to the process, and we anticipate that the Handbook will serve as an additional resource, as well...
as orientations for new OMERACTERs at meetings. For stakeholders that are new to the OMERACT process, there may be reluctance to speak up and participate. Group leaders should ensure that all participants are provided with opportunities and encouragement. If a particular stakeholder is not participating, it may be beneficial to discuss with them one-on-one to determine their reasons (e.g. not familiar with the process, not an expert with the content being discussed, not comfortable with voicing a contradictory position) so that any concerns may be addressed and reflected and assurance provided that their opinions are important.

3. Initiate and expand your list of domains

Get started on a literature review to identify what domains and outcome measures or instruments have been utilized for this condition in trials and other studies. Although you formally need to identify and agree a Core Domain Set, nevertheless it will be helpful even at this stage to work your way down the OMERACT Master Checklist. Along the way describe any readily available information on the validation and responsiveness of these measures. Often this work can form part of a higher degree or advanced career training and might be a suitable role for the Working Group Fellow.

Now expand your list to include as broad a range of ideas as possible. Potential domains may or may not have been used before or have valid measurement instruments available. The only criterion at this stage is that they are considered important to someone who has had is interested in the condition. One way this can be achieved is by sending the list to the wider OMERACT network or to a range of national groups such as professional rheumatology associations or patient organizations within the Working Groups members’ own countries. Another is to capture time in the OMERACT program for a SIG or organize other face-to-face meetings. Including contributions from patients in this process is essential for OMERACT. By following this OMERACT process consensus is more feasible - because every idea is in!

4. Refine your list of domains

The next step is to start refining the list. The initial list may have an unlimited number of domains, but in reality no clinical trial would be able to measure all aspects. A Delphi exercise asks participants to indicate how important a particular idea or domain is to them personally, and to allocate votes accordingly. Usually at least 2 rounds of Delphi must proceed, in which items with low voting scores are eliminated, those with high scores are retained, and those with intermediate scores are voted on again. OMERACT colleagues will take more notice of the results if the surveys include global representation – at least 3 continents – and include representatives from all relevant stakeholders. OMERACT has traditionally agreed that any domain achieving 70% of the responses as essential should be considered for further evaluation and inclusion in the core domain set.

Often several Delphi exercises are required using a combination of on-line and face-to-face sessions, whereby the list of ‘possible’ is reduced to a ‘reasonable’ number. Potential domains without clear consensus may form part of a research agenda to evaluate them further.

5. Moving on to the instruments

Once you have consensus on which domains to include, you need to find at least one applicable instrument for each. In practice, Working Groups and core set developers often find themselves looking for instruments at the same time as they are looking for domains. This is fine as long as the two processes are kept clear.
C. GUIDELINES FOR USING CONSENSUS GROUP METHODS

Developed by Maarten de Wit and Susan Humphrey-Murto

1. Please state the purpose of the overall study. State clear objective(s) described.
For example: For decision making or consultation? Generating items (domains) or ranking/rating? Assessing instruments?

2. What was the consensus method selected? If more than one method used, and rationale for selection?

3. Was there a clear outline/description of the overall process involved?
For example: Initial Delphi for item generation, followed by consensus conference. Was there a clear outline and description for each step? See below.

4. Was a literature review carried out? Was it described in appropriate detail?

5. How were the items generated and selected for inclusion in the initial Delphi questionnaire? Please justify the process used. Was it described in appropriate detail? How were results collated and selected for inclusion in subsequent rounds?

6. Please provide background information provided to participants?

7. How were items presented? Were items presented differently to different participants (e.g. Patients vs. medical doctors)? What was the order in which items were presented (eg random)?

8. Was feedback provided after each round? What type of feedback was provided?
For example: Qualitative data? Quantitative data with individual and group responses?)

9. Were domains assessed for different stakeholder groups? Was a rationale provided?

10. How many stakeholder/participant groups were involved for each step? Was a rationale provided for inclusion or exclusion? How was the stakeholder group defined and selected for each phase of the consensus process? How was the sample size of participant groups determined? Were the participant groups sufficiently representative to address the purpose of the study? Did the participants vary from round to round?

11. Can you provide an example of the response options? Whether ranking was used or was it a proportion of votes?

12. Was anonymity maintained?

13. Was iteration used? Was the number of rounds decided a priori? If not determined a priori, what were the criteria for terminating the process?

14. Was consensus defined a priori? How was it defined? How was the decision made to include terms/items on subsequent rounds? Was consensus forced?

15. Was the choice of facilitator described? Was the facilitator skilled in allowing different viewpoints to be expressed equally.
16. What is the response rate for each round of the consensus group method and for each stakeholder group (number of participants invited and number that responded). Did the authors list the items for each round (usually in appendix). Were all items listed in each round, or were items on which there was agreement dropped for subsequent rounds? Were participants allowed to add new items, and if so in which round?

17. If appropriate, were practical issues considered? For example: Did the consensus group consider issues surrounding implementation, acceptability? (relevant to domains?)

18. Did the discussion on the paper address potential methodological issues? Were the results credible?

RESULTS (see template below if group has a similar table representing endorsement & stakeholder group please use)

<table>
<thead>
<tr>
<th>Item</th>
<th>Round 1</th>
<th>Round 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient/Caregiver</td>
<td>Health Professional/Researcher</td>
<td>Patient/Caregiver Health Professional/Researcher</td>
</tr>
<tr>
<td>Proportion endorsed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D. GAINING CONSENSUS AT OMERACT MEETINGS

1. Concepts behind Breakout Group Sessions

Breakout group sessions are small group discussions (in practice, usually 12-24 participants) and form the heart of an OMERACT conference. These sessions allow in-depth discussions of the issues at hand, and are very productive. The purpose of the breakout sessions is to work towards achieving consensus and to identify lack of consensus where it exists. The objective is not to force consensus. If consensus emerges that is good, but it may in the end be even more helpful to identify the underlying reasons it does not do so. Any forced ‘consensus’ is unhelpful as disagreement will re-emerge later and in the end no overall supportive vote will be achieved. Described here is the general process, but the specific tasks for each activity/session are communicated by the breakout group facilitators (see below).

The technique used in the breakout sessions of OMERACT is a loose variant of the nominal group technique (see Appendix), derived from socio-psychological studies of decision conferences in several disciplines. Generation of ideas takes place individually – any issues mentioned and recorded in front of the group (usually on a flip chart for everyone to see and to act as a collective memory). Then a structured discussion of every item takes place.

2. Breakout group participants

Each group has a moderator whose role is to facilitate the group process, to allow everyone’s opinion to be heard and to help the reporter to summarize the group session. There is also a designated reporter whose role is to give a succinct summary of the group process and to present this at the plenary session. Where required, the reporter records priorities and votes in the group. Both moderator and reporter should be experienced in the OMERACT process, and selected for their role by the Working Group chair and the
OMERACT Executive Committee. The moderator and reporter prepare for their breakout session before the conference, attend a brief preparatory meeting at the conference and a brief meeting at the conference to help prepare the report which will later appear in print.

3. Preparing and running small group sessions – task and timing

Instructions for specific breakout group tasks are prepared by the organizers of the session with assistance from the OMERACT Executive. This may include some pre-conference reading material for all participants. Depending on the tasks, the breakout group may be based on brainstorming and be relatively unstructured; may be semi-structured towards a research agenda; or may be structured towards consensus. Tasks should be stated very explicitly, and moderators supplied with a timeline for the activity.

At the start of the breakout session, the moderator and reporter introduce themselves, followed by a brief round of introductions of each member of the group (name, profession, institution). The moderator will explain the purpose of their breakout group and the procedure that will be followed. He/she emphasizes the importance of every member fully participating in the discussion. The moderator then gives a brief summary of the problem and group tasks. The reporter writes the following summary information on the flip chart for all to see: objective of the session; time the breakout session begins and ends; specific steps to be followed (with time allocated to each step).

Most group tasks will produce a list of issues. Structured discussion about each issue will take place. The group leaders will ensure that each participant has contributed their views (if necessary by asking them directly). Most group discussions end with the group making a list of top priority issues, then voting on or giving points to the issues to summarize the overall view of the group. A brief discussion of the results and the overall process and views of the group (especially including any points of disagreement) helps the reporter to capture the essence of the group’s activities. A checklist for a typical breakout group process is provided in the Appendix.

4. Preparing the report back to the plenary session

Usually just before the report-back plenary session after a group breakout session there is a short coffee break. In this break the moderator and reporter finalize a one-page summary on a power-point slide to present to the plenary session. The summary from each breakout group is presented and briefly discussed at the plenary. When presenting, reporters should strive to identify redundancy, and not dwell on points already made by other groups/reporters. The chairs of the activity then engage all participants in a discussion that focuses on the key points identified in the groups. Interactive voting can sometimes be part of this process and if so formulating the questions should part of the group task. Only a very limited number of questions can be accommodated ‘on the fly’ during the report-back plenary session.

5. Preparing for the report back to the final conference ‘wrap-up’ session
Traditionally each OMERACT conference ends with a final plenary or ‘wrap-up’ session at which most formal votes of endorsement are held. Each activity during the meeting is allocated a time slot in the final conference session. This may be a very brief report (e.g. for a SIG or for a Workshop proposal that has already received strong endorsement at the plenary session of the Workshop) or a substantial opportunity (perhaps 30 min) to consider and then vote on proposals for a core set. The exact timing will be communicated at the conference. All activities can elect to formulate questions for interactive voting. The organizing group for each activity meets during the afternoon and evening breaks to formulate the conclusions for the final conference session, prepares any structured discussion that may be part of this report back (time is very limited at the final session) and formulates the questions for voting.

6. Preparing questions for voting

Anonymous voting using a key pad is central to the way OMERACT ensures everyone is a participant and that everyone’s views are noted. Formal votes are required to gain OMERACT endorsement for a core set proposal. Votes in support of further developments or research agendas can help guide the progress of Working Groups and can also be helpful in writing grants and seeking institutional support for activities undertaken by members of the Working Groups. One of the most common problems with voting is that the question is not well structured, is ambiguous, or does not cover all the options. Sometimes errors in the formulation of a voting question may have resulted in failure to gain the level of consensus required for OMERACT endorsement. For everyone’s benefit, it is worth considering carefully the structure and content of any voting question.

Mentors from the OMERACT Executive Committee or OMERACT Scientific Advisory Committee will be allocated to support core set developers prepare their final plenary questions. Often the questions will have been tested out as part of the SIG or workshop and developers will have become aware of deficiencies and corrected them before the final session.

7. Voting on core sets

When proposing the adoption of a Core Domain Set or a core outcome measurement set, participants should have been given all the information they need during the workshop or module sessions involved. A common occurrence is that during the workshop or module it becomes clear that some participants are not convinced by the arguments in favor of one particular domain or one particular measurement instrument. In these circumstances the developers might seek endorsement for all the other domains or instruments first: Do you agree that domain X should be included in the core domain set for this condition in this setting? (Then repeat the question for Y, Z, etc.) Then present briefly any further arguments or clarification about the contentious domain, W. Then ask: Do you agree that domain W should be included in the core domain set for this condition in this setting? If the vote is less than 70% have a supplementary question: Do you agree that domain W should be included in the research agenda for this condition in this setting?

8. Other types of vote

Advice about and endorsement of the way forward in specific areas can be sought. For example, at OMERACT 10 the Working Group concentrating on the best way of expressing quality adjusted life years (QALY) in rheumatologic clinical trials had recognized a substantially different approach among colleagues who measure QALY from a societal perspective and those who measure it from a patient perspective. After some time introducing and explaining the differences in approach and the implications of these differences...
at a mini-workshop, the Working Group leaders asked the plenary session for a ‘yes or no’ vote on the following questions:

a) Are preference from societal perspective (societal QALY) of interest for OMERACT research and inclusion in trials/observational studies?

b) Are preference from patient perspective (patient QALY) of additional value in outcome research in trials and observational studies and inclusion in trials/observational studies?

Only 38% of participants felt that continuing work on preference from a societal perspective (societal QALY) would be of interest to OMERACT researchers, but 81% felt preference from a patient perspective (patient QALY) would be of direct interest and relevance for OMERACT outcome research. As a result of the discussion and consensus at OMERACT 10, the Working Group refocused on the use of patient preferences in the elaboration of more robust outcome measures for patient reported outcome. - Are preference from patient perspective (patient QALY) of additional value in outcome research in trials and observational studies and inclusion in trials/observational studies. [3]

E. REFERENCES


F. APPENDIX

1. Check List For Breakout Group Consensus Process

1. Introductions
At the start of the breakout session, the moderator and reporter will introduce themselves, followed by a brief round of introductions of each member of the group (name, profession, institution).

2. Explanation of group task
The moderator will explain the purpose of their breakout group and the procedure that will be followed. He/she emphasizes the importance that every member fully participates in the discussion. Clear instructions for the task should be provided by the organizers before the conference. In this way, questions can be minimized when participants think about the group task. The reporter writes the following summary information on the flip chart for all to see.
   a. Objective of the breakout session.
   b. Time the breakout session begins and ends.
   c. Specific steps to be followed (with time allocated to each step).

3. Discussion starts
Most group tasks will produce a list of issues. Structured discussion about each issue will take place. The moderator will call on each participant. The reporter records the key points.

4. After discussion
Votes can be taken according to specific guidelines provided, and the results discussed briefly.

5. After the group session
Moderators and reporters of each breakout group will need to quickly prepare their report for the report-back plenary session, and the summary session at the end of the conference.

6. Process tips for moderator & reporter
   a. For those aspects of the process involving reporting of individual work, record each member’s ratings where all can see. Avoid comments on each individual’s rating. Address the ratings from a non-judgmental manner, respecting each individual rating. Again, your task as a moderator is not to force consensus, but to take care that the group understands the sources of disagreement. Consensus, if it arrives, will only have meaning once everyone understands the rationale behind other’s ratings.

   b. Be sure to stay on task. Reserve extraneous discussion to a time after the session is over. Do not get drawn into a discussion of the philosophy of the task and process; the organizers will be pleased to discuss the process with interested individuals after the session.

   c. Pay attention to patterns of disagreement. Patterns may emerge from a number of factors, including disciplinary group represented (e.g. statistician, industry), differences in country of origin, and differences in personality. Information on possible response patterns should be noted and shared with the other group leaders.

Equipment Supplied for the Breakout Sessions
LCD Projector/Screen – if requested. Flipcharts, flipchart easels, and markers.
2. A note on Nominal Group Technique

What is NGT?
Nominal Group Technique (NGT) is a weighted ranking method that enables a group to generate and prioritize a large number of issues within a structure that gives everyone an equal voice. The tool is called nominal because there is limited interaction between members of the group during the NGT process.

When should a team use NGT?
When a team needs to create a list of options and rank them, using NGT effectively neutralizes the domination of the loudest person, or the person with the most authority, over the decision-making process. This tool can also help a team achieve consensus about the relative importance of issues. The final result may not be everyone’s first priority, but they can live with it.

What are the benefits of Using NGT?
NGT is a good tool to use when dealing with controversial or emotional issues, or when a group is stuck. It is particularly useful when you need to reduce the number of issues for easier handling, get input from all group members and rank items in priority order.

What are the steps in NGT?
Part I - Define the Issue and Generate Ideas
• Define the issue
• Generate ideas
• Collect ideas
• Clarify ideas
• Combine ideas

Part II - Make the Selection
• Assign letters to ideas
• Rank ideas independently
• Collate the rankings
• Add the rankings
• Rewrite the list in priority order
• Perform a sanity check
CHAPTER 8 – PATIENT PARTNERS AND OMERACT

PART 1 – WORKING WITH PATIENT RESEARCH PARTNERS

A. INTRODUCTION

Capturing the patient perspective is an important part of research since the objective is to ultimately improve clinical outcomes for patients [1]. In order to effectively capture the patient perspective during the research phase, active collaboration between researchers and patients is essential. OMERACT has been involving patients in research as patient research partners (PRPs) since 2002, [2] to enable efficient inclusion of the patient perspective in the development of clinical outcome measures.

B. PATIENT RESEARCH PARTNERS (PRPs)

Patient Research Partners (PRPs) are defined as “persons with a relevant disease who operate as active research team members on an equal basis with professional researchers, adding to benefit of their experiential knowledge to a research project” [3]. In health research, several terms describe the specific role of patients in the context of research, including patient stakeholders, health care consumers, or patient partners. OMERACT has chosen to use the term PRPs to distinguish the active role of patients as collaborative partners from the role of patients as participants in a trial, a focus group, an interview or survey.

Although PRPs may themselves participate as subjects in research projects, their specific role as a PRP is one that reflects inclusion within the OMERACT work on an “equal basis”. This refers to equality in opportunities for full participation in the research process, to review all Working Group materials, and to vote in decision-making on the research process.

C. OVERARCHING PRINCIPLES OF PATIENT INVOLVEMENT IN OMERACT

OMERACT has adopted 3 overarching principles and 8 recommendations regarding patient involvement in research throughout the OMERACT process. These were developed through consensus and accepted at OMERACT 2014 [4]. A brief history of patient research partner involvement in OMERACT is given in the second part of this chapter.

1. OMERACT values the experiential knowledge of Patient Research Partners (PRPs)

The experiential knowledge of patients complements the evidence-based knowledge and clinical expertise of researchers and other stakeholders. Incorporating the patient perspective is an imperative for developing disease specific core-sets and patient reported outcomes. In such cases, patient participation is an unconditional requirement, independent of personal opinions or preferences.
2. Engaging patient research partners (PRPs) as integral stakeholders throughout the research process is a fundamental OMERACT principle.

Patients are an essential stakeholder in outcome research. Their involvement over the last decade has proven to add important values to the OMERACT research agenda and the conduct of outcome research. Patient participation has the advantage of aligning the focus of the scientific research with patient needs and priorities. The involvement of patients has also led to more empowered patients. They have taken co-ownership over research themes that are close to their own disease or to their daily life.[5]

By recognising the important role of patients, those in leadership roles within OMERACT should enable patients to contribute to the research process by providing appropriate psychosocial and practical support.

The OMERACT Executive recognizes that the level of involvement may vary depending on the scope and type of a particular research project. For example a statistical project might necessitate less patient involvement, as explained in the recommendations below.

OMERACT participants subscribe to the OMERACT values. These values include trust, respect, transparency, partnership, communication, diversity, confidentiality and co-learning with respect to patient involvement. The general values apply not only to those in leadership positions within OMERACT, but also to all OMERACT participants. We recognize that patients share their personal experiences and require personal support in order to undertake their role.

D. RECOMMENDATIONS FOR PUTTING THE OVERARCHING PRINCIPLES INTO PRACTICE

1. Working Group leadership and appropriate representation

While the Working Group leader should take primary responsibility, the entire research team has an active role in supporting patient involvement. The Working Group leader should take responsibility for the appropriate representation of the patient perspective in the research project The Working Group leader may delegate a specific support role to one of the other working group members, who would organize, support and facilitate the involvement of PRPs in the research project.

PRP involvement throughout the entire research project is expected. However, patient roles and tasks within an individual project or Working Group may vary according to the stage or content of the research project. In general, full PRP involvement will be expected in:

   a. All groups working on domain selection, identification or prioritization, including patient reported outcomes (PROs); alternative or composite uses of PROs; and construction of patient-reported scales;
   b. All attempts to define ‘Core Sets’;
   c. Groups considering the conceptual frameworks underpinning outcome measures;
   d. Classification of outcomes to aid clinical decision-making;
   e. Groups working on instruments or responsiveness.

Less direct PRP involvement may be appropriate in some circumstances:

   a. Projects which focus on instruments and responsiveness for outcomes which are not patient-reported (such as imaging techniques, blood tests and biomarkers). For example, defining the smallest detectable difference for detecting an erosion for an MRI or ultrasound may not
require high levels of patient input. However, determining a clinically relevant difference for the measurement as related to its consequential validity would require full patient involvement.

b. Projects which focus on methodology. For example, a discussion of the use of item-response theory or Rasch analysis to develop an interval scale for assessment may not require high levels of patient input (depending on the expertise of the PRPs). However, if a core patient-identified domain was excluded on the basis of its failure to meet measurement concerns, then this would require patient input to determine the face validity of the instrument and consideration of whether this area should remain a critical feature of a research agenda.

For each project submitted to OMERACT full PRP involvement is expected unless the OMERACT Executive has agreed otherwise. In order to receive an exception, Working Group leaders should discuss the issue with their Executive mentor and the Executive Patient Stream leaders in the first instance. Re-evaluation of the specific level of patient involvement should be considered over the life of a research project as needs and requirements may change.

All Working Groups will have PRPs attending their sessions during OMERACT conferences so that all groups will receive at least some PRP input.

Each group should involve at least 2 patient research partners. Two PRPs is the minimum number in each working group as recommended by other organisations [3]. An exception can be made in the case of some projects e.g. a statistical project where one PRP may be involved. Exception where only one PRP is recruited requires approval by the OMERACT Executive, and relates to the examples given in recommendation 1.

Not all PRPs involved in Working Groups will be able to attend OMERACT conference, where patient participation is limited to 10% of participants. The OMERACT Executive, working through the Patient Stream leaders, formally invites PRPs to attend the conference. Invitations take into account the range of activities taking place at the conference and the overall financial situation.

Working Groups running a Workshop will usually be expected to nominate 2 PRPs and those running a SIG will usually be expected to nominate 1 PRP. This limit on the number of PRPs a work group can nominate to attend an OMERACT conference does not preclude a much wider PRP involvement in the research process outside of the conference.

PRPs attending an OMERACT conference, in the same way as other participants, will also participate in and contribute to sessions not related to the particular research area of their nominating Working Group.

The Working Group should support the cost of PRP participation, for example by seeking to provide traveling expenses to attend meetings or reimbursing other incidental expenses. In addition to working group funds, some financial support (e.g. part of the travel and OMERACT meeting registration) for PRPs will be provided and coordinated through the OMERACT Executive Patient Stream leaders. Patients are not expected to fund the cost of their participation in OMERACT conferences and related activities.
2. **Patient research partners should be identified based on experiential knowledge and language skills, and personal interest**

These required characteristics are based on existing recommendations and guidelines available in the literature of PRP involvement in research projects [7,7,8]. Diversity is an important principle of OMERACT. Recruitment and selection of PRPs should therefore take into account differences in geographic origin, socioeconomic and cultural contexts, gender, age, disease duration, disease severity and disease impact, and potentially other disease, personal, or external characteristics. An attitude based on critical though constructive collaboration and a potential interest in research are important characteristics.

The involved PRPs in an OMERACT Working Group are not intended to represent inclusion of the entire patient perspective. The use of multiple, additional forms of data collection to capture the patient perspective, such as Delphi surveys, focus group interviews and surveys are likely to form part of the working group agenda, and should be performed appropriately.

Some participants may have overlapping roles as researchers and patients. This should be recognised so that roles can be appropriately defined. As with other OMERACT participants, potential conflict of interest needs to be disclosed, in particular financial interests that may be impacted by the person’s involvement in the research project. Such financial interests may include: stocks; bonds; ownership or partnership; consulting arrangements; grants or contracts; employment; and copyright on a specific measure or questionnaire [9]

3. **Patient Research Partners and the Working Group leadership should discuss the goals of the project and mutual expectations.**

Discussion of mutual goals and expectations before the start of the project, preferably during the first contact with a potential PRP, is good practice [10,11]. These expectations should be reviewed regularly throughout the process. Where possible, it is desirable to estimate the expected time PRPs are required to allocate for the project (e.g. 2 hours per month over 6 months) [11] with feasible timelines (e.g. feedback requested within 2 weeks).

4. **Patient Research Partners should be given the opportunity to be involved throughout the research process.**

PRPs should have the opportunity to be involved throughout the research process. [3,5,7,10,12,13,14] This includes the following stages: identifying the research question; reviewing and contributing to the study design; recruitment; data collection; analysis of findings; and dissemination of the results. PRPs should be consulted about and take part in decisions on implementation of the Working Group’s research agenda. Whenever possible PRPs should attend meetings of the Working Group or be connected by teleconferences etc.

Some PRPs may not wish or may not be available to participate in all phases, but they should be given the opportunity to do so. The frequency of involvement may differ, depending on the stage of the project: e.g. in core domain selection frequent involvement may be required, whereas data mining for discrimination may require less. This decision of inclusion or exclusion on any area of the project should be discussed by the PRPs and the working group leader.
5. The working group leadership should provide PRPs with timely and tailored support and information.

PRPs are full members of the working group and in order to contribute fully will require appropriate information and support. There is a wide range of support which can be provided, as indicated in the next section.

6. The nature of Patient Research Partner involvement should be reported throughout the OMERACT process

This recommendation encourages Working Groups to report the expected level of involvement and the names of the proposed PRPs in the initial research proposal. This ensures PRPs are involved early on in the process. To allow OMERACT participants to understand the extent of PRP involvement, sufficient detail should be reported in OMERACT documents such as proposals and pre-reading materials.

7. Involvement of Patient Research Partners should be recognized appropriately including co-chairing, co-presenting and co-authorship if applicable

Appropriate recognition can be enhanced by having PRPs who are willing and able to do so involved in facilitating discussion groups and reporting back at OMERACT conferences, presenting of data, and review of manuscripts. Recognition can also be provided by additional support to PRPs, such as arranging access to literature and libraries, offering thanks at special occasions or financial help to attend an educational symposium or international congress such as ACR and EULAR.

Acknowledgement of PRPs' contributions can take place through a text box at the end of the final research report or by offering co-authorship where the requirements of authorship are met. [3].

E. METHODS OF SUPPORTING PRP INVOLVEMENT

1. Information

PRPs are full members of the Working Group. Each team member is responsible for ensuring equality of all members in order to work effectively together. Each is responsible for creating a safe environment of open and honest interactions, that are sensitive to differences in culture, training and education of each member [6,10]. PRPs should receive appropriate and relevant information: for example: lay summaries and explanation of relevant statistics, research terms and disease features.

Open communication is important to all members of the Working Group, including PRPs. [6,10] Involvement in email exchanges, conference calls and corridor meetings at OMERACT conferences and other international congresses should be encouraged. Emails to the research team should either include the PRPs or a specific patient email should be sent at a similar frequency. Even when tasks or phases of the project may need little patient involvement a specific email or newsletter addressed to PRPs should be sent to keep PRPs informed.

PRPs should be offered the choice whether they would like to receive all information or whether they would like to receive less information relevant for the working group.
2. Invitations to meetings

While Working Groups do not generally provide financial support to members to attend, Working Group meetings held at international meetings such as ACR or EULAR, it may be appropriate to offer PRPs financial support if it is available. It will be the decision and efforts of each working group leadership to try to facilitate PRPs if their input is necessary. If PRPs are not present at the meeting, they should receive an update about what has been discussed at the meeting.

3. Support overall

Support refers to actions that encourage and promote PRPs to contribute with confidence throughout the research project by guaranteeing a positive and welcoming environment. Support includes tailored information and debriefings and encouragement of PRPs to speak up during meetings or any other interaction during the research process. (Here we are not referring to financial subsidies.)

4. Support during meetings and conferences

This refers to the provision of summaries of research in lay language ahead of the OMERACT meeting, as well as a list of abbreviations, terms and phrases relevant to the discussions and copies of appropriate outcome measures. It is also important to provide adequate time for PRPs to think about documents before responding. A rule of thumb is to allow at least one week, ideally 2, between distribution and feedback or subsequent teleconference. Moderation skills are important to enhance participation in group meetings.

5. Support between meetings

While involvement of PRPs between OMERACT meetings in Working Groups may vary according to the specific work phase PRPs should be kept informed. For some Working Groups, a separate PRP group may be established with a designated PRP leader who is a member of the working group steering committee. PRPs can have additional meetings and telecalls with PRPs to provide non-technical updates to keep the group informed and to receive additional input. Such separate discussions can be beneficial before or after regular telecalls of the working group to provide introductions and debriefing to material, discussions and issues for discussion.

There may be occasions on which PRPs are less comfortable in participating. For example, when a conference call will be discussing methods of statistical analysis for a PRO or outcome measure in which PRPs are an essential part of the group, but may be less able to contribute to the specific discussions or decision making. The starting point is the principle of inviting participation from PRPs throughout the process, regardless of the topic. However, it is recognized that “getting everyone on the same page” or to the same level of understanding may be difficult with certain technical aspects.

With a PRP as part of a group steering committee working with the co-chairs and other members of the steering committee, there will be opportunities to discuss the aspects of the work in which more or less PRP involvement may be appropriate.

6. Support from the OMERACT Executive Committee
The OMERACT Executive committee provides “generic” education and implements patient-centered ways to prepare interested patients to become capable PRPs in order to optimize research and results. During the OMERACT conference care is taken to maximize PRP participation within the overall constraints of the program. The patient stream leaders designated by the OMERACT Executive coordinate the patient group and patient involvement, working with the members of the Patient Board.

Examples of current initiatives include:

a. Patients arrive one day before the start of the formal conference
b. Summaries of research in lay language provided to patients ahead of the formal conference (prepared by Working Group leaders)
c. Addressing administration issues, special needs and concerns
d. A half-day PRP session at the beginning of the conference, including short education workshops to help understand some ‘terms’ and ‘methods’ of the research (eg explanation of statistics)
e. Daily sessions to explain each of the Working Group’s projects
f. Patient mentoring / buddy system (an experienced patient mentors with a less experienced patient)
g. Planned individual timetables (Personalized Patient Program) to prevent overburdening by pacing of time and energy of patients
h. Provision of an “on-call” physician to serve to triage medical concerns that may arise during the meeting
i. Glossary of medical and research terms updated before each conference
j. Personalized arrangements with conference hotels.
### 7. List for Working Group leaders

#### Working Group checklist for involvement of and support for patient research partners

<table>
<thead>
<tr>
<th>Paragraph</th>
<th>Text</th>
<th>Action</th>
<th>Check</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.0.1</td>
<td>The working group leadership should take responsibility for appropriate representation of the patient perspective in the research project.</td>
<td>4.0.1.1 There is one person among the working group leadership nominated to take responsibility for coordination and support of PRPs in the research project.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.1.2 Patient involvement is appropriate: patients are fully involved or the OMERACT Executive has given approval to a lower level of patient involvement.</td>
<td>☐</td>
</tr>
<tr>
<td>4.0.2</td>
<td>Each working group should involve at least 2 patient research partners. An exception may be some projects (e.g. a statistical project) where one patient research partner may be involved.</td>
<td>4.0.2.1 The working group includes at least 2 PRPs or the OMERACT Executive has given approval to a reduced extent of patient involvement.</td>
<td>☐</td>
</tr>
<tr>
<td>4.0.3</td>
<td>Patient research partners should be identified based on experiential knowledge and language skills, taking into account their personal interest in the topic.</td>
<td>4.0.3.1 The selected PRPs have good English language skills and have appropriate experiential knowledge.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.3.2 The PRPs fulfill the requirements set by the working group leadership.</td>
<td>☐</td>
</tr>
<tr>
<td>4.0.4</td>
<td>Patient research partners and the working group leadership should discuss the goal of the project and mutual expectations.</td>
<td>4.0.4.1 Goals and mutual expectations have been discussed with the PRPs prior to involvement.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.4.2 Goals and mutual expectations have been discussed with the PRPs at least once during the project, before the OMERACT conference.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.4.3 Goals and mutual expectations are discussed with the PRPs who are present at the OMERACT meeting, at the beginning of the OMERACT conference.</td>
<td>☐</td>
</tr>
<tr>
<td>4.0.5</td>
<td>Patient research partners should be given the opportunity to be involved throughout the research process.</td>
<td>4.0.5.1 Opportunity for PRPs to be involved in identifying the research question.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.5.2 Opportunity for PRPs to be involved in identifying the review and contributing to the study design.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.5.3 Opportunity for PRPs to be involved in recruitment.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.5.4 Opportunity for PRPs to be involved in data collection.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.5.5 Opportunity for PRPs to be involved in analysis of findings.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.5.6 Opportunity for PRPs to be involved in dissemination of results.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.5.7 Opportunity for PRPs to be involved in decisions on implementation of the working group, etc.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.5.8 PRPs attend meetings of the working group and are connected by teleconferences or other.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.5.9 Decisions about inclusion or exclusion of PRPs in any area of the project discussed by the PRPs and the working group leader.</td>
<td>☐</td>
</tr>
<tr>
<td>4.0.6</td>
<td>The working group leadership should provide PRPs with timely and tailored support and information.</td>
<td>4.0.6.1 PRPs should receive appropriate and relevant information, for example: lay summary of research results and opportunities for PRPs to be involved.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.6.2 Involvement in email exchanges, conference calls and corridor meetings at OMERACT conferences and other international congresses.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.6.3 Discussion help with PRPs relating to methods of communication.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.6.4 Consideration of financial and other support to attend meetings.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.6.5 Working group contributes to PRP support during OMERACT conferences.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.6.6 Consideration of additional PRP support measures.</td>
<td>☐</td>
</tr>
<tr>
<td>4.0.7</td>
<td>The nature of patient research partner involvement should be reported throughout the OMERACT process.</td>
<td>4.0.7.1 PRP involvement is reported in the initial proposal.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.7.2 PRP involvement is reported in OMERACT reports.</td>
<td>☐</td>
</tr>
<tr>
<td>4.0.8</td>
<td>Involvement of patient research partners should be recognized appropriately including co-chairing, co-presenting and co-authorship if applicable.</td>
<td>4.0.8.1 PRP recognition is enhanced by specific workgroup efforts.</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0.8.2 PRP contribution to publications acknowledged appropriately.</td>
<td>☐</td>
</tr>
</tbody>
</table>
F. DURING THE OMERACT MEETING

1. Patients opening session
Patients arrive the day before the meeting. On the morning of the meeting, they attend a 2 hour session preceding the main meeting program. This includes information about the diseases or conditions that are included in the conference programme; history and outputs of OMERACT; in particular the history of patient involvement and its consequences; the role of patients as key participants in the meeting consensus process; information about opportunities for participation in the process for reaching consensus including pre-conference, conference and post-conference time periods; an opportunity to ask experienced OMERACT patient and professional participants about issues or concerns.

2. Patients daily session
The purpose of the patients daily session is principally to prepare patients for the upcoming activities in the main programme. The first one takes place immediately after the patients opening session, and then each afternoon in preparation for the following day. In general, there are three workshop or module sessions requiring preparation. For 15 min each, leader(s) of each session come and explain how the session came to be in the programme, what the main issues are, and where patient input would be particularly useful. They can also address particular questions the patients may have.

3. Patients final session
On the day before the final plenary session, there will also be time to explain to patients how this works and how the OMERACT voting takes place. There is also usually time for the patient group to reflect on the overall conference programme and prepare feedback to the organisers.
PART 2 – A BRIEF HISTORY OF OMERACT PATIENT RESEARCH PARTNERS

1. Introduction
One member of the first group of patients to attend an OMERACT conference subsequently decided to study in detail how patients came to be involved in OMERACT, how the organisation changed in response to this, the lessons learned, the successes achieved and the challenges that lie ahead. These were the subject of his PhD thesis and some of the publications that derived from it [16,17,18,19] which provide an in-depth understanding of the process and outcome of these developments. The following section is extracted from Maarten de Wit’s thesis with minor modifications.

2. Initial decision to invite PRPs
During OMERACT 5 in 2000, participants discussed the concept of a Minimum Clinically Important Difference (MCID). Based on methodological arguments a growing interest in Patient Reported Outcomes (PROs) emerged, culminating in a spontaneous proposal at the final session to invite patients to the next conference. All participants voted in favor of this proposal [2]. The chair of the conference felt confident about the proposal because it had been discussed in the organizing committee before, although no decisions had been taken. Participants of the MCID module argued that patient perspectives should be explored further [20] and took responsibility for identifying 11 patients to join OMERACT 6 (2002) and to review the RA core set.

A document analysis revealed the unconditional positive reception of patient delegates at OMERACT conferences, and PRPs have confirmed during interviews that concerns regarding their involvement were misplaced [u]. They felt their reception was extremely welcoming [21]. Also the organizers were excited and called the patient involvement “a tremendous success” [22].

3. Subsequent role of PRPs
Between 2002 and 2012, a total of 57 PRPs with different rheumatic diseases had participated as full delegates with equal voting rights [23]. Their role and contributions had developed over time. At OMERACT 2002 they formed a homogeneous group of people with RA with little or no experience in scientific research. The level of involvement in the conference was limited; support was not organized, and the number of sessions patients attended was restricted. Contributions centered on participation in workshop discussions about the severity of fatigue and the definition of low disease activity, although there was a PRP keynote speech at the opening ceremony [21]. In contrast, by OMERACT 2014, PRPs were a heterogeneous group with different rheumatic conditions and different levels of experience and cultural backgrounds. They received a pre-conference information pack and were actively supported by a pre-conference dinner, a glossary of commonly used OMERACT research terms, training sessions and a buddy system. They carried out a variety of tasks similar to professionals such as giving plenary presentations, co-chairing and reporting back from breakout sessions, and preparing consensus statements. Several PRPs became co-authors of peer-reviewed publications and have presented aspects of group work at other conferences.

4. Patient contributions to OMERACT meetings and outcome research
In interviews about the OMERACT 2010 meeting, participants reported a variety of contributions made by PRPs during the conference, where they were an integral part of the deliberative and consensus-building process [24].

5. Five categories of contribution
Five main categories for PRP participation have been identified:

1) contributions to the research agenda;

2) the development of core sets;
3) the development of patient reported outcomes;

4) the culture of OMERACT; and

5) consequences outside OMERACT.

For example, PRPs have made contributions to the research agenda from the very beginning as they had significant influence by participating in workshops and small group discussions. They identified new outcome domains that are relevant from their perspective, but not included in existing core sets [25]. The first Patient Perspective Workshop, attended by 11 patient participants and 41 professionals, focused on the development of “valid outcome instruments that incorporate the perspective of the patient and to prepare the evidence and arguments for their inclusion in the (RA) core set” [26]. The preconference paper pointed out the methodological and political challenges: How to elicit and incorporate preferences of patients in RCT’s? [23] The workshop had been specifically arranged to support the PRP contributions including a pre- and post-workshop meeting. The workshop identified subjective experiences of RA, not encompassed in the RA core set but important aspects of the disease: a sense of well-being, fatigue, and disturbed sleep [26].

6. An added perspective

After the first conference attended by PRPs, it became apparent that perspectives of professionals and patients differ and more research was needed to articulate patients’ priorities [27,28,29,30]. PRPs emphasized the need for a holistic approach to people with arthritis [21]. The acknowledgement of the discordance of perspectives initiated new studies looking into the preferences, opinions and experiences of people with rheumatic diseases [31,32,33] and developing patient-derived core sets [12]. This made all OMERACT participants more aware of the relevance and importance of inclusion of the patient perspective. New topics emerged: remission, pain, flares, and foot problems. Contributions in the other areas identified are discussed in detail in a paper, whose first author was among the first OMERACT PRPs [24].
G. REFERENCES

15. de Wit MPT, Elberse JE, Broerse JEW, Abma TA. Do not forget the professional – the value of the FIRST model for guiding the structural involvement of patients in rheumatology research. Health Expectations 2013, online.
18. de Wit M, Abma TA, Koelewijn-van Loon M, Collins S, Kirwan J. What Has Been the Effect on Trial Outcome Assessments of a Decade of Patient Participation in OMERACT? J Rheumatol 2014;41:177–84; doi:10.3899/jrheum.130816)
CHAPTER 9 - THE OMERACT CONFERENCE: PLANNING TO PUBLICATION

A. THE OMERACT CONFERENCE

OMERACT conferences are held every 2 years in May. They rotate around the continents which contain the geographical distribution of most OMERACT organisers and participants – Europe, North America and Australasia – and are generally held in a well circumscribed site (such as a large conference facility) where participants can be in close interactive proximity for the whole of the meeting. The cost of the meeting is all-inclusive (which may make it look expensive on the surface) and participants just have to organize getting themselves there and back. There are generally between 180 and 220 places available at each conference. No more than approximately 10% of places are allocated to patient partners and a further 10% to industry stakeholders. All participants at OMERACT conferences come as individuals and do not represent any other organisation, society or company.

The meeting itself is more like an intensive 5-day workshop than a traditional scientific congress. The central component is the small group breakout sessions. These are facilitated by short introductory plenary presentations and followed by plenary feedback sessions. Special Interest Groups (SIGs) have 90 min sessions in the mornings or evenings running in parallel with other SIGs. Workshops have 2.5 hours, usually when there are no other sessions in progress. There are several hours in the middle of each day when there are no scheduled activities, but when participants meet to discuss the progress of their OMERACT work at the conference, share their data to help clarify the information available, work out how differences or misunderstandings can be resolved, find out how the OMERACT approach can be applied to their particular research interests, etc.

OMERACT endorsement to proposed areas of research, choices about domains and instruments, and agreement on proposed core outcome measurement sets takes place at a plenary session on the last day.

1. Membership

Membership of OMERACT is determined from the bottom up. Whoever registers and comes to the conference is a member! By 2013 there had been about 250 people who had attended many OMERACT conferences and a further 250 who had been to several, and a further 500 who had been to one or two. There may be another 500 who have worked as part of an OMERACT Working Group but who have never been to an OMERACT conference.

The OMERACT Executive grew out of the original OMERACT organizing group. It was given legitimacy by the readiness of OMERACT participants to follow the leadership of the Executive. Three of the current members have been members since the beginning. The group of 5 original members started to change from about 2008, and there has been a gradual process of individuals committed to OMERACT and interested to put in the effort required to take on an Executive role being invited to join and take on specific areas of interest. These include links to industry, patients, regulators and our small administrative team.

The Executive meets for several hours the day before or after the American College of Rheumatology (ACR) annual meeting and the European League Against Rheumatism (EULAR) annual meeting.
2. Planning the OMERACT conference

Planning starts 2 years ahead of the conference by deciding on which continent the meeting will be held and which Executive member will chair the OMERACT conference. The conference Chair gathers a conference group around him/her – including experienced Executive members who understand the finances, logistics, etc., and an organising company we employ to support the conference work.

On the day after each OMERACT conference there is an Executive meeting for several hours. An initial assessment of the strengths and weaknesses of the conference which has just finished is made, and lessons learned for the next meeting noted. If any big issues are likely to be carried forward to the next OMERACT then they are provisionally noted to be included in the programme.

3. Obtaining a time slot in the OMERACT conference

During the following year, while the organising group sort out the venue etc., the Executive will poll the Working Groups to find out how they are progressing and make some tentative enquiries about time allocation for the next meeting. Anyone wishing to set up a new Working Group will be allocated an Executive Mentor. Working Groups that have finished their work or have stopped moving forward will be taken off the list. By November (18 months before an OMERACT conference) the Executive will invite Working Groups (and potential Working Groups) to submit proposals for time allocation at the next meeting, either as a SIG, workshop or module. These proposals will include statements of the work the group will do leading up to the conference, and what they hope to achieve at the meeting. At least tentative identification of Patient Partners and a Fellow (see Chapter 11 Supporting participants with education and development at OMERACT) are required at this stage, and leadership roles and membership lists that look plausible are required, though they are often incomplete at this stage.

There are almost always more applications for sessions at the conference than there is time available. The Executive discusses and votes on the applications, and so generates a provisional programme for the meeting over 1 year ahead. Occasionally particularly strong case can be made for time in the programme after this and additional items can be included. Occasionally it becomes clear that a Working Group is not going to be able to deliver on its promised programme, and the item is removed.
B. OMERACT PUBLICATIONS

The proceedings of the OMERACT conference are published in the Journal of Rheumatology. OMERACT pays the page charges but there is a full peer review process and the Journal has the final decision on publication. All the usual Journal requirements apply.

C. OMERACT PUBLICATION POLICY

This guidance applies to manuscripts and not abstracts.

1. We expect all Working Groups publishing in the conference proceedings to use OMERACT in their publication title.

2. When publishing outside the conference proceedings, consider using OMERACT in the title of your manuscript if it is appropriate. One option is to consider including the term “OMERACT Working Group”, especially if all individual members are not being listed.

3. Please use OMERACT as a keyword.

4. We ask that OMERACT Working Groups wishing to publish outside the conference proceedings and making reference to OMERACT in the title or manuscript abstract, send their manuscript to the OMERACT secretariat admin@omeract.org to ensure alignment with OMERACT principles before submission. The Executive will ensure rapid turnaround of such manuscripts.

5. Authorship should reflect the breadth of individuals who participated in a particular working group or collaboration, including patient members, in accordance with journal authorship guidelines.

6. For those groups that present at the OMERACT biannual meeting, for inclusion in the conference proceedings we expect:

   a. A minimum of one ‘proceedings’ or ‘update’ paper with details of the projects, processes, outcomes and future research agenda of the working group. Such a manuscript is expected to summarise substantial work and must also include original research.

   b. Additional publication(s) will be considered by the Executive if they are original research and contain appropriate robust results. If there are no new findings, for example because results are being utilised in another manuscript, a manuscript will not be considered for publication. Decision on additional manuscripts will be based on abstracts submitted to the Executive pre or post-OMERACT meeting that describe the content of all OMERACT proceedings manuscripts coming from that group.

1. Pre-conference materials

All Working Groups which have a session at the conference must update and expand on their original application to produce an outline programme which is suitable for potential participants to see and obtain a feeling for the proposed content and conduct of the session. This is required by January of the year of the conference in order for potential participants to make their plans for attendance.

This statement, along with any other recommended reading, is required 2 months before the conference so that it can be collated and made available on the OMERACT web site. Working Groups may sometimes negotiate with the Executive that a separate, full preparatory paper is produced for
pre-conference reading, which in itself will be published with the conference proceedings. (For example, references 1 and 2.)

2. Conference session reports

For SIGs, the work of the Fellow in preparing for the meeting is combined with the meeting plan and the results of the discussions that arise in the SIG (and possibly any endorsement at the final plenary session) to write a short paper dealing with the work of the SIG. (Sometimes the Fellow will also have undertaken enough work so that a separate paper reporting that work in detail may also be submitted and published outside any arrangement with the conference publications.) The authors of this paper are expected to be the SIG organisers, the Fellow (who may be first author) and anyone who made presentations at the SIG, the gist of which is incorporated into the paper.

For Workshops there will often be a more substantive report, possibly even 2 papers, depending on the preparatory work and the content. It is expected that reports for these events will be authored by the session organisers, the Fellow, any speakers who give brief introductory presentations, and the reporters and facilitators from the discussion breakout groups. This makes for a long list of authors, but also gives credit to all those who have had intellectual input into producing the report.

D. FUNDING

1. Funding support

OMERACT conference costs for attendees are met by the registration fee. Patient partners do not pay a registration fee personally, but the OMERACT Patient Stream Coordinators – members of the Executive Committee – arranges a combination of central funding and funding obtained by Working Groups to enable about 20 patient participants at the conference (see Chapter 8 OMERACT Patient Partners).

In addition, some funds are made available to all Working Groups who have been chosen to organise sessions in the timetable. These funds can be used to support whatever is needed for the group to get its work done – paying travel and accommodation for their Fellow or their patient partner(s), organising a preparatory meeting, paying for some small preliminary research (such as a questionnaire survey), etc. The Executive Mentor attached to a Working Group can provide advice.

Fellows may be able to apply for Fellows Bursaries which have traditionally been made available through ARA and EULAR.

It is expected that Working Group organisers will seek other funds from a variety of sources to help ensure their patient partner(s) and Fellows are able to attend the meeting.

2. Managing the funds

It is up to the Working Group organisers to manage their funding. Allocations from central OMERACT funds are managed through their Executive Mentor, the Patient Stream Coordinators (who are the only persons who can formally invite patients to attend and guarantee them a place) and the managing company OMERACT is working with.
The last tranche of funds allocated to a Working Group in relation to an OMERACT conference is not delivered until the conference report paper has been submitted.

3. Availability of papers

All OMERACT proceedings are open access. The Journal of Rheumatology makes them available through its archive, and there is a full set of publications on the OMERACT web site. (http://www.omeract.org/conference_proceedings.html)

D. REFERENCES

CHAPTER 10 - SUPPORTING PARTICIPANTS WITH EDUCATION AND DEVELOPMENT AT OMERACT (FELLOWS & NEWBIES)

A. INTRODUCTION

The overall goal of OMERACT is to improve outcome measurement in rheumatology. This is a continuous process as new concepts, techniques and tools are developed; new theories generate new domains or instruments; and new generations of clinical researchers emerge. The key activity of OMERACT is to exchange ideas and reach consensus at the biennial meeting. A key principle of the OMERACT meeting is that all participants have an active role in the process. Education programs have been developed to enable participants to gain knowledge, skills, and attitudes that will help them to contribute. These occur while participants are currently contributing to the OMERACT process and to the development of future OMERACT activities.

B. OMERACT FELLOWS:

BACKGROUND

There is strong agreement with the overarching principal/philosophy to train Fellows in the tradition and methods of OMERACT recognizing that they are our solution to the sustainability of the excellent ground work established over the past 20 years.

There is strong agreement from participants and the OMERACT Executive Committee that the Fellows program is a valued part of the OMERACT Program (one-day meeting + daily sessions) and competing external programs will not be scheduled simultaneously (apart from New OMERACT Attendees and Patient Orientation Programs).

Therefore, as part of the OMERACT process, all Working Groups (WG) and special interest groups (SIGs) must nominate a Fellow to become registered with OMERACT.

However, young researchers, with genuine motivation to learn about outcome measures and OMERACT, are welcomed to apply to the Fellow program even if they are not linked to any OMERACT WG.

FELLOW ELIGIBILITY CRITERIA & ROLE

OMERACT accepts 2 categories of Fellows:

- WG-SPONSORED (i.e. already linked with a WG)
- WG-NOT PRE-SPONSORED (i.e. without any prior link with a WG)

Common rules and activities:

- A Fellow (WG-SPONSORED or WG-NOT PRE-SPONSORED) is anyone in a training scheme (an attestation letter from the supervisor or the institution secretary should be provided) or having already completed the training, who is “young” in his/her research career and not older than 40 years.

- We will not set a limit and we will accept as many as apply and if there are more than 20 Fellows, then extra senior OMERACTors will help run parallel orientation and daily sessions.

- Fellows must attend the first day meeting and present their work as a poster or oral presentation.

- Fellows must attend the scheduled daily Fellow sessions.
• Fellows will be invited to present the summary of the previous days SIGs at the beginning of each day to the main meeting.

• Fellows are expected to participate in the ENTIRE program, ALL sessions.

• A Fellow can only attend the OMERACT Fellow Daily program once; however, they are welcome at subsequent OMERACT meetings and will have an OMERACT Executive Committee member work with them.

A) WG-SPONSORED
• All WG-SPONSORED Fellows are required to submit an abstract and participate in the Fellow first day and daily session programs, which make them eligible for the subsidized registration rate.

• A WG-SPONSORED Fellow should expect to present for at least some part of the WG presentation in the main meeting of the group (SIG, WS, Module, etc.).

• The abstract work submitted to the Fellow Program should NOT be identical to the presentation in the main meeting, but may be a spin-off from that work or a more detailed version of what is going to be presented in the main OMERACT WG meeting (exceptions to this may be first time SIGs).

• WG leads are required to review and endorse the WG-SPONSORED Fellow’s abstracts prior to submission.

• WG-SPONSORED Fellows who are attending a second or subsequent OMERACT may still submit an abstract of their work for discussion and for receiving the reduced rate, if not already being supported by their WG. These Fellows would attend the Fellow First Day Session, but would not attend the Fellow Daily Sessions.

• All WGs presenting at the meeting must help support the WG-SPONSORED Fellow financially to attend the meeting.

B) WG-NOT PRE-SPONSORED
• All WG-NOT PRE-SPONSORED Fellows are required to submit an abstract and present a poster during the First Day Meeting, which make them eligible for the subsidized registration rate.

• WG-NOT PRE-SPONSORED will be accepted to participate to the Fellow program if they are funded by EULAR or ARA.

• WG-NOT PRE-SPONSORED Fellows who are not submitting an abstract or participating in the poster review session will not receive the reduced registration rate and cannot participate in the Fellow Program.

• A letter of support is required from current or most recent past supervisor at time of registration.

• The abstract should be on the spirit of OMERACT activities and feedback and revision will be supervised by a senior OMERACT panel.

• If WG-NOT PRE-SPONSORED Fellows have not found a WG of interest, one will be allocated before coming to the meeting (meaning that in some instances WGs will have more than one Fellow).

• Potential Fellows should contact the Working Group leader of their area of interest for further advice and information.
1. **Overall process at the conference**

All OMERACT Fellows are required to submit an abstract on their pre-conference work. This supports a poster presentation. Fellows arrive the day before the meeting starts and have a special Fellows programme interwoven into the overall programme. It is often helpful for Fellows to read this Handbook as part of their preparation.

2. **Fellows opening session**

Fellows spend the first half-day in a Fellows opening session. This includes some presentations and discussion about the philosophy of OMERACT; an introduction to OMERACT history, concept and process; and an introduction to the OMERACT general program. Fellows briefly introduce themselves and say a little about their background, and then all go on the poster tour. The submitted pre-conference works are presented as posters which remain on display throughout the meeting. Those Fellows who have been chosen to make an oral presentation have their poster commented on by experienced OMERACT members and briefly discussed by the whole group. Finally, the nature of the daily Fellows programme is explained.

3. **Fellows daily session**

Fellows attend each Daily Fellows session, at the end of the day. During this session, the Fellows (designated the day before) present what they have understood from the sessions attended during the day. Such presentation will be followed by a discussion with established investigators. One nominated fellow summarizes the session, and two designated others critique it. Supervision is ensured from one of the group chairs for modules and workshops. The nomination process for this reporting is at random, not by specific and personal domain of interest. The goal is to identify clear comprehension of the session, understanding of the OMERACT methods, identification of good process, and feedback to the organisers. The principle is learning by active involvement and reflection.

**What Fellows Deliver**

a) Be an active member of an OMERACT Working Group  
b) Attend all Fellows sessions  
c) Attend all voting sessions  
d) Present a 1 minute introduction in Fellows opening session  
e) Present a 3 minute ‘work highlights’ in Fellows opening session  
f) Present a session summary or a reflection/critique of the Fellows summary at least one OMERACT Module or Workshop session during a Daily Fellows session

**What you receive**

a) Sponsored by an OMERACT Working Group  
b) Mentored by OMERACT community  
c) Become familiar with the OMERACT process for reaching consensus  
d) Constructive feedback on academic work, presentation, reporting and reflection skills by senior experienced OMERACT participants  
e) The opportunity to provide feedback to develop and change the OMERACT process  
f) The opportunity to be an active participant in contributing to achieve international consensus on evidence based outcomes and measurement for rheumatology clinical trials
FIRST TIME OMERACT PARTICIPANTS

The New Participants (‘Newbies’) session was first introduced at OMERACT 11 in 2012. It includes a program based on the same educational principle of learning by active involvement.

1. New participants opening session

All new participants are encouraged to arrive early to learn about the unique OMERACT history and process. A 3-hour programme preceding the commencement of the main meeting program covers the history and outputs of OMERACT; information about opportunities for participation in the process for reaching consensus including pre-conference, conference and post-conference time periods; an opportunity to ask experienced OMERACT participants about issues or concerns.

2. New participants daily session

There is also a daily reflection session which first time participants are encouraged to attend. During this session, individuals present what they have understood from the sessions attended during the day and this is discussed among the group and with established investigators.
C. SMALL GROUP MODERATORS AND REPORTERS

The purpose of the breakout sessions is to work towards achieving consensus in the areas described in the specific group information, and identifying lack of consensus where it exists. The objective therefore is not to force consensus. Where consensus does not immediately exist, we will employ a systematic process to identify the underlying reasons. (For further elaboration see Chapter 03 Methods of achieving consensus.)

Moderators and reporters are supported in their role by the organisers of the session in which they are working, and by members of the Executive Committee. There is a checklist of requirements. We also recommend reading a paper [1] which describes in detail how poor management of the discussion group can result in effective disenfranchisement of some members.

D. WORKING GROUP LEADERS

Working Group leaders are almost always experts in the setting of the core set they are seeking to develop, but have different degrees of experience of and familiarity with the methodological issues in OMERACT. Every Working Group is allocated a Mentor from the Executive Committee or Scientific Advisory Committee, who helps and advises on the preparation of materials, establishing appropriate procedures, deciding on questions for voting, ideas about funding, and anything else the group finds helpful.

E. REFERENCES

THE OMERACT GLOSSARY
FOR PATIENT RESEARCH PARTNERS

EDITORS PAMELA RICHARDS & MAARTEN DE WIT
The OMERACT Glossary

Compiled by Maarten de Wit and Pamela Richards

Can be found here:
https://www.dropbox.com/s/argsziu6ddq1iam/OMERACT%2013%20Glossary%20April%202016.pdf?dl=0