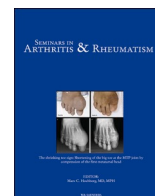




Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

Advancing composite outcome measures: Insights on weighting components from OMERACT 2023

George A Wells^{a,b,c,*}, Francis Guillemin^d, Peter A. Merkel^e, Maarten de Wit^f, Sarah Mackie^{g,h}, Lyn Marchⁱ, Gunnar Tómasson^{j,k}, Lauren K. King^{l,m}, Sam Michel Cembaloⁿ, Shawna Grosskleg^{o,p}, Lara J. Maxwell^q, Sara Monti^r, Kaitlin A. Quinn^s, Beverley J Shea^t, Peter Tugwell^{u,v,w,x}, Dorcas Beaton^y

^a Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Canada

^b School of Epidemiology and Public Health, University of Ottawa, Canada

^c Department of Medicine, University of Ottawa, Canada

^d Université de Lorraine, Inserm, INSPIRE, Nancy, France

^e Division of Rheumatology, Department of Medicine, Division of Epidemiology, Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA, USA

^f OMERACT Patient research partner, Netherlands

^g Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK

^h Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust, Leeds, UK

ⁱ Florence and Cope Professorial Department of Rheumatology, Royal North Shore Hospital and Kolling Institute, Faculty of Medicine and Health Sydney MSK, University of Sydney, Australia

^j Centre of Public Health Sciences, Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland

^k Internal Medicine and Emergency Services, Landspítali-The National University Hospital of Iceland, Reykjavik, Iceland

^l Department of Medicine, University of Toronto, Canada

^m Li Ka Shing Research Institute, St. Michael's Hospital, Unity Health Toronto, Canada

ⁿ Patient Research Partner, France

^o OMERACT, Toronto, Canada

^p Faculty of Medicine, University of Ottawa, Ottawa, Canada

^q Ottawa Hospital Research Institute, Centre for Practice Changing Research and Faculty of Medicine, University of Ottawa, Ottawa, Canada

^r IRCCS Istituto Auxologico Italiano, Immunorheumatology research laboratory, Milan, Italy

^s National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH, Bethesda, Maryland

^t Bruyère Research Institute, Senior Methodologist, Ottawa Hospital Research Institute, Adjunct Professor, Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Canada

^u University of Ottawa, Department of Medicine, Faculty of Medicine, K1H 8M5, Ottawa, Canada

^v Ottawa Hospital Research Institute, Clinical Epidemiology Program, K1Y 4E9, Ottawa, Canada

^w University of Ottawa, School of Epidemiology and Public Health, Faculty of Medicine, K1H 8M5, Ottawa, Canada

^x WHO Collaborating Centre for Knowledge Translation and Health Technology Assessment in Health Equity, Bruyère Research Institute, K1R 7G5, Ottawa, Canada

^y Senior Scientist, Institute for Work & Health, Toronto, ON

ARTICLE INFO

Keywords:

OMERACT

Composite

Outcome measure

Weighting

Arthritis

ABSTRACT

Objective: The OMERACT Composite Working Group hosted a workshop at OMERACT 2023 to explore the complexities of weighting components in the development of composite outcomes. This study presents the methodology and findings of this workshop, exploring the complexities of weighting the individual components of composite outcome measures.

Methods: The workshop featured a multifaceted program, beginning with a plenary session that introduced the concept of composite outcomes, shared a patient's journey with rheumatic disease through a narrative, illustrated a composite outcome for Osteoarthritis Flares, and outlined the five domains selected for this composite outcome. A breakout exercise engaged participants in ranking and assigning weights to these domains, followed by group discussions to reach a consensus on weights. The workshop concluded with another plenary session that

* Corresponding author at: University of Ottawa Heart Institute, Ottawa, Ontario Canada.

E-mail address: gawells@ottawaheart.ca (G.A. Wells).

<https://doi.org/10.1016/j.semarthrit.2024.152503>

Available online 30 June 2024

0049-0172/© 2024 Published by Elsevier Inc.

discussed various weighting approaches, including discrete choice and conjoint analysis from the ANCA-Associated Vasculitis working group, and outlined future directions for research on composite outcome methods. **Results:** The breakout exercise revealed the challenges in assigning relative importance to different domains, highlighting the variability in participant perspectives. Consensus discussions highlighted the diversity in approaches to weighting, the need for appropriate methods to determine domain weights and the impact of such weights on the interpretation of composite scores.

Conclusion: The OMERACT 2023 workshop underscored the significance of a systematic approach to weighting components in composite outcome development. It highlighted the complexity of achieving consensus on the importance of domains and the role of incorporating the perspectives of patient research partners in this process. Future research directions include refining weighting methodologies, moving composites through the OMERACT Filter and enhancing understanding of their implications for clinical trials. The findings contribute to the ongoing discourse on optimizing composite outcome measures in rheumatology and beyond, advocating for a balanced integration of scientific rigour and patient-centeredness in their development.

At OMERACT 2023, the OMERACT Composite Working Group led a workshop exploring the often overlooked stage of the weighting of components in the development of a composite outcome. The Workshop consisted of a plenary session that included an introduction to the concept of a composite, a patient story about their lived experience with rheumatic disease and why a composite outcome best represents their journey, an example of a composite outcome for Flare in Osteoarthritis (OA); and the selected five domains for the composite outcome for Flare in OA. This was followed by a breakout exercise involving all OMERACT participants to understand better the difficulty of weighting the relative importance of domains, an essential precursor step to combining the domains into a single composite score. The exercise involved each participant ranking the five domains of Flare in OA from the most important to the least important; each participant assigning weights of importance to each of the five Flare in OA domains; and then participants at each table sought to reach a consensus on the weights for each domain. At the end of the breakout exercise, a plenary session was again conducted, providing an overview of the different types of approaches to weighting, a presentation and a discussion about the use of discrete choice- or conjoint analysis to arrive at weights for composite criteria for response to treatment in ANCA-associated vasculitis (AAV); and the next steps and planning for future workshops on composite outcome methods at OMERACT.

Introduction to composites

A composite outcome is a measure that combines multiple domains into a single outcome measure to capture a broader concept. Fig. 1 presents the OMERACT method for developing a composite.

First, define the target concept, such as disease activity, flare, response to treatment, disease remission, disease impact, etc. If consensus is reached that multiple “components” are required to adequately capture that concept well, the need for a composite measure becomes apparent. Second, identify the domains. The components brought together are determined through the OMERACT way (method) for domain identification. These domains could be part of the “onion” core set, other layers of the “onion,” or outside the “onion.” Each domain must be defined in detail before entering the next step. Third, identify the instruments. A good quality instrument for measuring each domain is selected following OMERACT methods. The nature of the measurement for these instruments could be discrete (e.g., binary) or continuous. It is possible that one instrument measures more than one domain. Fourth, weight the domains (or components) of the composite. Multiple domains must be brought together for the composite to capture the target concept adequately. Such a combination of domains requires determining the relative importance (or weight) of the domains. The second and third steps for determining the domains and instruments follow the standard and familiar steps described in the OMERACT handbook[1]. However, the weighting step may be less familiar and determining the relative importance of the domains can be challenging. There are different methods to determine these weights, as identified in

the section on Approaches to Weighting. One key objective of this workshop was to provide insight into weighting issues. The final two steps are to put the composite through the OMERACT filter by gathering evidence of Truth, Feasibility, and Discrimination[2] and bring the final measure to a vote. Further details on composites and OMERACT are provided in the paper by Wells et al. 2012[3]. A well-known example of a composite is the American College of Rheumatology (ACR) response criteria for rheumatoid arthritis (Table 1). Together, these domains are believed to better capture the concept of response to treatment than using multiple instruments that measure individual domains in settings where the intent is to bring the domains together, for example, a single powered outcome of a clinical trial or in the outcome in a systematic review.

Composites from a patient’s perspective

Ideally, from the patient’s perspective, composites reflect the overall reality of a multifaceted disease and its impact. Patients expect that a composite score measures their disease state in a single score. However, because both patients and physicians are often not aware of the weighting of individual components in a composite, the score can be difficult for them to interpret. Therefore, patients also want insight into the individual components’ scores. A reliable composite that offers an overall score and insights into its contributing components can help patients and physicians monitor disease activity, formulate realistic treatment goals, and support shared decision-making.

Psoriatic arthritis (PsA) is an example of a multifaceted and heterogeneous disease. With substantial input from patient research partners (PRPs), important domains for PsA were agreed upon and are now part of the inner core of the OMERACT PsA core set (Fig. 2)[5].

In selecting instruments for the PsA core outcome measurement set, Musculoskeletal (MSK) disease activity was considered to need a composite to capture the multiple ways this domain is manifested. When reviewing available composites for disease activity, PRPs were involved in evaluating domain match and feasibility. Several composite instruments were developed for PsA; some were borrowed from rheumatoid arthritis. Like the disease, there is heterogeneity in the domains these composite measures capture, in the number of domains (range 3–7) and, when examined in detail, in the approaches to weighting (Table 2)[6,7]. This highlights why, when the target is not well defined, the involvement of PRPs in the assessment of domain match is imperative.

An important question from PRP’s perspective is: have patients with the condition of interest been involved in formulating a comprehensive definition of the target concept (e.g., disease activity) and in identifying relevant components? This is important because there are still domains that matter to patients, such as fatigue, that are not part of any existing PsA composite outcome. Another challenge for PRPs is the evaluation of the respective weights in the composite, particularly when PRPs try to assess the importance of domains that they may not have experienced, such as symptoms such as enthesitis or dactylitis. If pre-existing

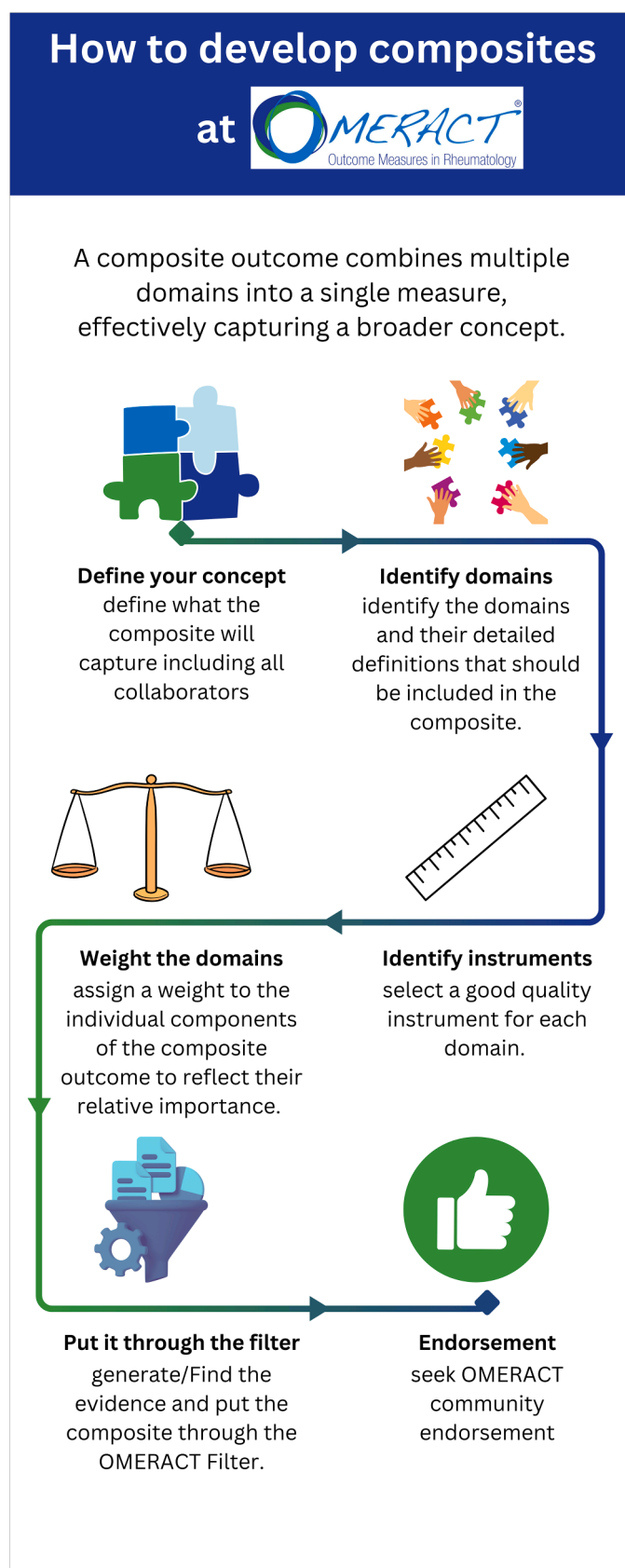


Fig. 1. The OMERACT Pathway on how to develop a composite.

Table 1

Example of a composite: the American College of Rheumatology disease activity response criteria for rheumatoid arthritis

Domain Components	Outcome Measurement Instrument	Weighted Multi-Outcome Domain Instrument (ACR20)*
Joint tenderness	Tender joint count (0–68)	≥20 % Improvement
Swollen joints	Swollen joint count (0–66)	≥20 % Improvement
Patient global	Patient assessment of disease activity (0–10)	
Physician global	Physician assessment of disease activity (0–10)	≥20 % Improvement in any three of these five measures
Pain	Pain assessment (0–10)	
Disability	HAQ disability index (0–3)	
Inflammation	Erythrocyte sedimentation rate (mm/hr)	

* ACR20: American College of Rheumatology response criteria for rheumatoid arthritis[4].

composites are used, explanations should be provided regarding the inclusion and exclusion of components and their weighting to PRPs, with the supportive data shown on the testing of the individual components and their contribution to the composite measure. There should also be continued recognition that all core domains will be assessed and reported regardless of whether a composite is used in a study.

Ultimately, the involvement of PRPs in developing or assessing composite measures would benefit from a clear definition of the target concept developed in collaboration with patients. Such a definition would make it easier for PRPs to assess concept match, enhancing the composite's face validity. Understanding the weighting method used in current composites will help PRPs contribute to working groups developing or assessing composites.

Identifying domains – Flare in Osteoarthritis

The exercise in weighting presented during the OMERACT workshop considered the domains of Flare in OA. The Flare in OA Working Group had previously defined their concept and identified five domains of flare [8]. Domains were identified and characterized through an inductive approach with patients and health professionals. Consensus processes were then used to finalize the domains and their definitions following OMERACT consensus methodology[9]. The five domains are:

1. Pain during flare: Pain that is more severe and lasts longer, that is particularly heightened with physical activity and persists with rest.
2. Swelling during flare: A new increase in size or feeling of fullness of the joint.
3. Stiffness during flare: Increased or prolonged stiffness of the joint that does not improve with movement.
4. Psychological aspects during flare: Alterations in mood, including depressive symptoms, greater anxiety, greater irritability, and/or low morale that are consequences of the symptoms during flare.
5. Impact of symptoms: A change in the ability to perform daily activities, requiring new adaptation and strategies due to the pain, swelling, and stiffness of the flare.

The next step is identifying good quality instruments for measuring each domain. Two literature reviews were conducted but no existing instruments measuring either all OA flare domains or any of its dimensions separately were identified such that when assembled in a composite measure would be of reasonable length for practical use[10, 11]. For that reason, the OA Flare Working Group voted for developing an instrument targeting good content validity[12].

The fourth step in reaching a composite for Flare in OA is to allocate different weights to each of the five domains according to their

The OMERACT – Onion Psoriatic Arthritis

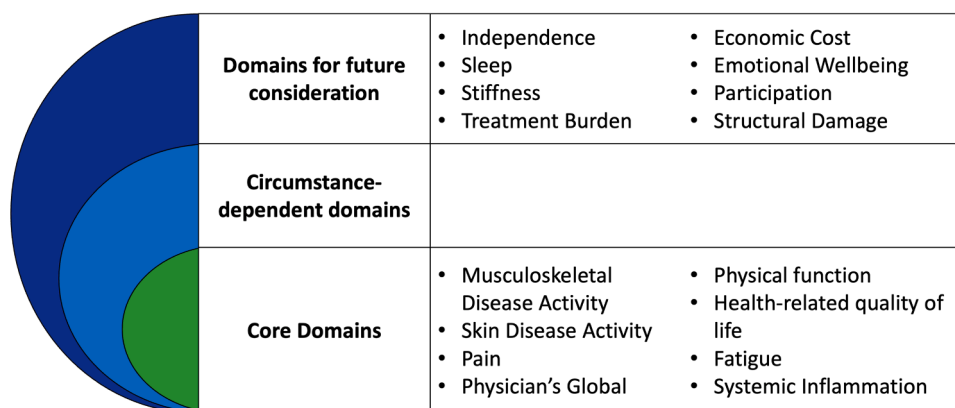


Fig. 2. The OMERACT Onion identifying the domains for psoriatic arthritis.

Table 2

Variable mixture of content across composites in Psoriatic Arthritis (adapted from Mease and Coates 2018[7]).

Outcome Domain	Composites Used in Psoriatic Arthritis						
	PsARC Psoriatic Arthritis Response Criteria	PsAJAI Psoriatic Arthritis Joint Activity Index	DAPSA Disease Activity for Psoriatic Arthritis	PASDAS Psoriatic Arthritis Disease Activity Score	GRACE Group for Research and Assessment of Psoriasis and Psoriatic Arthritis Composite Exercise	MDA Minimal Disease Activity	CPDAI Composite Psoriatic Arthritis Disease Index
MSK disease activity							
Peripheral arthritis	✓	✓	✓	✓	✓	✓	✓
Enthesitis				✓		✓	✓
Dactylitis				✓			✓
Spine disease							✓
Pain		✓	✓			✓	
Patient assessment	✓	✓	✓	✓	✓	✓	✓
Skin					✓	✓	✓
Physical Function		✓		✓	✓	✓	✓
Inflammation		✓	✓	✓			
Quality of Life					✓		
Physician assessment	✓	✓		✓			
Fatigue							

importance in characterizing the disease. Many of these were already in the OA Core Domain Set, but two (impact of symptoms and psychological aspects) arose as specific to the concept of Flare in OA. Formal weighting methods the Flare in OA Working Group used for weighing the Flare-OA questionnaire were published[13]. The following described exercise was not a formal process of weighting for Flare in OA. Only the domains formally identified were used in this exercise to illustrate some of the issues encountered in the process of determining such weights.

The weighting exercise

For a composite to adequately capture the target concept, multiple domains must be brought together, and the domains’ relative importance (or weights) need to be determined. Determining how much each domain should count in the total score is critical in developing a composite. This weighting can be done in many ways. The goal of the exercise at OMERACT 2023 was to provide insight into the issues of weighting by answering this simple question: What weights would you give to each of the five domains of Flare in OA?

Worksheets (Fig. 3) and beads were provided to each OMERACT participant sitting at one of 11 tables. Three tasks were successively carried out. First, each participant ranked the five domains (pain, swelling, stiffness, psychological aspects and impact of symptoms) based on their importance to the concept of Flare in OA and recorded their ranking (column 1 on the worksheet). Second, each participant was given beads of different colours for each domain, as identified on the worksheet and a cord. Working individually, each participant decided how much weight they would give each domain using 20 beads and recorded the number (column 2 on the worksheet). The finite and arbitrary number of 20 beads was used to help people allocate a reasonable number of beads that would still require choice across domains and allow comparability of weights in future steps of the exercise. When completed, the participant made a bracelet with these weights and the corresponding number of coloured beads. Third, participants at each table worked as a group to reach a consensus on one weighting system for the table. A second 20-bead bracelet was created based on the group’s consensus weights, and the result was recorded (column 3 on the worksheet). In reaching this consensus, participants at the table were asked to discuss the variability among the participants at their table and

DOMAIN WEIGHTING

Are you a PRP? Yes / No TABLE #: _____

Step #1: Number the domains 1 most important to 5 least important to reflect order of importance to the concept of Flare in OA.

Step #2: Think about how much weight should be given to each. 20 beads total

Step #3: Think about how much weight should be given to each as a group. 20 beads total

	Order of importance (1-5)		Individual # beads total = 20		Group # beads total = 20
Pain	<input style="width: 50px; height: 50px;" type="text"/>	●	<input style="width: 50px; height: 50px;" type="text"/>	●	<input style="width: 50px; height: 50px;" type="text"/>
Swelling	<input style="width: 50px; height: 50px;" type="text"/>	●	<input style="width: 50px; height: 50px;" type="text"/>	●	<input style="width: 50px; height: 50px;" type="text"/>
Stiffness	<input style="width: 50px; height: 50px;" type="text"/>	●	<input style="width: 50px; height: 50px;" type="text"/>	●	<input style="width: 50px; height: 50px;" type="text"/>
Psychological Aspects	<input style="width: 50px; height: 50px;" type="text"/>	●	<input style="width: 50px; height: 50px;" type="text"/>	●	<input style="width: 50px; height: 50px;" type="text"/>
Impact of Symptoms	<input style="width: 50px; height: 50px;" type="text"/>	●	<input style="width: 50px; height: 50px;" type="text"/>	●	<input style="width: 50px; height: 50px;" type="text"/>

Make your bracelet!

Fig. 3. Worksheets provided to each OMERACT participant for the weighting exercise.

Table 3

Weights are assigned to each domain (a total of 20 points to be assigned across five domains) as agreed upon by each group.*

Group	Consensus Weights Assigned to Each Domain by Group (/20)					Illustrative examples – composite scores (/100) for two individuals	
	Pain	Swelling	Stiffness	Psychological Aspects	Impact of Symptoms	Person A (NRS ratings: 8, 0, 0, 1, 4)	Person B (NRS ratings 5, 3, 6, 1, 8)
1	6	3	2	3	6	37.5	51.0
2	8	2	3	2	5	43.0	53.0
3	7	1	2	3	7	43.5	54.5
4	6	3	4	3	4	33.5	49.0
5	6	3	3	3	5	35.5	50.0
6	5	5	4	2	4	29.0	49.0
7	6	3	3	3	5	35.5	50.0
8	6	2	3	3	6	37.5	52.5
9	6	5	1	3	5	35.5	47.0
10	6	1	4	3	6	37.5	54.0
11	8	3	3	3	3	39.5	47.0

Legend: NRS = Numeric Rating Scale.

* The group's method to reach consensus on weights was not assigned in the exercise.

what factors were considered in reaching agreement. After the weighing exercise, participants wrote down their thoughts about making a weighting decision individually and as a group. The results of the weighting exercise recorded on the worksheets were consolidated in a spreadsheet and analyzed descriptively (mean weights and frequency of different patterns of weights across domains).

Results

Working groups submitted their table's consensus weights resulting from the weighting exercise for discussion at the plenary. The weights (all needed to total 20 across the five domains) are presented in Table 3. The pain domain was consistently assigned high weights as a domain (average of 6.4, range 5–8) followed by impact of symptoms (5.1/20, range 3–7). The other domains each shared the same mean score of 2.9. However, there was variability between groups, that is important to recognize. Two groups, group 5 and group 7, assigned the same weights across the five domains (i.e., 6, 3, 3, 3, 5). No other group assigned the same weights to each domain. Some weights were quite distinct; for example, group 11 assigned 8 points to pain and 3 to each of the other domains. Groups weighted the domain "impact of symptoms" very differently (range 3 to 7), with the extremes representing a two fold weighting difference.

This exercise revealed discordance in the weighting of components by different groups. When translated into a composite measure, these differences in weighting would lead to different composite score for the same person. Two scenarios illustrate the issue. Person A has severe pain (8/10), no swelling or stiffness, some distress (1/10), and moderate impact (4/10). This person's scores are shown under scenario A.

In contrast, Person B is an individual with moderate pain (5/10), swelling (3/10) and stiffness (6/10) with little distress (1/10) and significant impact on life (8/10). The composite scores they would obtain across the different weighting systems are shown in Table 3. These fictitious scenarios demonstrate the variability in composite scores for the same scenario. Depending on the group consensus weights, Person A scores ranged from 29 to 43.5/100, and Person B's from 47 to 54.5/100.

Approaches to weighting

The informal weighting exercise above provided insights into issues related to determining the relative importance of the domains pain,

Table 4

Examples of methods to determine weights for components in a composite endpoint.

Method	Description	Example
Discrete choice experiments - patients	<ul style="list-style-type: none"> Participants are presented with a scenario that would make them eligible for the study Respondents chose among pairs of procedures that differed on the probability of outcome(s) Conjoint analysis derived relative weights for these attributes 	Weighting components using a patient discrete choice experiment (Tong et al., Ann Thorac Surg. 2012) (11)
Delphi panels - clinician investigator	<ul style="list-style-type: none"> External Delphi panel to determine the relative severity of individual components of the composite end point Net clinical outcome assessed through the incorporation of risk thresholds for events 	Weighting components using a clinician-investigator Delphi panel (Armstrong et al., Am Heart J. 2011) (12)
Disability-adjusted life years (DALY)	<ul style="list-style-type: none"> DALY values for the most common major endpoints derived using World Health Organization Global Burden of Disease Project methodology 	Weighting components using disability-adjusted life-years (Hong et al., Stroke 2011) (13)

swelling, stiffness, psychological aspects, and impact of symptoms that target Flare in OA.

In general, after deciding on the outcome domains and the instruments that will be used to measure them, the weight for each outcome domain is considered using more formal approaches. Several recent approaches for weighting the outcome domains can be considered [14–16]. These approaches include discrete choice experiments, Delphi panels, and disability-adjusted life years (DALY) (Table 4).

Aside from these, more explicit approaches to weighting a composite, such as the Clinical Disease Activity Index (CDAI) for rheumatoid arthritis, have an intrinsic or built-in "weighting" in the composite's score [17]. That is, the different range of scores used to assess each domain provides weights when the scores are summed with the domains using larger possible scores (e.g. 27 points for joint count and only 10 for pain) receiving greater weight in the sum. In using these built-in weights, it is assumed, rightly or wrongly, that the magnitude of the score ranges reflects the importance of each domain to the composite outcome. The implicit assumption is that the four outcome domains (joint tenderness, swollen joints, patient global sense of disease activity and provider global sense of disease activity) are all of equal importance and/or the instruments used to measure these outcome domains provide a weighting that reflects the relative importance of the outcome domains. If this is not the case, then a weighting scheme may be needed, as described in Table 4.

The Disease Activity Score-28 (DAS28) for rheumatoid arthritis is an example of a weighted instrument and provides an example of statistical approaches to determine "weighting" of the outcome domain instruments. The development of the DAS28 involved the following steps: principal component analysis resulting in five factors (laboratory measures, joint counts, functional status measures and subjective assessments by patients and physicians); canonical discriminant analysis was used to select the variables that best discriminate between high and low disease activity resulting in nine variables (pain, hemoglobin, erythrocyte sedimentation rate (ESR), grip strength, morning stiffness, 44 swollen joint counts, Ritchie Articular Index (RAI), and α_2 -globulin, β -globulin); further analysis concentrated on the variables identified in the original DAS (RAI, number of swollen joints, ESR and patient global assessment) based on factor analysis, discriminant analysis and multiple regression analysis; and for feasibility, 28-joint counts and the discriminant function replaced the 2 comprehensive joint counts were recalculated [18]. There are many statistical modelling procedures that can be considered when deriving a weighted instrument.

Example of development of a composite outcome measure: ANCA-associated vasculitis

Another example of weighting and composites is the work by the vasculitis working group regarding response to treatment in ANCA-associated vasculitis (AAV) [19]. Almost all clinical trials in AAV conducted to date used a primary endpoint that relies on dichotomous definitions of active disease and remission based on the Birmingham Vasculitis Activity Score (BVAS), a purely physician-based instrument. The goals of the response criteria for AAV are to quantify treatment response beyond the simple transition from active disease and remission and to assess additional benefits of treatment, such as alleviating symptoms of importance to patients and preventing irreversible organ damage.

The working group identified and selected domains of importance with respect to treatment response in AAV through an iterative process that included i) a systematic literature review, ii) a 3-round Delphi exercise, and iii) input from PRPs and investigators on the Steering Committee. The final draft criteria include a combination of disease activity measures, damage assessment, and patient-reported outcomes. The group then used a discrete choice exercise (conjoint analysis) to arrive at weights for each criterion by administering a survey using the 1000minds software to clinical experts and patient research partners.

Which scenario represents greater treatment response in AAV?

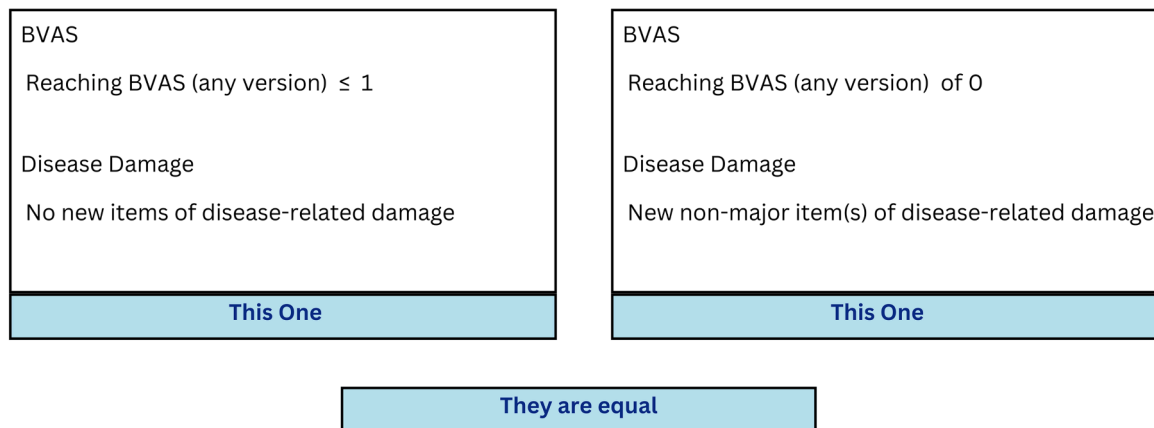


Fig. 4. Example scenario for the discrete choice exercise to arrive at weights for response criteria for ANCA-associated vasculitis.

This method involves comparing two scenarios simultaneously and determining which scenario represents the greater treatment response (or, to conclude, they show equal response) (Fig. 4). The figure shows an assessment of two responses to treatment in ANCA-associated vasculitis. The scenario on the left shows no progression of irreversible disease damage but could allow for some minimal disease activity, according to the BVAS instrument. In the scenario on the right, there is no detectable disease activity according to BVAS, but some irreversible organ damage occurred. When all possible combinations of treatment responses according to the criteria have been assessed pairwise, a relative weight can be calculated for each criterion.

Next steps: where do we go now?

At the end of the workshop, questions were posed to a panel comprised of investigators, fellow researchers, and PRPs conducting work on composite outcomes. Several important issues were raised and recorded. These included the importance of having the correct set of domains to cover the target concept of the composite and of getting all the right perspectives on board to ensure this coverage (e.g., patient perspectives, imaging experts); recognition of the complexities inherent in each method and therefore in the development of a composite; the many factors (e.g., order of presentation of domains) that could impact weighting exercises; the importance of having high-quality instruments to represent each domain within a composite; the need to keep a focus on the ultimate goal of having a composite index that can discriminate among responses in clinical trials, when weighting a composite; the need to report results for each domain as well as the overall composite; and the generation and availability of data to conduct well-powered data-driven analyses for the creation of composites.

Based in part on the issues raised during this session, the OMERACT Technical Advisory Committee (TAG) intends to investigate the many ways to weight; maintain a website to catalogue the various composites that currently exist or are in development; provide education related to the development of composite outcomes; and provide further guidance to mitigate the challenging issues that arise in the process of developing a composite outcome or moving such a composite through the OMERACT Filter for validity and reliability evidence.

Conclusion

As with all outcomes, there is always a What (what you want to measure) and How (how you will measure it). For composite outcome measures, those measures that gather heterogeneous domains into one score, what you want to measure is usually a higher-order domain such

as response, disease activity or flare that must be represented by more than one domain. The how you will measure it is with a well-developed composite index. This paper reviewed the content and structure of composites with a particular focus on weighting of the domains within a composite index. Weighting is a complicated task, with The OMERACT 2023 weighting exercise demonstrating just how varied weights can be and why a rigorous strategy is needed to help guide the weighting process.

Composite outcomes are not easy to create. This paper outlined and demonstrated the complexities related to even just one step in their development: weighting. However, composite outcomes continue to appeal to clinical trialists because the outcomes are efficiently represent multiple domains, and to patients, who experience their disease in a multifaceted manner. Composites are complex to construct, including the selection of domains, the weights assigned, and the degree of confidence associated with the final score. These challenges are worth the effort given the perspective a high-quality composite can provide.

CRediT authorship contribution statement

George A Wells: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Francis Guillemín:** Writing – review & editing, Writing – original draft, Conceptualization. **Peter A. Merkel:** Writing – review & editing, Writing – original draft. **Maarten de Wit:** Writing – review & editing, Writing – original draft, Conceptualization. **Sarah Mackie:** Writing – review & editing, Conceptualization. **Lyn March:** Writing – review & editing, Writing – original draft, Conceptualization. **Gunnar Tómasson:** Writing – review & editing, Writing – original draft. **Lauren K. King:** Writing – original draft, Conceptualization. **Sam Michel Cembalo:** Writing – review & editing. **Shawna Grosskleg:** Writing – review & editing, Writing – original draft, Project administration, Conceptualization. **Lara J. Maxwell:** Writing – review & editing. **Sara Monti:** Writing – review & editing. **Kaitlin A. Quinn:** Writing – review & editing. **Beverley J Shea:** Writing – review & editing. **Peter Tugwell:** Writing – review & editing, Conceptualization. **Dorcas Beaton:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Peter Tugwell receives consulting fees from Reformulary Group for providing independent medical consultation professional services to the firms listed in this section. Participation on a Data Safety Monitoring

Board or Advisory Board for UCB Biopharma GmbH & SPRL Parexel International Prahealth Sciences. Is an independent Committee Member for clinical trial Data Safety Monitoring Boards for FDA approved trials being conducted by:

UCB Biopharma GmbH & SPRL
Parexel International
Prahealth Sciences

Other financial or non-financial interests with Abbvie, Astra Zeneca, Aurinia, BMS, Centrexion, GSK, Horizon Pharma Inc, Janssen, Novartis, Pfizer & Sparrow. He is [unpaid] Chair of the Management Group of a registered non-profit independent medical research organization, OMERACT, whose goal is to improve and advance the health outcomes for patients suffering from musculoskeletal conditions. OMERACT receives arms-length funding from 8 companies.

Dorcus Beaton has an unpaid leadership role as a member of the Management team at OMERACT, co-chair of the methods group, and technical advisory group of OMERACT.

Maarten de Wit receives consulting fees from Evidera (Patient Partner Advisory Council), honoraria from UCB 2022, 2023 for facilitating a workshop.

Sara Monti received consulting fees from CLS-Vi for AstraZeneca

All other authors have no conflicts to declare.

References

- [1] Beaton D, Maxwell L, Grosskleg S, Shea B, Tugwell P., editors The OMERACT handbook. 2nd ed.; 2021.
- [2] Beaton DE, Maxwell LJ, Shea BJ, et al. Instrument selection using the OMERACT filter 2.1: the OMERACT methodology. *J Rheumatol* 2019;46(8):1028–35. <https://doi.org/10.3899/jrheum.181218>.
- [3] Wells GA, Tugwell P, Tomasson G, et al. Composite outcomes at OMERACT: multi-outcome domains and composite outcome domains. *Semin Arthritis Rheum* 2021;51(6):1370–7. <https://doi.org/10.1016/j.semarthrit.2021.11.001>.
- [4] Felson DT, Anderson JJ, Boers M, et al. The American college of rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. *Arthritis Rheum* 1993;36(6):729–40. <https://doi.org/10.1002/art.1780360601>.
- [5] Orbai AM, De Wit M, Mease PJ, et al. Updating the psoriatic arthritis (PsA) core domain set: a report from the PsA workshop at OMERACT 2016. *J Rheumatol* 2017;44(10):1522–8. <https://doi.org/10.3899/jrheum.160904>.
- [6] Coates LC, FitzGerald O, Merola JF, et al. Group for research and assessment of psoriasis and psoriatic arthritis/outcome measures in rheumatology consensus-based recommendations and research agenda for use of composite measures and treatment targets in psoriatic arthritis. *Arthritis Rheumatol* 2018;70(3):345–55. <https://doi.org/10.1002/art.40391>.
- [7] Mease PJ, Coates LC. Considerations for the definition of remission criteria in psoriatic arthritis. *Semin Arthritis Rheum* 2018;47(6):786–96. <https://doi.org/10.1016/j.semarthrit.2017.10.021>.
- [8] Guillemin F, Ricatte C, Barcenilla-Wong A, et al. Developing a preliminary definition and domains of flare in knee and hip osteoarthritis (OA): consensus building of the flare-in-OA OMERACT Group. *J Rheumatol* 2019;46(9):1188–91. <https://doi.org/10.3899/jrheum.181085>.
- [9] King LK, Epstein J, Cross M, et al. Endorsement of the domains of knee and hip osteoarthritis (OA) flare: a report from the OMERACT 2020 inaugural virtual consensus vote from the flares in OA working group. *Semin Arthritis Rheum* 2021;51(3):618–22. <https://doi.org/10.1016/j.semarthrit.2021.03.010>.
- [10] Cross M, Dubouis L, Mangin M, et al. Defining flare in osteoarthritis of the hip and knee: a systematic literature review - OMERACT virtual special interest group. *J Rheumatol* 2017;44(12):1920–7. <https://doi.org/10.3899/jrheum.161107>.
- [11] Queiroga F, Cross M, Thomas MJ, March L, Epstein J, Guillemin F. A scoping review of patient self-report measures of flare in knee and hip osteoarthritis (OA): a report from the OMERACT flares in OA working group. *Semin Arthritis Rheum* 2023;63. <https://doi.org/10.1016/j.semarthrit.2023.152281>. 152281-152281.
- [12] Queiroga F, Cembalo SM, Epstein J, et al. Assessing domain match and feasibility of candidate instruments matching with OMERACT endorsed domains to measure flare in knee and hip osteoarthritis. *Semin Arthritis Rheum* 2024;65. <https://doi.org/10.1016/j.semarthrit.2024.152371>. 152371-152371.
- [13] Queiroga F, Epstein J, Erpelding ML, et al. AB1678 construction of a composite score for patient self-report of flare in osteoarthritis: a comparison of methods with the flare-OA-16 questionnaire. *Ann Rheum Dis* 2023;82(Suppl 1):2076–7. <https://doi.org/10.1136/annrheumdis-2023-eular.1380>.
- [14] Tong BC, Huber JC, Ascheim DD, et al. Weighting composite endpoints in clinical trials: essential evidence for the heart team. *Ann Thorac Surg* 2012;94(6):1908–13. <https://doi.org/10.1016/j.athoracsur.2012.05.027>.
- [15] Armstrong PW, Westerhout CM, Van de Werf F, et al. Refining clinical trial composite outcomes: an application to the Assessment of the Safety and Efficacy of a New Thrombolytic-3 (ASSENT-3) trial. *Am Heart J* 2011;161(5):848–54. <https://doi.org/10.1016/j.ahj.2010.12.026>.
- [16] Hong KS, Ali LK, Selco SL, Fonarow GC, Saver JL. Weighting components of composite end points in clinical trials: an approach using disability-adjusted life-years. *Stroke* (1970) 2011;42(6):1722–9. <https://doi.org/10.1161/STROKEAHA.110.600106>.
- [17] Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI) : a review of their usefulness and validity in rheumatoid arthritis: quantitative clinical assessment of rheumatic diseases. *Clin Exper Rheumatol (Testo stampato)* 2005;23(5):S100–8.
- [18] van Riel PLCM. The development of the disease activity score (DAS) and the disease activity score using 28 joint counts (DAS28). *Clin Exp Rheumatol* 2014;32: S65–74.
- [19] Quinn KA, Monti S, Christensen R, et al. Developing a composite outcome tool to measure response to treatment in ANCA-associated vasculitis: a mixed methods study from OMERACT 2020. *Semin Arthritis Rheum* 2021;51(5):1134–8. <https://doi.org/10.1016/j.semarthrit.2021.07.001>.