Seminars in Arthritis and Rheumatism 51 (2021) 1370–1377

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

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

Composite outcomes at OMERACT: Multi-outcome domains and composite outcome domains

George A Wells[a,*], Peter Tugwell[b], Gunnar Tomasson[c], Francis Guillemin[d], Lara J Maxwell[e], Beverley J Shea[f], Shawna Grosskleg[g], Peter A Merkel[h], Lyn March[i], Dorcas E Beaton[j]

[a] School of Epidemiology and Public Health, University of Ottawa Heart Institute, University of Ottawa, Ottawa, Canada
[b] Division of Rheumatology, Department of Medicine, School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada
[c] Faculty of Medicine, University of Iceland, Reykjavik Iceland, Department of rheumatology, University Hospital, Iceland
[d] Université de Lorraine, APEMAC, Nancy, France
[e] Faculty of Medicine, University of Ottawa, Ottawa, Canada
[f] Ottawa Hospital Research Institute, Faculty of Medicine, University of Ottawa, Ottawa Canada
[g] OMERACT Secretariat, University of Ottawa, Ottawa, Canada
[h] Department of Rheumatology, Faculty of Medicine and Health, University of Sydney and the Royal North Shore Hospital, Australia
[i] Division of Rheumatology, University of Pennsylvania School of Medicine, United States
[j] Institute for Work & Health and Institute for Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

A B S T R A C T

The OMERACT Technical Advisory Group recognises that working groups during the process of creating a core outcome set may identify an outcome domain that would be best represented as a composite that encapsulates these component outcome domains by bringing them together into a single outcome. A multi-outcome domain (MOD) is a within-patient combination of component outcomes, and an individual patient's evaluation depends on the observation of all of the components in that patient with a single overall rating determined according to a specified rule; which is often applicable when we consider a disease activity score. A composite outcome domain (COD) consists of a number of component outcomes and is defined as the occurrence in a patient of one, some or all of these specified components; which is often applicable when we consider the risk of adverse events or remission criteria. We review the general benefits, challenges, reporting and interpretation of using MODs and CODs. The development of the MOD or COD instrument for an OMERACT core outcome measurement set is considered through four distinct steps: choosing relevant outcome domains; finding high quality instruments for each of these outcome domains; weighting the outcome domain instruments in the MOD/COD instrument; and putting MOD/COD instrument through the OMERACT Filter. Guidance and training are in preparation for working groups who will be completing the OMERACT Instrument Selection Algorithm (OFISA). As for other initiatives in OMERACT, we will seek feedback from OMERACT working groups who complete the development of their MOD/COD, which will then be incorporated into the refinement of the guidance and training.

Introduction

In studies of health interventions for a chronic disease, participants may experience several different outcome domains of interest. For a patient, the impact of the disease and the response to therapy is likely not just one of these outcome domains but actually a combination. An outcome that encapsulates all these outcome domains by bringing together each outcome domain into a single outcome may be of interest. This holistic approach to outcomes is appealing and it can work well in a field such as arthritis where a spectrum of outcomes and indicators of a multifaceted concept like disease activity or occurrence of a range of adverse events are of interest. This approach has been identified by regulatory agencies, among others, as one of the general approaches for handing 'multiple endpoints' [1].

The perspectives of the regulatory agencies on 'composites' can be instructive. Following the FDA guidance for industry on multiple endpoints [1], a 'composite endpoint' consists of a number of components and is defined as the occurrence in a patient of any one of the specified components. A different type of 'multi-component endpoint' is a within-patient combination of two or more components, and an individual patient's evaluation depends on the observation of

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

https://doi.org/10.1016/j.semarthrit.2021.11.001
0049-0172/

all of the components in that patient with a single overall rating determined according to a specified rule. Regarding the latter, the EMA guideline on clinical investigation for the treatment for rheumatoid arthritis [2] use the term 'composite disease activity scores'. However, using the word 'composite' leads one to thinking of the classical 'composite endpoint' (i.e., occurrence of any one of its components and the composite has occurred), which is often applicable when we consider the risk of adverse events. However, most of our combinations of interest at OMERACT are for efficacy (e.g., disease activity) and are more in alignment with the 'multi component endpoint'.

To avoid confusion with the word 'composite', we will use the terms 'composite outcome domain' (COD) and 'composite outcome domain instrument' when the interest is the occurrence in a patient of any one, some or all the component outcome domains; such as any one component occurring when considering the risk of adverse events or all components occurring by meeting threshold criteria when considering remission criteria. The terms 'multi-outcome domain' (MOD) and 'multi outcome domain instrument' will be used when an individual patient evaluation depends on the observation of all of the component outcome domains in that patient with a single overall score determined according to a specified rule. These latter terms are more suggestive of our work at OMERACT in that after pertinent outcome domains are identified, we may want to combine these outcomes domains into a single multi-outcome domain. Examples of these COD and MOD instruments are provided in Table 1.

Multi-outcome domain (MOD) instrument

More often than not, a MOD instrument will be considered when efficacy, such as disease activity, is of interest. For such a 'multi-component outcome' a within-patient combination of two or more outcome domains with a single overall score is determined according to specified rule. A MOD instrument is often referred to as a clinical index. When the outcome domains are measured on an ordinal or continuous numeric scale, one way of forming an overall score is to sum or average across the component outcome domain measures. For example, for rheumatoid arthritis (RA), the Composite Index of Disease Activity in RA (CDAI) is a MOD consisting of outcomes domains joint tenderness, swollen joints, patient global sense of disease activity and provider global sense of disease activity (Table 1) [3]. Since the MOD instrument is a simple sum of the instruments for the component outcome domains, it is an unweighted MOD instrument, although there is an internal weighting mechanism based on the scores of the instruments of the component outcome domains. As such, it may best be described as a self-weighted MOD instrument. The CDAI can be visualized as in Table 2.

Another MOD instrument used in RA is the DAS28 score, which is an example of a weighted MOD instrument (Table 1). The DAS28 consists of the component domain outcomes joint tenderness, swollen joints, general health status and erythrocyte sedimentation rate, which are incorporated in the DAS28 in a weighted fashion through a formula combining the domain outcomes (Table 1). Salaffi and

Table 1
Examples of multi-outcome domain and composite outcome domain instruments.

Multi-outcome domain instrument (MOD Instrument)	Composite outcome domain instrument (COD Instrument)
Multi-outcome domain instrument	**Composite outcome domain instrument**
Index - sum of the measurements of each of the outcome domains	One or more outcome domains occur vs no outcome domains occur
Example	**Example**
Composite Index of Disease Activity (CDAI) in rheumatoid arthritis	Major Adverse Cardiovascular Events (MACE)
Outcome domains	*Outcome domains*
Joint tenderness	Myocardial infarction
Swollen joints	Stroke
Patient global sense of disease activity	Death
Provider global sense of disease activity	*Outcome domain instruments*
Outcome domain instruments	Myocardial infarction: occur yes/no
TJC28: Tender joint count (0–28)	Stroke: occur yes/no
SJC28: Swollen joint count (0–28)	Death: occur yes/no
PtDA: Patient global sense of disease activity (0–10)	
PrDA: Provider global sense of disease activity (0–10)	*Composite outcome instrument*
Composite outcome instrument	1 if any one of myocardial infarction, stroke or death occurs
CDAI = TJC28 + SJC28 + PtDA +PrDA	0 if none occur
Multi-outcome domain instrument	**Composite outcome domain instrument**
Index - formula combining the measurements of each of the outcome domains in a weighted fashion	All outcome domains occur vs otherwise
Example	**Example**
Disease Activity Score in 28 joints (DAS28) in rheumatoid arthritis	ACR/EULAR Remission Criteria in rheumatoid arthritis
Outcome domains	*Outcome domains*
Joint tenderness	Joint tenderness
Swollen joints	Swollen joints
General health	Patient global sense of disease activity
Inflammation	Acute phase reactant
Outcome domain instruments	*Outcome domain instruments*
TJC28: Total joint count (0–28)	TJC28: Total joint count (0–28)
SJC28: Swollen joint count (0–28)	SJC28: Swollen joint count (0–28)
GH: General health status (0–100)	PtDA: Patient global sense of disease activity (0–10)
ESR: erythrocyte sedimentation rate	CRP: C-reactive protein
Composite outcome instrument	*Outcome domain indicators*
$DAS28 = 0.56 \times \sqrt{(TJC28)} + 0.28 \times \sqrt{(SJC28)} + 0.70 \times \ln(ESR) + 0.014 \times GH$	I_{TJC28}= 1 if TJC28≤1; 0 otherwise
	I_{SJC28}= 1 if SJC28≤1; 0 otherwise
	I_{PtDA}= 1 if PtDA≤1; 0 otherwise
	I_{CRP}= 1 if CRP≤1 mg/ml; 0 otherwise
	Composite outcome instrument (ACR/EULAR Remission)
	1 (yes) if $I_{TJC28} + I_{SJC28} + I_{PtDA} + I_{CRP} = 4$
	0 (no) if $I_{TJC28} + I_{SJC28} + I_{PtDA} + I_{CRP} < 4$

Table 2
Tabular Depiction of the Composite Index of Disease Activity (CDAI) in Rheumatoid Arthritis: A Self-weighted MOD Instrument.

Outcome Domain	Outcome Domain Instrument	Unweighted Multi-Outcome Domain Instrument
Joint tenderness	m(TJC28) =Tender joint count (0–28)	m(TJC28)
Swollen joints	m(SJC28) = Swollen joint count (0–28)	m(SJC28)
Patient global	m(PtDA) = Patient global sense of disease activity (0–10)	m(PtDA)
Provider global	m(PrDA) = Provider global sense of disease activity (0–10)	m(PrDA)
		MOD instrument = CDAI = m(TJC28) + m(SJC28) + m(PtDA) + m(PrDA)

Ciapett have provided a comprehensive catalogue of weighted MOD instruments used in assessing clinical disease activity in RA [4]. In their systematic review of randomized controlled trials in RA, Ibrahim and colleagues found that the MOD instruments most frequently reported were the ACR20 responder index, followed by the DAS28-ESR, DAS28-CRP and the ACR50 [5]. In general, the structure of a weighted MOD instrument can be viewed as in Table 3.

Composite outcome domain (COD) instrument

More often than not, a composite outcome instrument will be considered when a risk of adverse events is of interest, and the composite outcome occurs when any one of the components occurs. This is the situation that is usually associated with 'composites' in the literature. For example, in studies of RA, immunomodulatory agents have been associated with reduced cardiovascular (CV) events. Solomon and colleagues investigated if patients who had lower RA disease activity over time suffer fewer CV events, regardless of which immunomodulatory treatments they had received [6]. In this investigation, they considered the risk of a 'composite CV endpoint' MACE (major adverse cardiovascular events) that included myocardial infraction (MI), stroke and CV death and analyzed the time to MACE (i.e., first occurrence of any one of these events) (Table 1). Similarly, MACE was considered in an integrated long term safety analysis of tofacitinib for the treatment of rheumatoid arthritis (RA), psoriatic arthritis (PsA), ulcerative m colitis (UC) and psoriasis (PsO). [7] In terms of a COD instrument for this COD, it can be visualized as in Table 4.

These types of CODs have been considered for a variety of arthritic conditions when considering the risk of adverse events. For example, composites were considered to evaluate the gastrointestinal effects of nonsteroidal anti-inflammatory drugs [8, 9], and a COD 'basket of predefined designated endpoints in each organ system' [10] was proposed for primary safety outcomes based on the insights gained from the Vigor [11] and CLASS [12] coxib trials in which there were statistical power problems with the primary and unexpected endpoints, and it took years to withdraw rofecoxib from the market. Lampropoulos and colleagues [13] compared treatment related adverse events and infections in patients with RA treated with synthetic disease-modifying anti-rheumatic drugs (DMARDs) **or** biologic DMARDs in a real world setting. The primary outcome studied was the first adverse event encountered during follow-up within the hospital.

A composite outcome instrument can also be considered when the occurrence of some or all of the components occur, such as a disease remission criteria where each of the components is assessed if it meets a threshold criteria and remission is considered to have 'occurred' only if all the components meet their threshold. For example, the ACR/EULAR definition of remission in RA clinical trials, the remission criteria considered is a COD in which all components must occur for remission to be met [14]. They considered the four outcome domains joint tenderness, swollen joints, patient global sense of disease activity and acute phase reactant. For each outcome domain instrument a threshold of '$\leq 1'$ was defined and remission was met only if all instruments met their threshold (Table 1). In terms of a COD instrument for this COD, it can be visualized as in Table 5. In general, the structure of a COD instrument can be viewed as in Table 6.

What are the benefits of using a MOD or COD?

Several reasons have been identified for using a COD [15] that are also applicable to MOD, including: to avoid a misleading conclusion when an intervention reduces a less serious outcome by increasing a more serious outcome; to avoid unnecessary complexity when outcomes of identical significance to patients, but different pathophysiology, would otherwise be analyzed separately; to decrease the necessary sample size and duration of follow-up; to avoid the need for statistical adjustments for multiple testing of multiple outcomes; to estimate the net clinical benefit of an intervention; to improve understanding the effect of the interventions avoiding competing risks; and to avoid the need to choose a single primary outcome when many may be of equal importance.

What are the challenges of using a MOD or COD?

In contrast to these positive reasons for using a composite outcome, there are several challenges. Again, some of the challenges that have been identified for COD [15] apply as well to MOD. These include, the practical interpretation could be problematic when component outcome domains of the MOD/COD vary appreciable in patient and/or provider importance; the larger the number of components then the more the work to accurately ascertain the COD/MOD; excessive influence of the more imprecise subjective component outcome domains; and adjustment of the alpha error to draw confirmatory conclusions about the components when multiple statistical testing is involved. Other challenges that are primarily associated with COD include: possibility of biases secondary to competing risk; and a potential masking of an increase in a harmful effect associated with an intervention [15].

Table 3
Tabular schematic for the calculation of an unweighted and weighted multi-outcome domain instrument.

Outcome Domain	Outcome Domain Instrument	Unweighted Multi-Outcome Domain Instrument	Outcome Domain Instrument Weights	Weighted Multi-Outcome Domain Instrument
A	m(A)	m(A)	W_A	W_A m(A)
B	m(B)	m(B)	W_B	W_B m(B)
C	m(C)	m(C)	W_C	W_C m(C)
D	m(D)	m(D)	W_D	W_D m(D)
		MOD instrument = m(A) + m(B) + m(C) + m(D)		MOD instrument = W_A m(A) + W_B m(B) + W_C m(C) + W_D m(D)

Table 4
Tabular Depiction of the Major Adverse Cardiovascular Events (MACE): A COD Instrument.

Outcome Domain	Outcome Domain Instrument	Indicator for Occurrence of Outcome Domain Instrument
MI	m(MI) occurs yes/no	I_{MI}=1 if MI occurs; 0 if not occurs
Stroke	m(Stroke) occurs yes/no	I_{Stroke}=1 if stroke occurs; 0 if not occurs
CV death	m(CV death) occurs yes/no	$I_{CV\ death}$=1 if CV death occurs; 0 if not occurs COD instrument 1 (yes) if any one of myocardial infarction, stroke or death occurs = $I_{MI} + I_{Stroke} + I_{CV\ death} > 0$ 0 (no) if none occur = = $I_{MI} + I_{Stroke} + I_{CV\ death} = 0$

Table 5
Tabular Depiction of the ACR/EULAR Remission Criteria for Rheumatoid Arthritis: A COD Instrument.

Outcome Domain	Outcome Domain Instrument	Indicator for Occurrence of Outcome Domain Instrument
Joint tenderness	m(TJC28) =Tender joint count (0–28)	I_{TJC28} = 1 if m(TJC28)≤1; 0 otherwise
Swollen joints	m(SJC28) = Swollen joint count (0–28)	I_{SJC28} =1 if m(SJC28)≤1; 0 otherwise
Patient global	m(PtDA) = Patient global sense of disease activity (0–10)	I_{PtDA} = 1 if m(PtDA)≤1; 0 otherwise
Acute phase reactant	m(APR) = C-reactive protein (mg/ml)	I_{APR} = 1 if m(APR)≤1 mg/ml; 0 otherwise *(ACR/EULAR Remission)* 1 (yes) if $I_{TJC28} + I_{SJC28} + I_{PtDA} + I_{CRP} = 4$ 0 (no) if $I_{TJC28} + I_{SJC28} + I_{PtDA} + I_{CRP} < 4$

Table 6
Tabular schematic for the calculation of a composite outcome domain instrument.

Outcome Domain	Outcome Domain Instrument	Indicator for Occurrence of Outcome Instrument
A	m(A)	I(m(A))
B	m(B)	I(m(B))
C	m(C)	I(m(C))
D	m(D)	I(m(D)) COD instrument = F[I(m(A)), I(m(B)), I(m(C)), I(m(D))]

How should a MOD or COD be reported and interpreted?

A MOD/COD can be the primary or secondary outcome in a randomized controlled trial. Frequently, it is the primary outcome and is often associated with increased statistical efficiency. Systematic reviews have been constructed on problems in the defining and reporting of CODs [15, 16] that are applicable to the reporting of MODs. When reporting the MOD/COD it should be interpreted as a whole as the primary outcome and the component outcome domains should be considered and reported as secondary outcomes [2, 17]. This will help determine if there are any inconsistencies of the effects of the intervention across components and/or if any one of the component outcome domains dominates the MOD/COD [15]. Care must be taken to avoid the suggestion that individual components of the MOD/COD have been demonstrated to be effective (the effect on the components should be interpreted together rather than demonstrating efficacy of individual components). It can be informative if an a priori rank order from "worst" to "best" for the component outcome domains to be considered, and the "worst" outcome domain experienced reported [18]. Some suggest for a more fulsome consideration of the component outcome domains, and that all possible combinations of the components should be reported [19].

Some pertinent questions to consider to better interpret a COD include: are the component outcome domains of similar importance to participants; did the more and less important outcomes occur with similar frequency; are the components likely to have similar relative risk reductions; is the underlying biology of the components similar; are the point estimates of the relative risk reductions similar and the confidence intervals sufficiently narrow. The extent to which the answers to these questions are 'no' will determine whether you need to examine the component outcome domains separately.

What is the role of MOD and COD at OMERACT?

The OMERACT Technical Advisory Group (TAG) recognises that working groups may identify an outcome domain that would be best represented as a "multiple outcome" during the process of creating a core outcome set. The core domain set creation defines what the constructs or concepts (domains) will be involved in the core outcome set and the core outcome measure set defines how each of these domains in the core domain set will be measured by gathering evidence of 'truth discrimination and feasibility' of candidate MOD/COD instruments. After determining the relevant outcome domains and identifying an instrument for each domain, the next step may be to choose a scoring approach to federate these outcomes and then to put this MOD/COD instrument itself through the OMERACT Filter 2.1 to ensure it can be used to represent this MOD/COD in the proposed context of use. The purpose here is to reach a common understanding of the concepts and terms used for "composite outcomes" at OMERACT and to understand the key elements in the development and testing of these multiple outcome concepts of MOD and COD, which are described respectively in the next two sections.

Multi-Outcome domain (MOD) instrument development at OMERACT

Core domain sets at OMERACT may include a clinical phenomenon that is best captured by a MOD. Disease activity indices [(CDAI, SDAI), flare (flare RA, flare OA)], and responder indices (ACR20) are all examples of outcome domain sets that have been nominated for a core outcome domain set and have been encapsulated within a MOD.

Since the revision of the OMERACT Filter to version 2.1 in 2014, and going forward, OMERACT looks at the process of adopting a MOD instrument for a core domain set in a manner similar to any instrument. As a score representing the target domain (i.e., disease activity), it must pass the Filter of having Truth, Discrimination and Feasibility in the intended context of use.

Choosing relevant outcome domains

At OMERACT we consider multi-outcome domain to be a "higher order" domain that is itself comprised of outcome domains. In a MOD these outcome domains are brought together because they can describe a level of the target higher order domain like disease activity, or a response. The first decision in creating or working with a MOD is to decide on the outcome domains that need to go into the MOD (Fig. 1).

Outcome domains can come from the approved core domain set, or they may be unique outcome domains selected based on their relevance to the phenomenon of interest, but it is not something that would be mandated for every clinical trial, and therefore is not in the core domain set. If we take the example of Flare in Osteoarthritis, this higher order domain is made up of five outcome domains: pain, stiffness and swelling from the core domain set for osteoarthritis, but

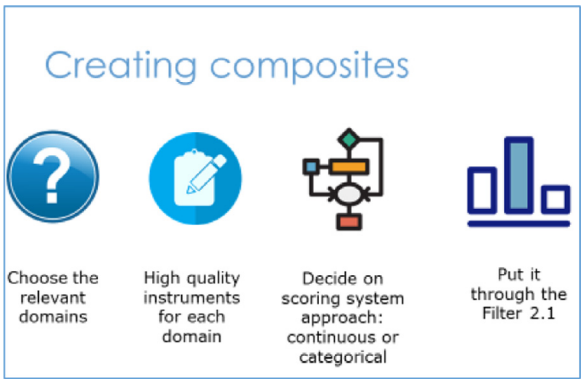

Fig. 1. Process of developing or considering a multi-outcome instrument for an OMERACT core outcome measurement set.

also psychological impact and social impact of the flare from outside of the core domain set.

Outcome domains are chosen by inductive processes to generate all relevant domains – group work or qualitative interviews, and decision making approaches such as a Delphi panel or nominal group processes to determine the final set that should be retained. As for many activities at OMERACT, we highly recommend the input of our patient research partners as well as the experience of clinicians/researchers and other stakeholders. The goal at this stage is to get a clear description of the outcome domains that should be considered for the MOD. For those coming from a core domain set, detailed definitions are likely available in the report for that core domain set. Feel free to supplement that information if necessary to capture the essence of the outcome domain within this specific context. For outcome domains that are outside the core domain set, but deemed relevant for the use of the MOD, we ask for working groups to develop a detailed definition of the outcome domain, including quotes from qualitative interviews. This detailed look at each of the outcome domains often provides information on the way it should be measured. For example, when looking at pain measures, it became clear to a working group at an OMERACT meeting that this could be the intensity of pain, or the frequency of pain or the impact of pain as seen through its effect on daily activities or life roles. This type of insight will make the next step easier.

As an example, the domains included in the CDAI are joint tenderness, swollen joints, patient global sense of disease activity and provider global sense of disease activity which are identified in column 1 of Fig. 2. Such a figure can be useful to report on the component outcome domains of a MOD and the other aspects associated with the multi-outcome domain.

Finding high quality instruments for each of these outcome domains

One of the greatest challenges in developing a multi-outcome domain is deciding on the instruments that will be used for each component outcome domain. An unreliable instrument will create imprecision in the multi-outcome domain score. Similarly choosing an instrument with poor evidence of validity for it to represent a particular domain outcome means that you will be misrepresenting one of the component domain outcomes. Given the challenges facing the interpretation of a multi-outcome domain, it is important that each of the component domain outcomes is measured accurately. Groups can choose instruments that have been passed through the OMERACT Filter as a starting point, and would want to make sure it also has validity evidence supporting its use to encapsulate the core outcome domains – be it disease activity, or flare or clinical response.

As an example, the instruments selected for the domains included in the CDAI are joint tenderness (28 joint count), swollen joints (28 joint count), patient global sense of disease activity (global scale 0–10) and provider global sense of disease activity (global scale 0–10), and are identified in column 2 of Fig. 2.

Weighting the outcome domain instruments in the multi-outcome domain instrument

Having decided on the outcome domains, and the instruments that will be used to measure them, the third column in Fig. 2 will need to be considered. This involves the weight for each outcome domain within the multi-outcome domain.

Composite Index of Disease Activity in Rheumatoid Arthritis - The CDAI (5)			
Outcome Domains	**Instrument used to measure this outcome domain**	**Weight of outcome domain within composite**	**Weighted score (patient score x weight)**
Joint tenderness	28 joint count (0-28)	1	
Swollen joints	28 joint count (0-28)	1	
Patient global sense of disease activity	Global scale (0-10)	1	
Provider global sense of disease activity	Global scale (0-10)	1	
Total score	---	----	SUM:

Fig. 2. Example of a multi-outcome domain: Composite Index of Disease Activity (CDAI) in rheumatoid arthritis.

Table 7
Examples of methods to determine weights for outcome domains in a multi-outcome domain.

Method	Description	Example
Discrete choice experiments - patients	• Patients 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) [20]
Delphi panels – clinician investigator	• 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) [21]
Disability–adjusted life years (DALY)	• 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) [22]

Several recent approaches for weighting the outcome domains in a multi-outcome domain can be considered [15]. These approaches include: patient discrete choice experiments; clinician–investigator Delphi panels; and disability–adjusted life years (DALY) (Table 7).

For a MODs such as the CDAI, there is built-in 'weight' that leads to the MOD score. In using these built-in weights, it assumed that they reflect the importance of each domain to the composite endpoint and the weights in Fig. 2 for CDAI are set to '1'. That is, the implicit assumption is 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 as described in Table 7 may be needed.

The DAS28 is an example of a weighted MOD instrument and provides an example of statistical approaches to determine a 'weighting' of the outcome domain instruments. The development of the DAS28 involved the following steps: principal component analysis was performed resulting in 5 factors (laboratory measures, joint counts, functional status measures, subjective assessments by the patient and globulins); canonical discriminant analysis was used to select the variables that best discriminate between high and low disease activity resulting in 9 variables (pain, hemoglobin, ESR, grip strength, morning stiffness, 44 swollen joint count, RAI, α_2-globuline, β-globuline); further analysis concentrated on the variables identified in the original DAS (Ritchie score, number of swollen joints, ESR and patient global assessment) based on factor analysis, discriminate analysis and multiple regression analysis; and for feasibility, the 2 comprehensive joint counts were replaced by 28-joint counts and the discriminant function was recalculated [23]. There are many such statistical modeling procedures that can be considered when deriving a weighted MOD instrument.

Putting multi-outcome domain instruments through the omeract filter

Multi-outcome domain instruments need evidence of their validity in adequately representing the target domain in the proposed patient population and context of use. At OMERACT this means that it should pass the OMERACT Filter (version 2.2) of have evidence supporting its Truth, Discrimination, and Feasibility in similar patients and in a similar setting. This process is analogous to the process for any other outcome domain instrument be it a patient reported outcome of physical functioning or a clinical observed outcome of joint count. The process involves moving through the four signaling questions in the OMERACT Filter Instrument Selection Algorithm (OFISA). First, ensuring it is a match to the desired target domain, such as disease activity. Second, to ensure it is feasible to use, that is not dependent on excessive costs, travels, burden to the patient or the clinicians/researchers. The third and fourth questions relate to gathering or creating enough relevant, high quality evidence of the MOD instrument's performance on key measurement properties to feel confident that the MOD instrument's score can represent the target domain in these patients. On balance, the evidence for the MOD instrument should show consistent, acceptable performance across low risk of bias studies, with no opposing evidence. To be able to reach this conclusion, a thorough or systematic review of the literature is required and any gaps in measurement property evidence should be identified and new evidence created by a measurement property study in order to fill the gap.

OFISA follows a stepwise decision making process across the four signaling questions and at each step a rating is given: RED = evidence against, stop; AMBER = some concerns, but go ahead; GREEN = met all criteria, go ahead or WHITE = no evidence available. If at any step a "red" rating is encountered (i.e., its' content does not match your conceptualization of the target domain and what should be found in an ideal MOD for that domain), the working group should consider stopping, setting aside that MOD and moving on to another MOD. AMBER and GREEN ratings encourage ongoing consideration and moving to the next signaling question albeit with some caution. WHITE means an absence of evidence, something that might be encountered in a MOD. In this case the group may decide to create evidence to fill in any gaps. In the end, the working group will have a rating for each measurement property and a completed summary of measurement properties. The OMERACT algorithm for the final rating for the instrument will be applied (as described in the handbook) and the results presented first to the Technical Advisory Group for review and then to the OMERACT community for ratification.

Composite outcome domain (COD) instrument development at OMERACT

For the development of the COD, the same four steps as for MOD should be followed. Step 1, on choosing relevant outcome domains, is similar with outcome domains being selected from the approved core domain set, or they may be unique outcome domains selected based on their relevance to the phenomenon of interest. Step 2, on finding high quality instruments for each of these outcome domains, is usually straightforward with the scoring simply being an indicator variable for each component outcome domain occurring versus not occurring, although the determination of the occurrence should follow a reliable and valid procedure.

Step 3, on weighting the outcome domain instruments in the COD instrument is a challenge. For a COD instrument, in which the binary distinction between patients experiencing one or more components of the composite outcome and those experiencing no components, is the practical interpretation when: the component outcome domains are dissimilar in patient importance; the composite outcome is driven by less important components which are observed more frequently and earlier; or the event rates or relative risk reduction vary appreciably across components. One solution is weighting the component outcome domains in the COD instrument which may improve

the relevance and interpretation of composite outcome analysis. For example, the COD Major Adverse Cardiac and Cerebrovascular Events (MACCE) consists of the outcome domains death, stroke, non-fatal myocardial infarction and repeat revascularization. The outcome d*omain instrument for each domain outcome was 1 (outcome occurred) and 0 (outcome did not occur).* In this setting, each outcome domain was given the same weight, and a binary composite outcome instrument was considered for MACCE taking the value 1 if any one of death, stroke, non-fatal myocardial infarction or repeat revascularization occurred or 0 if none occur. Using a discrete choice experiment and conjoint analysis, Tong and colleagues derived relative weights for the outcome domains and found that the risk of death was most important (relative weight 0.23), followed by stroke (0.18), myocardial infarction (0.14) and repeat revascularization (0.11). [20] The authors concluded that "using a weighted composite endpoint increases the validity of statistical analyses and trial conclusions." [20] The potential of weighting in this setting has been long recognized, but the availability of satisfactory procedures for determining the weights was also recognised. [24, 25, 26] More recent application of methods that are feasible have been proposed (Table 5). [20, 21, 22]

For the last, Step 4, of putting the COD instrument through the OMERACT Filter, assessing the content and feasibility of a COD may not fit easily with suggestions made for multi-item scales of a single domain. A framework for evaluating the COD instrument may be based on the clinical utility (part of Feasibility in OFISA) of a COD decided upon for medical decision-making. Three evaluation criteria were identified by Montori in 2005 in order to avoid misleading conclusions about composite outcomes. [27] These criteria have been expanded on over the years, and, in particular, is that by McCoy in 2018. [28] We adapt here these expanded criteria for use with binary and continuous composite outcome domains.

1 ***The component outcome domains of the COD must be of similar clinical importance to patients.*** Interpreting the meaning of a COD is complicated when component outcome domains with a wide variation in clinical importance to the patient are combined in a COD. The usefulness of the COD increases as the difference in importance to the patient between the most and least important component outcome domains decreases.
2 ***The frequency of the occurrence of the component outcome domains over the same time period must be similar; otherwise, the effect on the COD will be largely determined by the predominant event.*** Interpreting the meaning of a COD is complicated when component outcome domains occur with a wide variation in frequency between the most and least patient-important component outcome domains are combined. If the more important component outcome domains occur with far less frequency than the less important ones, the COD becomes less informative

1 **The effect of the intervention must be similar for each component outcome domain of the COD.** *The effect estimates of the component outcome domains are similar and precise*: Similar treatment effects among the component outcome domains leads to increased confidence in using the COD. The effect estimates under consideration will vary by the measurement level of the outcomes under consideration, namely: risk ratio, risk difference, odds ratio (for binary outcomes); mean ratio, mean difference, standardized mean difference (continuous outcomes); hazard ratio (time to event outcomes); rate ratios (count outcomes).

The underlying biology of the component outcome domains are similar enough that their effect estimates are expected to be similar: The stronger the biologic rationale for the reason an intervention should have a particular effect on the component outcome domains, then the more likely that the composite outcome accurately represents the overall effect of the intervention.

Discussion

The conception and development of MODs and CODs can be quite complex. We have taken the division of CODs, that primarily combine outcome domain instruments using Boolean logic, and MODs, that primarily use arithmetic approaches. However, there are many instances in which a MOD–COD hybrid approach may be needed. In particular, relapse is a composite construct which tends to take an AND structure in its outcome domain definition (for example it might be framed as: the patient has previously been in remission AND some measure of disease activity is higher than a certain level AND in the opinion of investigator that this is due to disease and not another cause). If a patient needs to have been in remission before they go into relapse, then remission needs to be defined, and remission also is a composite. So in fact relapse is a composite of a composite. This means that the concept of remission must be developed and validated before attempting relapse, if prior remission is one of the components of relapse. The trend towards complexity is a concern. In some instances, feasibility in using composites could be so complicated that special training may be needed on their scoring. The Systemic Lupus Erythematosus Disease Activity Instrument (SLEDAI) is such a complex composite and the Glucocorticoid Toxicity Index is a numerical scale which takes a weighted MOD structure based on a discrete-choice experiment among clinicians, and attempts to score both reversible and irreversible aspects of toxicity.

The OMERACT Technical Advisory Committee (TAG) intends to maintain a website that will catalogue the various MODs and CODs that currently exist and those that are subsequently develop. The breadth and depth of these examples will be instructive for future development of other composite outcomes and provide a platform on which a more in-depth framework for composite outcomes delineated.

The first steps in the development of MOD or COD is the identification of the core outcome domains and selection of the corresponding instruments assessing the outcome domains. The next important step is the determination of the appropriate construct of the instrument for the MOD or COD for combining the outcome domains. The resulting MOD or COD instrument must then be put through the OMERACT Filter. The final step is choosing the appropriate statistical and reporting approach. Although the primary analysis is the global assessment of the MOD or COD, the heterogeneity of treatment effects across the component outcome domains needs to be considered. When heterogeneity is detected or when inference on individual components is desired, individual component outcome domain analyses are needed. Strategies on reporting and statistical methods to meet the practical challenges of using MODs and CODs are available [2, 15, 16, 17, 18, 19]. Also, guides for interpreting and applying the results of studies that report composite endpoints in clinical practice are available [29]. Composite outcomes are not without their controversy. We have highlighted some of the challenges in their development, interpretation and reporting. Recently McKenna and Heaney have been highly critical in the use of composite outcomes in clinical trials, concluding that composite measures fail to apply measurement theory and as a result may produce invalid and misleading scores [30]. Care must be exercised to help ensure that proper development methodology is followed, transparency is maintained and misleading results are minimized. The TAG plans to provide further guidance on these issues and mitigating these pitfalls, as well as providing instruction and training related to the development of composite outcomes. Comprehensive guidelines will be made available in the OMERACT Handbook on these reporting, interpretation and statistical considerations related to MODs and CODs.

A number of OMERACT working groups have completed the core domain identification and selection and some are now considering the instrument selection phase. After choosing the relevant outcome domains and high quality instruments for assessing each outcome domain, the working group may be concerned that the impact of the

disease and the response to therapy may not just be one of these outcome domains but actually a combination. This is the role of a MOD and COD, bringing together scores on each outcome domain into a single outcome score. As this is a novel methodology being formally considered by OMERACT, the Technical Advisory Group will develop a more fulsome guidance on developing MODs and CODs and training materials in preparation for those working groups who will be completing OFISA. The guidance will include step-by-step methods for forming, analyzing and reporting the MOD or COD. The training will include an online information repository, videos, webinars and workshops. As for other initiatives in OMERACT, we will seek feedback from OMERACT working groups who complete the development of their MOD/COD, which will then be incorporated into the further development and refinement of the guidance and training.

Declaration of Competing Interest

GAW, DB, PT are members of the executive of OMERACT, an organization that develops outcome measures in rheumatology and receives arms-length funding from 8 companies. The other authors have no conflict of interest relevant to the content of this study.

References

1 Multiple Endpoints in Clinical Trials. Guidance for industry, draft guidance. U.S. Department of Health and Human Services, Food and Drug Administration; January 2017 http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

2 Guideline on clinical investigation of medicinal products for the treatment of rheumatoid arthritis. European medicines agency, CPMP/EWP/556/95 Rev. 2, committee for medicinal products for human use, December 2017.

3 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. Clin Exp Rheumatol 2005;23(Suppl. 39):S100–8.

4 Salaffi F, Ciapett A. Clinical disease activity assessments in rheumatoid arthritis. Int J Clin Rheumatol 2013;8(3):347–60.

5 Ibrahim F, Tom BDM, Scott DL, Prevost AT. A systematic review of randomised controlled trials in rheumatoid arthritis: the reporting and handling of missing data in composite outcomes. Trials 2016;17:272. doi: 10.1186/s13063-016-1402-5.

6 Solomon DH, Reed G, Kremer JM, Curtis JR, Farkouh ME, Harrold LR, Hochberg MC, Tsao P, Greenberg J. Disease activity in rheumatoid arthritis and the risk of cardiovascular events. Arthritis Rheumatol 2015;67(6):1449–55.

7 Burmester GR, Nash P, Sands BE, Papp K, Stockert L, Jones TV, Tan H, Madsen A, Valdez H, Cohen SB. Adverse events of special interest in clinical trials of rheumatoid arthritis, psoriatic arthritis, ulcerative colitis and psoriasis with 37 066 patient-years of tofacitinib exposure. RMD Open 2021;7:e001595.

8 Rostom A, Muir K, Dube C, Lanas A, Jolicoeur E, Tugwell P. Prevention of NSAID-related upper gastrointestinal toxicity: a meta-analysis of traditional NSAIDs with gastroprotection and COX-2 inhibitors. Drug Healthc Patient Saf October 2009.

9 Chan FKL, Cryer B, Goldstein JL, Lanas A, Peura DA, Scheiman JM, Simon LS, Singh G, Stillman MJ, Wilcox CM, Berger MF, Breazna A, Dodge W. A novel composite endpoint to evaluate the gastrointestinal (GI) effects of nonsteroidal antiinflammatory drugs through the entire GI tract. J Rheumatol 2010;37:167–74.

10 Tugwell P, Judd MG, Fries JF, Singh G, Wells GA. Powering our way to the elusive side effect: a composite outcome 'basket' of predefined designated endpoints in each organ system should be included in all controlled trials. J Clin Epidemiol 2005;58:785–90.

11 Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, Day R, Ferraz MB, Hawkey CJ, Hochberg MC, Kvien TK, Schnitzer TJ, VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR study group. N Engl J Med 2000;343:1520–8.

12 Silverstein FE, Faich G, Goldstein JL, Simon LS, Pincus T, Whelton A, Makuch R, Eisen G, Agrawal NM, Stenson WF, Burr AM, Zhao WW, Kent JD, Lefkowith JB, Verburg KM, Geis GS. Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: a randomized controlled trial. Celecoxib long-term arthritis safety study. JAMA 2000;284:1247–55 Bombardier.

13 Lampropoulos CE, Orfanos P, Bournia V-K, Karatsourakis T, Mavragani C, Pikazis D, Manoussakis MN, Tzioufas AG, Moutsopoulos HM, Vlachoyiannopoulos PG. Adverse events and infections in patients with rheumatoid arthritis treated with conventional drugs or biologic agents: a real world study. Clin. Exp. Rheumatol 2015;33:216–24 [G1].

14 Felson DT, Smolen JS, Wells G, et al. American college of rheumatology/European league against rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Ann Rheum Dis 2011;70:404–13.

15 Ferreira-González I, Permanyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coello P, Walter SD, Guyatt GH. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. J Clin Epidemiol 2007;60:651–7 discussion 658–662.

16 Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. Br Med J 2010;341:c3920.

17 Freemantle N, Calvert M. Weighing the pros and cons for composite outcomes in clinical trials. J Clin Epidemiol 2007;60:658–9.

18 Lubsen J, Kirwan BA. Combined endpoints: can we use them? Stat Med 2002;21:2959–70.

19 Chi GY. Some issues with composite endpoints in clinical trials. Fundam Clin Pharmacol 2005;19:609–19.

20 Tong BC, Huber JC, Ascheim DD, Puskas JD, Ferguson BT, Blackstone EH, Smith PK. Weighting composite endpoints in clinical trials: essential evidence for the heart team. Ann. Thorac. Surg. 2012;94:1908–13.

21 Armstrong PW, Westerhout CM, Van de Werf F, Califf RM, Welsh RC, Wilcox RG, Ja Bakal. 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;161:848–54.

22 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 2011;42:1722–9.

23 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(Suppl. 85):S65–74.

24 DeMets DL, Califf RM. Lessons learned from recent cardiovascular clinical trials: part I. Circulation 2002;106:746–51.

25 Hallstrom AP, Litwin PE, Weaver WD. A method of assigning scores to the components of a composite outcome: an example from the MITI trial. Control Clin Trials 1992;13:148–55.

26 Neaton JD, Gray G, Zuckerman BD, Konstam MA. Key issues in endpoint selection for heart failure trials: composite end points. J Card Fail 2005;11:567–75.

27 Montori VM, Permanyer-Miralda G, Ferreira-Gonzales I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite endpoints in clinical trials. BMJ 2005;330(7491):594–6.

28 McCoy E. Understanding the use of composite endpoints in clinical trials. Western J Emerg Med 2018;19(4).

29 Ferreira-Gonzales I, Montori VM, Busse JW, Sch€unemann HJ, Jaeschke R, Deveraux PJ, Permanyer-Miralda G, Guyatt G. Composite End Points (Chapter 12.4). In: Guyatt G, Rennie D, Meade MO, Cook DJ, editors. Users' guides to the medical literature: a manual for evidence-based clinical practice. 3rd ed USA: McGraw-Hill Education; 2015.

30 McKenna SP, Heaney A. Composite outcome measurement in clinical research: the triumph of illusion over reality? J Med Econ 2020;23(10):1196–204. doi: 10.1080/13696998.2020.1797755.