



Published in final edited form as:

*Pain*. 2021 July 01; 162(7): 1899–1905. doi:10.1097/j.pain.0000000000002188.

## Composite outcomes for pain clinical trials: considerations for design and interpretation

Jennifer S Gewandter<sup>1</sup>, Michael P McDermott<sup>2</sup>, Scott Evans<sup>3</sup>, Nathaniel P Katz<sup>4,5</sup>, John D Markman<sup>6</sup>, Lee S Simon<sup>7</sup>, Dennis C Turk<sup>8</sup>, Robert H Dworkin<sup>1</sup>

<sup>1</sup>Department of Anesthesiology and Perioperative Medicine, University of Rochester, Rochester, NY, United States.

<sup>2</sup>Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, United States.

<sup>3</sup>Department of Biostatistics and Bioinformatics, George Washington University, Rochville, MD, United States.

<sup>4</sup>Analgesic Solutions, Natick, MA, United States.

<sup>5</sup>Tufts University, Boston, MA, United States.

<sup>6</sup>Department of Neurosurgery, University of Rochester, Rochester, NY, United States.

<sup>7</sup>SDG, LLC, Cambridge, MA, United States.

<sup>8</sup>Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, WA, United States.

### 1. Introduction

Composite outcomes combine two or more constructs into a single measure that can be used to evaluate a treatment's efficacy, tolerability, or safety, or their combination (i.e., risk/benefit). The main potential advantage of such outcomes is that they can provide a more comprehensive assessment by including multiple domains that are important to the individuals being studied, rather than a single factor that may not adequately describe their experience. Such outcomes may better inform patient, clinician, researcher, regulator, and payor decisions. One way to think about this advantage is to compare 3 hypothetical participants in a trial that evaluates 2 domains: pain intensity and function (Figure 1). Participant 1 experiences a large improvement in pain intensity, but their function worsens, possibly due to treatment side effects. Participant 2 experiences a small improvement in pain intensity and a large improvement in function. Participant 3 experiences large improvements in both domains. If the primary outcome of the clinical trial was pain intensity, participants 1 and 3 would have the same outcome, but their experiences were quite different. This illustration suggests that the outcome for each participant might be better assessed using a composite measure that incorporates multiple important domains. In other words, the

composite outcome may be more clinically relevant than pain intensity alone. In addition, unlike standard separate analyses of multiple outcomes, composite outcomes can incorporate information about the relationships among the separate domains.

Composite outcomes can be particularly useful when the included domains assess both efficacy (i.e., benefits) and harm (i.e., risk) because the efficacy and harms of a treatment can be correlated within patients and represent positive and negative effects of the treatment (i.e., trade-offs). Typically in clinical trials, efficacy and harms are evaluated separately and the results of the separate outcomes are qualitatively combined to inform conclusions concerning the risk/benefit profile. Alternatively, a composite outcome could be used to assess the efficacy/harm trade-off within each participant and allow for a quantitative evaluation of the risk/benefit profile.

The approach used to construct the composite outcome will affect the degree to which it improves clinical relevance or is responsive to interventions. This article will review general considerations for developing composite outcomes, the advantages and limitations of various methods used to generate composite outcomes, and considerations for interpreting and disseminating trial results with composite outcomes. Examples of published composite outcomes will be used to illustrate these concepts.

## 2. General considerations

In order to ensure that a composite outcome is clinically relevant, it is imperative that previous research and patient and clinician input inform the selection of the included domains for a particular condition and “response” definitions or ranking of the joint outcomes, when applicable. This practice will minimize the potential for overestimation of treatment effects by composite outcomes that include less relevant domains that are strongly affected by a treatment. Concept elicitation interviews and focus groups can be used to obtain broad input for designing the outcome, and subsequent interviews or surveys should be used to further assess the content validity of the resulting outcome.[24] The inclusion of patients who have experienced improvement with pain treatments in these interviews and focus groups should be prioritized so that their opinions are based on actual rather than anticipated experiences. In addition, secondary analyses of previous trial data can be used to investigate the associations between the composite outcomes and global improvement measures. Secondary analyses can also be used to evaluate the relative responsiveness of different composite outcomes to effective treatments and should ideally use multiple data sets to evaluate consistency. The major goal of such secondary analyses would be to determine whether clinical relevance can be increased without sacrificing responsiveness to effective treatments.

Finally, when weighing the potential benefits of increased clinical relevance of a composite outcome with potential decreases in responsiveness compared to single domain outcomes, investigators should also consider the study phase and goal. For example, in early phase proof-of-concept trials with smaller sample sizes, it may be beneficial to prioritize responsiveness over maximum clinical relevance to avoid premature termination of a drug development plan.

Even with careful design of a composite outcome, interpretation of clinical trial results using these outcomes can be challenging. For example, suppose that there is a treatment-associated benefit on a subset of the domains included in a composite outcome, but there is a treatment-associated worsening on another domain. Many readers may interpret a positive result on the composite outcome to mean that the treatment positively affects ALL of the included domains. Thus, articles reporting the results of clinical trials with composite outcomes should be carefully worded to avoid the implication that a significant effect on the composite outcome implies a significant effect on all of the outcome domains in the composite, unless the composite is specifically designed for this purpose (e.g., requirement to improve on all domains).[10] Both the European Medicines Agency[6] and the U.S. Food and Drug Administration[23, 24] recommend analyzing the individual domains of the composite outcome to examine how each domain is affected by the treatment. Note that although these analyses are necessary to fully inform the interpretation of the results and should be presented in published trial reports, claims of significance related to these secondary analyses of individual outcomes are not justified unless they are incorporated in a pre-specified analysis plan that statistically adjusts for multiplicity.[22]

It is also important to recognize that even when included domains, method of outcome construction, and “response” definitions are based on previous research and stakeholder input, the priorities reflected in the outcomes will not necessarily apply to all participants. However, this is also the case for “response” definitions for single domain outcomes and when interpreting clinically meaningful differences for individual patients and between groups in clinical trials. Additionally, certain composite outcomes can be personalized for each participant at the beginning of the trial similarly to the approach suggested by the FDA for migraine trials in which participants pre-specify their most bothersome symptom to serve as their personalized primary outcome.[25] See Section 3.3 for an example of an approach to composite outcomes that can be personalized.

Finally, as with all outcomes, research to determine what would be considered a clinically meaningful group difference in a composite outcome is necessary to inform trial design (sample size) and the interpretation of the potential impact of treatments.[4]

### 3. Types of composite outcomes and implications for their interpretation

This article discusses composite outcomes that combine two or more different efficacy or safety domains assessed separately. This type of composite outcome is particularly beneficial when the domains are assessed using different modalities and for which a valid, single patient reported outcome (PRO) measure would be difficult to develop (e.g., self-report pain combined with objectively-assessed activity[21], adverse events, or rescue analgesic consumption). This article will not address composite PROs that include multiple items assessed using the same scale (e.g., the Brief Pain Inventory[3]).

#### 3.1. “Responder”-based methods

“Responder”-based composite outcomes group participants into discrete categories based on pre-specified criteria to define improvement or worsening (not necessarily response to an intervention[17]) during a trial. These types of outcomes are often dichotomous

(i.e., “responders” vs. “non-responders”), but can also be ordinal (e.g., worsening vs. no change/little improvement vs. moderate improvement vs. substantial improvement). As with most composite outcome methods, a “responder”-based composite outcome that requires a certain degree of improvement on only one of multiple domains can be driven by one domain alone (e.g., requires a 30% improvement in pain OR function). Such a result can be misinterpreted as an improvement on all domains of the composite outcome (see Section 2). On the contrary, a “responder”-based composite outcome that requires improvement on all included domains may not demonstrate improvement if a treatment has a clinically meaningful benefit on a subset of the included domains. For example, if a composite outcome requires 30% improvement in pain AND 30% improvement in physical activity, a treatment that is highly effective at improving pain might not affect the composite outcome unless the treatment was paired with an intervention focused on increasing activity. Thus, these endpoints could lead to underestimating the benefit of potentially effective treatments.

Composite “responder”-based outcomes can also be designed to require an improvement on at least one domain and no worsening on other domains. This approach would prevent a conclusion of benefit if any domain worsens for a particular participant, and would decrease the chances of missing an important therapeutic benefit on the most important domain because a second domain did not improve. It is important to avoid criteria to define “responders” that will rarely be achieved even with highly effective treatments because these outcomes might not be able to distinguish between effective treatments and placebo, causing underestimation of potential treatment benefits. One main limitation of dichotomous “responder”-based outcomes is that they may not capture potentially important finer gradations of improvement and tend to provide less statistical power than continuous outcomes.[20]

### 3.2. Z-score based methods

Standardized scores or z-scores are sometimes used to create measures that are unit-less and, hence, easier to combine across domains. A z-score relates a participant’s outcome on a given domain to the mean outcome for that domain in a sample. It is computed as  $Z = (\text{observed outcome} - \text{mean}) / \text{standard deviation (SD)}$ . The mean and SD are typically obtained from external sources (e.g., normative datasets of patients or healthy controls), but the baseline mean and SD of the study sample can also be used. The z-scores for each domain are then averaged, possibly with different weights being assigned to the different domains based on patient preferences or other clinical considerations, to generate a composite score. This approach is useful when individual domains cannot be assessed on the same scale. It does not require investigators to identify clinically meaningful cut-offs for the continuous variables, and it does not sacrifice information as do dichotomous “responder”-based outcomes. This approach is perhaps best suited for continuous outcomes that are interval-scaled and approximately normally distributed. To our knowledge, composite outcomes using this methodology have yet to be proposed for chronic pain. A composite measure for multiple sclerosis provides an example of this methodology.[5]

### 3.3. Rank-based methods

The O'Brien Rank-Sum Method ranks participants in each relevant domain and then calculates the sum of the ranks for each participant.[14] This approach has similar advantages to the z-score method in that it allows combination of outcomes measured on different scales. It is useful for instances when the individual outcome variables are not likely to be normally distributed. An adaptation of this approach that combines participant ranks based on their pain scores and their opioid consumption has been proposed by Silverman et al. for acute pain trials[18] and was utilized in a randomized clinical trial for post-hysterectomy pain.[11]

Desirability of Outcome Ranking (DOOR) is a rank-based method in which investigators use input from patients and clinicians to develop a prespecified ranking scheme of possible joint outcomes of the included domains. If the “perceived distance” between rankings is not uniform the DOOR outcome can be quantified using “partial credit” where the desirable category is scored 100 and the least desirable category is scored 0 with partial credit given to intermediate rankings.[7] This partial credit can be identified based on summarized input from stakeholders, including patients and clinicians, but could also be identified by each participant prior to the trial in order to personalize this outcome according to each participant’s priorities [26] (Figure 2A). A variation of the DOOR method can include continuous outcomes for a domain.[8] For example, a ranking scheme could involve first separating participants into 3 categories depending on whether their pain improved by < 30%, 30–49%, or ≥ 50%. Then within those groups, participants would be ranked based on physical activity levels as measured by an accelerometer. The prioritization of pain relief in this ranking structure implies that regardless of the change in a participant’s activity level, if their pain does not improve by at least 30%, the participant should be near the bottom of the ranking [8] (Figure 2B).The advantage of the DOOR method is it summarizes the experience of the person by joint outcomes that have been ranked *a priori* based on patient and clinician input or personalized by each participant in the trial. The DOOR approach provides the probability that a participant randomly selected from the active group has a more desirable outcome than a participant in a control group, providing a straightforward, meaningful interpretation.[7] This approach has yet to be developed for pain conditions.

## 4. Risk-benefit composite outcomes

Perhaps one of the most important concepts to assess using a composite outcome is the benefit-risk profile. Only by assessing the balance of benefits and risks within an individual participant can the outcome capture the relationship between the two outcome domains.[7, 8] Two groups have proposed benefit-risk measures that group participants in two-way tables based on the degree of pain improvement and degree (or presence) of adverse events. [2, 13] These outcomes could be adapted using a DOOR method to create finer gradations in categories with pre-specified prioritization of the benefit-risk trade-offs and potentially increase their ability to distinguish between treatments. A third group proposed a composite benefit-risk outcome defined as the number of days with at least 30% improvement in pain and no or mild adverse events.[12]

It is important to note that, when using a placebo control, composite outcomes that incorporate risk and benefit may be less responsive to treatments with considerable risks (e.g., opioids) than outcomes that include only pain intensity, but these outcomes may be better suited to evaluate the overall utility of such treatments. In addition, outcomes that incorporate risk and benefit may better detect differences between two treatments that provide similar pain relief, but have different degrees of adverse effects.

## 5. Progress and prospects

Table 1 presents examples of composite outcomes that have been developed for various pain conditions. The referenced studies provide good methodological examples for how to develop and evaluate such outcomes, including the steps taken to develop clinically relevant candidate outcomes and secondary analyses of existing trial data to assess the relative responsiveness and clinical relevance (e.g., via associations with global improvement measures) of potential composite outcomes.

## 6. Conclusions

Although randomized clinical trials of pain treatments assess many important outcome domains, pain intensity is often the only domain used in the primary efficacy assessment. A composite outcome is one approach that can be used to incorporate multiple relevant domains in the primary conclusion regarding whether a treatment provides benefit. Such outcomes arguably can be more clinically meaningful, especially when they assess competing domains that can affect one another. Considering the potential harms of common pain treatments and the multi-dimensional effects of pain on patients' lives, developing composite outcomes with good content and construct validity will advance pain research and facilitate more informed decisions regarding treatment utility. Efforts to develop novel composite outcomes should pay careful attention to how the design of such measures affects their clinical relevance, interpretation, and responsiveness in clinical trials. Finally, professional organizations should work toward obtaining consensus on accepted composite outcome measures to utilize across chronic pain trials.[16]

## Acknowledgements

This article was reviewed and approved by the Executive Committee of the Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks (ACTTION) public-private partnership with the United States Food and Drug Administration (FDA). Financial support for this project was provided by the ACTTION public-private partnership which has received research contracts, grants, or other revenue from the FDA, multiple pharmaceutical and device companies, philanthropy, and other sources. The views expressed in this article are those of the authors and no official endorsement by the FDA or the pharmaceutical and device companies that provided unrestricted grants to support the activities of the ACTTION public-private partnership should be inferred.

## Disclosures

In the past 36 months, Jennifer Gewandter has received research grants from the NIH and consulting income from MundiPharma, Disarm Therapeutics, Asahi Kasei Pharma, Magnolia Neurosciences, Orthogonal Neurosciences, Science Branding Communications, and SK Life Science. Michael P. McDermott has been supported in the past 36 months by research grants from NIH, FDA, NYSTEM, SMA Foundation, Cure SMA, Friedreich's Ataxia Research Alliance, Muscular Dystrophy Association, ALS Association, and PTC Therapeutics, has received compensation for consulting from Fulcrum Therapeutics, Inc. and NeuroDerm, Ltd., and has served on Data and Safety Monitoring Boards (DSMBs) for NIH, AstraZeneca, Eli Lilly and Company, Catabasis Pharmaceuticals, Inc., Vaccinex, Inc., Cynapsus Therapeutics, Voyager Therapeutics, and Prilenia Therapeutics Development, Ltd.



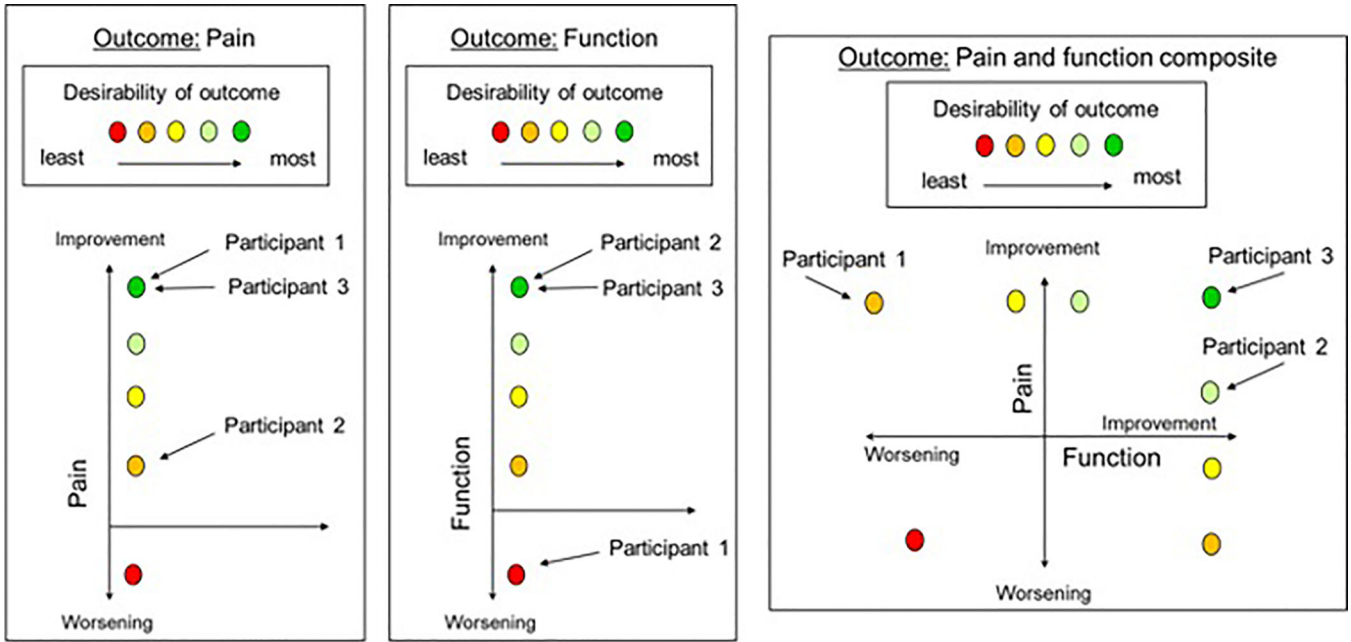
Scott Evans has no conflicts related to this work to declare. Nat Katz has no conflicts related to this work to declare. John Markman reports non-financial support from Pfizer Inc and Eli Lilly, during the conduct of the study; grants and other from Clexio, grants from Pfizer, other from Teva, compensation for consulting or advisory board participation from Quark, other from Biogen (Convergence), other from Nektar, other from ENDO, other from Immune Pharma, other from Chromocell, other from Collegium, other from Purdue, other from Novartis, grants and other from Depomed, other from Allergan, other from Sanofi, other from Aptinyx, other from Diaachi Sanyko, other from Plasmasurgical, other from Grunenthal, other from Clexio Bioscience, other from Editas Medicine, other from Trevena, other from Inspirion, other from Merck, other from Esteve Pharmaceuticals, other from Tremeau Pharmaceuticals, other from Sophren Pharmaceuticals, other from YellowBlack Corporation outside this work. Lee Simon is a drug development consultant, but has no conflicts to declare related to the topics discussed in this manuscript. In the past three years, Dennis C. Turk has received research grants and contracts from the US Food and Drug Administration and US National Institutes of Health and has received compensation for consulting on clinical trials and patient preferences from Eli Lilly and Company, GlaxoSmithKline, and Pfizer. Robert H. Dworkin, PhD, has received in the past 5 years research grants and contracts from the US Food and Drug Administration and the US National Institutes of Health, and compensation for serving on advisory boards or consulting on clinical trial methods from Abide, Acadia, Adynxx, Analgesic Solutions, Aptinyx, Aquinox, Asahi Kasei, Astellas, AstraZeneca, Biogen, Biohaven, Boston Scientific, Braeburn, Celgene, Centrexion, Chromocell, Clexio, Collegium, Concert, Coronado, Daiichi Sankyo, Decibel, Dong-A, Editas, Eli Lilly, Eupraxia, Glenmark, Grace, Hope, Hydra, Immune, Johnson & Johnson, Lotus Clinical Research, Mainstay, Medavante, Merck, Neumentum, Neurana, NeuroBo, Novaremed, Novartis, NSGene, Olatec, Periphagen, Pfizer, Phosphagenics, Quark, Reckitt Benckiser, Regenacy (also equity), Relmada, Sanifit, Scilex, Sennur, SK Life Sciences, Sollis, Spinifex, Syntrix, Teva, Thar, Theranexus, Trevena, Vertex, and Vizuri.

## References

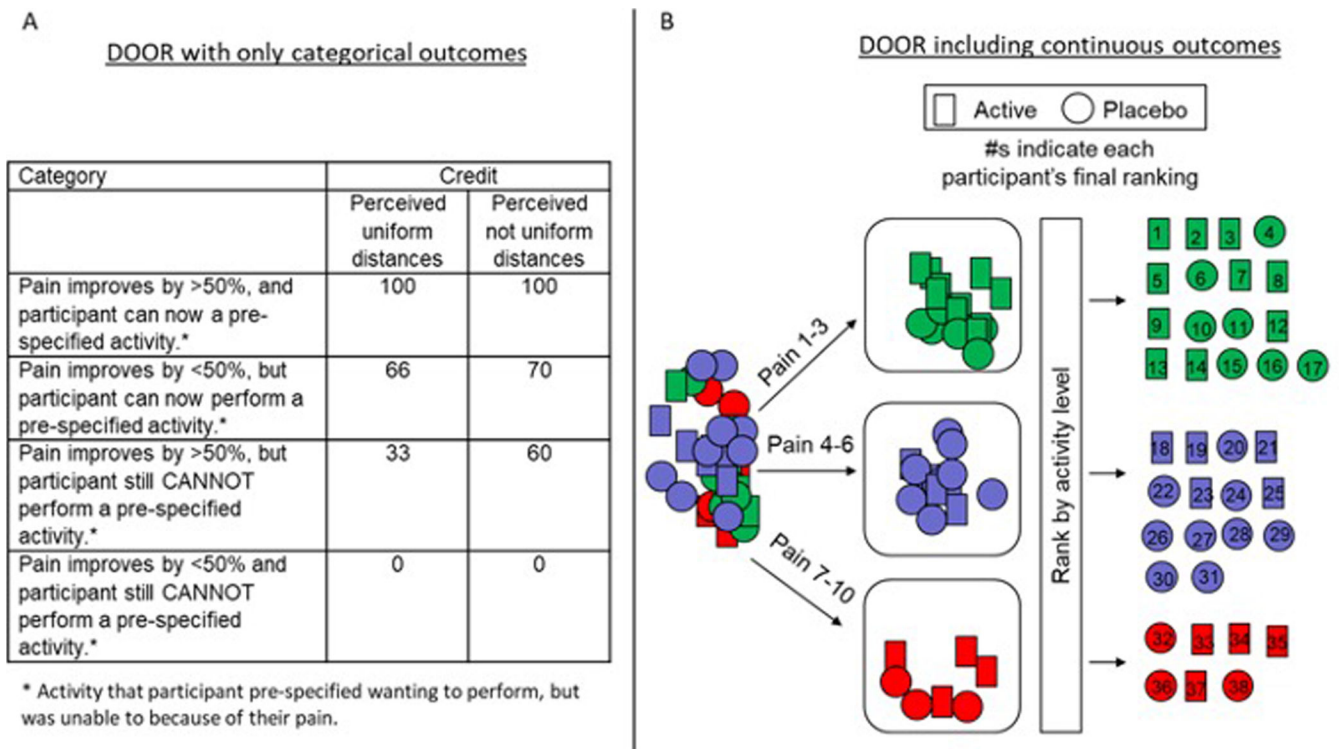
1. Arnold LM, Williams DA, Hudson JI, Martin SA, Clauw DJ, Crofford LJ, Wang F, Emir B, Lai C, Zabolocki R, Mease PJ. Development of responder definitions for fibromyalgia clinical trials. *Arthritis Rheum* 2012;64:885–94. [PubMed: 21953205]
2. Boers M, Brooks P, Fries JF, Simon LS, Strand V, Tugwell P. A first step to assess harm and benefit in clinical trials in one scale. *J Clin Epidemiol* 2010;63:627–32. [PubMed: 19800197]
3. Cleeland C Brief Pain Inventory, 1991.
4. Cook JA, Julious SA, Sones W, Hampson LV, Hewitt C, Berlin JA, Ashby D, Emsley R, Fergusson DA, Walters SJ, Wilson ECF, MacLennan G, Stallard N, Rothwell JC, Bland M, Brown L, Ramsay CR, Cook A, Armstrong D, Altman D, Vale LD. DELTA(2) guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *Bmj-British Medical Journal* 2018;363.
5. Cutter GR, Baier ML, Rudick RA, Cookfair DL, Fischer JS, Petkau J, Syndulko K, Weinshenker BG, Antel JP, Confavreux C, Ellison GW, Lublin F, Miller AE, Rao SM, Reingold S, Thompson A, Willoughby E. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain* 1999;122 (Pt 5):871–82. [PubMed: 10355672]
6. European Medicines Agency. Guideline on multiplicity issues in clinical trials. 2016; [Accessed 11-5-2020] [https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-multiplicity-issues-clinical-trials\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-multiplicity-issues-clinical-trials_en.pdf)
7. Evans SR, Follmann D. Using Outcomes to Analyze Patients Rather than Patients to Analyze Outcomes: A Step toward Pragmatism in Benefit:risk Evaluation. *Stat Biopharm Res* 2016;8:386–393. [PubMed: 28435515]
8. Evans SR, Rubin D, Follmann D, Pennello G, Huskins WC, Powers JH, Schoenfeld D, Chuang-Stein C, Cosgrove SE, Fowler VG Jr., Lautenbach E, Chambers HF. Desirability of Outcome Ranking (DOOR) and Response Adjusted for Duration of Antibiotic Risk (RADAR). *Clin Infect Dis* 2015;61:800–6. [PubMed: 26113652]
9. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, Katz LM, Lightfoot R Jr., Paulus H, Strand V, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;38:727–35. [PubMed: 7779114]
10. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA* 2003;289:2554–9. [PubMed: 12759327]
11. Gilon I, Orr E, Tu D, O’neill JP, Zamora JE, Bell AC. A placebo-controlled randomized clinical trial of perioperative administration of gabapentin, rofecoxib and their combination for

- spontaneous and movement-evoked pain after abdominal hysterectomy. *Pain* 2005;113:191–200. [PubMed: 15621380]
12. Katz NP, Mou J, Trudeau J, Xiang J, Vorsanger G, Orman C, Kim M. Development and preliminary validation of an integrated efficacy-tolerability composite measure for the evaluation of analgesics. *Pain* 2015;156:1357–65. [PubMed: 25867124]
  13. Moore RA, Derry S, Mcquay HJ, Straube S, Aldington D, Wiffen P, Bell RF, Kalso E, Rowbotham MC, Relief AWGOTISIGOSRIP. Clinical effectiveness: an approach to clinical trial design more relevant to clinical practice, acknowledging the importance of individual differences. *Pain* 2010;149:173–6. [PubMed: 19748185]
  14. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics* 1984;40:1079–87. [PubMed: 6534410]
  15. Patel KV, Allen R, Burke L, Farrar JT, Gewandter JS, Gilron I, Katz NP, Markman JD, Marshall SF, Resnick M, Rice ASC, Rowbotham MC, Smith SM, Vanhove GF, Wasan AD, Zhang S, Dworkin RH, Turk DC. Evaluation of composite responder outcomes of pain intensity and physical function in neuropathic pain clinical trials: an ACTION individual patient data analysis. *Pain* 2018;159:2245–2254. [PubMed: 30001225]
  16. Pham T, Van Der Heijde D, Lassere M, Altman RD, Anderson JJ, Bellamy N, Hochberg M, Simon L, Strand V, Woodworth T, Dougados M, Omeract O. Outcome variables for osteoarthritis clinical trials: The OMERACT-OARSI set of responder criteria. *J Rheumatol* 2003;30:1648–54. [PubMed: 12858473]
  17. Senn S, Julious S. Measurement in clinical trials: a neglected issue for statisticians? *Stat Med* 2009;28:3189–209. [PubMed: 19455540]
  18. Silverman DG, O'connor TZ, Brull SJ. Integrated assessment of pain scores and rescue morphine use during studies of analgesic efficacy. *Anesth Analg* 1993;77:168–70. [PubMed: 8317727]
  19. Simon LS, Evans C, Katz N, Bombardier C, West C, Robbins J, Copley-Merriman C, Markman J, Coombs JH. Preliminary development of a responder index for chronic low back pain. *J Rheumatol* 2007;34:1386–91. [PubMed: 17552065]
  20. Snapinn SM, Jiang Q. Responder analyses and the assessment of a clinically relevant treatment effect. *Trials* 2007;8:31. [PubMed: 17961249]
  21. Trudeau J, Van Inwegen R, Eaton T, Bhat G, Paillard F, Ng D, Tan K, Katz NP. Assessment of Pain and Activity Using an Electronic Pain Diary and Actigraphy Device in a Randomized, Placebo-Controlled Crossover Trial of Celecoxib in Osteoarthritis of the Knee. *Pain Practice* 2015;15:247–255. [PubMed: 24494935]
  22. Turk DC, Dworkin RH, Mcdermott MP, Bellamy N, Burke LB, Chandler JM, Cleeland CS, Cowan P, Dimitrova R, Farrar JT, Hertz S, Heyse JF, Iyengar S, Jadad AR, Jay GW, Jermano JA, Katz NP, Manning DC, Martin S, Max MB, Mcgrath P, Mcquay HJ, Quessy S, Rappaport BA, Revicki DA, Rothman M, Stauffer JW, Svensson O, White RE, Witter J. Analyzing multiple endpoints in clinical trials of pain treatments: IMMPACT recommendations. Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials. *Pain* 2008;139:485–93. [PubMed: 18706763]
  23. U.S. Health and Human Services, Food and Drug Administration. Multiple Endpoints in Clinical Trials: Guidance for Industry. 2017. [Accessed 11-5-2020] <https://www.fda.gov/media/102657/download>.
  24. U.S. Health and Human Services, Food and Drug Administration. Guidance for Industry: Patient-reported outcome measures: Use in medical product development to support labeling claims. 2009. [Accessed 11-5-2020] <https://www.fda.gov/media/77832/download>.
  25. U.S. Health and Human Services, Food and Drug Administration. Migraine: Developing Drugs for Acute Treatment Guidance for Industry. 2018; [Accessed 11-5-2020] <https://www.fda.gov/media/89829/download>.
  26. Van Duin D, Lok JJ, Earley M, Cober E, Richter SS, Perez F, Salata RA, Kalayjian RC, Watkins RR, Doi Y, Kaye KS, Fowler VG Jr., Paterson DL, Bonomo RA, Evans S, Antibacterial Resistance Leadership G. Colistin Versus Ceftazidime-Avibactam in the Treatment of Infections Due to Carbapenem-Resistant Enterobacteriaceae. *Clin Infect Dis* 2018;66:163–171. [PubMed: 29020404]





**Figure 1.** Composite outcomes provide a more complete overall assessment of each participant. The figure provides an illustrative example of how 3 participants would be ranked differently regarding the desirability of their outcome (i.e., different colors in the figure) using a composite outcome of pain intensity and physical function vs. each of its two components alone. Note that the color of the dots represents the relative desirability of the outcome for hypothetical participants **within** each outcome and that the desirability of one color, for example yellow, in the pain outcome is not necessarily similar to the desirability of yellow in the function outcome or the composite outcome.



**Figure 2.** Illustration of the DOOR method using (A) only categorical outcomes and (B) incorporating continuous outcomes. In (A) “Partial credit” can be used when the perceived distance between outcomes is not uniform. These credit values can be identified based on research with various stakeholders or identified at the beginning of the study by each participant to personalize the DOOR outcome. Group comparisons can be performed with respect to the partial credit scores (A) or the ranks (B) using an appropriate statistical test. Note that the examples used here are hypothetical (i.e., not informed by stakeholders) and serve only to illustrate the way in which the method is used.

Table.

Examples of published composite outcomes for pain

Composite outcome	Bases for selection of outcome domains or candidate composites	Criteria used to select final composite outcome(s)
<p><b>CLBP (Simon 2007)</b>[19] Response definition 30% improvement in pain <b>AND</b> 30% improvement in PGA <b>AND</b> 20% worsening in RDMQ</p>	<ul style="list-style-type: none"> <li>Literature review</li> <li>Focus groups with patients</li> </ul>	<p>Step 1: Selected from a development database based on (1) effect size, (2) high <math>\chi^2</math> value, and (3) placebo response &gt;25%. Step 2: Ranked using a validation database by <math>\chi^2</math> value and number of trials in which the outcome achieved statistical significance. Step 3: Composite candidate chosen because it was identified as the most clinically meaningful based on the included domains and showed &lt;25% responders in the placebo group of all evaluated clinical trials.</p> <ul style="list-style-type: none"> <li>Analyses were repeated in a second database to investigate consistency.</li> </ul>
<p><b>Neuropathic pain (DPN/PHN) (Patel 2018)</b>[15] Response definition 50% improvement in pain <b>OR</b> 20% improvement in pain <b>AND</b> 30% improvement in physical function</p>	<ul style="list-style-type: none"> <li>Literature review</li> <li>Discussions among co-investigators</li> </ul>	<p>Step 1: Selected outcomes that detected a statistically significant treatment effect in 2 datasets of merged pregabalin RCTs. Step 2: RR of active vs. placebo response were calculated using data from duloxetine and gabapentin studies and examined for consistency. Step 3: Final selection made based on which consistent outcomes from step 3 correlated best with PGIC.</p>
<p><b>Knee and Hip OA (Pham, 2003)</b>[16] Response definition 50% improvement in pain <b>OR</b> 50% improvement in pain function <b>OR</b> Meets at least 2 of the 3 following criteria:</p> <ul style="list-style-type: none"> <li>20% improvement in pain <b>AND</b> absolute change of 10</li> <li>20% improvement in function <b>AND</b> absolute change of 10</li> <li>20% improvement in PGA <b>AND</b> absolute change of 10</li> </ul>	<ul style="list-style-type: none"> <li>Literature review and presentations, discussions and polling of attendees * of OMERACT 3 conference</li> <li>Subsequent discussions with OMERACT scientific committee.</li> </ul>	<p>Step 1: Calculated the % responders in the active and placebo groups for each candidate composite outcome in 2 separate databases of merged OA RCT data. Step 2: Noted that the treatment effect sizes (i.e., % responders in the active group - % responders in the placebo group) were similar regardless of the outcome chosen. Step 3: At the OMERACT 6 conference, workshop participants were asked to vote on which of the outcomes was most appropriate given the fact that they all provided similar effect sizes (79% of participants agreed on the selected outcome).</p>
<p><b>Fibromyalgia (Arnold, 2012)</b>[1] (2 outcomes) Response definitions</p> <p><b>1</b> 30% improvement in pain <b>AND</b> 10% improvement in physical function <b>AND</b> 30% improvement in fatigue or sleep</p> <p><b>2</b> 30% improvement in pain <b>AND</b> 10% improvement in physical function <b>AND</b></p>	<ul style="list-style-type: none"> <li>Consensus at ACR and EULAR.</li> <li>Association of relevant domains with PGIC in RCTs of drugs that are effective for fibromyalgia.</li> </ul>	<p>Step 1: Selected outcomes with 15% response and significant difference between the placebo and active groups in all 12 trials. Step 2: Selected outcomes with RR of response to active drug vs. placebo with a lower limit of the 95% CI of 1.00 in the meta-analysis of trials of all drug types. Step 3: Identified outcomes with largest RRs and highest percentage responders in the active group.</p>

Composite outcome	Bases for selection of outcome domains or candidate composites	Criteria used to select final composite outcome(s)
<p>30% improvement in 2 of the following symptoms:</p> <ul style="list-style-type: none"> <li>Fatigue, sleep, depression, anxiety, cognition</li> </ul>	<ul style="list-style-type: none"> <li>Evaluation of risk-benefit</li> </ul>	<p>Step 1: Identify outcomes that detect a statistically significant difference between a mild and strong opioid (i.e., tapentadol ER and oxycodone CR, respectively).                      Step 2: Evaluate the magnitude of the SES between the 3 treatments (i.e., tapentadol, oxycodone, and placebo).                      Step 3: Evaluate correlation between the composite ETC score and PGIC and Pain intensity (i.e., original primary outcome of the trial) to evaluate the convergent validity of the outcome.</p>
<p><b>Efficacy-tolerability composite (ETC) (Katz, 2015)</b>[12]                      Percent of study days that the subject was classified as an ETC responder (i.e., 20% improvement in pain AND No or mild drug-related AEs)</p> <p><b>Rheumatoid arthritis (Felson, 1995)</b>[9]                      Response definition                      20% improvement in tender joint count AND 20% improvement in swollen joint count                      AND                      20% improvement in at least 3 of the following:                      Pain, PGA, self-assessed disability, acute-phase reactant (ESR or CRP)</p>	<ul style="list-style-type: none"> <li>Considered ACR core set of outcomes</li> <li>Proposed outcomes for which 80% of rheumatologists surveyed identified as representing an improvement when asked about real example participant cases from clinical trials.</li> </ul>	<p>Step 1: Identified outcomes with largest effect sizes (treatment vs. placebo). Step 2: Of the outcomes with the largest effect sizes, selected outcomes with the smallest placebo response.</p> <ul style="list-style-type: none"> <li>Steps 1 and 2 were repeated in data from a separate large clinical trial to assess consistency.</li> </ul> <p>Step 3: Prioritized final candidate outcomes based on which definitions were the easiest to use and which corresponded best with rheumatologists' impressions of improvement in a follow-up survey.</p>

\* Rheumatologists, epidemiologists, biostatisticians and members of industry and regulatory groups

Abbreviations: Adverse events (AEs); American College of Rheumatology (ACR); chronic low back pain (CLBP); C-reactive protein level (CRP); diabetic peripheral neuropathy (DPN); Efficacy-Tolerability Composite (ETC); Erythrocyte sedimentation rate (ESR); European league against rheumatism (EULAR); Outcome Measures in Rheumatology (OMERACT); Osteoarthritis (OA) Patient Global Assessment (PGA); Patient Global Impression of Change (PGIC); postherpetic neuralgia (PHN); Randomized clinical trial (RCT); Relative risk (RR); Rolland Morris Disability Questionnaire (RMDQ); Standardized effect size (SES).