

# It's Good to Feel Better But It's Better To Feel Good and Even Better to Feel Good as Soon as Possible for as Long as Possible. Response Criteria and the Importance of Change at OMERACT 10

VIBEKE STRAND, MAARTEN BOERS, LEANNE IDZERDA, JOHN R. KIRWAN, TORE K. KVIEN, PETER S. TUGWELL, and MAXIME DOUGADOS

**ABSTRACT.** The OMERACT patient reported outcomes (PRO) working group evaluated the methodologies for measuring responsiveness to change at the Outcome Measures in Rheumatology (OMERACT) 10 meeting. The outcome measures used in PRO studies are often expressed as continuous data at the group level (e.g., mean change in pain on a 0–100 visual analog scale). This is difficult to interpret and cannot easily be translated to the individual level of response. When interpreting scores at the individual level, it is important to take into account the following 4 main concepts: (1) improvement; (2) status of well-being; (3) onset of action; and (4) sustainability. Information from clinical trials on how many patients showed a response, what the level of response was, and how many patients are doing well, would be extremely useful for physicians. The objective of this article is to outline how continuous data may be reported in a clinically relevant manner. We will describe 5 techniques of reporting continuous variables in clinical studies and discuss the relevance of each. (*J Rheumatol* 2011;38:1720–7; doi:10.3899/jrheum.110392)

## Key Indexing Terms:

RESPONSE TO CHANGE      CLINICIANS      OUTCOMES      RESPONSIVENESS

Physicians manage patient treatment at the individual level within their clinics. Investigation, diagnosis, and treatment are often performed based on the experience of the physician. However, in the new millennium, such decisions should be based on the available evidence. Evidence is usually acquired by evaluating the clinical relevance of data obtained in research studies. In these studies, we need to use instruments that are clinically meaningful and responsive to changes in a patient's health. The Outcome Measures in Rheumatology (OMERACT) Patient Reported Outcomes (PRO) working group evaluated this issue at the OMERACT 10 meeting. This work is part of a larger PRO initiative that includes selecting domains and choosing instruments that are responsive to change. The outcome measures

used in PRO studies are often expressed as continuous data at the group level (e.g., mean change in pain on a 0–100 visual analog scale, VAS). This is difficult to interpret and cannot easily be translated to the individual level of response. To better understand the results of clinical trials, continuous variables can be translated into dichotomous variables, such as “therapeutic success (yes/no).” However, in order to turn a continuous variable into a dichotomous variable, the cut-off used for the dichotomization must be clinically relevant. Information from clinical trials on how many patients showed a response, what the level of response was, and how many patients are doing well, would be extremely useful for physicians. This article aims to address these questions.

When interpreting scores at the individual level, 4 main concepts need to be taken into account: Improvement, status of well-being, onset of action, and sustainability. Improvement (to feel better) can be measured using the minimally clinically important difference (MCID) [or minimally important difference (MID)] or minimally clinically important improvement (MCII). MCID is defined as “the smallest difference in a score that is considered to be worthwhile or important.” This necessitates a clinical change in a patient's health status<sup>1</sup>. MCII is similar; however, it is only concerned with positive improvements. With the development of more effective therapies in RA over the past decade, mean improvements within treatment groups have generally

*From Department of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto, CA, USA; Department of Clinical Epidemiology, VU University Medical Centre, Amsterdam, The Netherlands; Centre for Global Health, University of Ottawa; Academic Rheumatology Unit, Bristol Royal Infirmary, Bristol, UK; Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway; and Department of Rheumatology B, Hôpital Cochin, Paris, France.*

*V. Strand, MD, Department of Immunology/Rheumatology, Stanford University School of Medicine; M. Boers, MD, PhD, Department of Clinical Epidemiology, VU University Medical Centre; L. Idzerda, Centre for Global Health, University of Ottawa; J.R. Kirwan, MD, Academic Rheumatology Unit, Bristol Royal Infirmary; T.K. Kvien, MD, PhD, Department of Rheumatology, Diakonhjemmet Hospital; P.S. Tugwell, MSc, MD, Centre for Global Health, Department of Medicine, University of Ottawa; M. Dougados, MD, Department of Rheumatology B, Hôpital Cochin.*

*Address correspondence to Ms Idzerda; E-mail: lidzerda@uottawa.ca*

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2011. All rights reserved.

exceeded MCID, so that twice or 3 times the MCID in Health Assessment Questionnaire (HAQ) scores have frequently been reported. This has prompted definitions of “really important improvements”: for example, “moderate improvement,” which is defined as an improvement of greater than 30%, or “substantial improvement,” defined as an improvement greater than 50%.

Achievement of a status of well-being (to feel good) can be broken down into 3 separate states: the patient acceptable symptoms state (PASS); the low or minimal disease activity state (LDAS); and attainment of a normative state, such as attainment of the goal of the therapy or remission (Figure 1).

The PASS is defined as a state in which patients consider their condition to be satisfactory or acceptable, often interpreted as “feeling well.” LDAS is an intermediate state between activity of the disease and complete remission. Minimal or low disease activity (MDA) has been defined as the state of disease activity deemed a useful target of treatment by both the patient and the physician, given current treatment possibilities and limitations<sup>2,3</sup>. The attainment of “normative” values refers to whether a patient is able to attain the goals of therapy or alternatively enter into remission.

The third concept refers to the onset of action (to feel good/better as soon as possible), or time taken to achieve the therapeutic success. The fourth concept refers to sustainability (to feel good/better for as long as possible), or the duration of the therapeutic success.

The objective of this article is to outline how continuous data may be reported in a clinically relevant manner. We describe 5 techniques of reporting continuous variables in clinical studies and discuss the relevance of each of these.

## TECHNIQUES OF REPORTING SCORES AT THE INDIVIDUAL LEVEL

Continuous data can be reported in a number of ways. The 5 main techniques are defined below:

- i. Technique A, the conventional technique, presents data at the group level (i.e., mean and standard deviation).
- ii. Technique B presents data at an individual level and requires that a level of change, or response, is defined in a continuous dataset above which the patient can be categorized as a therapeutic success or a responder. Thus we need to know the minimum clinically important difference or minimum clinically important improvement.
- iii. Technique C also presents data at the individual level but

instead defines a state of being. This represents the level of the continuous variable below which a patient can be categorized as a therapeutic success, e.g., PASS or LDAS or remission.

iv. Technique D takes into account the concept of “onset of action.” Data can be expressed as either the percentage of patients achieving therapeutic success or the median time to achieve therapeutic success. This technique could use life-table analysis. In this analysis, the “event” is defined as the first visit during which the therapeutic success is observed, regardless of the values observed during consecutive visits.

v. Technique E takes into account the concept of sustainability. This technique could also use life-table analysis. In this analysis, the event is defined by the first visit during which the therapeutic success is observed. This differs from technique D as the event is measured during consecutive visits. Alternatively, one can describe the characteristics of the time spent in the state of success (e.g., mean, minimum, maximum, etc.) or assess the number of transitions in and out of the state. The extent to which a period is uninterrupted can be weighted positively. An intuitive method to do this is the ConRew (continuity rewarded) score<sup>4</sup>.

### Technique A.

The conventional techniques of reporting data using the mean and standard deviation have been discussed elsewhere in detail.

### Technique B1. Minimum Clinically Important Differences

MCID refers to the degree of improvement in PRO that would be perceptible to patients, on an individual basis, and would be considered clinically meaningful to them<sup>1,5,6,7</sup>. MCID is defined as “the smallest difference in score that patients perceive as beneficial and which would mandate a change in the patient’s management”<sup>8</sup>. Although these definitions are relevant only on an individual patient basis, when mean changes within a treatment group exceed such a value, it can be estimated that the majority of the group will have attained clinically important improvements. Alternatively, the percentage of patients who report improvements meeting or exceeding MCID is another way to indicate the clinical meaningfulness of treatment associated changes, and to translate group data to the level of individual impact. Next, a few scales commonly used in rheumatology trials will be presented with the defined MCID values.

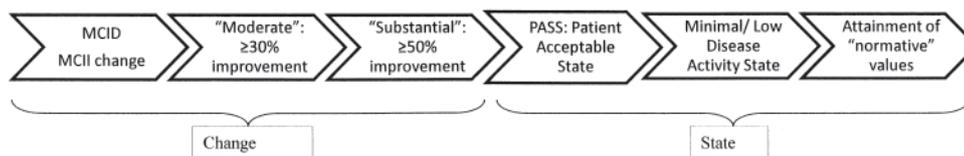


Figure 1. Conceptual diagram of the spectrum of improvements in patient reported outcomes.

*Pain and global assessment of disease activity by VAS.* It is generally accepted that a 10-point change on a 0–100 VAS corresponds to MCID for pain and global assessments.

*Western Ontario McMaster Universities Osteoarthritis Index Questionnaire (WOMAC).* Definitions of MCID are generally obtained by linking changes in PRO to improvements (or worsening) in anchored scores such as patient-reported global disease activity (VAS or Likert scales) or Guyatt feeling thermometer<sup>9,10</sup>. Following a variety of treatments in OA, including nonselective nonsteroidal antiinflammatory drugs (nsNSAID), cyclooxygenase 2 (COX-2) inhibitors, and physical therapy, it was demonstrated that a change of about 10 points on a 0–100 point VAS corresponds to MCID for total WOMAC and WOMAC pain and physical function subscores<sup>11,12,13,14,15</sup>. Functionally useful pain relief is frequently associated with improvements in physical function (so that function may be increased until limited by pain), and composite measures of response in chronic pain conditions such as osteoarthritis (OA), fibromyalgia (FM), and low back pain have included both pain and physical function, measured by WOMAC, Fibromyalgia Impact Questionnaire, Roland Morris Disability Questionnaire (RMDQ), or Medical Outcome Study Short Form-36 (SF-36) physical component score (PCS) as components.

*Health Assessment Questionnaire Disability Index (HAQ DI).* The HAQ DI queries patients' ability to perform activities of daily living as well as the need for help, or use of functional aids. Although a generic measure, it has predominantly been utilized in randomized controlled trials (RCT) in rheumatoid arthritis (RA), and has been shown to perform better in RA and gout than in psoriatic arthritis (PsA) or systemic lupus erythematosus (SLE)<sup>16,17,18</sup>. It is generally believed that the HAQ DI has its greatest item information content in more disabled populations and SF-36 physical function domain in more normalized ones, and that floor effects are more common with HAQ DI than SF-36 physical function domain. An improvement of –0.22 is considered to represent MCID in HAQ-DI<sup>5,6,19</sup>. As with all PRO, changes in HAQ have differentiated active from placebo treatment across recent RCT in RA<sup>20,21,22</sup>. It has also been recognized that improvements in physical function, measured by HAQ, are based on baseline status; higher scores prior to treatment reflect disease duration/severity and an “irreversible” component due to deformities or other irreversible reasons for loss of physical function, including muscle weakness, etc. Thus, HAQ has been recognized as a measure of “state” as well as “change”<sup>23</sup>. It should, however, be noted that the HAQ has scaling problems and is not in reality a linear or interval measure. Shorter versions of HAQ include the modified HAQ (MHAQ), Multidimensional HAQ (MDHAQ) among others<sup>24</sup>. MCID is generally recognized to be –0.25 for these instruments<sup>18</sup>.

*Roland Morris Disability Questionnaire.* In low back pain a change of –5 points on the RMDQ-24 scale has been defined as MCID<sup>25,26</sup>.

*Functional Assessment of Chronic Illness Therapy Fatigue Scale (FACIT) and Fatigue VAS.* FACIT-F scores range from 0 to 52, with higher scores representing less fatigue. The instrument has been validated in the general population and in patients with RA. MCID for FACIT-F in RA was determined to be  $\geq 4$ -point change from baseline<sup>27</sup>. Fatigue has also been assessed using a standard VAS scale, again with MCID considered to be  $\geq 10$  point improvement.

*Health related quality of life (HRQOL): SF-36.* Although a generic instrument, SF-36 has been validated in multiple languages and is sensitive to change across a wide variety of clinical conditions, including rheumatic diseases: RA, OA, SLE, systemic sclerosis (SSc), PsA, ankylosing spondylitis (AS), FM, and gout<sup>28,29,30,31,32,33</sup>. Further, comparisons to age- and gender-matched normative values in US, UK, Scandinavian countries, the Netherlands, and Turkey are available. Values for MCID in domain and summary scores (PCS and mental component score, MCS) of SF-36 have been derived, based on correlations with patient reported improvements in global disease activity or condition, on an individual patient basis. For example, Kosinski and Ware compared changes in HAQ DI and SF-36 domains and summary scores with patient global assessments and pain in 2 RCT comparing COX-2 selective agents to traditional NSAID in active RA and OA<sup>12,34,35,36</sup>. Mean changes in SF-36 domain scores corresponding to one point of improvement in patient global assessment of disease activity (by Likert) or 10 mm improvement (i.e., by VAS) ranged from 4.2 to 21.0 and 1.9 to 10.8, respectively.

Thumboo, Strand, Khanna, Singh, Choy, and others have shown that changes of 5–10 points in domain scores and 2.5–5 points in PCS and MCS summary scores can be considered to represent MCID in OA, RA, SLE, FM, and gout, based on correlations with Guyatt feeling thermometer and/or patient global assessments of disease activity<sup>11,12,13,14,15,18,23,37,38,39,40,41,42,43,44</sup>.

Values for MCID for deterioration are: –2.5 to –5.0 points in domain and –0.8 in PCS and MCS scores, indicating that patients perceive worsening with smaller changes than improvement<sup>12</sup>.

Another interpretation of clinical meaningful improvements is to focus on key questions in SF-36 to perform content-based analysis of HRQOL changes. Examples include assessing the percentage of subjects reporting improvements  $\geq$  MCID in performance of specific tasks or in answer to certain questions, such as climbing flights of stairs or walking a mile<sup>44</sup>.

## **B2. Minimally Important Differences**

Alternatively, statistical definitions, such as changes  $\geq 0.5$  standard deviations of the mean baseline, can be used to

reflect minimally important differences (MID), which are not specifically anchored to PRO<sup>45</sup>. As data have accumulated, it has become apparent that SF-36 MCID and MID values closely correspond and are remarkably consistent across disease states.

**Health Utilities Index Mark 3 (HUI3).** HUI3 is a validated generic HRQOL measure<sup>46,47,48</sup>. It includes a health-status classification system and preference-based scoring formula based on 8 attributes (vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain). There are 5 rating levels for speech, emotion, and pain [ranging from 1 (able to be understood completely, being pain free, etc.)] to 5 (unable to be understood, severe pain, etc.), and 6 rating levels for vision, hearing, ambulation, dexterity, and cognition (where 1 is positive and 6 is negative). Global HRQOL is calculated by translating these categorical data into a single attribute and overall utility scores (ranging from 0 = death to 1.00 = perfect health). The construction of the scale is one of preference or desirability. Score changes of 0.03 are considered to represent MID<sup>49</sup>.

**EuroQol (EQ-5D).** The EuroQol Group designed EQ-5D to be a simple, self-administered questionnaire measuring HRQOL and health utilities. It is valid, reproducible, and sensitive to change in RA, and has been translated into most major languages and cross culturally validated across a variety of clinical indications<sup>50,51</sup>. It consists of: (a) descriptive profile consisting of 5 dimensions, namely, mobility, self-care, usual activities, pain/discomfort and anxiety/depression, with each dimension rated from 1-3: 1 = no problems; 2 = moderate problems; and 3 = extreme health problems; and (b) a current general health status index, measured by a 20 cm VAS (EQ-VAS) with endpoints labeled "best imaginable health state" and "worst imaginable health state," anchored at 100 and 0, respectively. EQ-5D scores have been shown to differ according to disease duration, history of disease modifying antirheumatic drug use, and presence of probable depression or anxiety, in RCT as well as longitudinal observational studies (LOS) in RA<sup>47,52,53</sup>.

**SF-6D.** Brazier, *et al* developed the SF-6D, a preference-based measure of health utilities that used individual item responses from 11 questions of the SF-36 to derive 6 domain classifications of health states, generating 18,000 health states in total<sup>54,55</sup>. Scores range from 0.296 to 1 in which 0.296 = maximum impairment in all 6 domains and 1 = full health. Recently, a newer method for deriving SF-6D utilizing all 8 domain scores of SF-36 has been developed from published group mean data and validated against the traditional calculation method as well as EQ-5D in a cohort of 6350 patients with various diagnoses including OA<sup>56,57</sup>.

MID values for EQ-5D and HUI3 were 0.05 and 0.06, respectively, in a cohort of 222 RA patients with mean age 62 years, disease duration 14 years and baseline HAQ scores of 1.1, assessed 6 months apart<sup>45</sup>. Similarly, MCID or MID for SF-6D calculated using 6 domains is 0.048<sup>50,53,54,58</sup> and

with the recent revision of SF-6D, utilizing all 8 domains of SF-36 is 0.041<sup>55,56</sup>.

### **B3. Minimum Clinically Important Improvements**

Others think that even MCID and certainly clinically meaningful improvements should be defined not only by an absolute amount of change but also relative to baseline. This implies change in a positive direction, as well as more than minimally perceptible. Thus in patients with knee OA, both Tubach's definition of MCII (reduction in the WOMAC physical function subscale of  $\geq 26\%$ <sup>59</sup>) and achievement of a PASS (WOMAC physical function subscale score  $< 31$ <sup>59</sup>) would be required. This definition more closely corresponds to the OMERACT/OARSI definition of responders in OA, as determined by WOMAC<sup>60</sup>. OMERACT-OARSI criteria considered a patient to be a strict responder with improvements from baseline in WOMAC pain or physical function subscores  $\geq 50\%$  with absolute changes  $\geq 20$  mm. A responder reported improvements from baseline  $\geq 20\%$  with absolute changes  $\geq 10$  mm in 2 of 3 measures: WOMAC pain or physical function sub-scores; and/or patient global assessment of disease activity.

In low back pain, OMERACT response is defined as  $\geq 30\%$  improvement in pain and  $\geq 30\%$  improvement in patient global assessment, and no worsening of physical function (change  $\leq 2$ )<sup>61</sup>.

### **B4. Moderate, Substantial and Really Important Improvements**

With the development of more effective therapies in RA over the past decade, mean improvements within treatment groups have generally exceeded MCID — so that twice or 3 times MCID in HAQ scores have frequently been reported. This has prompted definitions of really important improvements: e.g., changes in HAQ DI of  $-0.50$  and  $-0.75$ <sup>62</sup>.

This is consistent with the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) recommendations that improvements in pain be assessed as not just by MCID but as moderate ( $\geq 30\%$ ) or substantial ( $\geq 50\%$ ) improvements<sup>63,64</sup>. These have been applied also to WOMAC pain and physical function subscores and are consistent with the OMERACT/OARSI Strict Responder definition.

### **Technique C. Attainment of Acceptable States**

During the past decade, a lot of work has been done in order to propose clinically relevant definitions of MCII, PASS, LDAS, and remission. The proposal of a value in a continuous variable permitting to define a clinically relevant improvement or an acceptable status has been done through different methodologies as described below:

#### **C1. Patient Acceptable States**

PASS are attainment of a state rather than a measure of

change — meaning that patients consider that level of improvement to be acceptable. They are based on changes observed in RCT in OA and longitudinal observational studies in RA and have been defined as<sup>58,65</sup>:

- WOMAC Physical Function score: < 31
- HAQ (RA): 1.04
- Patient global VAS (RA): 36
- Pain VAS (RA): 34
- Fatigue VAS (RA): 50
- SF-36 domains - physical function: 50; role physical: 41; general health: 45; vitality: 40; social functioning: 75; mental health: 68.

## C2. Low Disease Activity

This concept is similar to that of PASS but generally refers to a composite index evaluating the activity of a disease and comprising several domains. In the field of rheumatology, we have the examples of the DAS28-ESR (Disease Activity Score 28-erythrocyte sedimentation rate) for RA, and AS-DAS-CRP (C-reactive protein) for AS. These composite indices are usually presented as continuous variables, but several methodologies have been used to propose a threshold below which the condition of the patient is considered as acceptable. For example, in RA, a DAS28-ESR < 3.2 and in AS, AS-DAS-CRP < 2.1.

## C3. Attainment of Normative Values

As HAQ scores have decreased over time in longitudinal observational studies of newly diagnosed patients, and generally are higher with larger potential to show a reduction in subjects with active RA accrued into RCT, another high bar for comparison has been to assess the proportion of subjects who attain normative values of HAQ-DI scores  $\leq 0.5$ <sup>66</sup>.

SF-36 PCS and MCS scores are derived from z transformed and normative based data, such that 50 is considered the norm with SD of 10. More informative, however, is to compare individual domain scores at baseline and endpoint with age/gender matched normative data in subjects without arthritis, as a goal for treatment and a means to assess treatment associated improvements<sup>42</sup>.

## MCII and PASS

After the step of the elaboration/proposal of such cutoffs, different studies have been conducted to answer different questions such as:

- Should MCII and PASS be treatment-specific or the same whatever the treatment evaluates?
- What is the relation between MCII and PASS?
- What is the impact of various parameters (e.g., country location, gender, age) on the MCII and PASS estimates?

Determining treatment-specific MCII and PASS values [i.e., PASS for evaluating NSAID therapy vs PASS for evaluating anti-tumor necrosis factor- $\alpha$  (anti-TNF- $\alpha$ )] allows for taking into account the different levels of patients'

expectations for the treatment. Actually, whether patients consider a state (or a change) satisfactory independently of the treatment they receive (i.e., whether the PASS values are related to patients' expectations of the treatment) is not known. One may hypothesize, for instance, that patients expect stronger effects from a TNF- $\alpha$  antagonist than from NSAID therapy and thus would consider a lower level of symptoms as satisfactory with TNF- $\alpha$  antagonist therapy. This point should be investigated in further studies. The drawback of using treatment-specific PASS or MCII values is that these values should be regularly updated as treatment options and knowledge and expectations about them evolve.

Concerning the potential relation between MCII and PASS, the relative meaning of the MCII and PASS is unknown. Whether the concept of improvement or remission or both should be recommended was discussed, and this point was addressed in a survey following the special interest group session, reported below. The results on how the MCII and PASS interrelate in a study of hip and knee OA and acute rotator cuff syndrome were presented<sup>59</sup>. The MCII appears to be the needed change to achieve the PASS. It seems that patients consider that they experienced an important improvement only if this improvement allows them to achieve a satisfactory state. Consequently, it seems that what is important to patients is to feel acceptable or satisfactory (concept of PASS) rather than to feel better (concept of MCII).

At the disease level (e.g., disease activity of RA), several studies suggest also that the status after therapy is more important not only for the patient but also for reducing the risk of subsequent structural damage.

One could argue that the MCII and PASS estimates can vary according to different parameters. In the hip or knee OA study, the MCII was shown to vary greatly across tertiles of baseline scores and age. This impact of the baseline level of symptoms was only partly reduced when using relative change instead of absolute change. Gender and disease duration did not appear to affect the MCII value. The impact of the baseline value had been previously demonstrated in low back pain, in which the MCII varied between 3 and 13, depending on the baseline range of scores (from the Roland Morris Back Pain Questionnaire). Patients dealing with the most severe symptoms have to experience a greater change to consider themselves improved. Where the data on responsiveness are available, it may be possible to adjust or stratify MCII and PASS values for age and gender.

In the hip or knee OA study, the PASS was more constant across tertiles of the baseline score than the MCII; and age, gender and disease duration did not affect the PASS value. An important aspect of any desirable state is the time spent in that state. In the AS study<sup>31</sup>, the PASS was shown to be stable over 10 weeks. This key finding supports the use of the PASS values to describe patients achieving and maintaining such a state for a specified period of time. This point should be confirmed in a study with a longer followup.

## Improvements and States

In summary, one can develop a series of degrees of improvement on a spectrum from MCID through moderate and substantial changes — all best defined for pain — to attainment of state that may be PASS or reaching age/gender matched normative levels ( $\leq 0.5$  in HAQ,  $\leq 31$  in WOMAC, or age/gender matched normative values in SF-36 domain scores, as examples) (Figure 1).

In fact, PRO all conform to a spectrum ranging from minimal or just detectable to moderate, substantial, and really important differences. Reported improvements may also be looked at in parallel as attainments of state or normative values, even as goals of therapy.

The interpretation of improvements as clinically meaningful adds another dimension to statistically significant changes. In large RCT, treatment changes in HRQOL scores may be statistically different but not meaningfully different for most patients, and vice versa. These changes may also be used to calculate values for number needed to treat (NNT): the number of subjects required to receive a therapy to achieve an additional desired outcome. For example, NNT values may be based on the percentage of subjects who report improvements in PRO scores  $\geq$  MCID, moderate or substantial,  $\geq 50\%$  changes<sup>67</sup>. If 40% of subjects receiving active therapy versus 10% in placebo report clinically meaningful changes, then NNT would be calculated as  $1/(\text{relative response in active} - \text{relative response in control})$  or  $1/(0.40 - 0.10) = 3$ ; meaning that 3 patients must receive the treatment intervention to result in one more patient achieving "success," than would have been found with the control intervention. Similarly NNT estimates may be derived based on improvements  $\geq$  MCID in SF-36 domain or summary scores or health utility scores based on SF-6D or EQ-5D.

## Technique D: Onset of Action and E: Sustainability

Both the concepts of onset of action and sustainability have been recently recognized as important enough to be included in the EULAR/ACR collaborative recommendations for reporting results from clinical trials in RA<sup>68</sup>. These 2 concepts are also considered important from a patient's perspective<sup>69</sup>. However, there is still no consensus concerning how to report the analysis of such concepts. It was recently proposed that the life-table analysis approach be used to analyze the concept of onset of action (e.g., time to reach success) and of sustainability (e.g., time to reach a sustained success). When addressing the concept of sustainability, it is also possible to use other techniques such as the ConRew scoring system. Using prespecified observation periods (e.g., every month) the continuity rewarded (ConRew) score sums up periods in remission, and rewards extended periods by placing more value on uninterrupted periods than on interrupted periods.

## CONCLUSION

The advantages of the technique of reporting data at the individual level (e.g., percentage responder, percentage of patients in good condition) versus reporting at the group level (e.g., mean  $\pm$  standard deviation) are well understood. During the last decades, a huge amount of work has dramatically facilitated the awareness of these concepts by the entire medical community. Such recognition has been recently translated in the EULAR/ACR initiative aimed at providing recommendations for reporting disease activity of RA in clinical trials. Such recommendations considered the 4 concepts we have discussed, e.g., the presentation of the data at the individual level (importance of reporting both the responders and the good condition), but also the concepts of onset of action and sustainability/durability.

There remains, however, some discussion about the best way to define the cutoff values, and even the nomenclature that should be employed. The USA Food and Drug Administration, for some years, has required inclusion of graphs of cumulative distribution functions of all levels of responders in product labels for analgesics. The advantage of presenting data in this way is that each reader can decide for themselves which cutoff values are relevant. For many, however, this would add confusion compared to a simple statement such as  $> 90\%$  of the people taking this medication get more than 50% better. Recently it has been argued in treating chronic pain that: The choice of cut-points depends on the purpose of the study<sup>70</sup>. Further, clinically important improvement may depend upon the balance between adverse and beneficial effects, a point made at a previous OMERACT meeting<sup>71</sup> and reinforced by Dworkin at this one<sup>72</sup>.

At the final plenary session after these issues were discussed, participants voted on 2 questions: 1. Do participants agree that evaluation of relevant changes has to consider the specificities of the PRO (e.g., considering additional factors such as coping, adaptation over time, etc.)? 68% voted yes (13% no, 19% don't know). 2. Do participants agree that evaluation of PRO changes such as PASS, time to response, sustainability, etc., requires further standardization? 82% voted yes (5% no, 13% don't know). Thus, a research agenda centered around 2 areas of development has been delineated: the context of disease evaluation, including social and personal adaptation; and the development of agreed and standardized definitions, or possibly approaches to deriving definitions, of the nature and significance of different changes.

## REFERENCES

1. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol* 2002;14:109-4.
2. Wells GA, Boers M, Shea B, Brooks P, Simon L, Strand V, et al. Minimal disease activity for rheumatoid arthritis: a preliminary definition. *J Rheumatol* 2005;32:2016-24.
3. Boers M, Anderson JJ, Felson DT. Deriving an operational definition of low disease activity state in rheumatoid arthritis. *J Rheumatol* 2003;30:1112-4.

4. Boers M, Berkhof J, Twisk JW, Ader HJ, Bezemer D, Knol D, et al. A new graph and scoring system simplified analysis of changing states: disease remissions in a rheumatoid arthritis clinical trial. *J Clin Epidemiol* 2010;63:633-7.
5. Goldsmith C, Boers M, Bombardier C, Tugwell P; for the OMERACT Committee. Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. *J Rheumatol* 1993;20:561-5.
6. Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol* 1993;20:557-60.
7. Wells G. Minimally clinically important difference module: Introduction. *J Rheumatol* 2001;28:398-9.
8. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol* 1994;47:81-7.
9. Guyatt GH, Juniper EF, Walter SD, Griggith LE, Goldstein RS. Interpreting treatment effects in randomised trials. *BMJ* 1998;316:690-3.
10. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102-9.
11. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. *J Rheumatol* 2000;27:2635-41.
12. Zhao SZ, McMillen JI, Markenson JA, Dedhiya SD, Zhao WW, Osterhaus JT, et al. Evaluation of the functional status aspects of health-related quality of life of patients with osteoarthritis treated with celecoxib. *Pharmacotherapy* 1999;19:1269-78.
13. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. *Arthritis Care Res* 2001;45:384-91.
14. Strand V, Kelman A. Outcome measures in osteoarthritis: randomized controlled trials. *Curr Rheumatol Rep* 2004;6:20-30.
15. Strand V, Hochberg M, Ch. 18. Regulatory principles of drug discovery and clinical design. In: Moskowitz R, Goldberg V, Howell DS, Altman RD, Buckwalter J, eds. *Osteoarthritis: diagnosis and medical/surgical management*, 4th ed. Philadelphia: WB Saunders; 2006:313-26.
16. Alvarez-Hernandez E, Peláez-Ballesteros I, Vázquez-Mellado J, Terán-Estrada L, Bernard-Medina AG, Espinoza J, et al. Validation of the Health Assessment Questionnaire Disability Index in patients with gout. *Arthritis Rheum* 2008;59:665-9.
17. Taylor WJ, McPherson KM. Using Rasch analysis to compare the psychometric properties of the Short Form 36 physical function score and the Health Assessment Questionnaire disability index in patients with psoriatic arthritis and rheumatoid arthritis. *Arthritis Rheum* 2007;57:723-9.
18. Gladman DD, Urowitz MB, Ong A, Gough J. A comparison of 5 health status instruments in patients with SLE. *Lupus* 1996;5:190-5.
19. Tugwell P, Wells G, Strand V, Maetzel A, Bombardier C, Crawford B, et al. Clinical improvement as reflected in measures of function and health-related quality of life following treatment with leflunomide compared with methotrexate in patients with rheumatoid arthritis: sensitivity and relative efficiency to detect a treatment effect in a twelve-month, placebo-controlled trial. Leflunomide Rheumatoid Arthritis Investigators Group. *Arthritis Rheum* 2000;43:506-14.
20. Scott DL, Strand V. The effect of disease modifying antirheumatic drugs on the Health Assessment Questionnaire score: Lessons from the leflunomide clinical trials database. *Rheumatology* 2002;41:899-909.
21. Strand V, Cohen S, Crawford B, Smolen JS, Scott DL. Patient-reported outcomes better discriminate active treatment from placebo in randomized controlled trials in rheumatoid arthritis. *Rheumatology* 2004;43:640-7.
22. Cohen SB, Strand V, Aguilar D, Ofman JJ. Patient- versus physician-reported outcomes in rheumatoid arthritis patients treated with recombinant interleukin-1 receptor antagonist (anakinra) therapy. *Rheumatology* 2004;43:704-11.
23. Aletaha D, Strand V, Smolen JS, Ward MM. Treatment-related improvement in physical function varies with duration of rheumatoid arthritis: a pooled analysis of clinical trial results. *Ann Rheum Dis* 2008;67:238-43.
24. Zatarain E, Strand V. Monitoring disease activity of rheumatoid arthritis in clinical practice: contributions from clinical trials. *Nature Clin Prac Rheumatol* 2006;2:611-8.
25. Bombardier C, Hayden J, Beaton DE. Minimal clinically important difference. Low back pain: outcome measures. *J Rheumatol* 2001;28:431-8.
26. Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. *Best Pract Res Clin Rheumatol* 2005;19:593-607.
27. Cella D, Yount S, Sorensen M, Chartash E, Sengupta N, Grober J. Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. *J Rheumatol* 2005;32:811-9.
28. Schlenk EA, Erlen JA, Dunbar-Jacob J, McDowell J, Engberg S, Sereika SM, et al. Health-related quality of life in chronic disorders: a comparison across studies using the MOS SF-36. *Qual Life Res* 1998;7:57-65.
29. Thumboo J, Fong KY, Ng TP, Leong KH, Feng PH, Thio ST, et al. Validation of the MOS SF-36 for quality of life assessment of patients with systemic lupus erythematosus in Singapore. *J Rheumatol* 1999;26:97-102.
30. Husted JA, Gladman DD, Cook RJ, Farewell VT. Responsiveness of health status instruments to changes in articular status and perceived health in patients with psoriatic arthritis. *J Rheumatol* 1998;25:2146-55.
31. Singh JA, Strand V. Spondyloarthritis is associated with poor function and physical health related quality of life. *J Rheumatol* 2009;36:1012-20.
32. Khanna D, Furst DE, Clements PJ, Park GS, Hays RD, Yoon J, et al. Responsiveness of the SF-36 and the Health Assessment Questionnaire Disability Index in a systemic sclerosis clinical trial. *J Rheumatol* 2005;32:832-40.
33. Singh JA, Strand V. Gout is associated with more comorbidities, poorer Health Related Quality of Life and higher health care utilization in US Veterans. *Ann Rheum Dis* 2008;67:1310-6.
34. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE Jr. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. *Arthritis Rheum* 2000;43:1478-87.
35. Ehrich EW, Bolognese JA, Kong S, Watson DJ, Zeng K, Seidenberg BC. Improvements in SF-36 mental health domains with treatment of OA result of decreased pain and disability or independent mechanism [abstract]. *Arthritis Rheum* 1998;41 Suppl:S221.
36. Zhao SZ, Feichtner JI, Tindall EA, Dedhiya SD, Zhao WW, Osterhaus JT, et al. Evaluation of health related quality of life of rheumatoid arthritis patients treated with celecoxib. *Arthritis Care Res* 2000;13:112-21.
37. Strand V, Bombardier C, Maetzel A, Scott D, Crawford B. Use of minimum clinically important differences [MCID] in evaluating

- patient responses to treatment of RA [abstract]. *Arthritis Rheum* 2001;44 Suppl:S187.
38. Strand V, Cannon G, Cohen S, Ware J. Correlation of HAQ with SF-36: Comparison of leflunomide to methotrexate in patients with active RA [abstract]. *Arthritis Rheum* 2001;44 Suppl:S187.
  39. Strand V, Scott DL, Emery P, Kalden JR, Smolen JS, Cannon GW, for the Leflunomide Rheumatoid Arthritis Investigators Groups, et al. Physical function and health-related quality of life: analysis of 2-year data from randomized, controlled studies of leflunomide, sulfasalazine, or methotrexate in patients with active rheumatoid arthritis. *J Rheumatol* 2005;32:590-601.
  40. Strand V, Crawford B. Improvement in health-related quality of life in patients with systemic lupus erythematosus following sustained reductions in anti-dsDNA antibodies. *Expert Rev Pharmacoecon Outcomes Res* 2006;5:317-26.
  41. Thumboo J, Strand V. Health-related quality of life in patients with systemic lupus erythematosus: an update. *Ann Acad Med Singapore* 2007;36:115-22.
  42. Choy EH, Arnold LM, Clauw DJ, Crofford L, Glass JM, Simon LS, et al. Content and criterion validity of the preliminary core dataset for clinical trials in fibromyalgia syndrome. *J Rheumatol* 2009;36:2330-4.
  43. Strand V, Crawford B, Singh J, Choy E, Smolen JS, Khanna D. Viewpoint: Use of "Spydergrams" to present and interpret health related quality of life data across rheumatic diseases. *Ann Rheum Dis* 2009;68:1800-4.
  44. Strand V, Singh J. Newer biologic agents improve health related quality of life and productivity in rheumatoid arthritis. *Drugs* 2010;70:121-45.
  45. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003;41:582-92.
  46. Boyle MH, Furlong W, Feeny D, Torrance GW, Hatcher J. Reliability of the Health Utilities Index--Mark III used in the 1991 cycle 6 Canadian General Social Survey Health Questionnaire. *Qual Life Res* 1995;4:249-57.
  47. Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. *Pharmacoeconomics* 1999;15:141-55.
  48. Marra CA, Esdaile JM, Guh D, Kopec JA, Brazier JE, Koehler L, et al. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Med Care* 2004;42:1125-31.
  49. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI3): Concepts, measurement properties and applications. *Health Quality Life Outcomes* 2003;1:54.
  50. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. *Pharmacoeconomics* 2000;17:13-35.
  51. van Riel PL, Freundlich B, MacPeck D, Pederson R, Foehl JR, Singh A. Patient-reported health outcomes in a trial of etanercept monotherapy versus combination therapy with etanercept and methotrexate for rheumatoid arthritis: the ADORE trial. *Ann Rheum Dis* 2008;67:1104-10.
  52. Witney AG, Trehan GJ, Tavakoli M, Lyons AC, Vincent K, Scott DI, et al. The relationship of medical, demographic and psychosocial factors to direct and indirect health utility instruments in rheumatoid arthritis. *Rheumatology (Oxford)* 2006;45:975-81.
  53. Ariza-Ariza R, Hernandez-Cruz B, Carmona L, Ruiz-Montesinos MD, Ballina J, Navarro-Sarabia F. Assessing utility values in rheumatoid arthritis: a comparison between time trade-off and the EuroQol. *Arthritis Rheum* 2006;55:751-6.
  54. Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. *J Clin Epidemiol* 1998;51:1115-28.
  55. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002;21:271-92.
  56. Ara R, Brazier J. Deriving an algorithm to convert the eight mean SF-36 dimension scores into a mean EQ-5D preference-based score from published studies (where patient level data are not available). *Value Health* 2008;11:1131-43.
  57. Ara R, Brazier J. Predicting the Short Form-6D preference-based index using the eight mean short form-36 health dimension scores: estimating preference-based health-related utilities when patient level data are not available. *Value Health* 2009;12:346-53.
  58. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ* 2004;13:873-84.
  59. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Ann Rheum Dis* 2005;64:34-7.
  60. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. *Osteoarthritis Cartilage* 2004;12:389-99.
  61. Simon LS, Evans C, Bombardier C, West C, Robbins J. Preliminary development of a responder index for chronic low back pain. *J Rheumatol* 2007;34:1386-91.
  62. Wolfe F, Michaud K, Strand V. Expanding the definition of clinical differences: from minimally clinically important differences (MCID) to really important differences (RID). Analyses in 8,931 patients with rheumatoid arthritis. *J Rheumatol* 2005;32:583-9.
  63. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. Defining the clinically important difference in pain outcome measures. *Pain* 2000;88:287-94.
  64. Dworkin RH, Turk DC, Wyrwich KH, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain* 2008;9:105-21.
  65. Heiberg T, Kvein TK, Mowinckel P, Aletaha P, Smolen JS, Hagen KB. Identification of disease activity and health status cut off points for the symptom state acceptable to patients with rheumatoid arthritis. *Ann Rheum Dis* 2008;67:967-71.
  66. Krishnan E, Sokka T, Hakkinen A, Hubert H, Hannonen P. Normative values for the Health Assessment Questionnaire disability index: benchmarking disability in the general population. *Arthritis Rheum* 2004;50:953-60.
  67. Moore RA, Moore OA, Derry S, Peloso PM, Gammaitoni AR, Wang H. Responder analysis for pain relief and numbers needed to treat in a meta-analysis of etoricoxib osteoarthritis trials: bridging a gap between clinical trials and clinical practice. *Ann Rheum Dis* 2010;69:374-9.
  68. Aletaha D, Landewe R, Karonitsch T, Bathon J, Boers M, Bombardier C, et al. Reporting disease activity in clinical trials of patients with rheumatoid arthritis: EULAR/ACR collaborative recommendations. *Ann Rheum Dis* 2008;67:1360-4.
  69. Dougados M, Schmidely N, Le Bars M, Lafosse C, Schiff M, Smolen JS, et al. Evaluation of different methods used to assess disease activity in rheumatoid arthritis: analyses of abatacept clinical trial data. *Ann Rheum Dis* 2009;68:484-9.
  70. Farrar JT. Cut-points for the measurement of pain: The choice depends on what you want to study. *Pain* 2010;149:163-4.
  71. Kirwan JR. Minimum clinically important difference: The crock of gold at the end of the rainbow? *J Rheumatol* 2001;28:881-6.
  72. Kirwan J, Boonen A, Harrison MJ, Hewlett S, Wells G, Singh JA, et al. OMERACT 10 patient perspective virtual campus: Valuing health; measuring outcomes in RA fatigue, RA sleep, arthroplasty, and systemic sclerosis; and the clinical significance of changes in health. *J Rheumatol* 2011;38:1728-34.