



Workshop Minutes

Toward Consensus Development: Qualifying Endpoint Measures for Rheumatoid Arthritis Clinical Trials

Sponsored by:

**Critical Path Institute's Patient-Reported Outcome Consortium
Rheumatoid Arthritis Working Group**

**Sheraton Silver Spring Hotel
8777 Georgia Avenue ■ Silver Spring, MD 20910**

August 28, 2012 ■ 8:30 am to 4:00 pm



ATTENDEES

Clinical experts: Clifton O. Bingham (Johns Hopkins University), Maarten Boers (VU University Medical Center; EULAR), David Felson (Boston University School of Medicine), James R. O'Dell (ACR), Lee S. Simon (OMERACT), Jasvinder (Jas) Singh (University of Alabama at Birmingham School of Medicine), William St. Clair (ACR), and Vibeke Strand (OMERACT; Stanford University)

Patient representatives: Amye Leong and Brian Taylor

FDA: Laurie Burke (Study Endpoints and Label Development Team—SEALD), Badrul Chowdhury (Division of Pulmonary, Allergy, and Rheumatology Products—DPARP), Elektra Papadopoulos (SEALD), Keith M. Hull (DPARP), Larissa Lapteva (DPARP), Theresa Mullin (Planning and Informatics), Nikolay P. Nikolov (DPARP), Ashley F. Slagle (SEALD), Jessica Voqui (SEALD), Janet Woodcock (CDER), and Sarah (Okada) Yim (DPARP)

NIH-NIAMS: Susana Serrate Sztejn and Phil Tonkins

Rheumatoid Arthritis (RA) Working Group: Mallik Angalakuditi (Boehringer Ingelheim), Amy DeLozier (Eli Lilly), Carol Gaich (Eli Lilly), Amy Grahn (Horizon Pharma), Alison W. Greene (Roche), Chenglong Han (Johnson & Johnson), Azra Hassanali (Roche), Deborah Hogerman (UCB Pharma), Jeffrey Kent (Horizon Pharma), Chin Lee (Eli Lilly), April N Naegeli (Co-chair, Eli Lilly), Enkeleida Nikai (Co-chair, UCB Pharma), Dena Rosen Ramey (Merck), Jeffrey (Jeff) W. Sherman (Horizon Pharma), Irene Schubert (Novo Nordisk), Victor S. Sloan (UCB Pharma), and Susan Vallow (GlaxoSmithKline)

C-Path: Stephen Joel Coons, Theresa (T) Griffey, Lindsay Lehman, and Elisabeth (Liz) Piauult-Louis

Co-director of C-Path PRO Consortium: Ari Gnanasakthy (Novartis)

Moderator: Risa Hayes (Eli Lilly)

BACKGROUND: The workshop was convened by the Rheumatoid Arthritis (RA) Working Group within the Critical Path Institute's Patient-Reported Outcome (PRO) Consortium and was audio-recorded.

OBJECTIVES

The principal objective of this workshop was to identify RA-related symptoms and RA-defining decrements in physical functioning that could be investigated by the RA Working Group for use as patient-reported endpoints in clinical trials to support label claims.

The expected outcome of the workshop was a research agenda aimed at collecting evidence for the FDA qualification of one or more PRO instruments that capture concepts that are relevant to patients with RA. Once qualified, the PRO instruments will contribute to the assessment of treatment benefit in RA drug registration trials.

Qualification, as described in the FDA’s draft guidance for industry titled *Qualification Process for Drug Development Tools*, is based on an FDA review of evidence that supports the conclusion that a PRO instrument provides a well-defined and reliable assessment of a targeted concept in a specified context of use. Once qualified, the instruments must be made publicly available.

DISCUSSION TOPICS:

1. Drug Development Tool (DDT) Qualification process

Dr. Janet Woodcock MD, from the FDA, briefly discussed the DDT qualification program and indicated that it aims to foster the availability of adequate measures for use in drug development¹. A measure must demonstrate “fitness for use” for its intended use to be considered qualified. Because this is a new process and FDA has limited experience to date with the qualification of PRO instruments, the FDA is reaching out to clinical and measurement experts to determine the level of measurement performance that can be accepted as ‘good enough’ for instrument qualification. Dr. Woodcock cautioned that “the perfect can be the enemy of the good”. While many of the measures currently used to document treatment benefit have had very minimal assessments of their performance, these measures have proven to be effective in documenting treatment benefit.

In addition, Dr. Woodcock reminded the audience that the purpose of the qualification program is to qualify measures that are ‘very specifically related to the disease process and therefore that would be expected to be responsive to effective intervention.’ Once qualified, the measures are expected to support primary or secondary efficacy endpoints.

⇒ *Additional comment conveyed during the general discussion:*

Anyone, including OMERACT, could submit data for the qualification of a PRO instrument. Consortia, such as C-Path’s PRO Consortium, provide a framework that facilitates exchange of information and knowledge among relevant stakeholders to foster PRO instrument qualification.

2. Documentation of treatment benefit in RA: Current model and opportunities

2.1. Current model: The ACR response criteria

Sarah Yim, MD, from the FDA’s Division of Pulmonary, Allergy, and Rheumatology Products, provided a brief history of drug approvals in RA and indicated that the ACR response criteria is the ‘standard’ primary endpoint used to document treatment benefit in patients with RA. Further details are attached in Dr. Yim’s presentation.

¹ Guidance for Industry: *Qualification Process for Drug Development Tools (Draft 2010)*
<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf>

Furthermore, Dr. Woodcock briefly mentioned some concerns with the ACR response criteria: it is a combination of endpoints with multiple components and some of its PRO components were not developed according to modern standards of measurement, development, and interpretation. FDA indicated that they do not have much of a context of understanding what concepts legacy measures such as the components of the ACR response criteria are reflecting and the level of change that is considered appropriate/clinically meaningful for the patients. Dr. Woodcock insisted, regardless of those problems, the ACR response criteria is included in the “*FDA Guidance for Industry Clinical Development Programs for Drugs, Devices, and Biological Products for the Treatment of Rheumatoid Arthritis (RA), 1999*” and could be considered a benchmark as it has proven to be responsive in documenting treatment benefit in multiple clinical trials.

⇒ *Additional comments:* Throughout the day, several clinical experts expressed concerns regarding the perceived intent of the RA Working Group to modify the ACR response criteria. Experts recommended strongly against this for the following reasons:

- The PRO-based components of the ACR response criteria were developed using state of the art methods at that time. Their sensitivity to change and ability to support claims of treatment benefit have been demonstrated through extensive use in clinical trials.
- The ACR response criteria was developed using clinical trial data for each outcome measure that constitute the ACR composite endpoint. Any new measure that would be included in a revised endpoint model would need to provide information over and above what is currently captured by the ACR response criteria.
- Modifying the primary endpoint will compromise the ability to compare data across approved RA products.

⇒ *Consensus:* The RA Working Group does not intend to modify the ACR response criteria. Consensus was reached that the ACR response criteria is a well-established primary endpoint for RA trials. PRO instruments that will be considered for qualification will be included in the hierarchy of endpoints as secondary endpoints.

2.2. Opportunities

2.2.1. Definition of target population: FDA recommends the inclusion of a more refractory population in clinical trials including inadequate responders (after TNF inhibitors or DMARD) and early RA methotrexate-naïve patients.

2.2.2. Labeling claims: As per the Guidance for Industry on the clinical studies section of the label according the Physician’s Labeling Rule (PLR), the label is not intended to describe all available effectiveness data, but rather should present those endpoints that establish the effectiveness of the drug or show the limitations of effectiveness.² Historically, labels for products approved for RA in the US have included:

- Major clinical response, defined as ACR70 response for a continuous six-month period,
- Improved physical function (prevention of disability) using the mean change in the HAQ-DI from baseline to week 12 or longer, and
- Slowing or inhibition of the progression of structural damage using the total Sharp score or similar radiographic scores for a minimum of about 24 weeks.

In addition to the PRO components of the ACR response criteria, PRO instruments that have been used to support labeling claims of efficacy for currently approved medical products include an assessment of morning stiffness and scales of the SF-36 to document physical function.

2.2.3. Concept of measurement - Fatigue: The Agency acknowledged that fatigue is ‘extremely important to patients with RA’ and emphasized that fatigue is a multi-dimensional and multi-factorial concept. In addition, Dr. Yim reported that improvement in fatigue, as measured by some of the currently available fatigue PRO instruments, is consistently noted as an ancillary benefit when measured in clinical trials of effective DMARDS.

No fatigue claim has been granted to date by the FDA in RA, due in part to the uncertainty surrounding the interpretation of this outcome. It is challenging to determine whether a change in fatigue is representing a different treatment benefit than the benefit already measured in improvement in disease activity.

3. General discussion regarding RA qualification program

The goal of the workshop was discussed, and experts recommended against a discussion that would aim to identify the ‘core set of symptoms in RA.’ The rationale was that the core set of RA symptoms is captured in the ACR response criteria and RA-related damage is also evaluated through the Sharp score. Twenty years ago, when the ACR response criteria were defined [1990 – 1992], it was not clear that fatigue was an important symptom to measure. Because current treatments help manage most of the symptoms of the condition (e.g., joint pain), other symptoms like fatigue have emerged as important to RA patients.

Experts called for a more general discussion regarding the burden of disease in RA patients and identification of concepts that would provide an incremental difference in the assessment of treatment benefit (i.e., only PRO instruments that could ‘provide data over and above’ the information already collected in legacy instruments should be considered by the RA Working Group for PRO instrument qualification).

The RA Working Group co-chairs confirmed that the primary objective of the workshop was to identify concepts that are important to patients, are currently not well captured in the core set measures (i.e., ACR response criteria and Sharp score), and are treatment responsive. Evidence regarding the measurement properties of PRO instruments capturing those concepts will then be submitted to FDA through the qualification program. Qualified PRO measures would then be

implemented in clinical trials to ultimately ensure that these concepts will be visible in future product labels.

All participants agreed that patients are foundational, not only in identifying the concepts missing in the current core set measures, but also in defining the research agenda for PRO instrument qualification. Experts emphasized that the field of rheumatology has been at the forefront of using patient feedback in defining its standards for treatment evaluation. For instance, patients have been engaged in OMERACT since 2002. Since then fatigue, sleep quality, health-related quality of life (HRQL), and productivity/participation have been endorsed by OMERACT as valuable assessments in clinical research and practice.

4. Context of use: Identification of relevant patient subgroups

Lee S. Simon, MD from SDG LLC described the patient populations that are included in clinical trials and reviewed existing measures of disease activity. Below is a summary of the elements to be considered by the RA Working Group when determining their target population for the purpose of PRO qualification. Further details are attached in Dr. Simon's presentation.

Patients included in clinical trials for drug approval have moderate to severe disease activity and are not necessarily representative of "real life" patients. For instance, in the US, it is extremely difficult to recruit patients with severe disease activity.

The RA population is extremely heterogeneous (for instance, results from a post-hoc analysis of patients included in clinical trials indicated that 39 to 73 percent of patients with early RA develop hand and wrist erosion of one or more joints within five years). In addition, although similar selection criteria are applied in clinical trials, resulting baseline characteristics of enrolled patients vary from trial to trial.

In RA patients there is variability in terms of disability as well as in terms of radiographic damage. In addition, studies have demonstrated a lack of correlation between the damage as measured by imaging studies (e.g. x-ray imaging) and clinical outcomes.

Dr. Simon mentioned three instruments that could be used to determine a patient's level of disease activity for the purpose of recruiting the target population for studies to support PRO qualification. Disease activity indexes include: the Disease Activity Score (DAS 28), the Simple Disease Activity Index (SDAI) and the Clinical Simplified Disease Activity (CDAI). These instruments use similar assessments (e.g., Swollen-and tender joint counts of 28 joints, CRP, Patient global [DAS28]; SDAI: the previous 4 + MD global assessment; CDAI: SDAI minus CRP) but differ in their scoring algorithm.

⇒ *Additional comments regarding the definition of the target population for PRO instrument qualification:*

- The experts suggested looking across the spectrum of disease to ensure that the measure considered for qualification will have discriminatory power (i.e., able to differentiate between an active treatment and placebo) across the spectrum of disease and will not be

limited by floor or ceiling effects, as is the case with some of the current RA-specific PRO instruments.

- Experts discussed the limitations of considering disease activity to define the context of use for PRO instrument qualification:
 - From the patient's perspective, definition of disease activity changes on a daily basis and, is not a stable concept. ("It goes from mild, to moderate, to severe, or back and forth.")
 - Disease activity measurement by itself does not sufficiently account for remission and flares, and
 - Disease activity measurement does not account for the response shift that occurs over the course of rheumatoid arthritis in terms of self-management and in terms of treatment expectations. It was also suggested to consider disease duration when defining subgroups of the targeted population.

5. Concepts of measurement: RA-Defining Symptoms

5.1. – Experts perspective in Concepts of measurement: RA-Defining Symptoms

Clifton O. Bingham III, MD - Director, Johns Hopkins Arthritis Center, Co-Director Rheumatic Disease Research Core Center Associate Professor of Medicine, Divisions of Rheumatology and Allergy, Johns Hopkins University, described several qualitative studies aimed at eliciting from patients their most important symptoms to alleviate in RA or to define worsening in disease activity as well as the impact that RA has in patients' daily lives. Below is a summary of the concepts to be considered by the RA Working Group when selecting the concepts of measurement for the purpose of PRO instrument qualification. Further details are attached in Dr. Bingham's presentation.

5.1.1. RA-related fatigue and its current measurement

Fatigue is described as "something that is common, often severe, and has considerable life impact." There are multiple dimensions to it, and neither physical tiredness nor vitality adequately addresses the domain itself. Up to 97% of patients report fatigue, up to 69% report severe fatigue, and almost 50% report that fatigue occurs on a daily basis. Fatigue levels can differentiate between levels of RA disease activity; they can predict deterioration in health related quality of life (HRQL) and also affect employment and participation.

Sarah Hewlett and colleagues found that 23 instruments, including the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT- F), the Multidimensional Assessment of Fatigue (MAF), the Profile of Mood States (POMS), and a Visual Analogue Scale (VAS) for fatigue, have been used to measure fatigue in rheumatoid arthritis. Yet, authors demonstrated that some of the concepts elicited through qualitative research that were deemed important for patients with RA experiencing fatigue, were not comprehensively covered by any of the instruments that had previously

been used to assess RA-related fatigue. These concepts include: cognition, coping, emotion, energy, frequency, impact, planning, HRQL, impact on participation, sleep, and social life. In addition, none of the scales were standardized; in some cases, the same instrument that had been used in various studies used different anchors. Dr. Bingham noted another drawback is that results are reported as global scores, which precludes differentiation between the different aspects of fatigue that may be involved in rheumatoid arthritis (e.g., a patient could have an improvement in physical elements of fatigue but not in the mental aspects of fatigue).

The correlation between fatigue instruments with pain instruments is highly variable from one study to another and ranges from no correlation to good correlation. Variable results were found for fatigue and physical function. There were no associations between fatigue and inflammatory indices, cortisol levels, or anemia. Dr. Bingham stressed that limitations related to the study design (cross-sectional study) or the fatigue measure itself (use of global scores) might account for these variable results.

Dr. Bingham identified the Bristol RA Fatigue Multi-Dimensional Questionnaire (BRAFM-DQ) and the fatigue measure from PROMIS as potential candidates for PRO instrument qualification.

5.1.2. Stiffness

Morning symptoms include both stiffness and pain, and in many cases these cannot be easily distinguished. Morning symptoms have an impact on work life, showing that the ability to function in the morning can affect employment by increasing the rates of absenteeism. As pointed out by the patients and supported by experts during the general discussion, stiffness is broader than just morning stiffness.

Qualitative research reveals that the concept of “stiffness” has not been clearly defined yet. Patients use many words to describe stiffness.

Dr. Bingham indicated that there were currently no standardized methods for evaluating stiffness. The assessment measures that have been used, which are focused on duration of morning stiffness, are poorly responsive measures in RA. Measures assessing stiffness severity appear to more responsive.

5.1.3. Other concepts

Sleep: This is another important concept that was mentioned by patients and experts. The most appropriate instruments have not yet been identified but research is on-going.

Emotional health: this is a multi-factorial domain that is consistently identified by patients as another important concept. Some of these aspects are already covered by the SF-36. Additional research is needed for instruments such as Rheumatoid Arthritis Impact of Disease (RAID), FLARE questionnaire, or PROMIS.

Participation

Although productivity was initially suggested as a concept for measurement, the term ‘participation’ was recommended instead to remove any economic connotations. Participation was defined as the ability to participate in work inside and outside of the home including ability to participate in family, social or leisure activities. It was suggested that participation also encompasses the ability to function in social roles and to participate in valued life activities. It was further specified that participation is different from physical function and from ability to function in activities of daily living. Participation is part of the WHO’s ICF framework, and it has been an important concept for OMERACT due to its reported importance for patients.

⇒ *Additional comments:*

- The following instruments might include items relevant to participation: RAID (one item), the FLARE questionnaire, and finally PROMIS (social roles and activities domain, and satisfaction with social roles and participation domain).
- In addition, Dr. Strand suggested two instruments that could be useful for documenting participation: the Work Productivity Survey-RA which queries participation in family, social and leisure activities and work within the home as well as work outside the home and scales of the SF-36 [social functioning, role function-physical]. Another participant suggested the Late-Life Disability Instrument (LLDI) which is widely used in OA studies.
- Dr. Felson reported that in OA studies, participation outcomes were strongly related to physical function.

In conclusion, Dr. Bingham indicated that RA patients have consistently identified a number of important domains that are currently not well-assessed. These domains include fatigue, stiffness, sleep, participation, and emotional health. Dr. Bingham indicated that these concepts have not been studied sufficiently to understand how they impact patients’ ability to function and how they change in the natural history of the disease.

5.2. RA patient perspective

Amye L Leong, MBA and Brian Taylor, patient representatives introduced themselves as RA patient representatives and described their experience with RA. The following concepts were briefly discussed; more details are in the attached presentation by Ms. Leong and Mr. Taylor.

5.2.1. Pain

- Related to surgery (knees, hips, shoulder), to inflammation, and secondary to deformation (referred to as ‘damage pain’)
- Of very high intensity (“number one symptom to be relieved”)

- Pervasive (impact on ability to think, sleep, eat, or have social interactions for instance)

5.2.2. Joint stiffness

- After a prolonged period of rest
- In the whole body (e.g., back, shoulder)
- Amye noted that “morning stiffness” was an inadequate term to describe her symptom.

5.2.3. Poor sleep quality including waking up at night because of the pain, or secondary to RA medication (gain in weight leading to sleep apnea). This symptom further leads to a diminished ability to function the next day. Amye indicated that patients could decipher between disease-related poor sleep quality and poor sleep quality related to environmental factors.

5.2.4. Fatigue

- Described as being physically exhausted at the end of the day (Brian); all day long (‘my fatigue starts at the moment I wake up’) and gets worse through the day (Amye)
- Very distinguishable from fatigue due to physical activities.

5.2.5. Flare fatigue precursor of flare (Amy), flu-like symptoms

5.2.6. Swelling (e.g., knees, ankles)

5.2.7. Physical function

- Upper body mobility (could not raise hands to eat, turn neck, open jaw)
- Lower body mobility (being wheelchair bound, could not walk ten feet, run, sit on low chairs)

5.2.8. Participation

- Defined as “what you do with other people” including work inside and outside the home.
- Patients’ satisfaction with their level of participation and patients’ ability to participate in activities with family, or outside the house were both considered when defining participation.

5.2.9. Lack of independence (hygiene, house chores)

5.2.10. Deformities (lifted shoes, scars from surgeries)

5.2.11. In addition, Amye and Brian emphasized that patients were not in one state of **disease activity** (mild, moderate, or severe) as their symptoms were evolving on the daily basis (“I’d say it depends on the day. It depends on the hour.”). Although a current treatment goal in RA is to reach remission, Amye and Brian clarified that for

patients the ultimate goal is to minimize flare. Brian and Amye identified pain, vasculitis on the tips of finger, infection, flu-like symptoms including fatigue, and severe swelling as the prodromal symptoms associated with flares.

Amye concluded the presentation by reiterating that patient inclusion in discussions regarding RA measure development was critical.

5.3. General discussion for a research agenda regarding RA-related symptoms

5.3.1. Consensus was reached that a large amount of foundational research has already been performed in the realm of RA-related symptoms and that the RA Working group should not ‘recreate the wheel,’ rather, take advantage of the work that has been done and the expertise that is available. There was also agreement that ‘additional, well-designed research was certainly needed.’

5.3.2. Fatigue: Experts agree that fatigue was a very common disabling symptom in patients with rheumatoid arthritis and that it should be measured in clinical trials.

Dr. Felson cautioned that fatigue was not specific of RA and that many patients’ experience of fatigue might be related to other co-morbidities (e.g., fibromyalgia). Dr. Felson questioned how one will ensure that the fatigue experienced by patients is from their rheumatoid arthritis and not just due to other environmental factors. Dr. Felson proposed two important questions for discussion:

- Are these proposed symptoms (and more specifically fatigue) bringing any incremental value to document treatment benefit (i.e., is fatigue providing any additional information on the efficacy of therapies that is not adequately captured with the current endpoints?). What incremental information or value is being provided by newer or other PRO measures?
- Is fatigue or stiffness occurring in a large enough population of RA patients and on regular basis, to be considered a ‘universal’ symptom? Dr. Felson reminded the participants that morning stiffness was initially included in the RA trials as primary endpoint but since 30-40% of patients didn’t experience this symptom it was removed from the primary endpoints.

Brian and Amye indicated that patients could not be expected to distinguish RA fatigue from fatigue from other disease-related fatigue. They can however, distinguish RA fatigue from fatigue related to activity that causes tiredness. In addition, Amye reiterated that patients usually refer to the impact of their symptoms (how the symptom affects our lives every moment of the day) rather than the symptom itself.

Dr. Woodcock also indicated that randomization should help with regards to fatigue attribution to RA or other co-morbidities or life-related events.

For qualitative research, it was suggested to enroll patients in both the early disease state of RA when the number of joints involved is limited and in more advanced

disease state and document whether fatigue is discernible as fatigue that is related to the inflammatory condition.

Experts emphasized that to date, the incremental value of assessing fatigue when documenting treatment benefit was unknown and that it should be considered for inclusion in the RA Working Group research agenda.

Certain experts expressed concern that an existing instrument, namely the patient's global assessment of disease activity included in the ACR response criteria, might already capture patients' experience of fatigue.

Experts reiterated that data regarding fatigue has now been collected in a large number of biologic trials in rheumatoid arthritis. Experts suggested pooling the data to compare the sensitivity to change of fatigue instruments with the sensitivity to change of other measures, and analytically address whether the assessment of fatigue adds to the core set information. ACR indicated that they could contribute a fatigue sub-study of large biologic trials. The group was cautioned that some of the instruments have not been reviewed/accepted by the FDA as well-defined and reliable to assess RA-related fatigue; so the validity of the conclusions that could be drawn from such analyses might be questionable. The working group agreed to explore the possibility of collating the existing data and conducting such analyses. . OMERACT and the ACR Classification Response Criteria Committee offered to host such collaboration.

5.3.3. Stiffness: The concept of stiffness has not been studied enough to be considered for inclusion in the label. More work is needed to define stiffness in a way that comprehensively captures patients' experience.

5.3.4. Participation: Some experts and patient representative expressed interest in moving an agenda forward for the qualification of a PRO instrument assessing participation:

- “measures that look at returning to work and everything as a functional thing, not as an economic thing would be very useful in looking at the value of the therapies to the patients, as well as to society.”
- “I think it’s very interesting that, from a patient’s perspective, because I’ve been at an FDA meeting where something was given an extra indication for physical function, and from my perspective it didn’t matter. I didn’t really care. But if I reverse that, and said it was an extra indication for increased participation, now you’ve got my attention.”

Dr. Simon indicated that a lot of work still needs to be done to appropriately conceptualize participation. Stephen Coons cautioned the group that participation can be impacted by the core set of RA symptoms but can also be impacted by many

factors external to treatment (e.g., social opportunities). Laurie Burke indicated that conceptualization of participation should not include “participation in things that are so far downstream that they’re irrelevant to the impact of treatment, and measuring treatment benefit.”

6. Concepts of Measurement: RA- defining decrements

6.1. Concept of physical functioning.

Jasvinder Singh, MD, MPH, from the University of Alabama at Birmingham School of Medicine introduced the concept of physical functioning as the ability to function without limitations in the course of daily life. It incorporates aspects of strength, mobility, freedom of movement, balance and coordination.

Vibeke Strand, MD, FACP, FACR, who is Adjunct Clinical Professor in the Division of Immunology/Rheumatology at Stanford University School of Medicine, emphasized the importance of physical function in RA by reporting the results of an internet survey (n=2,000 female RA patients). Patients were asked to provide a definition of a good day with RA; responses included being free of pain; free of fatigue and having energy; and able to do everyday things easily.

Dr. Singh further described a randomized clinical trial comparing methotrexate with leflunomide, where patients (n=482 patients) identified ‘Do [ing] chores,’ ‘Stand[ing] from a chair,’ ‘Dressing self,’ ‘Get in/out of bed,’ and ‘Get[ing] down 5lb bag’ as their most affected activity among the 20 suggested physical activities. Dr. Strand indicated that all of them are queried in the HAQ-DI or PF domain of SF-36.

6.2. Health Assessment Questionnaire Disability Index

Dr. Singh reviewed the development and documentation of the psychometric performance of the HAQ-DI. Further details are attached in Dr. Singh’s presentation.

Content validity: The HAQ-DI queries activities of daily living but does not cover instrumental and discretionary activities such as shopping or climbing flights of stairs. At OMERACT workgroups, RA patients have confirmed that the items are relevant. Over time, the HAQ-DI has become the gold standard to assess physical function in RA.

Construct validity: Evidence for the construct validity of the HAQ-DI was provided by moderate correlations with patient reported VAS scores of pain and global assessment of disease activity. Joint counts, gender, ESR, age, and disease duration are significant predictors of higher HAQ scores, which reflects more impairment in physical function. Ceiling and floor effects have been shown in patients with mild disease severity (i.e., patients report a HAQ score of zero but still have some limitations in instrumental ADLs; this represents approximately ten percent of the RA patient population). As HAQ scores increase with longer disease duration, some losses in physical function may become irreversible.

Responsiveness: In RA trials, the HAQ-DI has shown to be responsive to treatment; it discriminates well between traditional as well as biologic DMARDs and control treatment, with moderate to large effect sizes, as well as between monotherapy versus combination therapy in both early RA and established RA.

⇒ *Additional comment:* One of the criticisms of the HAQ-DI is related to the inclusion of aid devices and its weighting at the item level. Yet, as patients are moving towards remission and are not using aids or not asking for help in performing physical activities, this has become less of an issue.

6.3. The 36-item Short Form Health Survey (SF-36) Physical Function domain

Vibeke Strand, MD was tasked to review the 10-item physical function (PF) scale of the SF-36 as a scale of interest to document physical function in patients with RA. Dr. Strand indicated that she was concerned with using only a part of a well-validated instrument and strongly recommended against utilizing only one domain of the SF-36. Below is a summary of the properties of the SF-36 and its PF scale; further details are attached in Dr. Strand's presentation.

The differences between the HAQ-DI and SF-36 PF scale are fairly clear.

- The self-care activities of daily living are covered by the HAQ-DI; The PF scale covers limitations of physical function, and includes instrumental and discretionary activities including vigorous activity, such as running, climbing several flights of stairs, walking a mile, lifting heavy objects, participating in strenuous sports; moderate activity, such as moving a table, pushing a vacuum cleaner, bowling or playing golf and lifting or carrying groceries. The PF scale is well-correlated with the HAQ, but these are moderate correlations indicating that they are not redundant instruments.
- The SF-36 is a generic “legacy” instrument that has been widely used to assess health status not only in RA but in most rheumatic diseases as well as diabetes, cardiovascular and pulmonary diseases, and many others.
- Normative population scores exist for most countries (i.e., RA patients can compare their actual health status with normative values).
- Unlike the HAQ, the SF-36 doesn't appear to have ceiling effects. The lower the domain scores at baseline, the more improvement is typically seen in RA with DMARD therapy, biologic and synthetic.

Responsiveness of the SF-36 PF scale and SF-36: Moderate to large effect sizes in treatment-associated changes in RA across populations with different disease durations, treatment failures and ‘damage’ have been documented.

⇒ *Additional consideration:* Dr. Strand pointed out that the SF-36 summary scores, PCS and MCS, are based on positive and negative weighting of all eight SF-36 domains and thereby “combine” multiple constructs; furthermore their scores are not independent of each other.

6.4. General discussion for a research agenda regarding RA-defining decrements in physical functioning

- 6.4.1.** Consensus was reached that the HAQ-DI should not be modified by the RA Working Group at this time because other efforts are underway to address unmet measurement needs as discussed below.
- 6.4.2.** Consensus was reached that physical functioning was not comprehensively captured by the HAQ-DI, particularly as we move toward lower disease activity states and patients with lower levels of disability and damage are included in clinical trials. There is an unmet need for instruments that could show responsiveness to treatment in patients with a more normal range of function, including patients who have been adequately treated and are still in the normal range (referred to as ‘undamaged arthritis patients’).
- 6.4.3.** Consensus was reached that there was not a need to develop a new measure to document decrements in physical functioning.

It was suggested that other measures, such as the SF-36 PF and relevant domains of PROMIS could be included in clinical trials as secondary endpoints. It was noted that PROMIS actually includes most of the questions from the HAQ and the physical function domain of SF-36.

Experts suggested that revised versions of the HAQ-DI, primarily the new PROMIS HAQ-DI using computer-adaptive testing (CAT), are likely to be more sensitive to change than legacy instruments. It should provide more information about experiences of RA patients with respect to their functional improvement, including patients at the lower disability end of the disease spectrum. For these patients, the HAQ didn’t discriminate well among levels of functional disability because it didn’t have enough questions in that range.

Laurie Burke commented that FDA was completely on board with using new measurement approaches of item response theory and other theories that can show quantitatively that an instrument is covering the full range of the targeted population and therefore, is a sensitive instrument across the entire spectrum of the clinical trial population in terms of detecting treatment effectiveness. However, Laurie indicated that use of CAT approaches as endpoint measures has not been reviewed at FDA to date and cannot be recommended at this time for use in clinical trials to support labeling claims of treatment benefit.

As a path forward, the experts indicated that candidate instruments such as the PROMIS HAQ and PROMIS physical function (PF) items need to be tested alongside legacy instruments in randomized trials and in other settings to see which of these will have incremental value in assessing any unmeasured aspect of physical function limitation in patients on either side of the disease activity spectrum. Dr.

Felson also stated that the use of the PROMIS PF in clinical research was moving forward without need of involvement of the RA Working Group.

7. Closing remarks

Throughout the day, fatigue, participation, stiffness and decrements in physical functioning were discussed as potential concepts to target in qualifying new or existing PRO measures.

- Workshop participants agreed that the concept of decrements in physical function is not comprehensively captured by the HAQ-DI. However, since other PRO instruments (PROMIS Physical Function and PROMIS HAQ-DI) are currently being investigated to determine their incremental value to support endpoints in RA, there was agreement that the RA Working Group remove this concept from its research agenda.
- Workshop participants agreed that stiffness is an important concept of measurement that requires further investigation; however, considering that other researchers (e.g., OMERACT) are currently collecting data on stiffness, there was agreement that this concept should not be a top priority for the RA Working Group at this time.
- Workshop participants agreed that both fatigue and participation are important concepts of measurement to pursue qualification of PRO instruments. While both are important, final comments by several of the experts and a show of hands indicated that a majority of workshop participants supported the RA Working Group moving forward with fatigue. It was agreed that the RA Working Group should make use of the existing body of work available on fatigue, including trial datasets of registered products, and should focus on development of a measure for FDA qualification of the concept of fatigue to support a secondary endpoint to document treatment benefit.