

# The Assessment of Disease Activity in Rheumatic Diseases

Guest Editors: Thurayya Arayssi, Zahi Touma, Mandana Nikpour,  
and Lilian Ghandour





---

# **The Assessment of Disease Activity in Rheumatic Diseases**

International Journal of Rheumatology

---

**The Assessment of Disease  
Activity in Rheumatic Diseases**

Guest Editors: Thurayya Arayssi, Zahi Touma, Mandana Nikpour,  
and Lilian Ghandour



---

Copyright © 2013 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in "International Journal of Rheumatology." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

---

## Editorial Board

Salvatore M.F. Albani, USA  
Ernest Brahn, USA  
Ruben Burgos-Vargas, Mexico  
Dirk Elewaut, Belgium  
Luis R. Espinoza, USA  
Barri J. Fessler, USA  
Piet Geusens, Belgium  
Hiroshi Hashimoto, Japan  
Tim Jansen, The Netherlands

Kenneth C. Kalunian, USA  
Shinichi Kawai, Japan  
Charles J. Malemud, USA  
Johanne Martel-Pelletier, Canada  
Terry L. Moore, USA  
Thomas Pap, Germany  
K. Pavelka, Czech Republic  
Proton Rahman, Canada  
Morris Reichlin, USA

Lisa Rider, USA  
Vicente Rodriguez-Valverde, Spain  
Bruce M. Rothschild, USA  
J. R. Seibold, USA  
Leonard H. Sigal, USA  
Malcolm Smith, Australia  
George Tsokos, USA  
Ronald F. van Vollenhoven, Sweden  
Muhammad B. Yunus, USA

# Contents

---

**The Assessment of Disease Activity in Rheumatic Diseases**, Thurayya Arayssi, Zahi Touma, Mandana Nikpour, and Lilian Ghandour  
Volume 2013, Article ID 275691, 2 pages

**The Utility of Outcome Measures in Total Knee Replacement Surgery**, Michelle M. Dowsey and Peter F. M. Choong  
Volume 2013, Article ID 506518, 8 pages

**Comparison of the Systemic Lupus Erythematosus Activity Questionnaire and the Systemic Lupus Erythematosus Disease Activity Index in a Black Barbadian Population**, Kim R. Quimby, Cindy Flower, Ian R. Hambleton, R. Clive Landis, and Anselm J. M. Hennis  
Volume 2013, Article ID 875369, 5 pages

**Disease Activity Measures in Paediatric Rheumatic Diseases**, Nadia J. Luca and Brian M. Feldman  
Volume 2013, Article ID 715352, 10 pages

**Detailed Joint Region Analysis of the 7-Joint Ultrasound Score: Evaluation of an Arthritis Patient Cohort over One Year**, S. Ohrndorf, B. Halbauer, P. Martus, B. Reiche, T. M. Backhaus, G. R. Burmester, and M. Backhaus  
Volume 2013, Article ID 493848, 9 pages

**Measuring Disease Activity in Psoriatic Arthritis**, Priscilla C. H. Wong, Ying-Ying Leung, Edmund K. Li, and Lai-Shan Tam  
Volume 2012, Article ID 839425, 10 pages

## Editorial

# The Assessment of Disease Activity in Rheumatic Diseases

**Thurayya Arayssi,<sup>1</sup> Zahi Touma,<sup>2</sup> Mandana Nikpour,<sup>3</sup> and Lilian Ghandour<sup>4</sup>**

<sup>1</sup> Weill Cornell Medical College in Qatar, Doha 24144, Qatar

<sup>2</sup> Centre for Prognosis Studies in the Rheumatic Diseases, University of Toronto, Toronto, ON, Canada M5T 2S8

<sup>3</sup> The University of Melbourne, St. Vincent's Hospital, 41 Victoria Parade Fitzroy, Melbourne, VIC 3065, Australia

<sup>4</sup> Faculty of Health Sciences, American University of Beirut, Beirut 1107 2020, Lebanon

Correspondence should be addressed to Thurayya Arayssi; [tha2002@qatar-med.cornell.edu](mailto:tha2002@qatar-med.cornell.edu)

Received 23 October 2013; Accepted 23 October 2013

Copyright © 2013 Thurayya Arayssi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disease activity in the rheumatic diseases can be defined as a reversible state, manifested by clinical, laboratory, or radiographical features. Disease activity occurs principally as a result of immunologic and inflammatory processes involving specific organs at a specific point in time. The multifaceted nature of clinical presentations in adults as well as in pediatric rheumatic diseases makes the assessment of disease activity challenging. Nevertheless, this assessment is fundamental in patient care and based on the use of valid, reliable, and interpretable instruments. The use of disease activity instruments enables clinicians, patients, and researchers to quantify and evaluate disease activity in a standardized way. The application of these instruments in a clinical care and research setting presents several challenges: namely, the administrative burden of the instrument, including the preparedness and skillfulness of the assessor, the mode of administration, the time required to complete the instrument, and the complexity of scoring. All of these factors need to be taken into consideration when choosing an instrument applicable in a particular setting.

In the last decade, we have witnessed the emergence of a plethora of instruments developed to measure disease activity in the rheumatic diseases. This special issue contains five papers: four related to the assessment of activity in adult rheumatic diseases and one paper with a focus on pediatric rheumatic disease.

Wong et al. "Measuring disease activity in psoriatic arthritis" provide a detailed review of the assessment of disease activity in psoriatic arthritis. Their paper highlights the currently available tools for the assessment of the various

rheumatologic and dermatologic aspects of the disease and discusses the composite indices that have been developed, or are still under development.

Dowsey and Choong "The utility of outcome measures in total knee replacement surgery" discuss the utility of outcome measures for total knee replacement surgery, most commonly performed to treat osteoarthritis. While the majority of patients have a substantial improvement in symptoms following this procedure, a proportion of patients report ongoing pain and poor function. The authors remind us that accurate and reproducible measurement of residual pain and functional impairment in patients undergoing arthroplasty forms an important basis for optimizing outcomes of this procedure. The paper by Dowsey et al. challenges our perception of disease "activity" in the rheumatic diseases and allows us to consider the overlap with the broader concept of "severity," particularly in osteoarthritis, which does not traditionally follow a relapsing-remitting course.

In an original research article, Ohrndorf et al. "Detailed joint region analysis of the 7-joint ultrasound score: evaluation of an arthritis patient cohort over one year" use ultrasound technique, specifically the 7-joint ultrasound (US7) score, to evaluate disease activity and the presence of erosions over one year in a cohort of patients with predominantly rheumatoid and psoriatic arthritis. They demonstrate a high synovitis score at baseline in the majority of patients, which diminishes over one year of treatment, corresponding with a decline in the DAS28 score. The responsiveness over time of ultrasound-based assessment of disease activity in inflammatory arthritis, compared with clinical-laboratory based assessment, and the

prognostic significance of mild persistent synovitis detectable on ultrasound are yet to be conclusively demonstrated and quantified. However, the study by Ohrndorf et al. highlights the potential usefulness of imaging modalities in assessment of activity in the rheumatic diseases.

Quimby et al. "*Comparison of the systemic lupus erythematosus activity questionnaire and the systemic lupus erythematosus disease activity index in a black barbadian population*" evaluate the use of the Systemic Lupus Activity Questionnaire (SLAQ), vis-à-vis the internationally accepted and validated Systemic Lupus Erythematosus Disease Activity Index (SLEDAI), and the physician global assessment (PGA). They conclude that SLAQ results in over-reporting of symptoms and is an inadequate disease monitoring tool. They underscore the use of laboratory measurements and highlight the need for a modified version that would be valid, feasible, and available to countries most burdened by SLE, yet least equipped to diagnose and manage the disease.

Finally, Luca's and Feldman's "*Disease activity measures in paediatric rheumatic diseases*" provide a comprehensive review of assessment of disease activity in the pediatric population. Their paper explains the theoretical framework required for the development of disease activity measures citing the challenges encountered in this particular patient population. In addition, they summarize the most common disease activity measures used for Juvenile Idiopathic Arthritis (JIA), Juvenile Systemic Lupus Erythematosus (JSLE), and Juvenile Dermatomyositis (JDM) reminding us that more research is needed to determine the most appropriate measures to be used in clinical practice and the research environment.

The assessment of disease activity in rheumatic diseases is very challenging but essential. Significant achievements have been made in improving the measurement properties of the instruments. Further research is needed to identify the measurement tools that are of the highest quality and can be conveniently used in clinics and in research studies.

## **Acknowledgments**

We would like to thank all the authors for contributing to this special issue and all the reviewers for volunteering their valuable time in reviewing the papers.

*Thurayya Arayssi  
Zahi Touma  
Mandana Nikpour  
Lilian Ghandour*

## Review Article

# The Utility of Outcome Measures in Total Knee Replacement Surgery

**Michelle M. Dowsey and Peter F. M. Choong**

*The University of Melbourne, Department of Surgery, Saint Vincent's Hospital, 41 Victoria Parade, Fitzroy, Melbourne, VIC 3065, Australia*

Correspondence should be addressed to Peter F. M. Choong; [pchoong@unimelb.edu.au](mailto:pchoong@unimelb.edu.au)

Received 9 June 2013; Accepted 15 September 2013

Academic Editor: Lilian Ghandour

Copyright © 2013 M. M. Dowsey and P. F. M. Choong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Total knee replacement (TKR) is the mainstay of treatment for people with end-stage knee OA among suitably “fit” candidates. As a high cost, high volume procedure with a worldwide demand that continues to grow it has become increasingly popular to measure response to surgery. While the majority who undergo TKR report improvements in pain and function following surgery, a significant proportion of patients report dissatisfaction with surgery as a result of ongoing pain or poor function. Poor outcomes of TKR require care that imposes on already overburdened health systems. Accurate and meaningful capture and interpretation of outcome data are imperative for appropriate patient selection, informing those at risk, and for developing strategies to mitigate the risk of poor results and dissatisfaction. The ways in which TKR outcomes are captured and analysed, the level of follow-up, the types of outcome measures used, and the timing of their application vary considerably within the literature. With this in mind, we reviewed four of the most commonly used joint specific outcome measures in TKR. We report on the utility, strengths, and limitations of the Oxford knee score (OKS), knee injury and osteoarthritis outcome score (KOOS), Western Ontario and McMaster Universities osteoarthritis index (WOMAC), and knee society clinical rating system (KSS).

## 1. Background

Total knee replacement is a major surgical procedure that requires multidisciplinary input prior to and after surgery to ensure the best possible outcome. Recovery from surgery is optimized with the inclusion of rehabilitation programs which are tailored to restore mobility and independence [1]. Time to recovery can vary following TKR, and most patients will report substantial gains between 3 and 6 months after surgery [2, 3]. Overall, a continuing pattern of improvement can be observed up to 12 months following surgery [4, 5]. While a majority of patients report improvements in pain and function following total knee replacement [6, 7], a substantial number of individuals do not meet the level of improvement expected at 12 months or more after surgery [8, 9].

A number of individual characteristics are known to influence pain and function after surgery [10]. Individual risk factors which impact on patient outcomes after TKR

include age and gender [7, 11, 12], antecedent diagnosis [13], body mass index [14, 15], ethnicity [16], psychological distress [13, 17], baseline pain and functional disability [7, 13], comorbidity profile [10, 18], socioeconomic status [19], and radiographic osteoarthritis severity [7, 20]. Some of these, such as obesity and psychological distress, are potentially modifiable, making accurate and meaningful capture and interpretation of outcome data imperative for both informing those at risk and for developing strategies to mitigate the risk of poor results and dissatisfaction.

Rates of ongoing knee pain and functional impairment following TKR vary considerably in the literature, ranging from 14% to 44% of individuals reporting persistent pain [7, 9, 21, 22] and from 20% to 50% of individual was reporting functional impairment [7, 22, 23] in the first 12 to 24 months following surgery. Of note the way in which data is captured and analysed, the level of follow-up, the types of patient-reported outcome measures (PROMs) used and the timing

of their application also vary considerably between these studies. Numerous instruments for measuring the outcomes of TKR exist; however, not all of them contain the necessary attributes of a “good” outcome measure. When selecting which measure to use, consideration should be given to whether the measure is appropriate for use specific to the procedure being assessed. A good outcome measure should be accessible, have demonstrated reliability and validity, place minimal burden on responders, and be responsive to change [24]. High floor and ceiling effects indicate insensitivity for detecting a change of symptoms and the maximum cut-off for floor and/or ceiling effects should be no more than 15% [24]. With this in mind, we reviewed the four most commonly used joint specific outcome measures in TKR and report on their utility, strengths, and limitations.

## 2. Oxford Knee Score (OKS)

The OKS is a knee joint specific 12-item questionnaire originally developed and validated in 1998 for use in randomised controlled trials in total knee replacement (Table 1) [25]. The OKS has 12 items, 5 for assessing pain and 7 for assessing function. Each item is worth equal weighting (1 to 5) for a total possible score ranging from 12 to 60. A lower score indicates a better outcome. The OKS is freely available at [http://phi.uhce.ox.ac.uk/ox\\_scores.php](http://phi.uhce.ox.ac.uk/ox_scores.php) and widely used in cohort studies and by some joint replacement registries [9, 26, 27]. A scoring manual, list of translations, and licensing information can be found via <http://www.isis-innovation.com/outcomes/orthopaedic/oks.html>.

The OKS is designed specifically for measuring outcomes in knee replacement. The OKS has also been used to evaluate pharmacological and conservative interventions and other knee surgery procedures in knee osteoarthritis (OA) [28]. Cross-cultural adaptations in Thai, British, Swedish, Portuguese, Dutch, German, Italian, Japanese Chinese, French, and Korean languages have been validated [29, 30]. Given the simplicity and brevity of the questionnaire, higher response rates have been reported than for other PROMs [25]; however, this is not always consistent [31].

Completion and scoring of the OKS is simple; each of 12 questions carries equal weighting (1 to 5) to provide an overall score between 12 and 60 [25]. An updated scoring method is also used, whereby each item is scored between 0 (worst outcome) and 4 (best outcome), to provide an overall score between 0 and 48 [28]. The OKS is patient administered and should take about 5 minutes to complete, and responses are based on symptoms in the preceding 4 weeks. Two missing values are accepted, and where this occurs should be replaced by the mean score for the missing item [25]. The outcome categories for the OKS have been reported based on the following cut points: excellent (>41), good (34–41), fair (27–33), and poor (<27) [32, 33]. However, these categories have not been validated and are neither commonly used nor recommended [28].

The minimum clinically important difference (MCID) estimates for the OKS as reported by Murray et al. [28]

are between 3 and 5 points. These estimates are based on half the standard deviation of change in OKS scores which Murray et al. report to be is between 6 and 10 points for joint replacement studies. This interpretation is based on a systematic review of health-related quality of life instruments by Norman et al. who concluded that, in most circumstances, the threshold of discrimination for changes in health-related quality of life for chronic diseases appears to be approximately half a standard deviation of the change in outcome score [34]. In a recent study, Judge et al. reported that an 11-point or more absolute change in the OKS at 6 months after TKR discriminated the best between patients' satisfaction and a 6-month OKS  $\geq 30$  points identified the highest level of satisfaction [35]. A weak floor effect (7%) has been reported for the OKS prior to TKR [36]; however, ceiling effects were reported at 6 months, (14%) and 12 months, (22%) following surgery, but this was attributed to patients attaining an optimal outcome rather than a limitation of the OKS [37].

## 3. Knee Injury and Osteoarthritis Outcome Score (KOOS)

The KOOS is a knee joint specific questionnaire developed in 1998 originally for the purpose of evaluating short-term and long-term symptoms and functioning in subjects with knee injury and osteoarthritis (Table 1). It was originally validated in patients undergoing anterior cruciate ligament ACL reconstruction [38]. The KOOS is a 42-item survey designed to assess people's opinions about the difficulties they experience with activity due to problems with their knees. A higher score indicates a better outcome. The questionnaire, scoring instructions, and translations are freely available at <http://www.koos.nu/>. The KOOS is widely used in younger and/or more active patients with knee injury and knee osteoarthritis [39].

The KOOS has been validated for measuring outcomes in TKR [39], ACL reconstruction [38], and posttraumatic knee OA [40]. The KOOS has also been used to evaluate other OA interventions including minor knee surgery procedures [41], conservative treatments [42, 43], and nutritional [44] and pharmacological interventions [45], and population-based reference data has been published [46]. High response rates have been reported for studies of TKR in the short term: 92% at 6 months and 86% at 12 months [47]. A short-form version (KOOS-PS) which is a 7-item questionnaire derived from the original KOOS has been validated for evaluating physical function in individuals with knee OA undergoing TKR [48]. The KOOS was originally concurrently developed in English and Swedish, and numerous cross-culturally validated and translated versions exist [49]. Translations include Austrian-German, Chinese, Croatian, Czech, Danish, Dutch, Estonian, French, German, Hindi (India), Italian, Japanese, Korean, Latvian, Lithuanian, Norwegian, Persian, Polish, Portuguese, Russian, Singapore English, Slovakian, Slovenian, Spanish (Peru), Spanish (US), Thai, Turkish, and Ukrainian.

Completion of the survey is straightforward; each of the 42 items carries equal weighting (0–4). There are 5 subscales, each measuring a specific outcome: pain (9 items), symptoms

TABLE 1: Patient outcome measures in knee osteoarthritis severity.

PROM	Scoring	Response criteria	Validated for use	Translation
OKS [25] Dawson et al. 1998	12 items (pain; 5 items, function; 7 items), each item worth 0–4 points. Range 0–48 points Categories: excellent (>41), good (34–41), fair (27–33), and poor (<27)	MCID 3–5 points change in score [28] ≥30 points at 6 months after TKR [35]	TKR	Thai, British, Swedish, Dutch, Portuguese, German, Italian, Japanese, Chinese, Korean, and French
WOMAC [50] Bellamy et al. 1988	24 items (pain; 5 items, stiffness; 2 items, and physical function; 17 items) Range from 0 to 96 points.	MCID 15 points [63] Responder criteria [52]: Improvement in pain or function ≥ 50% and absolute change ≥ 20 Or 2 of the 3 improvements as follows: (i) pain ≥ 20% and absolute change ≥ 10, (ii) function ≥ 20% and absolute change ≥ 10, (iii) global assessment ≥ 20% and absolute change ≥ 10.	TKR OA Post traumatic OA	>80 languages Arabic, Chinese, Dutch, Finnish, German, Hebrew, Italian, Japanese, Korean, Moroccan, Persian, Singapore, Spanish, Swedish, Thai, and Turkish
KSS [64] Insall et al. 1989 2011-KS [68] Scuderi et al. 2012	KSS-2 components: a knee rating (0–100 points) and function (0–100 points) Categories: excellent (≥80), good (70–79), fair (60–69), and poor (<60) 2011 KSS-5 subscales: knee score (0–50), pain score (0–25), satisfaction (0–40), expectation (0–15), and function (0–100)	MCID 34.5 points improvement in functional subscale [81]	2011 KS TKR UCKR	KSS-Portuguese, Spanish 2011-KSS-Italian, Japanese, Mandarin Chinese, Dutch, Portuguese, and Spanish
KOOS [38] ROOS et al. 1998	42 items, 5 subscales: pain (9 items), symptoms (5 items), activities of daily living (17 items), sports and recreation (5 items), and quality of life (4 items). Each item worth 0–4 points. Scores for each subscale are calculated separately and then transformed into a score between 0 and 100.	MCID 8–10 point change in score	TKR ACL Reconstruction Posttraumatic OA	Austrian-German, Chinese, Croatian, Czech, Ukrainian, Estonian, French, German, Hindi (India), Italian, Thai, Japanese, Korean, Latvian, Lithuanian, Norwegian, Persian, Polish, Portuguese, Russian, Dutch, Singapore, English, Slovakian, Spanish, Slovenian, Turkish, and Danish

(5 items), activities of daily living (17 items), sports and recreation function (5 items), and knee-related quality of life (4 items). Scores for each subscale should be calculated separately and then transformed into a score between 0 and 100 [38]. Scoring instructions and calculators are available at <http://www.koos.nu/>. The KOOS is patient administered and should take approximately 10–15 minutes to complete, and responses are based on symptoms in the preceding week. The process for managing missing values has been recently (2012) revised (see website), and the mean score for each subscale can be derived from a minimum response of pain (5 of 9 items), symptoms (4 of 5 items), activities of daily living (9 of 17 items), sports and recreation function (3 of 5 items), and knee-related quality of life (2 of 4 items) (<http://www.koos.nu/>). There are no categorical equivalents,

scoring of each outcome should be reported separately, and using an aggregate score is neither recommended nor valid.

The MCID estimates for the KOOS have not been established for patients undergoing TKR. However, the minimal important change (MIC) is currently suggested to be 8–10 according to the website details, while cautioning that there are a number of patients and related factors that may impact on the MIC. Floor and ceiling effects have been reported for studies of TKR in some domains of the KOOS [39]. Preoperatively, the percentage of patients undergoing TKR with the worst possible score have reached 48% for the sports and recreation domain of the KOOS. Ceiling effects at 6 months have also been reported (15% for pain scores and 16% for sports and recreation) and at 12 months (22% for pain scores and 17% for quality of life scores).

#### 4. Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)

The WOMAC was initially developed in 1982 and was first validated for the purpose of evaluating response to treatment in patients with hip and knee OA in 1998 (Table 1) [50]. The WOMAC underwent multiple subsequent revisions and refinements between 1996 and 1999 [51]. The WOMAC is a 24-item questionnaire with 3 subscales measuring pain (5 items), stiffness (2 items), and physical function (17 items). A lower score indicates a better outcome. The questionnaire, licensing information, scoring instructions, and translations are available at <http://www.womac.org/>.

Numerous validation studies have been conducted using the WOMAC [51]. The WOMAC has been validated for measuring outcomes in clinical trials of TKR [52] and for measuring treatment response of pharmacological interventions for knee OA [53]. It has also been used to evaluate many knee OA interventions, both surgical and conservative [54]. Response rates reaching 90% at 1 year for epidemiological studies in TKR have been reported [55, 56]. A short-form version (WOMAC-SF) which is a 7-item questionnaire derived from the physical function subscale of the WOMAC has been validated for assessing function in knee OA and TKR [57, 58]. The WOMAC is available in more than 80 languages and has been cross-culturally validated in Arabic, Chinese, Dutch, Finnish, German, Hebrew, Italian, Japanese, Korean, Moroccan, Persian, Singapore, Spanish, Swedish, Thai, and Turkish [24, 59].

Completion of the survey is straightforward; each of the 24 items has 5 possible responses for a possible score of 0 to 4 for each response. There are 3 subscales, each measuring a specific outcome: pain (5 items), stiffness (2 items), and physical function (17 items). Alternative versions are available using a visual analogue scale or numeric rating [60]. A total WOMAC score is calculated by summing the items for all 3 subscales, for a total score between 0 and 96 [61]. The WOMAC is patient administered and should take approximately 10 minutes to complete, and responses are based on symptoms in the preceding 48 hours. The method of managing missing values for the WOMAC is a variant of a standard mean imputation method. Scores of the nonmissing items for each case should be added, and the mean value is used for the missing values. However, if the patient has not replied to more than one of the 5 pain or 2 stiffness items or more than 4 of the 17 physical function items, then response for that scale is considered invalid and should not be included in the analyses [62].

The MCID estimates for the WOMAC as reported by Escobar et al. are 15 points [63]. In addition, Escobar et al. have validated the Outcome Measures in Rheumatology and Osteoarthritis Research Society International (OMERACT-OARSI) set of responder criteria in total joint replacement [52]. Patients are deemed responders or nonresponders based on a combination of absolute and relative changes of pain, function, and global patient's assessment. The criteria are as follows: (i) an improvement in pain or function  $\geq 50\%$  and an absolute change  $\geq 20$ , then the patient is considered a responder, and (ii) if the level of improvement does not

reach these criteria but improvement in at least two of the three following criteria, the patient will also be considered a responder, (a) pain  $\geq 20\%$  and absolute change  $\geq 10$ , and (b) function  $\geq 20\%$  and absolute change  $\geq 10$ , (c) global assessment  $\geq 20\%$  and absolute change  $\geq 10$ .

Minimal floor effects for the WOMAC have been reported with the exception of the quality of life subscale which was reported at 14% by Roos and Toksvig-Larsen [39]. Ceiling effects have been reported for TKR at both 6 months; 27% for the pain subscale and 15% for the stiffness subscale and at 12 months; 17% for the quality of life subscale, 30% for the pain subscale, and 27% for the stiffness subscale [39, 63].

#### 5. Knee Society Clinical Rating System (KSS)

The KSS is a knee joint specific questionnaire originally developed and validated in 1989 for use in assessing the outcome of total knee replacement (Table 1) [64]. The KSS has 2 components: a knee rating (0–100 points) and function (0–100 points) worth a total of 200 points. The knee rating is divided into pain (0–50 points) and a knee score which assesses range of motion, stability, and alignment (0–50 points). A higher score indicates a better outcome. The KSS is freely available at <http://www.kneesociety.org/web/index.html> and widely used in outcome studies for partial and total knee replacement. As a “clinician completed” scoring system aspects of its validity have been questioned by some authors [65–67]. In response to these criticisms, a revised knee society scoring system (2011-KS Score) has recently been developed [68] and validated [69] for measuring outcomes in TKR. A scoring manual, list of translations, and licensing information for both the KSS and 2011-KS can be found via <http://www.kneesociety.org/web/index.html>.

Despite validity issues, the KSS remains one of the most popular questionnaires amongst clinician researchers for measuring outcomes in knee replacement [68]. The KSS includes range of motion and alignment measurements, and this may in part contribute to its popularity. The importance of coronal alignment in TKR in terms of implant survival and functional outcomes has been well established in the literature [70, 71], and knee range of motion is an important marker for many activities of daily living [72]. The KSS has also been used to evaluate outcomes in other orthopaedic procedures such as high tibial osteotomy [73] and patellofemoral arthroplasty [74]. Linguistically translated versions of the KSS include Spanish [75] and Portuguese [76] and for the 2011-KS, Italian, Japanese, Mandarin Chinese, Portuguese, and Spanish. A Dutch version of the 2011-KS has also recently been validated [77]. Despite a “clinician” scoring system, the pain and function subscales of the KSS have been offered to patients to complete with high response rates reported at 12 months or more [7, 78].

Completion and scoring of the KSS are simple; the function subscale (0–100) is based on walking distance (0–50) and ability to climb stairs (0–50) with deductions for use of a gait aid (0–20). The pain subscale is (0–50) and the knee rating (0–50) is based on range of motion (0–25) and knee stability (0–25) with deductions made

dependent on the existence and severity of flexion contracture (0–15), extension lag (0–15), and malalignment (0–20) [64]. A negative score is possible and should be converted to zero. A web-based calculator is available at [http://www.orthoscores.com/scorepages/knee\\_society\\_score.html](http://www.orthoscores.com/scorepages/knee_society_score.html). Categories for the KSS have been established but not validated [79]. Cut points for each of the 2 subscales are excellent ( $\geq 80$ ), good (70–79), fair (60–69), and poor ( $< 60$ ). The KSS is clinician administered and should take less than 10 minutes to complete. The pain and function subscales should take about 5 minutes to complete whether by clinician or patient, and responses are based on symptoms in the preceding 4 weeks. No instruction on managing missing items could be found.

The 2011-KS scoring manual and instructions can be requested via the knee society website above. The 2011-KS expands on the KSS and includes subscales for patient satisfaction (5 items, 0–40 points), expectation (3 items, 0–15 points), and functional activities (19 items, 0–100 points), which is divided into functional activities (5 items, 0–30 points), standard activities (6 items, 0–30 points), advanced activities (5 items, 0–25 points), and discretionary knee activities (3 items 0–15 points) [68]. Satisfaction expectation and function should be reported as separate scores as a composite score is not recommended. The suggested method for managing missing values is to enter dummy values equal to the average of all of the other items in the same domain. This should be limited to instances where fewer than 50% of responses are missing [80].

The MCID estimates for the KSS and 2011-KS have not been identified for patients undergoing TKR. However, in a study by Jacobs and Christensen, a minimum change of 34.5 points at 3 months in the function subscale of the KSS was established as clinically important [81]. Ceiling effects have been reported for studies of TKR in both the knee (25%) and function (43%) subscales of the original KSS at 12 months [82]. Floor effects did not occur preoperatively and ceiling effects did not occur at 6 months after TKR in a Dutch study validating the KS-2011 [77].

## 6. Summary

Total knee replacement remains the mainstay of treatment for people with end-stage knee OA (among suitably “fit” candidates). As a high cost, high volume procedure with a worldwide demand that continues to grow, it is becoming increasingly important to understand the drivers behind response to surgery [24]. Poor outcomes of TKR require care that is an imposition on an already overburdened health system. Not only will there be a demand for ongoing outpatient specialist and community health consultations, persistent use of prescription medication, prolonged requirement for allied health services (physiotherapy and occupational therapy), and the possible need for repeated minor (arthroscopic) and major (revision joint replacement) surgery, these activities potentially deprive or delay other patients with untreated OA from receiving expeditious care.

Numerous instruments for measuring outcomes in TKR have been developed and validated over time [24] in an attempt to capture response to surgery and predict those who may be at risk of suboptimal results. We have presented a summary of the utility, strengths, and limitations of four of the most commonly used outcome measures for total knee replacement. Generic strengths among the four outcome measures included the relatively minimal burden required to complete each instrument and their design specific to measuring TKR outcomes, whereas ceiling and/or floor effects were a limitation to varying degrees for each of the four outcome measures, with the exception of the 2011-KS which requires further validation studies. No single outcome measure would be suitable for every foreseeable clinical situation or research activity. The individual strengths of each outcome measure may be useful in guiding the decision as to which measure is best suited for use, in any given situation. We noted a number of individual strengths amongst the outcome measures presented in this review.

The OKS is freely available and noted for its simplicity and brevity and appears to be the measure of choice for large data sets and joint registry’s [9, 26, 27]. The KOOS is also freely available and aside from TKR is valuable for measuring outcomes in younger and/or more active patients with knee injury and knee osteoarthritis. The KOOS is also used to measure outcomes following a range of both surgical and conservative interventions of the knee, both surgical and conservative, making it attractive for treatment comparisons [38–45]. WOMAC scores can also be derived from the KOOS. An important aspect of the utility of any outcome measure is the availability of responder definitions and cutoff points and that these are appropriately validated. The WOMAC is currently the only outcome measure that has validated responder definitions and cutoff points specifically for TKR [52]. Having a set of established and validated response criteria makes the WOMAC an excellent option for use in clinical trials that aim to measure response to TKR and other nonsurgical interventions [52, 53]. It also has the most extensive range of translations available. Despite validity issues, the KSS remains one of the most popular rating systems for measuring outcomes in TKR [68]. It is one of the few outcome measures that include assessment of clinical measures that are deemed important in terms of implant survival and functional outcomes [70, 71]. The 2011-KS also includes measures of patient expectation and satisfaction which are emerging as important adjuncts in measuring response to surgery [83].

While the utility of any one particular outcome measure over another continues to be debated and the number of available instruments continues to increase, we believe that there are 2 key factors that are essential in producing quality outcome data irrespective of the instrument used. Firstly, recording of baseline scores is essential for producing meaningful outcome data. It is well established that better baseline scores correlate with better outcome scores and those with the worst baseline scores demonstrate the greatest amount of improvement [13]. Therefore, at a minimum, data analyses should always be adjusted for baseline when either presenting outcome scores or measuring the change in scores.

Finally, individuals who do not respond to surveys report significantly poorer outcomes than those who do [78]. As such establishing a process for data collection that ensures the highest possible response rate such as those used by Bourne et al. will minimise nonresponder bias [83].

## Disclosure

Dr. Dowsey holds an NHMRC early career Australian clinical fellowship (APP1035810).

## References

- [1] C. A. Jones, L. A. Beaupre, D. W. Johnston, and M. E. Suarez-Almazor, "Total joint arthroplasties: current concepts of patient outcomes after surgery," *Clinics in Geriatric Medicine*, vol. 21, no. 3, pp. 527–541, 2005.
- [2] R. K. Shields, L. J. Enloe, and K. C. Leo, "Health related quality of life in patients with total hip or knee replacement," *Archives of Physical Medicine and Rehabilitation*, vol. 80, no. 5, pp. 572–579, 1999.
- [3] P. Rissanen, S. Aro, H. Sintonen, P. Slätis, and P. Paavolainen, "Quality of life and functional ability in hip and knee replacements: a prospective study," *Quality of Life Research*, vol. 5, no. 1, pp. 56–64, 1996.
- [4] N. F. Huang, M. M. Dowsey, E. Ee, J. D. Stoney, S. Babazadeh, and P. F. Choong, "Coronal alignment correlates with outcome after total knee arthroplasty: five-year follow-up of a randomized controlled trial," *Journal of Arthroplasty*, vol. 27, no. 9, pp. 1737–1741, 2012.
- [5] A.-. Nilsdotter, S. Toksvig-Larsen, and E. M. Roos, "A 5 year prospective study of patient-relevant outcomes after total knee replacement," *Osteoarthritis and Cartilage*, vol. 17, no. 5, pp. 601–606, 2009.
- [6] M. M. Dowsey, D. Liew, J. D. Stoney, and P. F. M. Choong, "The impact of obesity on weight change and outcomes at 12 months in patients undergoing total hip arthroplasty," *Medical Journal of Australia*, vol. 193, no. 1, pp. 17–21, 2010.
- [7] M. M. Dowsey, M. Nikpour, P. Dieppe, and P. F. Choong, "Associations between pre-operative radiographic changes and outcomes after total knee joint replacement for osteoarthritis," *Osteoarthritis and Cartilage*, vol. 20, no. 10, pp. 1095–1102, 2012.
- [8] H. M. K. Ghomrawi, N. F. Ferrando, L. A. Mandl, H. Do, N. Noor, and A. Gonzalez Della Valle, "How often are patient and surgeon recovery expectations for total joint arthroplasty aligned? Results of a pilot study," *HSS Journal*, vol. 7, no. 3, pp. 229–234, 2011.
- [9] V. Wylde, A. W. Blom, S. L. Whitehouse, A. H. Taylor, G. T. Pattison, and G. C. Bannister, "Patient-reported outcomes after total hip and knee arthroplasty. Comparison of midterm results," *Journal of Arthroplasty*, vol. 24, no. 2, pp. 210–216, 2009.
- [10] M. M. Dowsey and P. F. M. Choong, *Predictors of Pain and Function Following Total Joint Replacement*, edited by P. Kinov, Arthroplasty—Update, InTech, Rijeka, Croatia, 2013.
- [11] D. P. Williams, A. J. Price, D. J. Beard et al., "The effects of age on patient-reported outcome measures in total knee replacements," *Bone and Joint Journal*, vol. 95, no. 1, pp. 38–44, 2013.
- [12] S. S. Liu, A. Buvanendran, J. P. Rathmell et al., "A cross-sectional survey on prevalence and risk factors for persistent postsurgical pain 1 year after total hip and knee replacement," *Regional Anesthesia and Pain Medicine*, vol. 37, no. 4, pp. 415–422, 2012.
- [13] A. Judge, N. K. Arden, C. Cooper et al., "Predictors of outcomes of total knee replacement surgery," *Rheumatology*, vol. 51, no. 10, pp. 1804–1813, 2012.
- [14] M. M. Dowsey, D. Liew, J. D. Stoney, and P. F. Choong, "The impact of pre-operative obesity on weight change and outcome in total knee replacement: a prospective study of 529 consecutive patients," *Journal of Bone and Joint Surgery B*, vol. 92, no. 4, pp. 513–520, 2010.
- [15] M. J. McElroy, R. Pivec, K. Issa, S. F. Harwin, and M. A. Mont, "The effects of obesity and morbid obesity on outcomes in TKA," *The Journal of Knee Surgery*, vol. 26, no. 2, pp. 83–88, 2013.
- [16] M. M. Dowsey, M. L. Broadhead, J. D. Stoney, and P. F. Choong, "Outcomes of total knee arthroplasty in English- versus non-English-speaking patients," *Journal of Orthopaedic Surgery*, vol. 17, no. 3, pp. 305–309, 2009.
- [17] M. G. Paulsen, M. M. Dowsey, D. Castle, and P. F. Choong, "Preoperative psychological distress and functional outcome after knee replacement," *Australian and New Zealand Journal of Surgery*, vol. 81, no. 10, pp. 681–687, 2011.
- [18] G. A. Hawker, E. M. Badley, C. M. Borkhoff et al., "Which patients are most likely to benefit from total joint arthroplasty?" *Arthritis and Rheumatism*, vol. 65, no. 5, pp. 1243–1252, 2013.
- [19] N. D. Clement, P. J. Jenkins, D. MacDonald et al., "Socioeconomic status affects the Oxford knee score and short-form 12 score following total knee replacement," *Bone and Joint Journal*, vol. 95, no. 1, pp. 52–58, 2013.
- [20] M. M. Dowsey, P. Dieppe, S. Lohmander, D. Castle, D. Liew, and P. F. M. Choong, "The association between radiographic severity and pre-operative function in patients undergoing primary knee replacement for osteoarthritis," *Knee*, vol. 19, no. 6, pp. 860–865, 2012.
- [21] T. Czurda, P. Fennema, M. Baumgartner, and P. Ritschl, "The association between component malalignment and post-operative pain following navigation-assisted total knee arthroplasty: results of a cohort/nested case-control study," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 18, no. 7, pp. 863–869, 2010.
- [22] V. Wylde, S. Hewlett, I. D. Learmonth, and P. Dieppe, "Persistent pain after joint replacement: prevalence, sensory qualities, and postoperative determinants," *Pain*, vol. 152, no. 3, pp. 566–572, 2011.
- [23] J. A. Singh, M. O'Byrne, S. Harmsen, and D. Lewallen, "Predictors of moderate-severe functional limitation after primary total knee arthroplasty (TKA): 4701 TKAs at 2-years and 2935 TKAs at 5-years," *Osteoarthritis and Cartilage*, vol. 18, no. 4, pp. 515–521, 2010.
- [24] N. J. Collins and E. M. Roos, "Patient-reported outcomes for total hip and knee arthroplasty: commonly used instruments and attributes of a, "good" measure," *Clinics in Geriatric Medicine*, vol. 28, no. 3, pp. 367–394, 2012.
- [25] J. Dawson, R. Fitzpatrick, D. Murray, and A. Carr, "Questionnaire on the perceptions of patients about total knee replacement," *The British Editorial Society of Bone and Joint Surgery*, vol. 80, no. 1, pp. 63–69, 1998.
- [26] New Zealand Orthopaedic Association, *New Zealand Joint Registry 2012*, New Zealand Orthopaedic Association, 2013.
- [27] D. P. Williams, C. M. Blakey, S. G. Hadfield, D. W. Murray, A. J. Price, and R. E. Field, "Long-term trends in the Oxford knee score following total knee replacement," *Bone and Joint Journal*, vol. 95, no. 1, pp. 45–51, 2013.

- [28] D. W. Murray, R. Fitzpatrick, K. Rogers et al., "The use of the Oxford hip and knee scores," *Journal of Bone and Joint Surgery B*, vol. 89, no. 8, pp. 1010–1014, 2007.
- [29] I. S. Eun, O. G. Kim, C. K. Kim, H. S. Lee, and J. S. Lee, "Validation of the Korean version of the Oxford Knee Score in patients undergoing total knee arthroplasty," *Clinical Orthopaedics and Related Research*, vol. 471, no. 2, pp. 600–605, 2013.
- [30] J. Y. Jenny and Y. Diesinger, "Validation of a French version of the Oxford knee questionnaire," *Orthopaedics and Traumatology: Surgery and Research*, vol. 97, no. 3, pp. 267–271, 2011.
- [31] S. L. Whitehouse, A. W. Blom, A. H. Taylor, G. T. Pattison, and G. C. Bannister, "The Oxford knee score, problems and pitfalls," *Knee*, vol. 12, no. 4, pp. 287–291, 2005.
- [32] P. Moonot, G. A. Medalla, D. Matthews, Y. Kalairajah, and R. E. Field, "Correlation between the Oxford knee and American knee society scores at mid-term follow-up," *The Journal of Knee Surgery*, vol. 22, no. 3, pp. 226–230, 2009.
- [33] P. B. Pynsent, D. J. Adams, and S. P. Disney, "The Oxford hip and knee outcome questionnaires for arthroplasty," *Journal of Bone and Joint Surgery B*, vol. 87, no. 2, pp. 241–248, 2005.
- [34] G. R. Norman, J. A. Sloan, and K. W. Wyrwich, "Interpretation of changes in health-related quality of life the remarkable universality of half a standard deviation," *Medical Care*, vol. 41, no. 5, pp. 582–592, 2003.
- [35] A. Judge, N. K. Arden, A. Kiran et al., "Interpretation of patient-reported outcomes for hip and knee replacement surgery: identification of thresholds associated with satisfaction with surgery," *Journal of Bone and Joint Surgery B*, vol. 94, no. 3, pp. 412–418, 2012.
- [36] J. Y. Jenny and Y. Diesinger, "The Oxford knee score: compared performance before and after knee replacement," *Orthopaedics and Traumatology: Surgery and Research*, vol. 98, no. 4, pp. 409–412, 2012.
- [37] R. G. Marx, E. C. Jones, N. C. Atwan, R. F. Closkey, E. A. Salvati, and T. P. Sculco, "Measuring improvement following total hip and knee arthroplasty using patient-based measures of outcome," *Journal of Bone and Joint Surgery A*, vol. 87, no. 9, part 1, pp. 1999–2005, 2005.
- [38] E. M. Roos, H. P. Roos, L. S. Lohmander, C. Ekdahl, and B. D. Beynon, "Knee injury and osteoarthritis outcome score (KOOS)—development of a self-administered outcome measure," *Journal of Orthopaedic and Sports Physical Therapy*, vol. 28, no. 2, pp. 88–96, 1998.
- [39] E. M. Roos and S. Toksvig-Larsen, "Knee injury and osteoarthritis outcome score (KOOS)—validation and comparison to the WOMAC in total knee replacement," *Health and Quality of Life Outcomes*, vol. 1, article 17, 2003.
- [40] E. M. Roos, H. P. Roos, and L. S. Lohmander, "WOMAC osteoarthritis index—additional dimensions for use in subjects with post-traumatic osteoarthritis of the knee," *Osteoarthritis and Cartilage*, vol. 7, no. 2, pp. 216–221, 1999.
- [41] K. B. Hare, L. S. Lohmander, R. Christensen, and E. M. Roos, "Arthroscopic partial meniscectomy in middle-aged patients with mild or no knee osteoarthritis: a protocol for a double-blind, randomized sham-controlled multi-centre trial," *BMC Musculoskeletal Disorders*, vol. 14, article 71, 2013.
- [42] M. Saleki, T. Ahadi, M. Razi, G. R. Raeisi, B. Forough, and M. K. Ali, "Comparison of the effects of acupuncture and isometric exercises on symptom of knee osteoarthritis," *International Journal of Preventive Medicine*, vol. 4, supplement 1, pp. S73–S77, 2013.
- [43] G. A. Ghasemi, A. Golkar, and S. M. Marandi, "Effects of hata yoga on knee osteoarthritis," *International Journal of Preventive Medicine*, vol. 4, supplement 1, pp. S133–S138, 2013.
- [44] B. F. Riecke, R. Christensen, P. Christensen et al., "Comparing two low-energy diets for the treatment of knee osteoarthritis symptoms in obese patients: a pragmatic randomized clinical trial," *Osteoarthritis and Cartilage*, vol. 18, no. 6, pp. 746–754, 2010.
- [45] S. T. Skou, E. M. Roos, M. B. Laursen et al., "Efficacy of multimodal, systematic non-surgical treatment of knee osteoarthritis for patients not eligible for a total knee replacement: a study protocol of a randomised controlled trial," *BMJ Open*, vol. 2, no. 6, Article ID e002168, 2012.
- [46] P. T. Paradowski, S. Bergman, A. Sundén-Lundius, L. S. Lohmander, and E. M. Roos, "Knee complaints vary with age and gender in the adult population. Population-based reference data for the knee injury and osteoarthritis outcome score (KOOS)," *BMC Musculoskeletal Disorders*, vol. 7, article 38, 2006.
- [47] M. A. Peer and J. Lane, "The knee injury and osteoarthritis outcome score (KOOS): a review of its psychometric properties in people undergoing total knee arthroplasty," *Journal of Orthopaedic and Sports Physical Therapy*, vol. 43, no. 1, pp. 20–28, 2013.
- [48] A. M. Davis, A. V. Perruccio, M. Canizares et al., "Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis," *Osteoarthritis and Cartilage*, vol. 17, no. 7, pp. 843–847, 2009.
- [49] R. S. Gonçalves, J. Cabri, J. P. Pinheiro, and P. L. Ferreira, "Cross-cultural adaptation and validation of the Portuguese version of the knee injury and osteoarthritis outcome score (KOOS)," *Osteoarthritis and Cartilage*, vol. 17, no. 9, pp. 1156–1162, 2009.
- [50] N. Bellamy, W. W. Buchanan, C. H. Goldsmith, J. Campbell, and L. W. Stitt, "Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee," *Journal of Rheumatology*, vol. 15, no. 12, pp. 1833–1840, 1988.
- [51] N. Bellamy, "WOMAC: a 20-year experiential review of a patient-centered self-reported health status questionnaire," *Journal of Rheumatology*, vol. 29, no. 12, pp. 2473–2476, 2002.
- [52] A. Escobar, M. Gonzalez, J. M. Quintana et al., "Patient acceptable symptom state and OMERACT-OARSI set of responder criteria in joint replacement. Identification of cut-off values," *Osteoarthritis and Cartilage*, vol. 20, no. 2, pp. 87–92, 2012.
- [53] T. Pham, D. van der Heijde, R. D. Altman et al., "OMERACT-OARSI initiative: osteoarthritis research society international set of responder criteria for osteoarthritis clinical trials revisited," *Osteoarthritis and Cartilage*, vol. 12, no. 5, pp. 389–399, 2004.
- [54] N. J. Collins, D. Misra, D. T. Felson, K. M. Crossley, and E. M. Roos, "Measures of knee function: international knee documentation committee (IKDC) subjective knee evaluation form, knee injury and osteoarthritis outcome score (KOOS), knee injury and osteoarthritis outcome score physical function short form (KOOS-PS), knee outcome survey activities of daily living scale (KOS-ADL), lysholm knee scoring scale, Oxford knee score (OKS), Western Ontario and McMaster," *Arthritis Care and Research*, vol. 63, supplement 11, pp. S208–S228, 2011.

- [55] I. Papakostidou, Z. H. Dailiana, T. Papapolychroniou et al., "Factors affecting the quality of life after total knee arthroplasties: a prospective study," *BMC Musculoskeletal Disorders*, vol. 13, article 116, 2012.
- [56] V. Wylde, S. Dixon, and A. W. Blom, "The role of preoperative self-efficacy in predicting outcome after total knee replacement," *Musculoskeletal Care*, vol. 10, no. 2, pp. 110–118, 2012.
- [57] G. Baron, F. Tubach, P. Ravaud, I. Logeart, and M. Dougados, "Validation of a short form of the Western Ontario and McMaster Universities osteoarthritis Index function subscale in hip and knee osteoarthritis," *Arthritis Care and Research*, vol. 57, no. 4, pp. 633–638, 2007.
- [58] S. L. Whitehouse, E. A. Lingard, J. N. Katz, and I. D. Learmonth, "Development and testing of a reduced WOMAC function scale," *Journal of Bone and Joint Surgery B*, vol. 85, no. 5, pp. 706–711, 2003.
- [59] H. Nadrian, N. Moghimi, E. Nadrian et al., "Validity and reliability of the Persian versions of WOMAC osteoarthritis index and Lequesne algofunctional index," *Clinical Rheumatology*, vol. 31, no. 7, pp. 1097–1102, 2012.
- [60] N. Bellamy, Ed., *WOMAC Osteoarthritis Index User Guide*, Version 5, 2002.
- [61] ARHP Research Committee, Western Ontario and McMaster Universities Osteoarthritis Index, 2013, [http://www.rheumatology.org/Practice/Clinical/Clinicianresearchers/Outcomes\\_Instrumentation/Western\\_Ontario\\_and\\_McMaster\\_Universities\\_Osteoarthritis\\_Index\\_\(WOMAC\)/](http://www.rheumatology.org/Practice/Clinical/Clinicianresearchers/Outcomes_Instrumentation/Western_Ontario_and_McMaster_Universities_Osteoarthritis_Index_(WOMAC)/).
- [62] H. M. K. Ghomrawi, L. A. Mandl, J. Rutledge, M. M. Alexiades, and M. Mazumdar, "Is there a role for expectation maximization imputation in addressing missing data in research using WOMAC questionnaire? Comparison to the standard mean approach and a tutorial," *BMC Musculoskeletal Disorders*, vol. 12, article 109, 2011.
- [63] A. Escobar, J. M. Quintana, A. Bilbao, I. Aróstegui, I. Lafuente, and I. Vidaurreta, "Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement," *Osteoarthritis and Cartilage*, vol. 15, no. 3, pp. 273–280, 2007.
- [64] J. N. Insall, L. D. Dorr, R. D. Scott, and W. N. Scott, "Rationale of the knee society clinical rating system," *Clinical Orthopaedics and Related Research*, no. 248, pp. 13–14, 1989.
- [65] E. Ghanem, I. Pawasarat, A. Lindsay et al., "Limitations of the Knee Society score in evaluating outcomes following revision total knee arthroplasty," *Journal of Bone and Joint Surgery A*, vol. 92, no. 14, pp. 2445–2451, 2010.
- [66] E. A. Lingard, J. N. Katz, R. J. Wright, E. A. Wright, and C. B. Sledge, "Validity and responsiveness of the knee society clinical rating system in comparison with the SF-36 and WOMAC," *Journal of Bone and Joint Surgery A*, vol. 83, no. 12, pp. 1856–1864, 2001.
- [67] C. M. Bach, M. Nogler, I. E. Steingruber et al., "Scoring systems in total knee arthroplasty," *Clinical Orthopaedics and Related Research*, no. 399, pp. 184–196, 2002.
- [68] G. R. Scuderi, R. B. Bourne, P. C. Noble, J. B. Benjamin, J. H. Lonner, and W. N. Scott, "The new knee society knee scoring system," *Clinical Orthopaedics and Related Research*, vol. 470, no. 1, pp. 3–19, 2012.
- [69] P. C. Noble, G. R. Scuderi, A. C. Brekke et al., "Development of a new knee society scoring system," *Clinical Orthopaedics and Related Research*, vol. 470, no. 1, pp. 20–32, 2012.
- [70] P. F. Choong, M. M. Dowsey, and J. D. Stoney, "Does accurate anatomical alignment result in better function and quality of life? Comparing conventional and computer-assisted total knee arthroplasty," *Journal of Arthroplasty*, vol. 24, no. 4, pp. 560–569, 2009.
- [71] M. A. Ritter, K. E. Davis, P. Davis et al., "Preoperative malalignment increases risk of failure after total knee arthroplasty," *Journal of Bone and Joint Surgery*, vol. 95, no. 2, pp. 126–131, 2013.
- [72] K. N. Laubenthal, G. L. Smidt, and D. B. Kettelkamp, "A quantitative analysis of knee motion during activities of daily living," *Physical Therapy*, vol. 52, no. 1, pp. 34–43, 1972.
- [73] W. H. Jung, C. W. Chun, J. H. Lee, J. H. Ha, J. H. Kim, and J. H. Jeong, "Comparative study of medial opening-wedge high tibial osteotomy using 2 different implants," *Arthroscopy*, vol. 29, no. 6, pp. 1063–1071, 2013.
- [74] D. J. Sisto and V. K. Sarin, "Custom patellofemoral arthroplasty of the knee," *Journal of Bone and Joint Surgery*, vol. 88, no. 7, pp. 1475–1480, 2006.
- [75] O. Ares, E. Castellet, F. Maculé et al., "Translation and validation of "the knee society clinical rating system" into Spanish," *Knee Surgery, Sports Traumatology, Arthroscopy*, 2013.
- [76] A. L. P. Silva, M. K. Demange, R. G. Gobbi, T. F. C. da Silva, J. R. Pécora, and A. T. Croci, "Translation and validation of the knee society score—KSS for Brazilian Portuguese," *Acta Ortopédica Brasileira*, vol. 20, no. 1, pp. 25–30, 2012.
- [77] C. van der Straeten, E. Witvrouw, T. Willems, J. Bellemans, and J. Victor, "Translation and validation of the Dutch new knee society scoring system," *Clinical Orthopaedics and Related Research*, 2013.
- [78] J. Kim, J. H. Lonner, C. L. Nelson, and P. A. Lotke, "Response bias: effect on outcomes evaluation by mail surveys after total knee arthroplasty," *Journal of Bone and Joint Surgery A*, vol. 86, no. 1, pp. 15–21, 2004.
- [79] S. Asif and D. S. Choon, "Midterm results of cemented press fit condylar sigma total knee arthroplasty system," *Journal of Orthopaedic Surgery*, vol. 13, no. 3, pp. 280–284, 2005.
- [80] The Knee Society, The 2011 Knee Society Knee Scoring System: Licenced User Manual, 2012, [http://www.kneesociety.org/web/2011%20KSS%20User%20Manual\\_FINAL.12-2012.pdf](http://www.kneesociety.org/web/2011%20KSS%20User%20Manual_FINAL.12-2012.pdf).
- [81] C. A. Jacobs and C. P. Christensen, "Correlations between knee society function scores and functional force measures," *Clinical Orthopaedics and Related Research*, vol. 467, no. 9, pp. 2414–2419, 2009.
- [82] S. E. Na, C. W. Ha, and C. H. Lee, "new high-flexion knee scoring system to eliminate the ceiling effect," *Clinical Orthopaedics and Related Research*, vol. 470, no. 2, pp. 584–593, 2012.
- [83] R. B. Bourne, B. M. Chesworth, A. M. Davis, N. N. Mahomed, and K. D. Charron, "Patient satisfaction after total knee arthroplasty: who is satisfied and who is not?" *Clinical Orthopaedics and Related Research*, vol. 468, no. 1, pp. 57–63, 2010.

## Research Article

# Comparison of the Systemic Lupus Erythematosus Activity Questionnaire and the Systemic Lupus Erythematosus Disease Activity Index in a Black Barbadian Population

Kim R. Quimby,<sup>1</sup> Cindy Flower,<sup>2</sup> Ian R. Hambleton,<sup>1</sup>  
R. Clive Landis,<sup>1</sup> and Anselm J. M. Hennis<sup>1</sup>

<sup>1</sup> Chronic Disease Research Centre, Tropical Medicine Research Institute, The University of the West Indies (UWI), Jemmott's Lane, St. Michael, Bridgetown BB11115, Barbados

<sup>2</sup> Faculty of Medical Sciences, The University of the West Indies (UWI), Cave Hill Campus, St. Michael, Bridgetown BB11000, Barbados

Correspondence should be addressed to Kim R. Quimby; [kim.quimby@cavehill.uwi.edu](mailto:kim.quimby@cavehill.uwi.edu)

Received 24 June 2013; Accepted 10 September 2013

Academic Editor: Lilian Ghandour

Copyright © 2013 Kim R. Quimby et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In Barbados, use of the Systemic Lupus Erythematosus (SLE) Disease Activity Index (SLEDAI) is limited by the unavailability of serologic markers. The SLE Activity Questionnaire (SLAQ) excludes laboratory measurements and is therefore more accessible. Here, we investigate the agreement between the SLAQ, the SLEDAI, and the physician global assessment (PGA). A pilot of 32 participants completed the SLAQ and SLEDAI. The tools were compared (1) in their original format, (2) limited to common indices, and (3) limited to the same patient recall period. We compared the proportions of persons reporting disease activity and the concordance between calculated activity scores for SLAQ versus SLEDAI and for SLAQ versus PGA. Seventy-eight percent versus 59% of participants reported disease activity with the original SLEDAI versus SLAQ, respectively. The relationship was reversed to 22% versus 59% when the matched item tools were compared. Concordance was 0.62 (95% CI 0.42–0.81) between the original scores, 0.70 (0.57–0.83) when restricted by matched items, and 0.72 (0.59–0.84) when further restricted by recall period. Concordance between the SLAQ and PGA was 0.56 (0.32–0.80). Reversal of the disease activity percentage in the matched items comparison highlights the inadequacy of tools that exclude laboratory measurements and suggests that the subjective nature of SLAQ may contribute to over-reporting. Further work is needed to produce a robust disease activity tool apt for resource-constrained environments.

## 1. Introduction

Systemic Lupus Erythematosus (SLE) is a multisystem autoimmune disease characterized by periods of clinical quiescence punctuated by acute disease flares. A five-year population-based study highlighted the striking disease excess among women of African descent, with crude incidence rates (per 100,000) of 0.4 for White males, 3.5 for White females, 0.7 for African-American males, and 9.2 for African-American females [1]. Mortality rates follow a similar pattern, with several studies indicating higher rates of lupus deaths among Black women compared to their Caucasian counterparts [2–5]. Some of the regions with the highest disease burden and mortality rates are also the least equipped to

diagnose and manage the disease [6]. Barbados, a Caribbean nation with a population of 288,000 persons [7], 93% of African origin, has one of the highest documented incidence rates of SLE among women (12.21 per 100,000 person-years; 95% CI 10.46–14.18) [5]. Similar to the experience of African-Americans, SLE in Barbadian patients has been reported to run a clinically aggressive course, with a 5-year survival rate of 79.9% (95% CI 69.6–87.1). In this Barbados cohort, 47 percent of patients developed lupus nephritis, which reduced survival in this subpopulation to 68% (95% CI 51–80) [5]. Despite the documented clinical impact of lupus in Barbados, routine monitoring using internationally accepted disease activity indices such as the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) is not feasible for most patients,

as two principal markers of disease activity, the complement levels and anti-dsDNA titres, are not routinely available.

Recommendations for SLE management developed by the European League Against Rheumatism (EULAR) and the American College of Rheumatology (ACR) focus on international best practice. However, the financial resources and available infrastructure in many countries do not allow for these guidelines to be achieved in practice [8, 9]. A report of an international symposium at the 9th International Congress on SLE in 2010 highlighted the disparities in the management of SLE worldwide and sought to establish a consensus on the minimum best practice guidelines which may be employed in resource-poor clinical settings [10]. The Systemic Lupus Activity Questionnaire (SLAQ) was considered a suitable screening tool which could guide the more in-depth assessment by the clinician.

The SLAQ is a self-administered questionnaire which has been developed to monitor disease activity in populations with financial constraints and has been compared to the Systemic Lupus Activity Measure (SLAM) with promising results [11]. This contrasts with its reported poor correlation with the SLEDAI and with the Physician Global Assessment (PGA) which is considered the best evaluation of disease activity [12]. Differing statistical methods used in previous questionnaire comparisons and the structural differences between questionnaires (SLAQ and SLEDAI, for example, use different patient recall times) hamper the ability to compare studies. In view of the possible application of the SLAQ, but its varying agreement with the other instruments, our aim was to assess the ability of the SLAQ to measure disease activity compared to the SLEDAI and the PGA, utilizing a single statistical measures of agreement and providing a simple sensitivity analysis of agreement by standardizing questionnaire items and questionnaire recall time.

## 2. Materials and Methods

Ethical approval was obtained from the Ministry of Health/University of the West Indies Ethics Review Board and the Queen Elizabeth Hospital Ethics Committee.

**2.1. Participants.** The Barbados Lupus Registry was established in 2007. In December 2009, there were 226 persons alive with definite SLE (ACR  $\geq$  4 criteria) of whom 98% were of Black ancestry and the majority (94%) were female. The 32 participants involved in this pilot were a convenience sample from this cohort selected for a companion investigation. Each participant completed the SLAQ, and a thorough systematic examination including administration of the SLEDAI was performed by the study rheumatologist.

### 2.2. Disease Activity Indices

**2.2.1. The Systemic Lupus Activity Questionnaire (SLAQ).** The SLAQ is a self-reporting tool that assesses the presence and severity of twenty-four clinical indices over the previous three months [11]. It carries a weighting regime identical to the SLAM with 0 points awarded for absent disease, 1 for mild,

2 for moderate, and 3 for severe disease, yielding a range from 0–44 points [11]. In addition, there is a single question assessing the presence and severity of lupus activity with a score ranging from 0 (no flare) to 3 (severe flare) and a single numerical rating of disease activity from 0 (no activity) to 10 (most activity).

**2.2.2. Systemic Lupus Erythematosus Disease Activity Index (SLEDAI).** The SLEDAI is a physician-administered instrument accounting for the preceding 10 days. It assesses 16 clinical features and 8 laboratory indices. The weight which has been applied to each index gives this tool a range of 0–105 points. All participants in this study completed the full SLEDAI including the dsDNA titres and the complement 3 levels.

**2.2.3. SLAQ versus SLEDAI.** There are considerable similarities between the SLAQ and SLEDAI with 15 of the 24 SLAQ and 12 of the 24 indices of SLEDAI overlapping. These were (SLEDAI descriptor/SLAQ question) seizure/seizure, organic brain syndrome/forgetfulness, lupus headache/unusual headaches, CVA/stroke, vasculitis/white fingers or toes in cold, arthritis/joint pain or swelling or stiffness, myositis/muscle pain or weakness, new rash/other rash, alopecia/bald patches on scalp, mucosal ulcers/rash in mouth or nose, pleurisy/shortness of breath or chest pain on deep breathing, and fever/fever (Table 1).

**2.2.4. PGA.** The PGA is based on the physicians' overall assessment of disease status. It is a composite of disease activity tools, other clinical or laboratory markers not included within the tools, and the physician's knowledge of the patient disease history. It carries the score of 0 for no disease, 1 for mild disease, 2 for moderate disease, and 3 for severe disease.

**2.3. Statistical Analysis.** Agreement between SLAQ and SLEDAI was assessed using the concordance coefficient for approximately continuous scores (SLAQ score, SLEDAI scores) or using Kendall's W agreement for ordinal scores (PGA, SLAQ flare). Agreement was assessed under various data restrictions, each designed to increase the clinical comparability of the two tools. Firstly, the tools were compared in their original state. Then, in a simple sensitivity analysis, agreement was further assessed (a) using only indices that were included in both SLAQ and SLEDAI—this reduced each tool to 12-items—and (b) analyzing responses from participants exhibiting symptoms within the SLEDAI 10-day participant recall window.

Agreement was also assessed between the SLAQ flare rating (i.e, the single question assessing the presence and severity of lupus activity) and the PGA.

## 3. Results

Fifty-nine percent of the 32 participants reported disease activity using the original SLAQ, compared to 78% identified by the SLEDAI. Based on the matched item scale, only 22% of

TABLE 1

#	Descriptor	SLEDAI		SLAQ	
		Definition	#	Symptom	
1	Seizure	Recent onset. Exclude metabolic, infectious, or drug cause.	2p	Seizure	
3	Organic brain syndrome	Altered mental function with impaired orientation, memory, or other intelligent function, with rapid onset fluctuating clinical features. Include clouding of consciousness with reduced capacity to focus and inability to sustain attention to environment, plus at least two of the following: perceptual disturbance, incoherent speech, insomnia or daytime drowsiness, or increased or decreased psychomotor activity. Exclude metabolic, infectious, or drug causes.	2r	Forgetfulness	
6	Lupus headache	Severe persistent headache may be migrainous, but must be nonresponsive to narcotic analgesia.	2t	Unusual headaches	
7	CVA	New onset of cerebrovascular accident(s). Exclude arteriosclerosis.	2q	Stroke	
8	Vasculitis	Ulceration, gangrene, tender finger nodules, periungual, infarction, splinter hemorrhages, or biopsy or angiogram proof of vasculitis.	2m	Fingers/toes turning dead white or very pale in the cold	
9	Arthritis	More than 2 joints with pain and signs of inflammation (i.e., tenderness, swelling, or effusion).	2w 2x	Pain or stiffness in the joints Swelling in the joints	
10	Myositis	Proximal muscle aching/weakness, associated with elevated creatinine phosphokinase/aldolase or electromyogram changes or a biopsy showing myositis.	2u 2v	Muscle pain Muscle weakness	
15	New Rash	New onset or recurrence of inflammatory type rash.	2f	Other rash	
16	Alopecia	New onset or recurrence of abnormal, patchy, or diffuse loss of hair.	2i	Bald patches on scalp	
17	Mucosal Ulcers	New onset or recurrence of oral/nasal ulcerations.	2d	Sores in mouth/nose	
18	Pleurisy	Pleuritic chest pain with pleural rub or effusion or pleural thickening.	2k 2l	Shortness of breath Chest pain on deep breath	
19	Fever	>38°C. Exclude infectious cause.	2c	Fevers (>101°F, 38.5°C) taken by thermometer	

SLAQ versus SLEDAI: the description and definition of SLEDAI indices and the matched SLAQ symptoms.

participants reported activity with the SLEDAI whereas the SLAQ result did not change (59%).

Concordance between the original activity scores was 0.62 (95% CI: 0.42 to 0.81) (Figure 1(a)). Restricting the tools to common items increased the concordance to 0.70 (0.57 to 0.83) (Figure 1(b)). Two participants experienced disease flares within the SLAQ but not the SLEDAI timeline. Reanalysis with the removal of the scores from these 2 patients further increased the agreement between the SLAQ and SLEDAI instruments to 0.72 (0.59 to 0.84) (Figure 1(c)).

Agreement between the SLAQ flare rating and the physician global assessment was 0.56 (0.32 to 0.80).

#### 4. Discussion

The SLEDAI is widely accepted as a tool for monitoring SLE activity both in clinical practice and in research. However, the SLEDAI includes laboratory measures not routinely available in resource-constrained settings as experienced in Barbados, and therefore, the resulting clinical assessments are often incomplete. The SLAQ is a self-administered questionnaire

that is inexpensive but gives variable results when compared to other instruments measuring lupus activity.

Our comparison of the SLAQ and SLEDAI instruments presented several challenges. Firstly, although both tools monitor lupus activity, they assess different disease domains and are therefore not expected to be in total agreement. For example, psychosis, visual disturbance, cranial nerve disorder, pericarditis, and laboratory markers are assessed by SLEDAI but not by SLAQ, whereas weight loss, fatigue, malar rash, lymphadenopathy, and photosensitivity are assessed by SLAQ and not by SLEDAI. To partly overcome this, an altered questionnaire which mapped only indices common to both was used (Table 1).

Secondly, the timeframe for patient symptom recall varies between instruments. The SLEDAI takes into account the previous 10 days whereas the SLAQ covers a 3-month period. Two participants experienced flares between the 10-day and 3-month window (measured by the SLAQ but not the SLEDAI). Exclusion of the scores of these 2 patients resulted in better agreement between the scores of the SLAQ and the SLEDAI (Figure 1(c)).

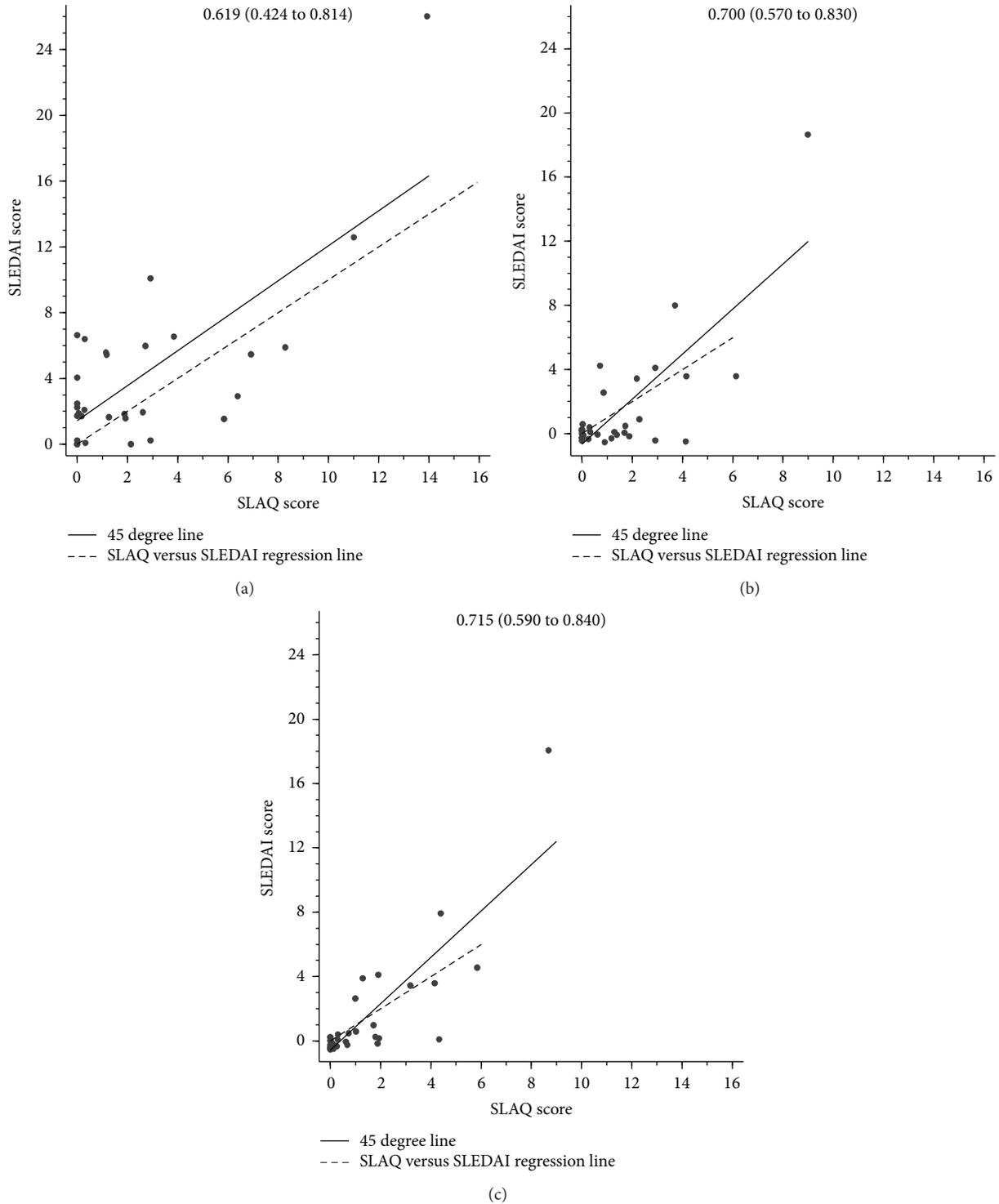


FIGURE 1: The concordance between (a) the original activity scores, (b) the matched-items score, and (c) the scores further restricted by recall time.

In spite of the usefulness of the disease activity measures, the PGA is still considered the best assessment of disease activity. Concordance between the SLAQ flare rating and the PGA was 0.56 (0.32 to 0.80), significantly lower than the SLAQ-SLEDAI scores concordance of 0.72 (0.59 to 0.84),

indicating that a collection of multiple indices is more accurate than a single self-reported measure of disease activity over the previous three months.

Fifty-nine percent of participants reported disease activity with the SLAQ compared to 78% identified by the SLEDAI.

This seemed to dispel the notion that the subjective nature of SLAQ leads to the over-reporting of symptoms; however, this difference could be attributed to the disparity in reporting laboratory indices. Laboratory indices are only assessed by the SLEDAI, and it is reasonable to expect that more indicators of disease would be identified by this method. To adjust for the inequity in the indices being assessed, we investigated these summary measures using the matched item scale. The SLAQ results did not change, with 59% of participants reporting disease activity. However, the proportion reporting disease activity now decreased to 22% (from 78%) with the reduced item SLEDAI. This indicated two things; firstly, it highlighted the inadequacy of disease activity indices that do not include laboratory measures; in this case, there was a 56% point reduction in the disease markers identified. All cases of proteinuria, which is critical to diagnosis of nephritis which affects nearly half of Barbadian patients, would have been missed. The mandatory inclusion of basic laboratory measures, for example, full blood count and urinalysis in the monitoring of SLE suggested during the 9th International Congress on SLE in 2010, was aimed at rectifying this gap [10]. Secondly, it underscored the difference in outcome between physician-rated and patient-reported conclusions, even when matched clinical indices are being compared. In this pilot, more disease activity was recounted with the SLAQ suggesting that its subjective nature might lead to over-reporting by patients who do the reporting. With the support of the Barbados Lupus Cohort [5], we intend to examine these potential deficiencies in a longitudinal comparison of the SLAQ and SLEDAI questionnaires, where both tools will be administered to participants at predetermined intervals. This follow-up study will allow for further investigations to identify particular domains of the SLAQ which promote over-reporting. With the additional statistical power in this longitudinal investigation, we will examine the agreement of individual matched questionnaire items from SLAQ and SLEDAI. Those SLAQ indices that exhibit low agreement with the matched SLEDAI indices would be flagged for deletion or substitution. Any SLAQ alterations would then be subjected to further validation.

Construction of this tool is even more relevant now with the thrust towards the treat-to-target initiative, which has proven to be beneficial in the management of rheumatoid arthritis [13, 14]. Using this method, clinical targets for each follow-up time-point is established, and at the time of follow-up, therapy is adjusted based on the realisation (or not) of the predefined disease activity goal. Measuring the identified target requires a relevant disease activity tool.

In conclusion, the current SLAQ is inadequate as a disease monitoring tool because of its lack of laboratory measurements and its patient-based subjective reporting which might lead to inadvertent over-reporting of symptoms. This leaves us still in need of a SLE disease monitoring tool which is accessible to resource-limited populations. The aim of our future study is to transform the current SLAQ into a clinically relevant but accessible tool by adding basic, cheap laboratory measurements and reducing over-reporting of symptoms.

## Acknowledgments

The authors would like to thank the members of the Barbados Lupus Study for their participation. This work was supported by the Caribbean Health Research Council.

## References

- [1] D. J. McCarty, S. Manzi, T. A. Medsger Jr., R. Ramsey-Goldman, R. E. LaPorte, and C. K. Kwok, "Incidence of systemic lupus erythematosus: race and gender differences," *Arthritis and Rheumatism*, vol. 38, no. 9, pp. 1260–1270, 1995.
- [2] S. Bernatsky, J. Boivin, L. Joseph et al., "Mortality in systemic lupus erythematosus," *Arthritis and Rheumatism*, vol. 54, no. 8, pp. 2550–2557, 2006.
- [3] M. C. Hochberg, "Mortality from systemic lupus erythematosus in England and Wales, 1974–1983," *The British Journal of Rheumatology*, vol. 26, no. 6, pp. 437–441, 1987.
- [4] "Trends in deaths from systemic lupus erythematosus—United States, 1979–1998," *Morbidity and Mortality Weekly Report*, vol. 51, no. 17, pp. 371–374, 2002.
- [5] C. Flower, A. J. M. Hennis, I. R. Hambleton, G. D. Nicholson, and M. H. Liang, "Systemic lupus erythematosus in an Afro-Caribbean population: incidence, clinical manifestations and survival in the Barbados national lupus registry," *Arthritis Care and Research*, vol. 64, no. 8, pp. 1151–1158, 2012.
- [6] A. Vasudevan and A. N. Krishnamurthy, "Changing worldwide epidemiology of systemic lupus erythematosus," *Rheumatic Disease Clinics of North America*, vol. 36, no. 1, pp. 1–13, 2010.
- [7] US Census Bureau DIS, International Programs, International Data Base [Internet], <http://www.census.gov/population/international/data/idb/region.php>.
- [8] "Guidelines for referral and management of systemic lupus erythematosus in adults," *Arthritis and Rheumatism*, vol. 42, no. 9, pp. 1785–1796, 1999.
- [9] M. Mosca, C. Tani, M. Aringer et al., "European league against rheumatism recommendations for monitoring patients with systemic lupus erythematosus in clinical practice and in observational studies," *Annals of the Rheumatic Diseases*, vol. 69, no. 7, pp. 1269–1274, 2010.
- [10] C. H. Flower, A. J. M. Hennis, and M. H. Liang, "Report of an international symposium on narrowing the Gap in the treatment and study of SLE worldwide: minimum best practices in the management and monitoring of moderate to severe SLE and improving outcomes in constrained environments," *Current Reviews in Musculoskeletal Medicine*, vol. 4, no. 3, pp. 105–112, 2011.
- [11] E. W. Karlson, L. H. Daltroy, C. Rivest et al., "Validation of a systemic lupus activity questionnaire (SLAQ) for population studies," *Lupus*, vol. 12, no. 4, pp. 280–286, 2003.
- [12] M. Jolly, M. Ishimori, I. Moldovan, E. Katsaros et al., "What does patient-reported systemic lupus activity questionnaire (SLAQ) measure?" *Arthritis and Rheumatism*, vol. 60, supplement 10, p. 295, 2009.
- [13] M. F. Bakker, J. W. G. Jacobs, S. M. M. Verstappen, and J. W. J. Bijlsma, "Tight control in the treatment of rheumatoid arthritis: efficacy and feasibility," *Annals of the Rheumatic Diseases*, vol. 66, supplement 3, pp. iii56–iii60, 2007.
- [14] M. Mosca, D. T. Boumpas, I. N. Bruce et al., "Treat-to-target in systemic lupus erythematosus: where are we today?" *Clinical and Experimental Rheumatology*, vol. 30, no. 4, supplement 73, pp. S112–S115, 2012.

## Review Article

# Disease Activity Measures in Paediatric Rheumatic Diseases

**Nadia J. Luca<sup>1</sup> and Brian M. Feldman<sup>1,2</sup>**

<sup>1</sup> *Division of Rheumatology, Department of Paediatrics, University of Toronto and The Hospital for Sick Children, 555 University Avenue, Toronto, ON, Canada M5G 1X8*

<sup>2</sup> *Departments of Medicine, Institute of Health Policy Management and Evaluation and the Dalla Lana School of Public Health, University of Toronto, 155 College Street Toronto, ON, Canada M5T 3M7*

Correspondence should be addressed to Brian M. Feldman; [brian.feldman@sickkids.ca](mailto:brian.feldman@sickkids.ca)

Received 9 June 2013; Accepted 2 August 2013

Academic Editor: Zahi Touma

Copyright © 2013 N. J. Luca and B. M. Feldman. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disease activity refers to potentially reversible aspects of a disease. Measurement of disease activity in paediatric rheumatic diseases is a critical component of patient care and clinical research. Disease activity measures are developed systematically, often involving consensus methods. To be useful, a disease activity measure must be feasible, valid, and interpretable. There are several challenges in quantifying disease activity in paediatric rheumatology; namely, the conditions are multidimensional, the level of activity must be valued in the context of treatment being received, there is no gold standard for disease activity, and it is often difficult to incorporate the patient's perspective of their disease activity. To date, core sets of response variables are defined for juvenile idiopathic arthritis, juvenile systemic lupus erythematosus, and juvenile dermatomyositis, as well as definitions for improvement in response to therapy. Several specific absolute disease activity measures also exist for each condition. Further work is required to determine the optimal disease activity measures in paediatric rheumatology.

## 1. Introduction

Measurement of a health state is the cornerstone of clinical practice and medical research. As stated in 1883 by Lord Kelvin, “when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot, your knowledge is of a meagre and unsatisfactory kind” [1]. When caring for children and youth with rheumatic diseases, standardized and validated definitions of disease activity, as well as of change in disease activity in response to treatment, are critical for making decisions in clinical practice as well as for conducting clinical research studies. Use of standardized measures can also facilitate comparisons between trials through meta-analyses.

Over the past several years, an international collaborative effort has been made to develop and validate measures of disease activity, treatment response, inactive disease, and flare of disease in several paediatric rheumatic conditions [2]. These measures are numerous, and several new ones are emerging on an annual basis. This paper will discuss the theoretical basis of disease activity measures, methods

for their development and validation, and an overview of available measures within paediatric rheumatology.

## 2. Definition of Disease Activity in Paediatric Rheumatic Diseases

In order to measure disease activity, it is necessary to define it and to identify its measurable components. “Disease activity” may be defined as aspects of a patient's disease that are potentially reversible [3–5]. This is in contrast to “disease damage,” which refers to the irreversible (i.e., due to scarring) manifestations of the disease or its treatment. Both disease activity and damage contribute to “disease severity,” which is the total effect of the disease on the individual and often relates to prognosis [5]. Disease activity may fluctuate widely during a patient's disease course. In the context of rheumatic diseases, a disease activity measure typically attempts to quantify the inflammatory process of the disease [3]. Consolaro et al. define disease activity in juvenile idiopathic arthritis (JIA) as measuring “signs and

symptoms related to inflammation,” which “should separated from measures of disease severity, such as those that assess...structural damage” [6]. Evaluating disease activity may include measurement of biochemical markers of inflammation (e.g., erythrocyte sedimentation rate (ESR)), quantifying the amount of inflammatory tissue (e.g., number of swollen joints), and assessing the clinical consequences of inflammation (e.g., degree of morning stiffness). Generally, disease activity in most paediatric rheumatic diseases is multidimensional, and therefore this concept is measured in several domains [7]. For example, a measure of disease activity in juvenile dermatomyositis (JDM) incorporates muscle strength, skin disease, internal organ manifestations, and any other disease components in patients with JDM. Disease activity measures may take the form of observed values (e.g., muscle enzymes level), rating scales completed by health care providers, parent- or patient-reported outcomes (PROs), or any combination of these. When measuring disease activity, its impact on the biological, individual, and social perspectives of health is often considered, as described in the International Classification of Functioning, Disability, and Health (World Health Organization) [8, 9]. This system takes into account that disease activity interacts with internal and external factors to determine the levels of biological, physical, and social functioning. For example, a child with JIA who has swelling and reduced motion of a joint may have difficulty running and may not participate in gym class because of fear of injuring herself. Interventions can be aimed at each level: medical treatment to reduce joint inflammation, physiotherapy to improve running, and accommodations at school to allow the child to participate at their own pace.

### 3. Challenges in Measuring Disease Activity in Paediatric Rheumatic Diseases

Given the multidimensionality of disease activity in paediatric rheumatic disease, a challenge exists when some components suggest that the disease is more active, while others suggest that it is less active. Thus there is appeal to a pooled and appropriately weighted composite index that can give one answer. For example, the American College of Rheumatology (ACR) core outcome variables for JIA were combined into a definition to dichotomously divide patients into those who improve by a clinically important amount and those who do not [10]. Variables in the core set include: (1) physician global assessment of disease activity, (2) parent/patient assessment of overall wellbeing, (3) functional ability, (4) number of joints with active arthritis, (5) number of joints with limited range of motion (ROM), and (6) ESR. Consensus methods were used to define improvement; for example, an ACR Pedi30 response requires  $\geq 30\%$  improvement from baseline in 3 of 6 variables, with  $\leq 1$  remaining variable worsening by  $>30\%$ . Similar core sets and definitions for improvement have also been developed for JDM and juvenile systemic lupus erythematosus (JSLE) [11–14].

The ACR Pedi response criteria are defined relative to each patient's baseline parameters but do not enable the quantification of absolute disease activity or comparison

of absolute responses amongst patients. For instance, these response criteria cannot distinguish between a patient who has improved by 30% starting with 30 active joints (still has 21 active joints) and one who started with 3 active joints (still has 2 active joints). This prompted the development of a scale that quantifies the absolute level of disease activity, the Juvenile Arthritis Disease Activity Score (JADAS), which includes 4 items: (1) physician global assessment of disease activity, (2) parent/patient global assessment of wellbeing, (3) number of active joints, and (4) ESR [6]. The JADAS is scored on a numerical scale, which allows patients to be compared directly. Absolute disease activity measures have also been developed for JSLE (e.g., European consensus lupus activity measurement (ECLAM), systemic lupus erythematosus disease activity index (SLEDAI)) and JDM (e.g., disease activity score (DAS), myositis disease activity assessment tool (MDAAT)) and are discussed further in Section 6.

An additional factor to consider is how to value clinical improvement in the context of the treatment regimen of the patient. For example, if one polyarticular JIA patient is managed on a nonsteroidal anti-inflammatory drug and achieves inactive disease, is his/her “disease activity” equivalent to another patient whose disease is quiescent while taking methotrexate and an anti-tumor necrosis factor alpha agent? Criteria for inactive disease and clinical remission in JIA do distinguish between patients who are on or off medical therapy. Wallace et al. define “inactive disease” as when a patient on medical therapy achieves the following criteria: (1) no joints with active arthritis, (2) no fever, rash, serositis, splenomegaly, or generalized lymphadenopathy, (3) no active uveitis, (4) normal ESR and/or C-reactive protein (CRP), and (5) physician global assessment indicates no disease activity [15]. If patients meet the inactive disease criteria for 6 continuous months while taking medication they are classified as being in “clinical remission on medication,” and if they meet the criteria for 12 continuous months off medication they are considered to be in “clinical remission off medication” [15]. Each of these health states is considered to be a distinct outcome. However, these criteria do not take into consideration the relative intensity of the treatment required to control the disease.

Another significant challenge when measuring disease activity in pediatric rheumatic disease is that expert clinical judgment (i.e., opinion of a paediatric rheumatologist) is commonly used for validation of disease activity instruments [3], and this can be ambiguous and subjective. Thus, a limitation is that there is no truly objective “gold” standard with which to validate disease activity measures, although some promising biomarkers are being identified [16]. In addition, it is important that “circularity of reasoning” is avoided [17]. That is, if the group of clinicians that derives the criteria set is the same group that classifies disease activity, there will be a strong correlation between the criteria set and the so called “gold” standard. Lastly, the use of a physician-based measure alone for comparison entirely excludes the patient perspective of the disease burden.

Parent or patient global assessments are included in many paediatric rheumatic disease activity measures. Interestingly, studies show concordance with physician assessments of JIA

TABLE 1: Steps in the development of disease activity measures (adapted from [17, 24, 26, 27]).

Step	Comments
Statement of purpose of measure	Should be explicitly stated a priori
Methods of item generation	From the literature, expert opinion, patient opinion, consensus methods Item list should be comprehensive
Method to identify number of factors or constructs	For example, principal component and factor analysis
Derivation sample	Across spectrum of disease activity, academic and community clinics, ideally $n = 200$
Clinicians involved	Clinicians deriving criteria should be different than those providing cases/controls
Definition of gold standard	Validated set of criteria or expert opinions
Technique for selection of final set of items	For example, sensitivity/specificity, ROC curve, and regression models
Assessment of psychometric properties	Feasibility, face validity, content validity, reliability, construct validity, and responsiveness
External validation of criteria set	Distinct validation sample

disease activity in only about 40% of cases [18–20]. Sztajn bok et al. report that physicians tend to give higher global ratings compared to parents [18]. Conversely, parents gave scores of zero in only 65% of cases who met the Wallace definition of inactive disease in JIA (listed above) [19]. Parents are more likely to score higher disease activity if their child has a shorter disease duration, is taking second-line drug therapy, has increased reported pain, or has functional impairment. Alternatively, physicians consistently rate disease activity more highly than parents in the presence of any active joints [18, 19]. In JDM patients, Rider et al. found that physician and parent global ratings of disease activity were not collinear and that the nonredundancy may be the result of each evaluating different aspects of the disease [21]. This may also be because parents are often asked about overall “wellbeing” rather than specifically “disease activity.” Thus, the parent or patient perspective adds information on the impact of disease activity on overall health.

For this reason, several new composite disease assessment measures have included more PROs. The juvenile arthritis parent assessment index (JAPAI) and the juvenile arthritis child assessment index (JACAI) include: (1) parent/child rating of overall wellbeing, (2) parent/child rating of pain intensity, (3) assessment of physical function, and (4) assessment of HRQOL [22]. Similarly, the juvenile arthritis multidimensional assessment report (JAMAR) incorporates parent- or patient-reported physical functioning, social participation, and HRQOL, among other items [23].

Although it can be challenging to validate PRO measures as they are inherently subjective, they provide rich information on the patient’s perspective of their disease activity. However, it should be noted that paediatric rheumatologists developed these measures, and so they include items that health care providers judge as significant to patients but may not address all of the perceptions, values, and preferences important to patients.

#### 4. Development of Disease Activity Measures

Measurement instruments are developed systematically, and the process is often iterative [7, 24, 25]. An approach for developing outcome measures in paediatric rheumatic diseases has

been described in detail by Brunner and Ravelli [25]. Here we present a brief overview of the process (summarized in Table 1).

*4.1. Purpose of the Measure.* Along with defining what is to be measured and in which population, the purpose of the disease activity measure should be explicitly determined at the outset. Kirshner and Guyatt [45] describe three main purposes of health status measures. Firstly, “discriminative” measures are used to distinguish between individuals or groups with a certain attribute (e.g., criteria for inactive disease in JIA [15]). “Predictive” measures classify individuals into a set of predefined risk categories when a gold standard is available. This type of index is generally used as a screening or diagnostic instrument (e.g., dipstick proteinuria in lupus nephritis). Lastly, “evaluative” measures are used to assess the magnitude of longitudinal change in the dimension of interest (e.g., manual muscle testing (MMT) in JDM [46]). Disease activity measures typically fall into the evaluative category as they typically measure clinically significant changes in disease activity over time and in response to interventions.

*4.2. Item Generation and Reduction.* The process of item generation creates a list of all possible items for inclusion, generally thorough the literature review and opinions from important stakeholders [47]. Ideally, patients and parents should be included in the process of measure development [48]. Item generation is followed by item reduction, with the goal of including all the important items but minimizing redundancy. Consensus methods have commonly been used for this phase of measure development (see below). Ideally, it is also important to determine the number of different aspects of disease activity, or “factors,” being measured. Principal component and factor analysis identify the number of factors within the scale and which of these explain the majority of the overall variance [26, 49]. At the item level, item-scale correlations and measures of internal consistency (Cronbach’s alpha) can indicate which individual items within the factors may be redundant and possibly unnecessary [26].

*4.3. Consensus Methods.* In recent years, consensus methods such as Delphi surveys and nominal group techniques

TABLE 2: Definitions of domains of measurement properties, adapted from COSMIN taxonomy [28].

Domain	Measurement property	Definition
Reliability	Internal consistency	The degree of interrelatedness among items
	Reliability	The proportion of the total variance in measurements which is due to “true” differences among patients
	Measurement error	The systematic and random errors of a patient’s score that is not attributed to true changes in the construct being measured
Validity	Content validity	The degree to which the content of an instrument is an adequate reflection of the construct to be measured
	Face validity	The degree to which the items look to be an adequate reflection of the construct to be measured
	Construct validity	The degree to which the scores of an instrument are consistent with hypotheses (e.g., relationships to scores of other instruments) based on the assumption that the instrument validly measures the construct to be measured
	Criterion validity	The degree to which the scores of an instrument are an adequate reflection of a “gold standard”
Responsiveness		The ability of an instrument to detect change over time in the construct to be measured
Interpretability		The degree to which one can assign qualitative meaning to an instrument’s quantitative scores or change in scores

(NGTs) have commonly been used in the item generation and item reduction phases of measure development in paediatric rheumatology [50]. The definition of consensus should be stated a priori and utilized as criteria for termination of the process [51], although this is not always the case in practice [52]. The Delphi technique is an iterative multistage method that allows the use of anonymous written or online responses [53]. It involves a series of questionnaires, each of which is based on the results of the previous step, and the process stops when participants approach consensus. NGT utilizes a highly structured face-to-face meeting to gather information from relevant stakeholders [53]. Panelists rank, discuss, and then rerank a series of items related to the topic. The results are analyzed for agreement in prioritization, and generally 70–80% consensus is required [51]. These approaches have been used successfully to develop several disease activity measures in paediatric rheumatology such as the definition for improvement in JIA [10] and disease activity core set measures for JDM and JSLE [11].

**4.4. Derivation Study.** Once candidate sets of criteria are determined, they should be applied to a large and diverse group of patients with the disease from a variety of practices [17, 26]. A comparator group is chosen based on the intended use of the criteria and should represent patients from whom the criteria aim to distinguish. In order to avoid circularity of reasoning, the group of clinicians who create the list of candidate items should be separate from those who provide and classify the patients [17, 26]. The sample size required for derivation studies is approximately 100 per group [24].

**4.5. Selection of Final Set of Items.** The final set of items is often selected using statistical techniques such as comparing the sensitivity and specificity of candidate sets of items with

receiver operating characteristic curve analysis. This was performed by Brunner et al. to determine the best definition for flare in patients with JIA [33]. Logistic regression may also be used to determine what variables best discriminate between more and less severely affected patients as was employed in the development of criteria for minimal disease activity (MDA) in JIA [42].

**4.6. Assessment of Psychometric Properties.** Prior to carrying out formal psychometric testing of a measurement tool, it should be pretested in a small group of users to ensure comprehensibility and relevance [47]. Evaluation of face and content validity ensures the items are reasonable and cover all the aspects of the disease. Feasibility should also be assessed, with consideration given to ease of use and minimal burden on the patient and health care provider. Without these characteristics, the tool is unlikely to be accepted by medical practitioners or adopted into clinical care.

Once a tool is developed and piloted, its measurement properties must be evaluated in the target population. Most disease activity measures should be assessed with regards to their reliability, validity, responsiveness, and interpretability (Table 2) [28]. As a final step, measures should be applied to an independent sample to ensure external validation [24, 27].

## 5. Disease Activity Measures in JIA, JSLE, and JDM

A summary of the most commonly cited disease activity measures in JIA, JSLE, and JDM is presented in Table 3.

**5.1. Core Response Variables in JIA, JSLE, and JDM.** Composite disease activity measures comprising 5 or 6 core response variables (CRV) have been defined for each disease

TABLE 3: Summary of disease activity measures for juvenile idiopathic arthritis (JIA), juvenile systemic lupus erythematosus (JSLE), and juvenile dermatomyositis (JDM).

	JIA	JSLE	JDM
	Relative disease activity measures		
Core response variables (CRV)	<p>ACR [10]:</p> <ol style="list-style-type: none"> <li>(1) Physician global assessment of disease activity</li> <li>(2) Parent/patient assessment of overall wellbeing</li> <li>(3) Functional ability (CHAQ)</li> <li>(4) Number of joints with active arthritis</li> <li>(5) Number of joints with limited range of motion</li> <li>(6) ESR</li> </ol> <p>For SJIA patients: absence of spiking fever (<math>\leq 38^{\circ}\text{C}</math> during past week)</p>	<p>PRINTO [13]:</p> <ol style="list-style-type: none"> <li>(1) Physician global assessment of disease activity</li> <li>(2) Parent/patient assessment of overall wellbeing</li> <li>(3) Renal involvement: 24-hour proteinuria</li> <li>(4) Global disease activity tool (SLEDAI, ECLAM, or SLAM)</li> <li>(5) HRQL (CHQ)</li> </ol>	<p>PRINTO [14]:</p> <ol style="list-style-type: none"> <li>(1) Physician global assessment of disease activity</li> <li>(2) Parent/patient assessment of overall wellbeing</li> <li>(3) Muscle strength (CMAS, MMT)</li> <li>(4) Functional ability (CHAQ)</li> <li>(5) Global disease activity tool (DAS, MDAAT)</li> <li>(6) HRQL (CHQ)</li> </ol> <p>IMACS [12]:</p> <ol style="list-style-type: none"> <li>(1) Physician global assessment of disease activity</li> <li>(2) Muscle strength (MMT)</li> <li>(3) Functional ability (CHAQ, CMAS)</li> <li>(4) Muscle enzymes (<math>\geq 2</math> of CPK, aldolase LDH, AST, ALT)</li> <li>(5) Extramuscular disease (MDAAT)</li> </ol>
Definition of improvement	$\geq 30\%$ improvement from baseline in 3 of 6 CRVs, with $\leq 1$ CRV worsening by $>30\%$ [10, 29]	$\geq 50\%$ improvement from baseline in 2 of 5 CRVs, with $\leq 1$ CRV worsening by $>30\%$ [30]	<p>PRINTO [31]:</p> <p><math>\geq 20\%</math> improvement in 3 of 6 CRVs, with <math>\leq 2</math> worsening by <math>\geq 20\%</math> (muscle strength excluded)</p> <p>IMACS [32]:</p> <p><math>\geq 15\%</math> improvement in muscle strength and physical function, <math>\geq 20\%</math> physician and patient global and extramuscular assessments, <math>\geq 30\%</math> improvement in muscle enzymes</p>
Definition of flare	Worsening of 2 CRV by $\geq 40\%$ without improvement in $>1$ CRV by $\geq 30\%$ [33]		
	Absolute disease activity measures		
Global disease activity tools	<p>JADAS [6, 34]:</p> <ol style="list-style-type: none"> <li>(1) Physician global assessment of disease activity</li> <li>(2) Parent/patient global assessment of wellbeing</li> <li>(3) Count of joints with active disease</li> <li>(4) ESR</li> </ol> <p>MDA [42]:</p> <ol style="list-style-type: none"> <li>(1) Oligoarthritis: physician global assessment <math>\leq 2.5</math> cm and swollen joint count of 0</li> <li>(2) Polyarthritis: physician global assessment <math>\leq 3.4</math> cm, parent global assessment <math>\leq 2.1</math> cm, and swollen joint count <math>\leq 1</math></li> </ol> <p>Wallace criteria [15, 43]</p>	<p>SLEDAI [35]</p> <p>BILAG [36]</p> <p>SLAM [37]</p> <p>ECLAM [38]</p>	<p>DAS [39]</p> <p>MDAAT [40]</p> <p>CAT [41]</p>
Minimal disease activity (MDA)	<p>Wallace criteria [15, 43]</p> <p><i>Inactive disease:</i></p> <ol style="list-style-type: none"> <li>(1) No joints with active arthritis</li> <li>(2) No fever, rash, serositis, splenomegaly, or generalized lymphadenopathy attributable to JIA</li> <li>(3) No active uveitis</li> <li>(4) Normal ESR and/or CRP</li> </ol>		
Inactive disease/remission			<p>PRINTO [44]:</p> <p>Inactive disease on or off medication, <math>\geq 3/4</math> of:</p> <ol style="list-style-type: none"> <li>(1) CPK <math>\leq 150</math></li> <li>(2) CMAS <math>\geq 48</math></li> <li>(3) MMT <math>\geq 78</math></li> <li>(4) Physician global assessment <math>\leq 0.2</math></li> </ol>

TABLE 3: Continued.

JIA	JSLE	JDM
(5) Physician global assessment indicates no disease activity <i>Clinical remission:</i> (1) On medication—criteria for inactive disease met for minimum 6 continuous months while patient on medication (2) Off medication—criteria for active disease met for minimum 12 continuous months while off all arthritis and uveitis medications		

ACR: American College of Rheumatology, CHAQ: child health assessment questionnaire, ESR: erythrocyte sedimentation rate, SJIA: systemic juvenile idiopathic arthritis, PRINTO: Pediatric Rheumatology International Trials Organization, SLEDAI: systemic lupus erythematosus disease activity index, BILAG: British Isles lupus assessment group index, SLAM: systemic lupus activity measure, ECLAM: European consensus lupus activity measurement, HRQL: health-related quality of life, CHQ: child health questionnaire, IMACS: International Myositis Assessment and Clinical Studies group, JADAS: juvenile arthritis disease activity score, MMT: manual muscle testing, CMAS: childhood myositis assessment scale, DAS: disease activity score, MDAAT: myositis disease activity assessment tool, and CAT: cutaneous assessment tool.

by consensus methodologies involving members of the ACR, PRINTO, and/or International Myositis Assessment and Clinical Studies (IMACS) groups, as mentioned above. Note that there are 2 slightly different core sets for JDM developed by PRINTO and IMACS (Table 3). The common components for each disease include physician and parent or patient global assessments, which use 10 cm visual analogue scales (VAS) to evaluate the extent of active inflammation based on all information available [21]. A score of zero indicates inactive disease and 10 indicates extremely high activity. In addition, the JIA and JDM core sets include a measure of physical function (most typically the childhood health assessment questionnaire (CHAQ), which has been validated in both populations [54, 55]). The CHAQ is divided into disability and discomfort indices which assess function in 8 areas (score 0–3) and pain intensity and overall wellbeing, respectively [54]. The JSLE and JDM core sets include a generic measure of HRQL that has been validated in both populations, the child health questionnaire (CHQ). The CHQ contains 14 domains to evaluate the physical, emotional, and social components of health and provides physical and psychosocial summary scores [56]. A notable challenge is that measures of physical function and HRQL within the core sets may reflect disease damage in addition to disease activity. The remaining CRV are specific to each disease and are further discussed in the sections below.

Clinically meaningful improvement in the core set is often used in clinical research to define the primary endpoint. Criteria for improvement have been developed for each composite measure [10, 29–32], and a definition of disease flare has been proposed for JIA only [33]. As noted in Section 4, achieving improvement or flare is relative to the patient's baseline status, and thus the absolute degree of disease activity is different for each patient.

**5.2. Disease Activity Measures in JIA.** In addition to the items mentioned above, the ACR core set also includes a measure of systemic inflammation (usually ESR or CRP), the number

of joints with active arthritis (defined as joint effusion or limitation of motion accompanied by heat, pain, or tenderness), and the number of joints with limited ROM [10]. A systematic review of outcome measures in JIA compares the psychometric properties of the measures within the ACR core set, although it did not identify any validation studies of the composite core set itself [57]. The CHAQ has strong reliability and moderate correlations with other measures of disease activity but poor responsiveness. Active joint count and physician global assessment have the strongest correlations with disease activity and are most responsive to change.

The JADAS is a measure of absolute disease activity in JIA. It is similar to the core set in that it incorporates physician and parent/patient global assessment of disease activity, number of active joints, and ESR; however, it does not include functional status or number of joints with reduced ROM because these may be more reflective of disease damage [6]. Depending upon the number of joints assessed, the JADAS is scored on a continuous scale from 0–101 (71 joints), 0–57 (27 joints), or 0–40 (10 joints) and has good construct validity and responsiveness [6]. Changes in JADAS score are able to classify patients according to ACR Pedi response and have excellent ability to predict flare and inactive disease [34]. Recently, the “JADAS3,” which excludes ESR and avoids the necessity of a blood sample, was found to correlate with individual measures of disease activity as well as the original JADAS [58].

With improved success in the treatment of JIA, criteria for MDA, inactive disease, and remission have been defined. MDA is intended to define a state between high disease activity and remission that is acceptable to the physician and patient [42]; however, there is no data on the prognostic significance of maintaining this state. The MDA definitions have been validated in independent samples [42]. Similarly, the measurement properties of the Wallace criteria for inactive disease, remission on medication, and remission off medication have been assessed on data from 3 independent JIA clinics [43].

Systemic JIA (SJIA) differs from the other subtypes in that it is characterized by fever, rash, serositis, organomegaly, and lymphadenopathy, and these features must be considered in a disease activity measure. Bathish and colleagues have begun to develop an SJIA-specific disease activity measure using patient and parent interviews [48] and Delphi survey of health professionals to generate items [59]. The top items included fever, rash, increased CRP and ESR, and requirement for increasing medications. From here, the goal is to determine the best and most parsimonious list of items. At present, the ACR core set specifies the absence of spiking fever ( $\leq 38^{\circ}\text{C}$  during past week) [2], and the definition of inactive disease requires absence of SJIA symptoms [15] as an additional requirement for SJIA children.

**5.3. Disease Activity Measures in JSLE.** Several disease activity measures initially developed in adult SLE patients have been found to be sensitive to change in JSLE [60]. The SLEDAI [35], British Isles lupus assessment group index (BILAG) [36], systemic lupus activity measure (SLAM) [37], and ECLAM [38] consist of many attributes grouped into 8–10 organ systems. The SLEDAI and SLAM are scored by weighting and summing of each attribute to give a total score. There are several rules for scoring the ECLAM to give an integer between 0 and 10. The BILAG uses an alphabetical rating system for each organ system, although numerical conversion schemes have been suggested [60]. Of the measures described, the SLEDAI and SLAM are the most user-friendly, and the ECLAM appears to be the most sensitive to change in JSLE [61]. The BILAG is complicated and takes longer to score.

**5.4. Disease Activity Measures in JDM.** Rider et al. have recently published a very comprehensive review of all the disease activity measures available for JDM, including their psychometric properties and utility for both clinical and research purposes [62]. The core sets for JDM include a measure of myositis activity assessed either with MMT or the childhood myositis assessment scale (CMAS). A modified MMT of 8 proximal, distal, and axial muscle groups tested unilaterally is scored from 0 to 80, and was found to closely approximate a total MMT score of 26 muscle groups tested bilaterally and is more feasible for use in the clinic setting as it takes less than 5 minutes to perform [63]. The CMAS uses 14 maneuvers (score 0–52) to measure muscle strength, physical function, and endurance and has excellent measurement properties [46]. The IMACS cores set considers the CMAS a measure of physical function, along with the CHAQ [12]. Levels of muscle enzymes are included in the IMACS core set as they are felt to be clinically useful but not in the PRINTO core set because they were found to correlate poorly with myositis activity in a clinical trial setting [14]. Global disease activity in JDM is measured with the DAS and the MDAAT. The DAS assesses muscle, cutaneous, and vasculopathic features of JDM [39], while the MDAAT includes extramuscular manifestations and assesses disease activity in 7 organ systems [40]. The cutaneous assessment tool (CAT) assesses the full range of cutaneous manifestations

of JDM and provides both skin disease activity and skin damage scores [41].

Recently, criteria have been proposed for inactive disease in JDM (on or off therapy), requiring 3 or more of (1)  $\text{CPK} \leq 150 \text{ U/I}$ , (2)  $\text{CMAS} \geq 48$ , (3)  $\text{MMT} \geq 78$ , and (4) physician global assessment  $\leq 0.2$  [44]. These criteria should be externally validated prior to use in the clinical or research setting [64].

## 6. Conclusions

Over the last 20 years there has been an explosion of disease activity measures in paediatric rheumatology. The development of these tools is important as we aim to investigate and compare therapies for JIA, JSLE, and JDM. Challenges remain in determining which disease activity measures are optimal for use in the clinic and research settings. We are still working towards incorporating patient perspectives to obtain a comprehensive assessment of health in paediatric rheumatology patients.

## Abbreviations

JIA: Juvenile idiopathic arthritis  
 JSLE: Juvenile systemic lupus erythematosus  
 JDM: Juvenile dermatomyositis  
 NGT: Nominal group technique  
 PRO: Patient-reported outcome.

## Conflict of Interests

The authors declare that they have no conflict of interest.

## References

- [1] W. Thomson, "Electrical units of measurement," *Popular Lectures*, vol. 1, 1883.
- [2] N. Ruperto, A. Pistorio, A. Ravelli et al., "Criteria to define response to therapy in paediatric rheumatic diseases," *European Journal of Clinical Pharmacology*, vol. 67, no. 1, supplement, pp. S125–S131, 2011.
- [3] D. P. M. Symmons, "Disease assessment indices: activity, damage and severity," *Bailliere's Clinical Rheumatology*, vol. 9, no. 2, pp. 267–285, 1995.
- [4] G. Valentini, A. J. Silman, and D. Veale, "Assessment of disease activity," *Clinical and Experimental Rheumatology*, vol. 21, no. 3, pp. S39–S41, 2003.
- [5] T. A. Medsger Jr., S. Bombardieri, L. Czirjak, R. Scorza, A. Della Rossa, and W. Bencivelli, "Assessment of disease severity and prognosis," *Clinical and Experimental Rheumatology*, vol. 21, no. 3, pp. S42–S46, 2003.
- [6] A. Consolaro, N. Ruperto, A. Bazso et al., "Development and validation of a composite disease activity score for juvenile idiopathic arthritis," *Arthritis Care and Research*, vol. 61, no. 5, pp. 658–666, 2009.
- [7] H. de Vet, C. B. Terwee, L. B. Mokkink, and D. L. Knol, *Measurement in Medicine: A Practical Guide*, Cambridge University Press, Cambridge, UK, 2011.
- [8] Organization WH, "Towards a Common Language for Functioning, Disability and Health: the International Classification

- of Functioning, Disability and Health,” ICF Beginner’s Guide, Geneva, Switzerland, 2002.
- [9] M. Arkela-Kautiainen, J. Haapasaaari, H. Kautiainen et al., “Functioning and preferences for improvement of health among patients with juvenile idiopathic arthritis in early adulthood using the WHO ICF model,” *Journal of Rheumatology*, vol. 33, no. 7, pp. 1369–1376, 2006.
- [10] E. H. Giannini, N. Ruperto, A. Ravelli, D. J. Lovell, D. T. Felson, and A. Martini, “Preliminary definition of improvement in juvenile arthritis,” *Arthritis and Rheumatism*, vol. 40, no. 7, pp. 1202–1209, 1997.
- [11] N. Ruperto, A. Ravelli, K. J. Murray et al., “Preliminary core sets of measures for disease activity and damage assessment in juvenile systemic lupus erythematosus and juvenile dermatomyositis,” *Rheumatology*, vol. 42, no. 12, pp. 1452–1459, 2003.
- [12] F. W. Miller, L. G. Rider, Y.-L. Chung et al., “Proposed preliminary core set measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies,” *Rheumatology*, vol. 40, no. 11, pp. 1262–1273, 2001.
- [13] N. Ruperto, A. Ravelli, R. Cuttica et al., “The pediatric rheumatology international trials organization criteria for the evaluation of response to therapy in juvenile systemic lupus erythematosus: prospective validation of the disease activity core set,” *Arthritis and Rheumatism*, vol. 52, no. 9, pp. 2854–2864, 2005.
- [14] N. Ruperto, A. Ravelli, A. Pistorio et al., “The Provisional Paediatric Rheumatology International Trials Organisation/American College of Rheumatology/European League Against Rheumatism disease activity core set for the evaluation of response to therapy in juvenile dermatomyositis: a prospective validation study,” *Arthritis Care and Research*, vol. 59, no. 1, pp. 4–13, 2008.
- [15] C. A. Wallace, N. Ruperto, and E. H. Giannini, “Preliminary criteria for clinical remission for select categories of juvenile idiopathic arthritis,” *Journal of Rheumatology*, vol. 31, no. 11, pp. 2290–2294, 2004.
- [16] D. Foell, N. Wulffraat, L. R. Wedderburn et al., “Methotrexate withdrawal at 6 vs 12 months in juvenile idiopathic arthritis in remission a randomized clinical trial,” *Journal of the American Medical Association*, vol. 303, no. 13, pp. 1266–1273, 2010.
- [17] S. R. Johnson, O.-N. Goek, D. Singh-Grewal et al., “Classification criteria in rheumatic diseases: a review of methodologic properties,” *Arthritis Care and Research*, vol. 57, no. 7, pp. 1119–1133, 2007.
- [18] F. Sztajn bok, D. L. Coronel-Martinez, A. Diaz-Maldonado et al., “Discordance between physician’s and parent’s global assessments in juvenile idiopathic arthritis,” *Rheumatology*, vol. 46, no. 1, pp. 141–145, 2007.
- [19] A. Consolaro, R. Vitale, A. Pistorio et al., “Physicians’ and parents’ ratings of inactive disease are frequently discordant in juvenile idiopathic arthritis,” *Journal of Rheumatology*, vol. 34, no. 8, pp. 1773–1776, 2007.
- [20] M. L. Miller, J. Ruprecht, D. Wang et al., “Physician assessment of disease activity in JIA subtypes. Analysis of data extracted from electronic medical records,” *Pediatric Rheumatology*, vol. 9, article 9, 2011.
- [21] L. G. Rider, B. M. Feldman, M. D. Perez et al., “Development of validated disease activity and damage indices for the juvenile idiopathic inflammatory myopathies—I. Physician, parent, and patient global assessments,” *Arthritis and Rheumatism*, vol. 40, no. 11, pp. 1976–1983, 1997.
- [22] A. Consolaro, N. Ruperto, A. Pistorio et al., “Development and initial validation of composite parent- and child-centered disease assessment indices for juvenile idiopathic arthritis,” *Arthritis Care and Research*, vol. 63, no. 9, pp. 1262–1270, 2011.
- [23] G. Filocamo, A. Consolaro, B. Schiappapietra et al., “A new approach to clinical care of juvenile idiopathic arthritis: the juvenile arthritis multidimensional assessment report,” *Journal of Rheumatology*, vol. 38, no. 5, pp. 938–953, 2011.
- [24] J. A. Singh, D. H. Solomon, M. Dougados et al., “Development of classification and response criteria for rheumatic diseases,” *Arthritis Care and Research*, vol. 55, no. 3, pp. 348–352, 2006.
- [25] H. I. Brunner and A. Ravelli, “Developing outcome measures for paediatric rheumatic diseases,” *Best Practice and Research*, vol. 23, no. 5, pp. 609–624, 2009.
- [26] D. T. Felson and J. J. Anderson, “Methodological and statistical approaches to criteria development in rheumatic diseases,” *Bailliere’s Clinical Rheumatology*, vol. 9, no. 2, pp. 253–266, 1995.
- [27] M. Dougados and L. Gossec, “Classification criteria for rheumatic diseases: why and how?” *Arthritis Care and Research*, vol. 57, no. 7, pp. 1112–1115, 2007.
- [28] L. B. Mokkink, C. B. Terwee, D. L. Patrick et al., “The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes,” *Journal of Clinical Epidemiology*, vol. 63, no. 7, pp. 737–745, 2010.
- [29] N. Ruperto, A. Ravelli, F. Falcini et al., “Performance of the preliminary definition of improvement in juvenile chronic arthritis patients treated with methotrexate,” *Annals of the Rheumatic Diseases*, vol. 57, no. 1, pp. 38–41, 1998.
- [30] N. Ruperto, A. Ravelli, S. Oliveira et al., “The Pediatric Rheumatology International Trials Organization/American College of Rheumatology Provisional criteria for the evaluation of response to therapy in juvenile systemic lupus erythematosus: Prospective validation of the definition of improvement,” *Arthritis Care and Research*, vol. 55, no. 3, pp. 355–363, 2006.
- [31] N. Ruperto, A. Pistorio, A. Ravelli et al., “The Paediatric Rheumatology International Trials Organisation provisional criteria for the evaluation of response to therapy in juvenile dermatomyositis,” *Arthritis Care & Research*, vol. 62, no. 11, pp. 1533–1541, 2010.
- [32] L. G. Rider, E. H. Giannini, M. Harris-Love et al., “Defining clinical improvement in adult and juvenile myositis,” *Journal of Rheumatology*, vol. 30, no. 3, pp. 603–617, 2003.
- [33] H. I. Brunner, D. J. Lovell, B. K. Finck, and E. H. Giannini, “Preliminary definition of disease flare in juvenile rheumatoid arthritis,” *Journal of Rheumatology*, vol. 29, no. 5, pp. 1058–1064, 2002.
- [34] S. Ringold, R. Bittner, T. Neogi, C. A. Wallace, and N. G. Singer, “Performance of rheumatoid arthritis disease activity measures and juvenile arthritis disease activity scores in polyarticular-course juvenile idiopathic arthritis: analysis of their ability to classify the american college of rheumatology pediatric measures of response and the preliminary criteria for flare and inactive disease,” *Arthritis Care and Research*, vol. 62, no. 8, pp. 1095–1102, 2010.
- [35] C. Bombardier, D. D. Gladman, M. B. Urowitz, D. Caron, and C. H. C. Chi Hsing Chang, “Derivation of the SLEDAI: a disease activity index for lupus patients,” *Arthritis and Rheumatism*, vol. 35, no. 6, pp. 630–640, 1992.
- [36] E. M. Hay, P. A. Bacon, C. Gordon et al., “The BILAG index: a reliable and valid instrument for measuring clinical disease

- activity in systemic lupus erythematosus," *Quarterly Journal of Medicine*, vol. 86, no. 7, pp. 447–458, 1993.
- [37] M. H. Liang, S. A. Socher, M. G. Larson, and P. H. Schur, "Reliability and validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus," *Arthritis and Rheumatism*, vol. 32, no. 9, pp. 1107–1118, 1989.
- [38] C. Vitali, W. Bencivelli, D. A. Isenberg et al., "Disease activity in systemic lupus erythematosus: report of the Consensus Study Group of the European Workshop for Rheumatology Research. II. Identification of the variables indicative of disease activity and their use in the development of an activity score," *Clinical and Experimental Rheumatology*, vol. 10, no. 5, pp. 541–547, 1992.
- [39] R. K. Bode, M. S. Klein-Gitelman, M. L. Miller, T. S. Lechman, and L. M. Pachman, "Disease activity score for children with juvenile dermatomyositis: reliability and validity evidence," *Arthritis Care and Research*, vol. 49, no. 1, pp. 7–15, 2003.
- [40] S. M. Sultan, E. Allen, C. V. Oddis et al., "Reliability and validity of the myositis disease activity assessment tool," *Arthritis and Rheumatism*, vol. 58, no. 11, pp. 3593–3599, 2008.
- [41] A. M. Huber, E. M. Dugan, P. A. Lachenbruch et al., "The cutaneous assessment tool: development and reliability in juvenile idiopathic inflammatory myopathy," *Rheumatology*, vol. 46, no. 10, pp. 1606–1611, 2007.
- [42] S. Magni-Manzoni, N. Ruperto, A. Pistorio et al., "Development and validation of a preliminary definition of minimal disease activity in patients with juvenile idiopathic arthritis," *Arthritis Care and Research*, vol. 59, no. 8, pp. 1120–1127, 2008.
- [43] C. A. Wallace, A. Ravelli, B. Huang, and E. H. Giannini, "Preliminary validation of clinical remission criteria using the OMERACT filter for select categories of juvenile idiopathic arthritis," *Journal of Rheumatology*, vol. 33, no. 4, pp. 789–795, 2006.
- [44] D. Lazarevic, A. Pistorio, E. Palmisani et al., "The PRINTO criteria for clinically inactive disease in juvenile dermatomyositis," *Annals of the Rheumatic Diseases*, 2012.
- [45] B. Kirshner and G. Guyatt, "A methodological framework for assessing health indices," *Journal of Chronic Diseases*, vol. 38, no. 1, pp. 27–36, 1985.
- [46] A. M. Huber, B. M. Feldman, R. M. Rennebohm et al., "Validation and clinical significance of the childhood myositis assessment scale for assessment of muscle function in the juvenile idiopathic inflammatory myopathies," *Arthritis and Rheumatism*, vol. 50, no. 5, pp. 1595–1603, 2004.
- [47] G. H. Guyatt, C. Bombardier, and P. X. Tugwell, "Measuring disease-specific quality of life in clinical trials," *Canadian Medical Association Journal*, vol. 134, no. 8, pp. 889–895, 1986.
- [48] M. Batthish, R. Schneider, A. V. Ramanan, C. Achonu, N. L. Young, and B. M. Feldman, "What does 'active disease' mean? Patient and parent perceptions of disease activity in the systemic arthritis form of juvenile idiopathic arthritis (SO-JIA)," *Rheumatology*, vol. 44, no. 6, pp. 796–799, 2005.
- [49] D. L. Streiner and G. R. Norman, "Chapter 18. Principal components and factor analysis: fooling around with factors," in *Health Measurement Scales: A Practical Guide to Their Development and Use*, Oxford University Press, 2003.
- [50] N. Ruperto, S. Meiorin, S. M. Iusan, A. Ravelli, A. Pistorio, and A. Martini, "Consensus procedures and their role in pediatric rheumatology," *Current Rheumatology Reports*, vol. 10, no. 2, pp. 142–146, 2008.
- [51] R. Nair, R. Aggarwal, and D. Khanna, "Methods of formal consensus in classification/diagnostic criteria and guideline development," *Seminars in Arthritis and Rheumatism*, vol. 41, no. 2, pp. 95–105, 2011.
- [52] I. R. Diamond, B. M. Feldman, P. B. Pencharz, S. C. Ling, A. M. Moore, and P. W. Wales, "Defining consensus: a systematic review of Delphi studies," Personal Communication.
- [53] J. Jones and D. Hunter, "Consensus methods for medical and health services research," *British Medical Journal*, vol. 311, no. 7001, pp. 376–380, 1995.
- [54] G. Singh, B. H. Athreya, J. F. Fries, and D. P. Goldsmith, "Measurement of health status in children with juvenile rheumatoid arthritis," *Arthritis and Rheumatism*, vol. 37, no. 12, pp. 1761–1769, 1994.
- [55] A. M. Huber, J. E. Hicks, P. A. Lachenbruch et al., "Validation of the childhood health assessment questionnaire in the juvenile idiopathic myopathies," *Journal of Rheumatology*, vol. 28, no. 5, pp. 1106–1111, 2001.
- [56] J. M. Landgraf, E. Maunsell, K. Nixon Speechley et al., "Canadian-French, German and UK versions of the child health questionnaire: methodology and preliminary item scaling results," *Quality of Life Research*, vol. 7, no. 5, pp. 433–445, 1998.
- [57] H. A. Van Mater, J. W. Williams Jr., R. R. Coeytaux, G. D. Sanders, and A. R. Kemper, "Psychometric characteristics of outcome measures in juvenile idiopathic arthritis: a systematic review," *Arthritis Care and Research*, vol. 64, no. 4, pp. 554–562, 2012.
- [58] F. McErlane, M. W. Beresford, E. M. Baildam et al., "Validity of a three-variable Juvenile Arthritis Disease Activity Score in children with new-onset juvenile idiopathic arthritis," *Annals of the Rheumatic Diseases*, 2012.
- [59] A. V. Ramanan, R. Schneider, M. Batthish et al., "Developing a disease activity tool for systemic-onset juvenile idiopathic arthritis by international consensus using the Delphi approach," *Rheumatology*, vol. 44, no. 12, pp. 1574–1578, 2005.
- [60] H. I. Brunner, B. M. Feldman, C. Bombardier, and E. D. Silverman, "Sensitivity of the Systemic Lupus Erythematosus Disease Activity Index, British Isles Lupus Assessment Group Index, and Systemic Lupus Activity Measure in the evaluation of clinical change in childhood-onset systemic lupus erythematosus," *Arthritis and Rheumatism*, vol. 42, pp. 1354–1360, 1999.
- [61] H. I. Brunner, E. D. Silverman, C. Bombardier, and B. M. Feldman, "European Consensus Lupus Activity Measurement is sensitive to change in disease activity in childhood-onset systemic lupus erythematosus," *Arthritis Care and Research*, vol. 49, no. 3, pp. 335–341, 2003.
- [62] L. G. Rider, V. P. Werth, A. M. Huber et al., "Measures of adult and juvenile dermatomyositis, polymyositis, and inclusion body myositis: physician and Patient/Parent Global Activity, Manual Muscle Testing (MMT), Health Assessment Questionnaire (HAQ)/Childhood Health Assessment Questionnaire (C-HAQ), Childhood Myositis Assessment Scale (CMAS), Myositis Disease Activity Assessment Tool (MDAAT), Disease Activity Score (DAS), Short Form 36 (SF-36), Child Health Questionnaire (CHQ), physician global damage, Myositis Damage Index (MDI), Quantitative Muscle Testing (QMT), Myositis Functional Index-2 (FI-2), Myositis Activities Profile (MAP), Inclusion Body Myositis Functional Rating Scale (IBMFRS), Cutaneous Dermatomyositis Disease Area and Severity Index (CDASI), Cutaneous Assessment Tool (CAT), Dermatomyositis Skin Severity Index (DSSI), Skindex, and Dermatology Life Quality Index (DLQI)," *Arthritis Care and Research*, vol. 63, supplement 11, pp. S118–S157, 2011.

- [63] L. G. Rider, D. Koziol, E. H. Giannini et al., "Validation of manual muscle testing and a subset of eight muscles for adult and juvenile idiopathic inflammatory myopathies," *Arthritis Care and Research*, vol. 62, no. 4, pp. 465–472, 2010.
- [64] N. J. Luca and B. M. Feldman, "Paediatric rheumatic disease: defining clinically inactive disease in juvenile dermatomyositis," *Nature Reviews Rheumatology*, vol. 8, pp. 638–639, 2012.

## Research Article

# Detailed Joint Region Analysis of the 7-Joint Ultrasound Score: Evaluation of an Arthritis Patient Cohort over One Year

S. Ohrndorf,<sup>1</sup> B. Halbauer,<sup>1</sup> P. Martus,<sup>2</sup> B. Reiche,<sup>1</sup> T. M. Backhaus,<sup>3</sup>  
G. R. Burmester,<sup>1</sup> and M. Backhaus<sup>1</sup>

<sup>1</sup> Department of Rheumatology and Clinical Immunology, Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany

<sup>2</sup> Department of Applied Biostatistics, UKT, University of Tübingen, 71070 Tübingen, Germany

<sup>3</sup> Department of Medicine III, Nephrology and Rheumatology, Universitätsklinikum UKE, 20246 Hamburg, Germany

Correspondence should be addressed to M. Backhaus; [marina.backhaus@charite.de](mailto:marina.backhaus@charite.de)

Received 3 April 2013; Accepted 7 June 2013

Academic Editor: Mandana Nikpour

Copyright © 2013 S. Ohrndorf et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Objective.** The main objective of this study was to evaluate the 7-joint ultrasound (US7) score by detailed joint region analysis of an arthritis patient cohort. **Methods.** The US7 score examines the clinically most affected wrist, MCP and PIP II, III, MTP II, and V joints for synovitis, tenosynovitis/paratenonitis, and erosions. Forty-five patients with rheumatoid arthritis (RA) (84.4%) and spondyloarthritis with polyarticular peripheral arthritis (PsA 13.3%; AS 2.2%) with a median disease duration of 6.5 yrs (range 7.5 mths–47.6 yrs) were included and examined at baseline and 3, 6, and 12 months after starting or changing therapy (DMARD/biologic). In this study, detailed US7 score joint region analysis was firstly performed. **Results.** The joint region analysis performed at baseline disclosed synovitis in 95.6% of affected wrists in the dorsal aspect by greyscale (GS) US where Grade 2 (moderate) was most often (48.9%) detected. Palmar wrist regions presented Grade 1 (minor) capsule elevation in 40% and Grade 2 (moderate synovitis) in 37.8%. Tenosynovitis of the extensor carpi ulnaris (ECU) tendon was found in 40%, with PD activity in 6.6%. Most of the erosions in MCP II were detected in the radial (68.9%), followed by the dorsal (48.9%) and palmar (44.4%) aspects. In MTP V, erosions were seen in 75.6% from lateral. **Conclusions.** Synovitis in GSUS was more often detected in the wrist in the dorsal than in the palmar aspect. ECU tendon involvement was frequent. Most erosions were found in the lateral scan of MTP V and the medial (radial) scan of MCP II.

## 1. Introduction

Objective imaging modalities are needed to detect the inflammatory and destructive processes in arthritic diseases such as rheumatoid arthritis (RA) and seronegative spondyloarthritis (e.g., psoriatic arthritis). In recent years, there have been numerous studies reporting early detection of soft tissue and bone processes in arthritic diseases and a high level of sensitivity in musculoskeletal ultrasonography (US) [1–8]. This imaging method allows disease activity and therapeutic response to be detected objectively and for immunosuppressive therapy to be adapted accordingly. As a result, better rheumatic disease outcomes might be achieved and structural damage prevented at earlier stages [9–13]. Due to rapid technical improvements, US has become the “extended diagnostic finger” in the rheumatologist’s daily

practice with high patient acceptability. Therefore, accurate assessment of joint inflammation such as synovitis and bone processes is extremely important and standardization is, therefore, essential. Recently, a novel 7-joint US (US7) score for use in daily rheumatologic practice has been developed which includes examination of the clinically most affected wrist, MCP II, III, PIP II, III, MTP II, and V, that is, the joints that are most frequently involved in RA [14, 15]. They are assessed for synovitis, tenosynovitis/paratenonitis, and erosions according to the EULAR criteria [16] and the OMERACT definition [17] including greyscale (GS) and power Doppler (PD) US. Synovitis and synovial/tenosynovial vascularity are scored semiquantitatively (grade 0–3) by PDUS according to Szkudlarek et al. [18]. Synovitis (effusion and synovial hypertrophy combined) in GSUS is analyzed semiquantitatively as described by Scheel et al. [19].

Tenosynovitis/paratenonitis as well as erosions in GSUS are registered as being absent (0) or present (1). The first publication describes the implementation of the US7 score in a nationwide project in order to prove its value and feasibility in daily rheumatologic practice. One hundred and twenty patients with RA (91%) and PsA (9%) were evaluated at three visits (baseline and after 3 and 6 months) using the US7 score. Clinical data (DAS28) and laboratory parameters (CRP and ESR) were also evaluated. All parameters were significantly reduced after three (except for the PDUS synovitis and erosions score) and six (except for the erosions score) months of therapy or change in immunosuppressive therapy (DMARDs and/or TNF $\alpha$  inhibitors versus DMARDs alone). The study demonstrated that the novel US7 score is not only a feasible score for monitoring disease activity in daily rheumatologic practice but also represents therapeutic response and is therefore sensitive to change [14]. This could also be presented by another US7 score validation study of a larger cohort of  $n = 432$  RA patients [15].

The aim of the present study was to further validate this sum scoring system by detailed joint region analysis. To this end, one-year data of a monocenter subgroup of an arthritis patient cohort were analyzed. Primarily, the occurrence of pathologic findings such as synovitis, tenosynovitis/paratenonitis, and erosions in each joint region included in the US7 score was analyzed over the period of one year under certain therapies. Secondly, each feature was evaluated for its predictive value for later erosions. Besides, US7 score data were compared to clinical and laboratory parameters.

## 2. Patients and Methods

Forty-five patients (77.8% female) suffering from RA (84.4%) and seronegative spondyloarthritis such as psoriatic arthritis (PsA; 13.3%) and ankylosing spondylitis (AS) with peripheral joint involvement (2.2%) with a median age (in years) of 56.0 (range 22–75 {95% CI: 45.5–65.0}) and median year of disease duration of 6.5 (range 7.5 mths–47.6 yrs {95% CI: 4.0–10.5}) were recruited from the Rheumatologic Outpatient Department of the Charité-Universitätsmedizin Berlin, Germany, from February 2007 to March 2009. The patient cohort consists of a monocenter subgroup from a large nationwide study. This study was approved by the ethical committee of Tuebingen, Germany (no. 199/2007BO2), and all included patients gave their informed consent. Required for inclusion was the indication for start or change of therapy to DMARDs and/or biologics (because of either disease activity or medical side effects), as well as a minimum age of 18 years and a statement of agreement. 57.8% of the included patients were positive for rheumatoid factor IgM (RF IgM) and 55.6% for anticitrullinated antibodies (ACPA). Clinical, laboratory, and US data were evaluated before (baseline) starting or changing therapy (DMARD and/or biologica) and after 3, 6, and 12 months. Following the baseline examination, 57.8% of the patients received DMARD and TNF $\alpha$  inhibitor combination therapy, 26.7% DMARDs only, and 15.5% TNF $\alpha$  inhibitor monotherapy.

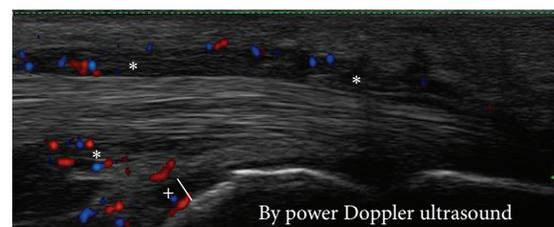


c = caput of the 2nd metacarpus;  
b = basis of the 2nd proximal phalanx

FIGURE 1: Erosion in the 2nd MCP joint from dorsal.



(a)



(b)

FIGURE 2: Tenosynovitis (\*) of the 2nd flexor tendon and synovitis (+) of the 2nd MCP joint—longitudinal scan.

**2.1. Clinical and Laboratory Assessment.** For current disease activity evaluation, the disease activity score 28 (DAS28) was assessed at each patient's visit. Furthermore, erythrocyte sedimentation rate (ESR; normal level < 20 mm/h) and C-reactive protein (CRP; normal level < 5 mg/L) levels were measured at baseline and after 3, 6, and 12 months.

**2.2. Imaging Assessment.** For the *ultrasound* examination, the novel US7 score [14, 15] was applied at each visit. This score includes US evaluation of the following joints of the clinically most affected hand and forefoot: wrist, MCP II, III, PIP II, III, MTP II, and V which are assessed for synovitis, tenosynovitis/paratenonitis, and erosions. Synovitis and synovial/tenosynovial vascularity are scored semiquantitatively (grade 0–3) by GSUS and PDUS and tenosynovitis as well as erosions for their presence (0/1). In this study, the *wrist* was examined in the dorsomedian, ulnar, and palmar aspects for synovitis and tenosynovitis in GSUS and PDUS and for erosions. The finger joints *MCP II* and *III* were assessed in the dorsal aspect for synovitis in PDUS, for paratenonitis in GSUS and PDUS, also for erosions (see an example of an erosion in the 2nd MCP joint in Figure 1), then in the medial (radial) aspect for erosions (only MCP II), and in the palmar aspect for synovitis and tenosynovitis in GSUS and PDUS (an example of tenosynovitis and synovitis in the 2nd MCP joint detected by GSUS and PDUS is given in Figure 2) as well as for erosions. *PIP joints II* and *III* were

examined in the dorsal aspect for synovitis in PDUS and for erosions and in the palmar aspect for synovitis in GSUS and PDUS as well as for erosions. The toe joints, *MTP II and V*, were examined in the dorsal aspect for synovitis in GSUS and PDUS and for erosions and in the plantar and lateral (only *MTP V*) aspects for erosions. Sum scores for synovitis, tenosynovitis/paratenonitis, and erosions were composed for each time of US assessment. The scoring range for the GS synovitis score was 0–27, for the PD synovitis score 0–39, for the GS tenosynovitis score 0–7, for the PD tenosynovitis score 0–21, for the erosions score 0–14 excluding wrist examination, and 0–17 including wrist examination. Sum scores excluding the forefoot joints examination (“US5” score) were also performed in order to determine whether the scoring system, without the forefoot joints *MTP II and V*, is sensitive to change.

Furthermore, a detailed joint region analysis was performed at baseline and after 3, 6, and 12 months calculating the amount of pathologic findings in each joint region included. Two expert sonographers (MB, SO) performed the US7 score examination. The ultrasonographers were both aware of the treatment and the treatment decision. There was no blinding to treatment in this study.

At baseline and after 12 months, conventional *radiographic scans* of both hands and forefeet in two planes were performed according to German recommendations. The presence of erosions was qualitatively assessed (0/1) by the Steinbrocker score (Steinbrocker score  $\leq 1 = 0$ , Steinbrocker  $> 1 = 1$ ) [20].

**2.3. Statistical Analysis.** The statistical calculation was carried out using the statistical software program SPSS 18.0 (SPSS, Chicago, Illinois, USA). Median values and interquartile ranges as well as the amount of pathologic findings (%) were calculated, and changes were subjected to the 2-sided exact Wilcoxon test. The longitudinal differences of laboratory, clinical, and US parameters were correlated using Spearman nonparametric correlation coefficients. For the calculation of predictive values, linear regression analysis was performed, and predictors for dependent variables (US7 erosion score) were calculated. For the calculation of the prediction of later erosions by conventional radiography, logistic regression was used; fit was assessed by the Hosmer-Lemeshow test. Variable selection was applied with inclusion probability assessed by the Rao Score test but significance within the model by the Wald test. Statistical significance ( $P$ ) was set at the  $\alpha \leq 0.05$  level. No adjustment for multiple testing was applied.

### 3. Results

**3.1. Detailed Joint Region Analysis.** Baseline joint region analysis disclosed synovitis in 95.6% affected wrists in the dorsomedian aspect by GSUS, where Grade 2 (moderate) was detected most often (48.9%). Power Doppler activity was found in 64.4% with Grade 2 (35.6%) being detected most often. Furthermore, erosions in this joint region were seen in 68.9% of the patients. Tenosynovitis in GS mode was detected in 17.8% in this joint region with PD activity in

4.4%. Palmomedian wrist regions presented Grade 1 (minor) capsule elevation in 40%, and Grade 2 (moderate synovitis) in 37.8% with PD activity of 11.1%. Erosions in this region were detected in 57.8% of the cases. Tenosynovitis in this joint region was seen in only 6.7% without any PD activity. Tenosynovitis of the extensor carpi ulnaris (ECU) tendon detected by GSUS was found in at least 40% of the joints examined, with PD activity of 6.6%. Erosive changes were seen in 44.4% of the ulnar wrist regions.

Regarding *finger joint* examination, synovitis was determined in 95.6% both in MCP joint II and III in the palmar aspect by GSUS, whereby Grade 2 was detected most often in MCP II (33.3%) and Grade 1 (51.1%) in MCP III. In both joints, PD activity was more frequently observed in the dorsal aspect than in the palmar aspect (MCP II: 24.4% versus 17.8%; MCP III: 24.4% versus 4.4%). Most erosions in MCP II were detected in the medial (radial) aspect (68.9%), followed by the dorsal aspect (48.9%) and the palmar aspect (44.4%). In MCP III, erosions were seen in 44.4% of the joints in the dorsal aspect and in 35.6% in the palmar aspect. Tenosynovitis was detected in 15.6% in MCP II and in 6.7% in MCP III, and paratenonitis was (only) found in 4.4% in MCP II and in 2.2% in MCP III. In the PIP joints included, the following observations were made: PIP II synovitis by GSUS was detected in 81.8% of the joints in the palmar aspect where Grade 1 with 43.2% was detected most often. PIP III was affected by GS synovitis in 68.9% of the cases with Grades 1 and 2 reaching the same incidence (24.4%). In both joints, PD activity was more frequently seen in the dorsal than in the palmar aspect (PIP II: 14.3% versus 13.6%; PIP III 13.3% versus 11.4%). Furthermore, erosions were detected more often in the dorsal aspect than in the palmar aspect (PIP II 51.1% versus 42.2%; PIP III 46.7% versus 31.1%).

In the *toe joints*, *MTP II* was affected by GS synovitis in 84.4% of the cases with PD activity in 33.3% (dorsal examination). Erosions in this joint were more often detected in the dorsal aspect (46.7%) than in the palmar (22.2%). In *MTP V*, synovitis was seen in 51.1% of the joints with PD activity in only 8.9% (dorsal examination). Erosions in *MTP V* were most frequently detected in the lateral aspect (75.6%), followed by the dorsal (57.8%) and palmar aspects (44.4%). See Figure 3 for the semiquantitative analysis of each joint region included in the US7 score (baseline). The one-year joint region analysis is presented in Figure 4.

The following joint regions were the ones which were most significantly changing under new treatment regime over one year: PD positive synovitis of the dorsomedian wrist and GSUS synovitis in the palmar MCP III joint region and in the palmar PIP III region (each  $P < 0.001$ ).

**3.2. Prediction of Erosions after One Year Based on the US7 Score and Laboratory and Clinical Parameters.** For the prediction of erosions in both hands and forefeet detected by conventional radiography after one year, certain predictors were approved. Considering the US7 score features, the synovitis score by GSUS at baseline was a significant predictor ( $P < 0.05$ ) for erosions on radiographs after one year. Furthermore, initial erosions in the US7 score (excluding

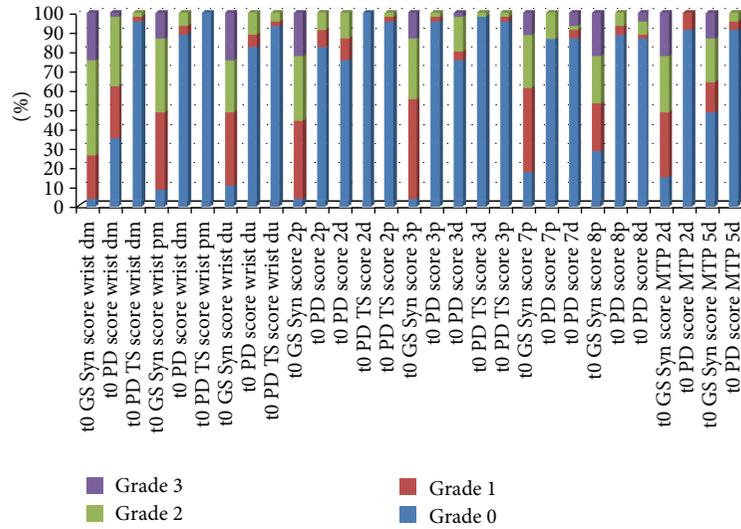


FIGURE 3: Semi-quantitative analysis of each joint region included in the US7 score (without erosions)—baseline (t0). *Legend:* Syn = synovitis; TS = tenosynovitis/paratenonitis; GS = greyscale; PD = power Doppler; dm = dorsomedian; pm = palmomedian; du = dorsoulnar; 2p = MCP II palmar; 2d = MCP II dorsal; 3p = MCP II palmar; 3d = MCP III dorsal; 7p = PIP II palmar; 7d = PIP II dorsal; 8p = PIP III palmar; 8d = PIP III dorsal.

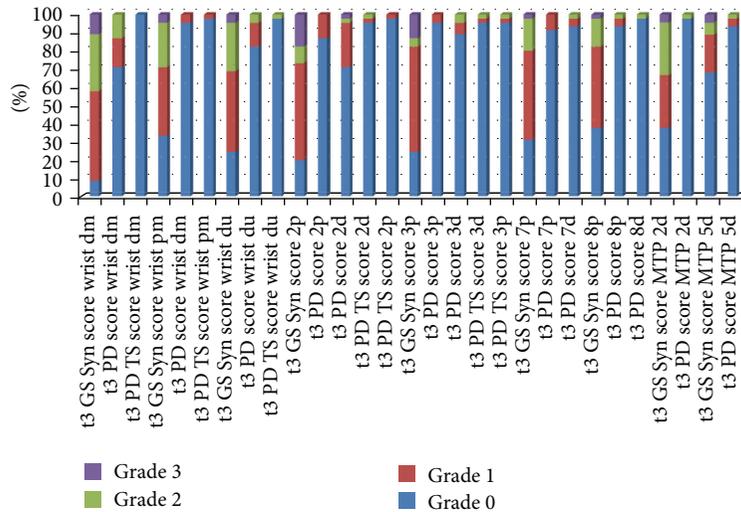


FIGURE 4: Semi-quantitative analysis of each joint region included in the US7 score (without erosions)—one-year followup (t3). *Legend:* Syn = synovitis; TS = tenosynovitis/paratenonitis; GS = greyscale; PD = power Doppler; dm = dorsomedian; pm = palmomedian; du = dorsoulnar; 2p = MCP II palmar; 2d = MCP II dorsal; 3p = MCP II palmar; 3d = MCP III dorsal; 7p = PIP II palmar; 7d = PIP II dorsal; 8p = PIP III palmar; 8d = PIP III dorsal.

and including wrist examination) were highly significant predictors for erosions in conventional radiography after one year ( $P < 0.001$ ). In multivariate analysis, baseline US5 erosions sum score ( $P < 0.001$ , odds ratio = 2.31 {95% CI: 1.46–3.65}) and US7 GS tenosynovitis/paratenonitis sum score (Rao score test  $P = 0.021$ , Wald test  $P = 0.056$ , odds ratio = 2.45, {95% CI: 0.98–6.15}), which was not significant in univariate analysis, were included by the variable selection procedure. The Hosmer-Lemeshow test revealed a good fit of the final model ( $P = 0.32$ ). Other US7 score features such as the synovitis score in PDUS and

tenosynovitis/paratenonitis score did not significantly predict erosions by radiographs after one year. Laboratory (CRP, ESR) and the clinical parameter DAS28 were not significant predictors of erosions in conventional radiography either.

For the prediction of erosions detected by the US7 score (erosion score including the wrist) after one year, the following significant baseline predictors were evaluated: DAS28 ( $P < 0.05$ ), US7 synovitis score in GSUS ( $P = 0.001$ ), and PDUS ( $P < 0.05$ ), of which the synovitis score in GSUS was the only multivariately significant predictor. Laboratory data (CRP, ESR) and antibodies (RF IgM and

ACPA positivity) as well as tenosynovitis/paratenonitis scores by GSUS and PDUS did not significantly predict erosions in ultrasonography. The fit of this linear model was satisfactory; quadrat terms were not significant and residuals approximately normally distributed.

For the prediction of the erosions score excluding wrist examination, similar results were obtained as follows: US7 synovitis score in GSUS ( $P = 0.002$ ) and PDUS ( $P = 0.001$ ). The DAS28 was not predictor for significant erosions if excluding the wrist examination for erosions ( $P = 0.102$ ).

**3.3. Laboratory, Clinical, and US7 Score Data over One Year.** Laboratory (ESR, CRP), clinical (DAS28), and US7 (also without forefoot joints examination—"US5") score data were evaluated at each assessment time, and changes to baseline were examined (Tables 1 and 2).

The laboratory parameters, ESR and CRP, significantly decreased 3 months after starting or changing therapy. After 6 and 12 months, median ESR and CRP did not change significantly versus baseline examination. Median DAS28 significantly decreased from 4.8 at baseline to 4.1 (after 3 months), then from 3.7 (after 6 months) to 3.8 (after 12 months;  $P < 0.05$  in each case to baseline). A significant reduction of the median synovitis score in GSUS was also observed at each assessment time from 13.0 initially to 9.0 (30.8%;  $P < 0.05$ ) 12 months later. The median synovitis score in PDUS decreased significantly from 2.0 initially to 1.0 after 6 months (50%;  $P < 0.05$ ). The median tenosynovitis/paratenonitis score in GSUS (initially 1.0) was significantly reduced after 6 and 12 months (each 0.0;  $P < 0.05$ ) of US examination. The median tenosynovitis/paratenonitis score in PDUS remained the same over the period of one year. According to the erosions score, both excluding and including wrist examination, a significant reduction was observed from 7.0 initially (excluding the wrist) and 8.0 (including the wrist) to 5.0 and 6.0 after 12 months, respectively ( $P < 0.05$  in each case; US7 score data, Table 1). Excluding the forefoot (MTP II and V) US examination, the same statistical analysis compared to the US7 score analysis was performed ("US5" score data, Table 2). The GSUS synovitis score significantly decreased at each assessment time from 11.0 initially to 7.0 (36.4%,  $P < 0.05$ ) after 12 months. A significant reduction of the median synovitis score in PDUS was first seen after 6 months from 2.0 to 1.0 (50.0%;  $P < 0.05$ ), then up to 1.7, but still significantly reduced to baseline (15.0%;  $P < 0.05$ ). Furthermore, the erosions score for the hand, both excluding and including wrist examination, was significantly reduced from 4.0 initially (excluding the wrist) and 6.0 (including the wrist) to 2.9 and 5.0, respectively ( $P < 0.05$  in each case).

At baseline, 49% of the patients had erosions in both hands and 35% in both feet in conventional radiographs. After 12 months, 51.1% of the patients had erosions in both hands and 26.7% in both feet.

**3.4. Correlations between the US7 Score and Laboratory and Clinical Parameters over 12-Month Followup.** There was a significant correlation between changes in the US7 score

obtained by GSUS and the ESR through 12 months of followup (GSUS7/ESR:  $r = 0.31$ ;  $P < 0.05$ ). No other significant correlations were detected between the US7 score obtained by GSUS and PDUS compared to laboratory and clinical parameters. No significant positive correlation coefficients were found between the "US5" score (excluding the forefoot joints) data and clinical and laboratory parameters.

## 4. Discussion

The aim of the present study was to further validate the US7 score by a detailed joint region analysis. For this purpose, the incidence of pathologic findings such as synovitis, tenosynovitis/paratenonitis, and erosions in each included joint region were evaluated and analyzed over the period of one year. Secondly, different components of the US7 sum score were examined for their ability to predict later erosions. Furthermore, one-year data of the US7 score evaluated in this subgroup of patients were compared to clinical (DAS28) and laboratory (CRP, ESR) parameters.

Regarding the detailed joint region analysis, the most affected joints/joint regions affected by synovitis (detected by GSUS) were the wrist in the dorsal aspect and the MCP II in the palmar aspect, whereby the wrist was more severely affected. Furthermore, power Doppler activity  $>$  Grade 1 was mostly seen in the dorsal wrist joint. It was also one of the joint regions, which was most significantly changing under new treatment regime. Recently, Ellegaard et al. presented a study in which they proposed that a standardized color Doppler US examination of the wrist joint as the only target joint was very helpful in the detection of disease activity with a high correlation to CRP, ESR, swollen joint count, and DAS28 [21].

For the US7 score evaluation, synovitis in grey scale was only evaluated in the palmar aspect for the finger joint examination. In view of the findings of the study by Scheel et al. [19] in which synovitis was most often presented in the palmar proximal side of the finger joints (86%), we decided to exclude the dorsal compartment examination by GSUS, especially in the case of time period. This phenomenon was recently determined in a study by Vlad et al. in which the palmar finger joint regions (MCP and PIP II-V) showed better correlation to clinical evaluation (CDAI and SDAI) than the dorsal examination [22]. Furthermore, Dougados et al. were not able to demonstrate that musculoskeletal US was more sensitive than clinical examination [23]. This may be due to the fact that synovitis was only examined in the dorsal aspect in this study; a region which can easily be evaluated by clinical examination. However, for the detection of synovitis by PDUS, the dorsal side of the finger joints should be included, as PD activity was found there more often.

Most erosions were found in the ulnar scan of the MTP V and the radial scan of the MCP II joint. This distribution of erosions was shown by a French US working group before presenting new semiquantitative erosions score with good correlation to radiography. This group also found erosions primarily in MTP V, followed by MCP II [24].

TABLE 1: Laboratory, clinical, and US7 score data over one year.

<i>n</i> = 45 patients	Baseline	After 3 months	After 6 months	After 12 months
ESR (mm/h)	24.5 (5–82) {95% CI: 14.3–36.0}	14.0 (3–74)* {95% CI: 8.0–29.5}	19.0 (3–120) {95% CI: 7.5–38.3}	14.5 (3–94) {95% CI: 9.8–36.8}
CRP (mg/L)	4.1 (0.2–119) {95% CI: 2.0–9.2}	3.0 (0.3–34.8)* {95% CI: 1.1–5.9}	3.0 (0.2–38.1) {95% CI: 1.0–6.8}	3.6 (0.0–65) {95% CI: 1.4–7.5}
DAS28	4.8 (1.6–8.0) {95% CI: 4.1–6.1}	4.1 (1.4–7.6)* {95% CI: 2.9–4.9}	3.7 (1.1–6.8)* {95% CI: 2.5–5.2}	3.8 (1.3–7.1)* {95% CI: 2.8–4.7}
Synovitis score in GSUS (0–27)	13.0 (7–25) {95% CI: 0.5–6.5}	10.0 (3–22)* {95% CI: 6.5–15.0}	10.0 (2–18)* {95% CI: 7.0–13.5}	9.0 (0–20)* {95% CI: 6.0–12.0}
Synovitis score in PDUS (0–39)	2.0 (0–16) {95% CI: 0.5–6.5}	1.0 (0–20) {95% CI: 0.0–5.0}	1.0 (0–18)* {95% CI: 0.0–3.0}	2.0 (0–9)* {95% CI: 0.0–2.5}
Tenosynovitis/paratenonitis score in GSUS (0–7)	1.0 (0–5) {95% CI: 0.0–1.0}	0.0 (0–5) {95% CI: 0.0–1.0}	0.0 (0–3)* {95% CI: 0.0–1.0}	0.0 (0–5) {95% CI: 0.0–1.0}
Tenosynovitis/paratenonitis score in PDUS (0–21)	0.0 (0–8) {95% CI: 0.0–0.0}	0.0 (0–8) {95% CI: 0.0–0.0}	0.0 (0–5) {95% CI: 0.0–0.0}	0.0 (0–6) {95% CI: 0.0–0.0}
Erosions score excluding the wrist (0–14)	7.0 (0–14) {95% CI: 4.0–9.0}	6.0 (1–14) {95% CI: 4.0–9.0}	6.0 (0–14) {95% CI: 4.0–9.0}	5.0 (0–11)* {95% CI: 3.0–7.0}
Erosions score including the wrist (0–17)	8.0 (0–17) {95% CI: 5.0–11.0}	7.0 (2–17) {95% CI: 6.0–11.5}	8.0 (0–17) {95% CI: 5.0–11.5}	6.0 (1–14)* {95% CI: 5.0–9.0}

\* *P* value < 0.05 to baseline examination by 2-sided exact Wilcoxon test.

TABLE 2: “US5” (without forefoot results) score data over one year.

<i>n</i> = 45 patients	Baseline	After 3 months	After 6 months	After 12 months
Synovitis score in GSUS (0–21)	11.0 (7–19) {95% CI: 8.0–13.0}	9.0 (2–18)* {95% CI: 6.0–12.5}	9.0 (1–16)* {95% CI: 5.5–10.5}	7.0 (0–20)* {95% CI: 5.0–10.0}
Synovitis score in PDUS (0–33)	2.0 (0–15) {95% CI: 0.5–6.0}	1.0 (0–17) {95% CI: 0.0–5.0}	1.0 (0–18)* {95% CI: 0.0–3.0}	1.7 (0–7)* {95% CI: 0.0–2.5}
Erosions score excluding the wrist (0–9)	4.0 (0–9) {95% CI: 2.0–6.0}	3.0 (0–15) {95% CI: 2.0–6.0}	4.0 (0–9) {95% CI: 2.0–6.0}	2.9 (0–7)* {95% CI: 1.0–4.5}
Erosions score including the wrist (0–12)	6.0 (0–12) {95% CI: 3.0–8.5}	5.0 (1–12) {95% CI: 4.0–9.0}	6.0 (0–12) {95% CI: 3.0–8.5}	5.0 (0–10)* {95% CI: 3.0–7.0}

\* *P* value < 0.05 to baseline examination by 2-sided exact Wilcoxon test.

The predictive value of the US7 erosions score after one year was most significant for the synovitis score in GSUS and less, but still significant, for the synovitis score in PDUS and the DAS28 clinical score (if detecting erosions in the wrist). The predictive value for radiographic erosions after one year was most significant for the US7 erosions score at baseline indicating that the initial US erosions score did have the highest predictive power for the disease outcome. Furthermore, the synovitis score in GSUS did significantly predict later erosions on radiography. Therefore, permanent reduction of synovitis needs to be the most important aim of the therapeutic concept in order to prevent the destructive process of the disease. In the present study, tenosynovitis, especially ECU tenosynovitis, was not predictive for later erosions. However, this could recently be presented by Lillegraven et al. [25] who found ECU tenosynovitis to be the only independent predictive value for later erosions in the hands in MRI. The present calculation probably results from the fact that these arthritis patients already had a long disease history at study onset, and the erosions were only detected by

conventional radiography. It is a known fact that MRI is more sensitive in the detection of erosions than radiography.

One-year data of the US7 score compared to clinical (DAS28) and laboratory (CRP, ESR) parameters disclosed a significant reduction in the one-year data in most of its components; that is, this novel sum score is sensitive to change under certain therapies. Even *without* examination of the forefoot joints MTP II and V (“US5” score), this scoring system still worked well as the same statistically significant changes were calculated between followup and baseline. Consequently, additional time was saved by omitting the forefoot US examination which would make this score even more feasible. However, correlation analysis of changes to baseline only presented significant coefficients between the synovitis score in GSUS and ESR when the forefoot examination (US7 score) was included but not without it (“US5” score). Therefore, inclusion of the forefoot (MTP II and V) examination may contribute to higher sensitivity of this composite scoring system. Furthermore, for the individual patient it might be very important to include these joints even

so the average statistical comparison of populations and/or time points suggest to shorten the examination accordingly.

Regarding the US7 erosion score, a statistically significant reduction over the examination period of one year was detected. Therefore, a potential “healing effect” of bone lesions beyond immunosuppressive therapy needs to be discussed. This phenomenon was already described by Rau et al. for radiographic erosions [26, 27] and was recently evaluated by Finzel et al. for erosions detected by micro-computed tomography ( $\mu$ CT) [28]. In the Finzel et al. paper, bone erosions in RA patients receiving either tumor necrosis factor inhibitors or methotrexate were assessed by micro-CT imaging. After one year, patients taking TNF $\alpha$  inhibitors showed partial recovery in terms of a decrease in the mean depth of erosions while the mean width remained the same. In contrast, patients taking only methotrexate demonstrated an increase in the main depth and width of the erosions. To our knowledge, a reduction of erosions has not yet been described for musculoskeletal ultrasonography. One reason for the reduction of erosions might be due to the fact that, with less synovitis in the follow-up examinations, the delay path to bone surface is reduced so that the erosions are no longer readily detectable. Therefore, the erosions would only seem to be reduced but no longer be reproduced. However, interobserver reliability for the US7 erosions score was  $\kappa = 0.45$  [29], at least, which means that there is moderate agreement concerning the detection of this pathology. Furthermore, in a recent study by Døhn et al. it was shown that erosions evaluated by US are true erosions compared to computer tomography [30]. Most of the patients participating in the present study also received TNF $\alpha$  inhibitor therapy (57.8% in combination with DMARDs and 15.5% alone), a fact that makes the results even more plausible. Therefore, the healing effect of erosions is also detectable by musculoskeletal US, a new aspect for this imaging modality.

Summarizing the findings, it could be said that both the US7 and the “US5” scores are feasible sum scoring systems for use in daily rheumatologic practice. But because paratenonitis was a rare finding, it might not be a necessary component in the US7 scoring system and could therefore be excluded.

Further studies, especially with a homogenous group of early RA/arthritis patients, need to follow in order to examine the role of the additive value of the US7 score compared to conventional clinical and serological parameters, especially with regard to the outcome parameters (e.g., its value as a predictor of later erosions or to characterize patients who do not respond to certain therapies (i.e., TNF $\alpha$  inhibitors), etc.). Besides, the question concerning the meaning of subclinical disease activity detected by musculoskeletal US should be further discussed, for example, in case of therapeutic escalation. Furthermore, thresholds for the different components of the US7 score need to be analyzed and defined in order to standardize this composite scoring system more thoroughly.

Regarding the observed positive overall effect comparing followup with baseline, we have to point out that this was not a controlled therapeutic study. Especially we cannot exclude that patients entered the observation at a peak of disease burden, and in the followup, a regression to the mean was

observed. Furthermore, the factor of a “mix” of different longstanding arthritis patients, though mainly RA patients, is the main limitation of the study.

## Abbreviations

US7 score:	7-joint ultrasound score
“US5” score:	5-joint ultrasound score (without forefoot)
RA:	Rheumatoid arthritis
PsA:	Psoriatic arthritis
AS:	Ankylosing spondylitis
DMARD:	Disease modifying antirheumatic drugs
GSUS:	Grayscale ultrasound
PDUS:	Power Doppler ultrasound
MCP:	Metacarpophalangeal
PIP:	Proximal interphalangeal
MTP:	Metatarsophalangeal
EULAR:	European League Against Rheumatism
OMERACT:	Outcome measure in rheumatology clinical trial
ESR:	Erythrocyte sedimentation rate
CRP:	C-reactive protein
DAS28:	Disease activity score of 28 joints
CDAI:	Clinical disease activity index
SDAI:	Simplified disease activity index
CT:	Computer tomography
MRI:	Magnetic resonance imaging.

## Conflict of Interests

There were no competing interests in the study’s concept.

## Authors’ Contribution

All authors were involved in drafting the paper or revising it critically for important intellectual content, and all authors approved the final version to be published. No medical writer was involved in the preparation of the paper.

## Acknowledgments

The authors acknowledge Dr. Ekkehart Dietz for any statistical advice. Furthermore, the authors thank Abbott GmbH & Co. KG, Max-Planck-Ring 2a, Wiesbaden, Germany, for financially supporting this project. Abbott financially supported this project but did not exert any influence on the statistical analysis or preparation of the paper.

## References

- [1] M. Backhaus, T. Kamradt, D. Sandrock et al., “Arthritis of the finger joints: a comprehensive approach comparing conventional radiography, scintigraphy, ultrasound, and contrast-enhanced magnetic resonance imaging,” *Arthritis and Rheumatism*, vol. 42, no. 6, pp. 1232–1245, 1999.
- [2] R. J. Wakefield, W. W. Gibbon, P. G. Conaghan et al., “The value of sonography in the detection of bone erosions in patients with rheumatoid arthritis: a comparison with conventional

- radiography," *Arthritis and Rheumatism*, vol. 43, no. 12, pp. 2762–2770, 2000.
- [3] M. Backhaus, G. R. Burmester, D. Sandrock et al., "Prospective two year follow up study comparing novel and conventional imaging procedures in patients with arthritic finger joints," *Annals of the Rheumatic Diseases*, vol. 61, no. 10, pp. 895–904, 2002.
  - [4] M. Szkudlarek, E. Narvestad, M. Klarlund, M. Court-Payen, H. S. Thomsen, and M. Østergaard, "Ultrasonography of the metatarsophalangeal joints in rheumatoid arthritis: comparison with magnetic resonance imaging, conventional radiography, and clinical examination," *Arthritis and Rheumatism*, vol. 50, no. 7, pp. 2103–2112, 2004.
  - [5] A. K. Scheel, K. G. A. Hermann, S. Ohrndorf et al., "Prospective 7 year follow up imaging study comparing radiography, ultrasonography, and magnetic resonance imaging in rheumatoid arthritis finger joints," *Annals of the Rheumatic Diseases*, vol. 65, no. 5, pp. 595–600, 2006.
  - [6] N. J. McHugh, C. Balachrishnan, and S. M. Jones, "Progression of peripheral joint disease in psoriatic arthritis: a 5-yr prospective study," *Rheumatology*, vol. 42, no. 6, pp. 778–783, 2003.
  - [7] D. Kane, "The role of ultrasound in the diagnosis and management of psoriatic arthritis," *Current Rheumatology Reports*, vol. 7, no. 4, pp. 319–324, 2005.
  - [8] A. L. Tan and D. McGonagle, "Psoriatic arthritis: correlation between imaging and pathology," *Joint Bone Spine*, vol. 77, no. 3, pp. 206–211, 2010.
  - [9] R. Maini, E. W. St Clair, F. Breedveld et al., "Infliximab (chimeric anti-tumour necrosis factor  $\alpha$  monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial," *The Lancet*, vol. 354, no. 9194, pp. 1932–1939, 1999.
  - [10] J. S. Smolen, C. Han, M. Bala et al., "Evidence of radiographic benefit of treatment with infliximab plus methotrexate in rheumatoid arthritis patients who had no clinical improvement: a detailed subanalysis of data from the anti-tumor necrosis factor trial in rheumatoid arthritis with concomitant therapy study," *Arthritis and Rheumatism*, vol. 52, no. 4, pp. 1020–1030, 2005.
  - [11] F. C. Breedveld, M. H. Weisman, A. F. Kavanaugh et al., "The PREMIER study: a multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment," *Arthritis and Rheumatism*, vol. 54, no. 1, pp. 26–37, 2006.
  - [12] D. van der Heijde, L. Klareskog, V. Rodriguez-Valverde et al., "Comparison of etanercept and methotrexate, alone and combined, in the treatment of rheumatoid arthritis: two-year clinical and radiographic results from the TEMPO study, a double-blind, randomized trial," *Arthritis and Rheumatism*, vol. 54, no. 4, pp. 1063–1074, 2006.
  - [13] R. Landewé, D. van der Heijde, L. Klareskog, R. van Vollenhoven, and S. Fatenejad, "Disconnect between inflammation and joint destruction after treatment with etanercept plus methotrexate: results from the trial of etanercept and methotrexate with radiographic and patient outcomes," *Arthritis and Rheumatism*, vol. 54, no. 10, pp. 3119–3125, 2006.
  - [14] M. Backhaus, S. Ohrndorf, H. Kellner et al., "Evaluation of a novel 7-joint ultrasound score in daily rheumatologic practice: a pilot project," *Arthritis Care and Research*, vol. 61, no. 9, pp. 1194–1201, 2009.
  - [15] T. M. Backhaus, S. Ohrndorf, H. Kellner et al., "The US7 score is sensitive to change in a large cohort of patients with rheumatoid arthritis over 12 months of therapy," *Annals of the Rheumatic Diseases*, vol. 72, no. 7, pp. 1163–1169, 2013.
  - [16] M. Backhaus, G. R. Burmester, T. Gerber et al., "Guidelines for musculoskeletal ultrasound in rheumatology," *Annals of the Rheumatic Diseases*, vol. 60, no. 7, pp. 641–649, 2001.
  - [17] R. J. Wakefield, P. V. Balint, M. Szkudlarek, and G. A. W. Bruyn, "Musculoskeletal ultrasound including definitions for ultrasonographic pathology," *Journal of Rheumatology*, vol. 32, no. 12, pp. 2485–2487, 2005.
  - [18] M. Szkudlarek, M. Court-Payen, S. Jacobsen, M. Klarlund, H. S. Thomsen, and M. Østergaard, "Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis," *Arthritis and Rheumatism*, vol. 48, no. 4, pp. 955–962, 2003.
  - [19] A. K. Scheel, K. A. Hermann, E. Kahler et al., "A novel ultrasonographic synovitis scoring system suitable for analyzing finger joint inflammation in rheumatoid arthritis," *Arthritis and Rheumatism*, vol. 52, no. 3, pp. 733–743, 2005.
  - [20] O. Steinbrocker, C. H. Traeger, and R. C. Batterman, "Therapeutic criteria in rheumatoid arthritis," *The Journal of the American Medical Association*, vol. 140, no. 8, pp. 659–662, 1949.
  - [21] K. Ellegaard, S. Torp-Pedersen, L. Terslev, B. Danneskiold-Samsøe, M. Henriksen, and H. Bliddal, "Ultrasound colour Doppler measurements in a single joint as measure of disease activity in patients with rheumatoid arthritis—assessment of concurrent validity," *Rheumatology*, vol. 48, no. 3, pp. 254–257, 2009.
  - [22] V. Vlad, F. Berghea, S. Libianu et al., "Ultrasound in rheumatoid arthritis—volar versus dorsal synovitis evaluation and scoring," *BMC Musculoskeletal Disorders*, vol. 12, article 124, 2011.
  - [23] M. Dougados, S. Jousse-Joulin, F. Mistretta et al., "Evaluation of several ultrasonography scoring systems for synovitis and comparison to clinical examination: results from a prospective multicentre study of rheumatoid arthritis," *Annals of the Rheumatic Diseases*, vol. 69, no. 5, pp. 828–833, 2010.
  - [24] J. Sommier, S. Michel-Batot, N. Sauliere et al., "Structural lesions in RA: proposition for a new semiquantitative score (ScUSSE: Scoring by UltraSound Structural erosion)," *Arthritis and Rheumatism*, vol. 54, supplement 209, 2006.
  - [25] S. Lillegraven, P. Bøyesen, H. B. Hammer et al., "Tenosynovitis of the extensor carpi ulnaris tendon predicts erosive progression in early rheumatoid arthritis," *Annals of the Rheumatic Diseases*, vol. 70, no. 11, pp. 2049–2050, 2011.
  - [26] R. Rau, S. Wassenberg, G. Herborn, W. T. Perschel, and G. Freitag, "Identification of radiologic healing phenomena in patients with rheumatoid arthritis," *Journal of Rheumatology*, vol. 28, no. 12, pp. 2608–2615, 2001.
  - [27] R. Rau, G. Herborn, and S. Wassenberg, "Healing of erosive changes in rheumatoid arthritis," *Clinical and Experimental Rheumatology*, vol. 22, no. 5, supplement 35, pp. S44–S49, 2004.
  - [28] S. Finzel, J. Rech, S. Schmidt et al., "Repair of bone erosions in rheumatoid arthritis treated with tumour necrosis factor inhibitors is based on bone apposition at the base of the erosion," *Annals of the Rheumatic Diseases*, vol. 70, no. 9, pp. 1587–1593, 2011.
  - [29] S. Ohrndorf, I. U. Fischer, H. Kellner et al., "Reliability of the novel 7-joint ultrasound score (US7): results from an inter- and intra-observer study performed by rheumatologists," *Arthritis Care and Research*, vol. 64, no. 8, pp. 1238–1243, 2012.

- [30] U. M. Døhn, B. J. Ejlertsen, M. Court-Payen et al., "Are bone erosions detected by magnetic resonance imaging and ultrasonography true erosions? A comparison with computed tomography in rheumatoid arthritis metacarpophalangeal joints," *Arthritis Research and Therapy*, vol. 8, no. 4, article R110, 2006.

## Review Article

# Measuring Disease Activity in Psoriatic Arthritis

**Priscilla C. H. Wong, Ying-Ying Leung, Edmund K. Li, and Lai-Shan Tam**

*Department of Medicine and Therapeutics, Prince of Wales Hospital, Chinese University of Hong Kong, Shatin, New Territories, Hong Kong*

Correspondence should be addressed to Lai-Shan Tam, [lstam@cuhk.edu.hk](mailto:lstam@cuhk.edu.hk)

Received 3 September 2012; Accepted 17 November 2012

Academic Editor: Zahi Touma

Copyright © 2012 Priscilla C. H. Wong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Over the past decade, the assessment of the disease activity in psoriatic arthritis (PsA) has rapidly evolved in view of the need for valid, feasible, and reliable outcome measures that can be ideally employed in longitudinal cohorts, clinical trials, and clinical practice as well as the growing paradigm of tight disease control and treating to target in the management of PsA. This paper reviews the currently available measures used in the assessment of the disease activity in PsA. The composite measures for PsA that are under development are also discussed.

## 1. Introduction

Psoriatic arthritis (PsA) is a heterogenous multifaceted inflammatory arthritis associated with psoriasis. In addition to peripheral arthritis, patient with PsA may develop spondylitis, dactylitis, enthesitis, and nail disease as well as extra-articular features common to the spondyloarthropathies (SpA). The assessment of disease activity in PsA should therefore evaluate each of these clinical domains carefully. An accurate measurement of disease activity is essential to guide the medical therapy and monitor the treatment response.

Over the past decade, significant progress on the development and validation of instruments for the measurement of disease activity in PsA has been achieved by the Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) and the Outcome Measures in Rheumatology Clinical Trials (OMERACT). In 2007, the GRAPPA-OMERACT achieved consensus on 6 core domains that should be included in randomized controlled trials and longitudinal observational cohorts of subjects with PsA [1]. These included peripheral joint activity, skin activity, pain, patient global assessment (PGA), physical function, and health-related quality of life. Several other domains (spinal disease, dactylitis, enthesitis, fatigue, nail disease, radiography, physician global assessment, and acute-phase reactants)

were considered important, not mandatory, but preferably to be assessed at some point in a clinical trial development program. This paper reviews the currently available tools for the clinical measurement of the disease activity in PsA. The composite measures for PsA that take different disease domains into account will also be reviewed.

A collected tools are used in the validation of different patient outcomes measures. These include reliability, validity, and responsiveness. The definition of these tools are first summarized here. In general, assessments of reliability indicate the consistency of responses within a scale or the extent to which a response of score is free from random error (precision). Internal consistency tests the extent that the items of a scale measure the same underlying construct or theme. Intraobserver reliability assesses the likelihood that someone administering a scale to make the same ratings or judgements on repeated administrations. Interobserver reliability assesses the likelihood that two observers to make the same ratings or judgements about the same individual. Validity has been considered to be an expression of the extent to which a question or measure assesses what it is intended to measure. Content validity assesses whether the instrument covers the concepts it is intended to measure. Construct validity assesses whether the measure correlates with measures of other variables in hypothesized ways. Responsiveness

TABLE 1: Clinical assessment of psoriasis outcome measures.

	Erythema	Induration	Desquamation	BSA	Psychosocial impact	Measured by
BSA				Yes		Physician
PASI	Yes	Yes	Yes	Yes		Physician
PGA	Yes	Yes	Yes			Physician
LSPGA	Yes	Yes	Yes	Yes		Physician
SAPASI	Yes	Yes	Yes	Yes		Patient
SPI	Yes	Yes	Yes	Yes	Yes	Physician

BSA: Body Surface Area; PASI: Psoriasis Area and Severity Index; PGA: Physician's Global Assessment; LGPGA: Lattice System Physician's Global Assessment; SAPASI: Self-Administered PASI; SPI: Salford Psoriasis Index.

or sensitivity to change tests how well the scores on the instrument change to reflect changes in the criterion measure.

## 2. Literature Search

Literature search was conducted using the PubMed database up to November 2012. Different combinations of the following search terms were used: "psoriatic arthritis," "psoriasis," "disease activity," "outcome measures," "peripheral arthritis," "dactylitis," "enthesitis," "nail," "spondylitis," "patient global," "quality of life," "fatigue" and "composite measures," with limits set to include humans and written in English. The initial search yielded more than 3000 abstracts, which were reviewed to include only studies relevant to this paper. This yielded 81 studies, of which the full articles were then reviewed by the authors.

## 3. Peripheral Joint Assessment

Moll and Wright described five clinical patterns among patients with PsA: distal interphalangeal (DIP), asymmetrical oligoarticular, symmetric polyarticular, spondylitis, and arthritis mutilans [2]. Unlike rheumatoid arthritis (RA), the pattern of joint involvement in PsA is usually asymmetric and frequently involves the DIP joints. Peripheral joints are assessed for tenderness and swelling. There is no validated measure to assess peripheral joint in PsA. The measure used is the American College of Rheumatology (ACR) joint count initially developed in 1949 for the assessment of patients with RA [3]. The ACR joint count ranges from 28, 44, 68, and 78 for tenderness; 28, 44, 66, and 76 for swelling (excluding hips from assessment of swelling). The reduced joint count with the 28 joints or 44 joints count that does not assess the DIP joint or the feet does not have content validity for measuring peripheral arthritis in PsA. Discussions at GRAPPA and OMERACT meetings recommended that the ACR joint count of 68 tender and 66 swollen joints count be used, as it includes a majority of joints affected in PsA, and it can be readily performed in a clinic visit [4]. It was decided not to include the distal joints of the feet (78 tender joints count) as it may be difficult to distinguish proximal interphalangeal (PIP) joint from DIP joint inflammation in the toes. It has been suggested that if either the PIP or DIP of the toe is involved it should be marked as a PIP. The ACR joint

count was proven to be a reliable measure of peripheral joint activity in PsA in 3 different studies [5–7]. Gladman et al. showed that the 68 tender and 66 swollen joints count had minimal intraobserver and interobserver variation [6]. The 68 tender and 66 swollen joints count includes the temporomandibular, sternoclavicular, acromioclavicular, shoulder, elbow, wrist (including the carpometacarpal and intercarpal joints as one unit), metacarpophalangeal (MCP), PIP, DIP, hip, knee, talotibial, midtarsal (including subtalar), metatarsophalangeal, and interphalangeal joints of the toes (proximal and distal joints of each toe counted as one unit).

## 4. Skin Assessment

PsA disease activity is commonly not mirrored by active skin disease. A wide variety of scoring systems have been proposed to evaluate severity in psoriasis. A systemic review of all clinical studies (prospective and retrospective) investigating the severity of psoriatic patients was published in 2010 [8]. Based on methodological validation and quality criteria, six clinical severity scores were selected and analyzed (Table 1). They included Body Surface Area (BSA) [9], Psoriasis Area, and Severity Index (PASI) [10], the Physician's Global Assessment (PGA) [11], the Lattice System Physician's Global Assessment (LS-PGA) [11], the Self-Administered PASI (SAPASI) [12], and the Salford Psoriasis Index (SPI) [13]. It appeared that none of the severity scores used for psoriasis met all of the validation criteria required for an ideal score. However, it was concluded that the PASI score was the most extensively studied psoriasis clinical severity score and the most thoroughly validated. PASI was demonstrated to be reliable and reproducible. Its internal consistency [11], intraobserver reliability [11, 12, 14] and interobserver reliability [11, 13, 14] are good and its sensitivity to change is acceptable [12]. It has a good content validity. Evidence-based recommendations to assess psoriasis severity stated that the PASI can be used in everyday clinical practice in the management of adult patients with plaque-type psoriasis, in particular, if a systemic treatment is considered [15]. However, PASI has a number of drawbacks. Its construct validity and its acceptability have not been evaluated. It has poor sensitivity to change and responsiveness when skin psoriasis is less than 10% BSA involvement. The correlation with quality of life measures is poor [16]. Spuls et al. therefore suggested drawing on multiple measurement tools to fully

characterize disease severity and responsiveness to therapy [17]. The PASI and LS-PGA, for example, complement each other and provide a representative picture of disease severity [17].

## 5. Dactylitis

Dactylitis is characterized by swelling of a whole digit and represents a combination of synovitis and inflammation of tendon and ligament insertions. It is a hallmark feature of PsA, occurring in 16–48% of reported cases [18]. It can be further characterized as acute dactylitis where the digit is erythematous, swollen, hot, and tender to touch; or as chronic dactylitis where the digit is swollen but without redness and tenderness. There is a study that demonstrated that digits with dactylitis are associated with a greater degree of radiological damage than those which occur in digits not affected by dactylitis [19].

The Leeds Dactylitis Instrument (LDI) was developed in response to the need for a clinical, objective, validated outcome measure for dactylitis [18]. It measures the ratio of the circumference of the affected digit to the circumference of the digit on the contra-lateral hand or foot: a minimum difference of 10% is used to define a dactylitic digit. If the contralateral digit is also dactylitic, a table of normative values based on population averages is used to provide the comparison. The ratio of circumference is multiplied by a tenderness score, originally based on the Ritchie index (graded 0–3), but a later modification amended this to a binary score (0 for nontender, 1 for tender—this later modification is referred to as the LDI basic). The results from each digit with dactylitis are then summed to produce a final score. The aim of the LDI is to provide a quantification of both the size of the swollen digit and the tenderness so that the score can differentiate between tender and nontender dactylitis. Both the LDI and LDI basics have demonstrated good inter- and intraobserver reliability [18]. The LDI was tested in the International Multicenter Psoriasis and Psoriatic Arthritis Reliability Trial (IMPART) and showed good agreement among rheumatologists but not dermatologists [20].

## 6. Enthesitis

Enthesitis is defined as inflammation at the site of tendons, ligaments, or joint capsule fibre insertion into bone. It is a unique and important clinical feature of spondyloarthropathy. Most clinical studies have reported the frequency of enthesitis in PsA cohorts in the 30–50% range [21, 22], but this may be an underestimate, as now imaging studies such as ultrasonography or magnetic resonance imaging (MRI) have demonstrated enthesopathy not appreciated clinically.

In 1987, Mander et al. published the first instrument to investigate enthesitis in ankylosing spondylitis (AS), the Mander enthesitis index (MEI) [23]. 66 enthesal sites which were accessible to clinical examination were defined (Figure 1). These sites were to be examined for tenderness and the intensity of pain was graded on a 0 to 3 scale. The MEI is sensitive to change in clinical state associated with

nonsteroidal antirheumatic drug therapy, and one study demonstrated good intraobserver reliability [24]. On the other hand it was not possible to demonstrate any treatment group differences in a large placebo-controlled trial of infliximab in AS using the MEI, suggesting that discrimination was poor because of low interobserver reliability [25]. MEI is generally regarded as being nonfeasible in clinical use because it is time consuming and not all enthesitic sites are readily identifiable on physical examination. It has also been criticized for potentially causing distress to patients, and not adequately distinguishing enthesitic sites from fibromyalgia tender points.

The Maastricht Ankylosing Spondylitis Entheses Score (MASES) was developed to modify MEI to produce a less time consuming index with similar validity [26]. The grading of tenderness score from 0 to 3 in MEI was removed and substituted with a dichotomous 0/1 score for tenderness in MASES. The number of entheses index was reduced as concise as possible. After exclusion of entheses difficult to localize or near to other sites, the 66 entheses were reduced to 13 most specific and sensitive sites. These include the bilateral first and seventh costochondral joints, the anterior and posterior superior iliac spines, the iliac crests, the fifth lumbar spinous process, and the proximal insertion of Achilles tendon (Figure 1). MASES is a more feasible instrument but it has not been assessed in other diseases of the SpA including PsA. The fact that it does not score one of the main enthesitic sites (plantar fascia insertions into the calcaneum) also gives rise to some concern.

Gladman et al. have reported on the performance of investigators from the Canadian Spondyloarthropathy Group in their ability to reliably assess four enthesitis areas: rotator cuff insertion at the shoulder, tibial tuberosity at the knee, Achilles tendon, and plantar fascia insertions in the calcaneus [6]. The reliability was “fair” in the assessment of rotator cuff enthesitis, “moderate” for tibial tuberosity and Achilles enthesitis, and “moderate to substantial” for plantar enthesitis. This may have been due to difficulties in the anatomic localization of rotator cuff entheses.

In a study on infliximab in patients with AS, Braun et al. used an enthesitis index composed of 12 entheses that are reported to be commonly affected in the inflammatory process in AS (major enthesitis index) [27]. This index included the iliac crests, the great trochanters of the femur, the medial and lateral condyles of the humerus, the proximal insertion of the Achilles tendon, and insertion of the plantar fascia to the calcaneus. This index did not perform better compared to the MASES in the above-mentioned study. It has not been widely used or studied.

The Spondyloarthritis Research Consortium of Canada (SPARCC) created a new outcome measure for enthesitis in SpA using information from ultrasound and MRI studies [28]. The selection of entheses was based on two published findings from power Doppler ultrasound and MRI [29, 30]. Sixteen sites were selected: the bilateral greater trochanter, quadriceps tendon insertion into the patella, patellar ligament insertion into the patella and tibial tuberosity, Achilles tendon insertion, plantar fascia insertion, medial and lateral epicondyles, and the supraspinatus insertion (Figure 1).

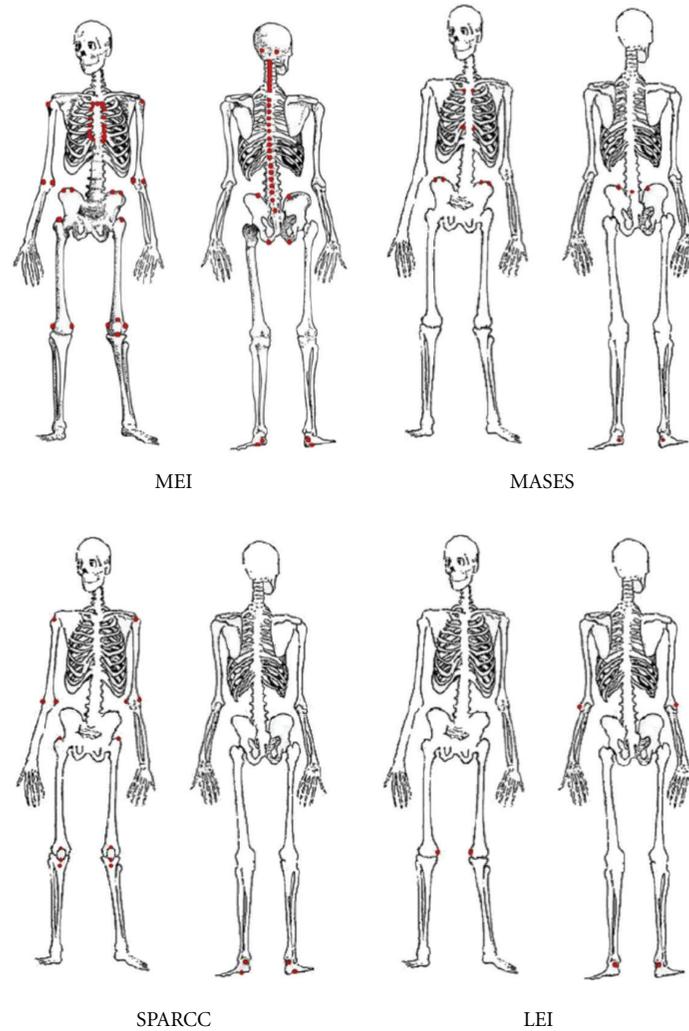


FIGURE 1: Enthesitis sites recorded by the principle enthesitis indices [32]. MEI: Mander Enthesitis Index; MASES: Maastricht Ankylosing Spondylitis Enthesitis Score; SPARCC: Spondyloarthritis Research Consortium of Canada Enthesitis Index; LEI: Leeds Enthesitis Index.

Tenderness at each site was quantified on a dichotomous basis: 0 means nontender and 1 means tender. Interobserver reliability was good and a substantial correlation was seen between the total enthesitis score and other disease activity measures [28].

The MEI, the MASES, the Gladman index, and the major enthesitis index were all developed and validated for patients with AS. The SPARCC score was developed using a full spectrum of SpA patients, but the validation was only done in patients with AS [28]. The Leeds Enthesitis Index (LEI) which was published in 2008 is the only measure developed specifically for PsA [31]. The LEI consisted of 6 sites: bilateral Achilles tendon insertions, medial femoral condyles, and lateral epicondyles of the humerus (Figure 1). Tenderness at each site was quantified on a dichotomous basis: 0 means nontender and 1 means tender. This index was compared with other enthesial indices including the MEI, MASES, the Gladman index and the major enthesitis index in an open-label longitudinal study. The LEI was able to distinguish between patients with active disease and those

without. It showed strong correlation with other disease activity measures, a large effect size, and the lowest floor effect. Floor effect means not picking up cases with low index score. A low floor effect means that it will be likely to detect the large majority of patients with active enthesitis.

The reproducibility of enthesitis assessment among patients with PsA with spinal involvement was investigated in the International Spondyloarthritis Interobserver Reliability Exercise—the INSPIRE study [7]. Enthesitis sites included in the MASES, SPARCC, LEI, and other enthesitis scoring systems were investigated. The results showed that there was excellent agreement among the observers with regard to the number of active enthesitis sites per individual patient. The individual indices provided substantial to excellent agreement for all patients. The LEI performs well in PsA.

## 7. Nail Assessment

Nail involvement is common in patients with psoriasis and PsA and can be severe and disfiguring. Nail psoriasis occurs

in as many as 50% of psoriatic patients [33] and has been reported from 63% to 83% in patients with PsA [34, 35].

Psoriatic nail disease can be broadly divided into psoriasis affecting the nail matrix and the nail bed. Involvement in the nail matrix results in changes to the nail plate. Characteristic features of psoriasis affecting the nail matrix include pitting, leukonychia, red spots in the lunula, and nail plate crumbling. Characteristic features of psoriasis affecting the nail bed include oil-drop discoloration, onycholysis, nail bed hyperkeratosis, and splinter hemorrhages.

In 2003, a group of dermatologist had developed the Nail Psoriasis Severity Index (NAPSI) as a psoriatic nail grading instrument [36]. This instrument is used to evaluate the severity of nail matrix psoriasis and nail bed psoriasis by area of involvement in the nail unit. The nail is divided with imaginary horizontal and longitudinal lines into four quadrants. Each quadrant of the nail is evaluated by presence of any of the nail matrix and nail bed psoriasis features. The score range from 0 to 4 according to the number of quadrants with any of the features present. Each nail gets a nail matrix score (0–4) and a nail bed score (0–4), and the total nail score is the sum of these two (0–8). The sum of the scores from all nails is 0–80 or 0–160 if toenails are included. If a target nail scale is desired, the same technique can be used to evaluate all eight parameters (pitting, leukonychia, red spots in lunula, crumbling, oil drop, onycholysis, hyperkeratosis, and splinter hemorrhages) in each quadrant of the nail, giving that one nail scores 0–32. An informal survey of the NAPSI score suggested that the NAPSI score was reproducible among dermatologists grading the same nails. However, the NAPSI scores was not validated.

In 2007, a group of rheumatologists with a goal to validate a psoriatic nail disease assessment instrument to assess disease severity and response to treatment in clinical trials had modified the original NAPSI in an attempt to enhance its reliability and face validity [37]. First, the division of the nail into quadrants was eliminated because quadrants were felt to be difficult to precisely quantify and varied among observers. Second, a more quantitative aspect was added to the scoring of several features in order to increase the sensitivity of the overall grading. Nail pitting was scored 0–3 depending on the number of pits present. Crumbling and onycholysis were scored 0–3 depending on the percentage of the nail involved. Splinter hemorrhages, leukonychia, red spots in the lunula, oil-drop dyschromia, and nail bed hyperkeratosis were individually scored as 1 if they were present and 0 if they were absent. Oil-drop dyschromia was regarded as part of the same pathologic process as onycholysis, and therefore oil-drop dyschromia and onycholysis were graded together. In the end, the range of possible scores using the modified NAPSI (mNAPSI) was 0–14 for each fingernail, or 0–140 for all 10 fingernails. In addition to the mNAPSI, global nail psoriasis severity ratings from both patients and physicians were added using a visual analogue scale (VAS) (0–10). The mNAPSI scores showed excellent internal consistency and interobserver reliability [37]. Its construct validity was shown by the correlation between mNAPSI scores and global nail severity VAS scores and by correlation between

the physician and patient global assessments of nail disease activity.

## 8. Spine Assessment

Spondylitis has been reported in 40–51% of PsA patients [38]. Sacroiliitis has been reported in up to a quarter of patients in several series [39–41]. Unlike AS, where axial involvement is present in all patients and tends to be more severe both clinically and radiologically, axial PsA is more heterogenous and less severe than that in AS. Up till now, there is no consensus on the definition of axial PsA. The assessment for spinal involvement has been borrowed from the Assessment of Ankylosing Spondylitis (ASAS) working group. The ASAS working group has recommended the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) to measure the disease activity [42], the Bath Ankylosing Spondylitis Function Index (BASFI) to evaluate the functional ability [43], and the Bath Ankylosing Spondylitis Metrology Index (BAMI) to assess spinal mobility [44]. The International Spondyloarthritis Interobserver Reliability Exercise (INSPIRE) study showed that the axial measures of spinal mobility used in AS perform well with respect to interobserver reliability and are equally reproducible when applied to PsA patients with axial involvement [45]. There are studies showing BASDAI correlated highly with patient perception of disease activity but there was no significant effect of the pattern of disease (axial or peripheral) [46, 47], suggesting that the BASDAI does not differentiate between axial and peripheral disease activity.

## 9. Patient Global Assessment

PGA of disease activity is important because it enhances the patient physician interaction to become more patient-centered by highlighting the global influences of PsA on the individual patient's well-being. The PGA is very dependent on the wording of the question poses to patient. In PsA, when the patients are being evaluated for PGA, they (and even the clinicians) may get confused whether they should relate the assessment to joint involvement, or skin involvement, or both. To address this issue, the GRAPPA organized a multi-center study to assess the reliability of the PGA, measured by means of 0–100 mm VAS, and the additional utility of separate VAS scales for joints (PJA) and skin (PSA) [48]. The specific question for PGA was "In all the ways in which your psoriasis and arthritis, as a whole, affect you, how would you rate the way you felt over the past week?" Results showed that PGA with a single question addressing both joint and skin disease is a reliable and responsive measure in assessing patient in totality. Because joint and skin disease often diverges, it is suggested that in some circumstances, such as study of a drug that improves the joints but not the skin, both PJA and PSA are also assessed.

## 10. Health-Related Quality of Life

The most commonly used measure of health-related quality of life include the Health Assessment Questionnaire (HAQ),

the Medical Outcomes Study Short Form 36 (SF-36), the Psoriatic Arthritis Quality of Life (PsAQoL), Dermatology Life Quality Index (DLQI), and the EuroQol 5-domain (EQ-5D).

The HAQ was originally developed for the assessment of disability in patients with RA [49]. It focuses on 2 dimensions of health status, physical disability (8 categories), and pain. The 8 categories, reviewing a total of 20 specific functions, evaluate patient's difficulty with activities of daily living over the past week. The 8 categories include dressing and grooming, arising, eating, walking, hygiene, reaching, gripping, and errands and chores. It also identifies specific aids or devices utilized for assistance, as well as help needed from another person. The HAQ has been modified for spondyloarthropathies (HAQ-S), which includes 2 spinal domains (SPAR1 and SPAR2) [50]. It has also been further modified for psoriasis (HAQ-SK) [51]. Both the HAQ-S and HAQ-SK scores were shown to perform similarly to the original HAQ score [52], suggesting that neither the spine or the skin related questions enhance the assessment of health status provided by the original HAQ.

The SF-36 is a generic health assessment questionnaire intended to measure general health concepts not specific to any age, disease, or treatment group [53]. It measures 8 health domains: physical functioning, pain, vitality, social functioning, psychological functioning, general health perceptions, and role limitations due to physical and emotional problems. It also can be subdivided into two summary scores, the physical component summary and the mental component summary scores. This instrument has been validated in PsA [54]. It was found to be reliable in patients with PsA and could be used to distinguish PsA patients from the general population.

PsAQoL is a 20-item, PsA-specific health-related QOL instrument. It has shown reliability and construct validity, but its use in clinical trial has not yet been published [55].

DLQI is a 10-item questionnaire developed as a measure of disability for a wide range of dermatological conditions [56]. It has been validated in assessment of psoriasis and shows discrimination and responsiveness in PsA trials [57–59].

The EQ-5D is comprised of a 5-dimension set of health status measures and a VAS [60]. The 5 dimensions are mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 3 levels: no problems, some problems, and extreme problems. The VAS records the respondent's self-rated health on a 20-centimeter vertical VAS where the endpoints are labelled "the best imaginable health state" and "the worst imaginable health status." The EQ-5D has shown discrimination and responsiveness in PsA trials [60].

Currently, there is no single generally agreed upon definition or conceptual model of health-related quality of life. The choice of the different measures of health-related quality of life in PsA depends on its content, respondent burden, administrative burden, translation and adaptations, acceptability, reliability, validity, and ability to detect change. PsAQoL is the only measure specific to PsA. It is now being

used for randomized controlled trial, from which more will be learned about its performance characteristics.

## 11. Fatigue

Fatigue in varying degrees is a frequent, often debilitating problem that significantly affects patients' lives. Fatigue can be constant and persistent, or fluctuating and unpredictable. Several scales have been used to assess fatigue in rheumatic diseases, including the Fatigue Severity Scale (FSS) [61], the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) [62], the Multidimensional Assessment of Fatigue (MAF) [63], the Multidimensional Fatigue Inventory (MFI) [64], the Profile of Fatigue (ProFF) [65], the Short Form 36 Vitality Subscale (SF-36 VT) [53] and the Visual Analog Scales for Fatigue [66]. A modified version of the FSS has been validated in PsA [67]. The FACIT-F was validated in a Toronto PsA cohort study. It correlated well with the modified FSS, showing high internal consistency, test-retest reliability, criterion, and construct validity [68].

Currently no single measure is favoured in use for PsA patients. The area of fatigue measure in PsA still needs to be further studied.

## 12. Composite Measures

A composite measure is a one way of assessing all relevant clinical outcomes in a single instrument. It incorporates several dimensions of disease status, often by combining these different domains into a single score. Given with the large number of domains that may be affected in a single patient with PsA, developing a comprehensive composite measure is a major challenge. The GRAPPA and the society OMERACT are working on the development of composite measures of disease severity and responses to therapy that take into account most of the disease domains.

Disease Activity for Psoriatic Arthritis (DAPSA) was adapted from the Disease Activity Index for Reactive Arthritis (DAREA) [69], a score developed and validated for reactive arthritis. DAPSA was developed from a clinical cohort [70] and validated using clinical trial data [71]. It comprises 68 tender and 66 swollen joints count, patient global, pain, and C-reactive protein level (Table 2). The composite score is a simple sum of the scores. Skin assessment was excluded because it did not reach statistical significance.

The Psoriatic Arthritis Joint Activity Index (PsAJAI) was developed from pooled data from randomized clinical trials of antitumor necrosis factor agents in PsA [72, 73]. A response is defined as a 30%-improvement in six measures with weights of 2 given to tender and swollen joint counts, C-reactive protein, and physician global assessment of disease activity (Table 2). Weights of 1 are given to pain, patient global assessment of disease activity, and the HAQ.

Composite Psoriatic Disease Activity Index (CPDAI) was a domain-based measure [74]. Disease involvement is assessed in up to 5 domains: peripheral arthritis, skin, enthesitis, dactylitis, and spinal manifestations (Table 2). For each

TABLE 2: Clinical domains included in composite measures in PsA.

	Peripheral arthritis	Pain	Patient global assessment	Physician global assessment	Skin	Enthesitis	Dactylitis	Spine disease	HAQ	CRP
DAPSA	Yes	Yes	Yes							Yes
PsAJAI	Yes	Yes	Yes	Yes					Yes	Yes
CPDAI	Yes		Yes		Yes	Yes	Yes	Yes	Yes	

DAPSA: Disease Activity for Psoriatic Arthritis; PsAJAI: Psoriatic Arthritis Joint Activity Index; CPDAI: Composite Psoriatic Disease Activity Index; HAQ: Health Assessment Questionnaire; CRP: C-reactive protein.

domain, instruments are used to assess both the extent of disease activity and the effect of involvement in that domain on patients' function and health-related quality of life. Domains are scored from 0 to 3, with empirical cutoffs for disease severity/activity proposed in each one, largely based on the literature. Individual domain scores are summed to give an overall composite score (range 0–15) [74].

The CPDAI has been validated in a large clinical trial dataset Psoriasis Randomized Etanercept Study in Subjects with Psoriatic Arthritis (PRESTA) [75]. In PRESTA, 752 patients were randomized to a double-blind, 2-period study that evaluated the safety and efficacy of 2 doses of etanercept on skin and musculoskeletal disease. Joint responses were equally determined by both the CPDAI and the DAPSA composite scores; but it was only the CPDAI, which also encompasses other domains including skin, enthesitis, and dactylitis, being able to discern the global treatment response between the 2 etanercept doses. This demonstrates that the CPDAI is a more sensitive instrument to detect changes in the different domains of disease activity in PsA.

### 13. GRACE Project

In an attempt to develop new composite outcome measures for PsA, the GRAppa Composite Exercise (GRACE) project has been developed following the GRAPPA annual meeting in 2008. This is a long-term project with longitudinal observational data which are being collected on a large cohort of PsA patients at multiple centres internationally. To date, baseline information on 503 patients with PsA has been collected [76, 77]. Recently 2 novel indices, Psoriatic Arthritis Disease Activity Score (PASDAS) and Arithmetic Mean of Desirability Functions (AMDF) were developed using multiple linear regression and physician-defined cutoffs for disease activity, respectively [76]. It is anticipated that further testing in other datasets including comparison with existing measures will be done to validate these new instruments.

### 14. Minimal Disease Activity

Minimal disease activity is a concept that has been defined by the OMERACT as a state of disease activity deemed a useful target of treatment by both the patient and physician. Coates et al. had developed minimal disease activity (MDA) criteria for PsA using data on 40 patients with the disease and the expert opinions of 60 rheumatologists and dermatologists [78]. The goal of the development of this instrument is to

TABLE 3: Minimal disease activity (MDA) criteria in psoriatic arthritis.

A patient is classified as in MDA when 5 of the following 7 criteria are met:

- Tender joint count  $\leq 1$
- Swollen joint count  $\leq 1$
- PASI  $\leq 1$  or BSA  $\leq 3$
- Patient pain VAS  $\leq 15$
- Patient global activity VAS  $\leq 20$
- HAQ  $\leq 0.5$
- Tender enthesial points  $\leq 1$

PASI: Psoriasis Area and Severity Index; BSA: Body Surface Area; VAS: Visual Analog Scale; HAQ: Health Assessment Questionnaire.

“treat to target,” to achieve disease remission or low disease activity state. For a patient to achieve MDA, 5 of the 7 criteria must be met (Table 3). The MDA criteria has been validated in an observational cohort and 2 infliximab studies of PsA [79, 80]. They showed patients with active PsA achieving MDA with effective therapy having a significant reduction in joint damage radiographic progression.

### 15. Conclusion

Given with the complexities of the disease nature of PsA, assessment of disease activity needs to take into account the core set of domains in order to assess their impact on the patient and the response to treatment. There are still considerable controversial issues about the content and performance of composite measures. Efforts by the OMERACT/GRAPPA working group are underway to determine if composite measures that capture both disease activity and response to therapy can be developed that effectively encompass all of the domains. Currently there is great heterogeneity in PsA assessment, even since publication of the OMERACT core set. Better consensus on instruments to assess each domain of disease activity in PsA is anticipated.

### References

- [1] D. D. Gladman, “Consensus on a core set of domains for psoriatic arthritis,” *The Journal of Rheumatology*, vol. 34, no. 5, pp. 1167–1170, 2007.
- [2] J. M. H. Moll and V. Wright, “Psoriatic arthritis,” *Seminars in Arthritis and Rheumatism*, vol. 3, no. 1, pp. 55–78, 1973.

- [3] Cooperating Clinics Committee of The American Rheumatism Association, "A seven-day variability study of 488 patients with peripheral rheumatoid arthritis," *Arthritis and Rheumatism*, vol. 8, pp. 302–334, 1965.
- [4] D. D. Gladman, P. J. Mease, P. Healy et al., "Outcome measures in PsA," *The Journal of Rheumatology*, vol. 34, no. 5, pp. 1159–1166, 2007.
- [5] D. D. Gladman, R. J. Cook, C. Schentag et al., "The clinical assessment of patients with psoriatic arthritis: results of a reliability study of the spondyloarthritis research consortium of Canada," *The Journal of Rheumatology*, vol. 31, no. 6, pp. 1126–1131, 2004.
- [6] D. D. Gladman, V. Farewell, D. Buskila et al., "Reliability of measurements of active and damaged joints in psoriatic arthritis," *The Journal of Rheumatology*, vol. 17, no. 1, pp. 62–64, 1990.
- [7] D. D. Gladman, R. D. Inman, R. J. Cook et al., "International spondyloarthritis interobserver reliability exercise—the INSPIRE study: II. Assessment of peripheral joints, enthesitis, and dactylitis," *The Journal of Rheumatology*, vol. 34, no. 8, pp. 1740–1745, 2007.
- [8] E. Puzenat, V. Bronsard, S. Prey et al., "What are the best outcome measures for assessing plaque psoriasis severity? A systematic review of the literature," *Journal of the European Academy of Dermatology and Venereology*, vol. 24, supplement 2, pp. 10–16, 2010.
- [9] B. Ramsay and C. M. Lawrence, "Measurement of involved surface area in patients with psoriasis," *British Journal of Dermatology*, vol. 124, no. 6, pp. 565–570, 1991.
- [10] T. Fredriksson and U. Pettersson, "Severe psoriasis—oral therapy with a new retinoid," *Dermatologica*, vol. 157, no. 4, pp. 238–244, 1978.
- [11] R. G. Langley and C. N. Ellis, "Evaluating psoriasis with psoriasis area and severity index, psoriasis global assessment, and lattice system physician's global assessment," *Journal of the American Academy of Dermatology*, vol. 51, no. 4, pp. 563–569, 2004.
- [12] S. R. Feldman, A. B. Fleischer Jr., D. M. Reboussin et al., "The self-administered psoriasis area and severity index is valid and reliable," *The Journal of Investigative Dermatology*, vol. 106, no. 1, pp. 183–186, 1996.
- [13] B. Kirby, D. G. Fortune, M. Bhushan, R. J. G. Chalmers, and C. E. M. Griffiths, "The salford psoriasis index: an holistic measure of psoriasis severity," *British Journal of Dermatology*, vol. 142, no. 4, pp. 728–732, 2000.
- [14] J. Berth-Jones, K. Grotzinger, C. Rainville et al., "A study examining inter- and intrarater reliability of three scales for measuring severity of psoriasis: psoriasis area and severity index, physician's global assessment and lattice system physician's global assessment," *British Journal of Dermatology*, vol. 155, no. 4, pp. 707–713, 2006.
- [15] C. Paul, P. A. Gourraud, V. Bronsard et al., "Evidence-based recommendations to assess psoriasis severity: systematic literature review and expert opinion of a panel of dermatologists," *Journal of the European Academy of Dermatology and Venereology*, vol. 24, supplement 2, pp. 2–9, 2010.
- [16] F. Sampogna, F. Sera, D. Abeni, and IDI Multipurpose Psoriasis Research on Vital Experiences (IMPROVE) Investigators, "Measures of clinical severity, quality of life, and psychological distress in patients with psoriasis: a cluster analysis," *The Journal of Investigative Dermatology*, vol. 122, no. 3, pp. 602–607, 2004.
- [17] P. I. Spuls, L. L. A. Lecluse, M. L. N. F. Poulsen, J. D. Bos, R. S. Stern, and T. Nijsten, "How good are clinical severity and outcome measures for psoriasis: quantitative evaluation in a systematic review," *The Journal of Investigative Dermatology*, vol. 130, no. 4, pp. 933–943, 2010.
- [18] P. S. Helliwell, J. Firth, G. H. Ibrahim, R. D. Melsom, I. Shah, and D. E. Turner, "Development of an assessment tool for dactylitis in patients with psoriatic arthritis," *The Journal of Rheumatology*, vol. 32, no. 9, pp. 1745–1750, 2005.
- [19] J. E. Brockbank, M. Stein, C. T. Schentag, and D. D. Gladman, "Dactylitis in psoriatic arthritis: a marker for disease severity?" *Annals of the Rheumatic Diseases*, vol. 64, no. 2, pp. 188–190, 2005.
- [20] V. Chandran, A. Gottlieb, R. J. Cook et al., "International multicenter psoriasis and psoriatic arthritis reliability trial for the assessment of skin, joints, nails, and dactylitis," *Arthritis Care and Research*, vol. 61, no. 9, pp. 1235–1242, 2009.
- [21] J. L. Fernández-Sueiro, A. Willisch, S. Pérttega-DÍAZ et al., "Evaluation of ankylosing spondylitis spinal mobility measurements in the assessment of spinal involvement in psoriatic arthritis," *Arthritis Care and Research*, vol. 61, no. 3, pp. 386–392, 2009.
- [22] D. D. Gladman and V. Chandran, "Observational cohort studies: lessons learnt from the University of Toronto psoriatic arthritis program," *Rheumatology*, vol. 50, no. 1, pp. 25–31, 2011.
- [23] M. Mander, J. M. Simpson, A. McLellan, D. Walker, J. A. Goodacre, and W. C. Dick, "Studies with an enthesitis index as a method of clinical assessment in ankylosing spondylitis," *Annals of the Rheumatic Diseases*, vol. 46, no. 3, pp. 197–202, 1987.
- [24] G. R. Auleley, K. Benbouazza, A. Spoorenberg et al., "Evaluation of the smallest detectable difference in outcome or process variables in ankylosing spondylitis," *Arthritis Care and Research*, vol. 47, no. 6, pp. 582–587, 2002.
- [25] D. van der Heijde, B. Dijkmans, P. Geusens et al., "Efficacy and safety of infliximab in patients with ankylosing spondylitis: results of a randomized, placebo-controlled trial (ASSERT)," *Arthritis and Rheumatism*, vol. 52, no. 2, pp. 582–591, 2005.
- [26] L. Heuft-Dorenbosch, A. Spoorenberg, A. van Tubergen et al., "Assessment of enthesitis in ankylosing spondylitis," *Annals of the Rheumatic Diseases*, vol. 62, no. 2, pp. 127–132, 2003.
- [27] J. Braun, J. Brandt, J. Listing et al., "Treatment of active ankylosing spondylitis with infliximab: a randomised controlled multicentre trial," *The Lancet*, vol. 359, no. 9313, pp. 1187–1193, 2002.
- [28] W. P. Maksymowych, C. Mallon, S. Morrow et al., "Development and validation of the spondyloarthritis research consortium of Canada (SPARCC) enthesitis index," *Annals of the Rheumatic Diseases*, vol. 68, no. 6, pp. 948–953, 2009.
- [29] M. A. D'Agostino, R. Said-Nahal, C. Hacquard-Bouder, J. L. Bresseur, M. Dougados, and M. Breban, "Assessment of peripheral enthesitis in the spondylarthropathies by ultrasonography combined with power Doppler: a cross-sectional study," *Arthritis and Rheumatism*, vol. 48, no. 2, pp. 523–533, 2003.
- [30] R. G. W. Lambert, S. S. Dhillon, G. S. Jhangri et al., "High prevalence of symptomatic enthesopathy of the shoulder in ankylosing spondylitis: deltoid origin involvement constitutes a hallmark of disease," *Arthritis Care and Research*, vol. 51, no. 5, pp. 681–690, 2004.
- [31] P. J. Healy and P. S. Helliwell, "Measuring clinical enthesitis in psoriatic arthritis: assessment of existing measures and development of an instrument specific to psoriatic arthritis," *Arthritis Care and Research*, vol. 59, no. 5, pp. 686–691, 2008.

- [32] L. C. Coates and P. S. Helliwell, "Disease measurement—enthesitis, skin, nails, spine and dactylitis," *Best Practice and Research Clinical Rheumatology*, vol. 24, no. 5, pp. 659–670, 2010.
- [33] E. M. Farber and L. Nall, "Nail psoriasis," *Cutis*, vol. 50, no. 3, pp. 174–178, 1992.
- [34] O. Elkayam, J. Ophir, M. Yaron, and D. Caspi, "Psoriatic arthritis: interrelationships between skin and joint manifestations related to onset, course and distribution," *Clinical Rheumatology*, vol. 19, no. 4, pp. 301–305, 2000.
- [35] L. Williamson, N. Dalbeth, J. L. Dockerty, B. C. Gee, R. Weatherall, and B. P. Wordsworth, "Extended report: nail disease in psoriatic arthritis—clinically important, potentially treatable and often overlooked," *Rheumatology*, vol. 43, no. 6, pp. 790–794, 2004.
- [36] P. Rich and R. K. Scher, "Nail psoriasis severity index: a useful tool for evaluation of nail psoriasis," *Journal of the American Academy of Dermatology*, vol. 49, no. 2, pp. 206–212, 2003.
- [37] S. E. Cassell, J. D. Bieber, P. Rich et al., "The modified nail psoriasis severity index: validation of an instrument to assess psoriatic nail involvement in patients with psoriatic arthritis," *The Journal of Rheumatology*, vol. 34, no. 1, pp. 123–129, 2007.
- [38] D. D. Gladman, B. Brubacher, D. Buskila, P. Langevitz, and V. T. Farewell, "Psoriatic spondyloarthritis in men and women: a clinical, radiographic, and HLA study," *Clinical and Investigative Medicine*, vol. 15, no. 4, pp. 371–375, 1992.
- [39] D. D. Gladman, R. Shuckett, and M. L. Russell, "Psoriatic arthritis (PSA)—an analysis of 220 patients," *Quarterly Journal of Medicine*, vol. 62, no. 238, pp. 127–141, 1987.
- [40] J. C. T. Alonso, A. R. Perez, J. M. A. Castrillo, J. B. Garcia, J. L. R. Noriega, and C. L. Larrea, "Psoriatic arthritis (PA): a clinical, immunological and radiological study of 180 patients," *British Journal of Rheumatology*, vol. 30, no. 4, pp. 245–250, 1991.
- [41] D. Veale, S. Rogers, and O. FitzGerald, "Classification of clinical subsets in psoriatic arthritis," *British Journal of Rheumatology*, vol. 33, no. 2, pp. 133–138, 1994.
- [42] S. Garrett, T. Jenkinson, L. G. Kennedy, H. Whitelock, P. Gaisford, and A. Calin, "A new approach to defining disease status in ankylosing spondylitis: the bath ankylosing spondylitis disease activity index," *The Journal of Rheumatology*, vol. 21, no. 12, pp. 2286–2291, 1994.
- [43] A. Calin, S. Garrett, H. Whitelock, J. O'Hea, P. Mallorie, and T. Jenkinson, "A new approach to defining functional ability in ankylosing spondylitis: the development of the bath ankylosing spondylitis functional index," *The Journal of Rheumatology*, vol. 21, no. 12, pp. 2281–2285, 1994.
- [44] T. R. Jenkinson, P. A. Mallorie, H. C. Whitelock, L. G. Kennedy, S. L. Garrett, and A. Calin, "Defining spinal mobility in ankylosing spondylitis (AS). The bath AS metrology index," *The Journal of Rheumatology*, vol. 21, no. 9, pp. 1694–1698, 1994.
- [45] D. D. Gladman, R. D. Inman, R. J. Cook et al., "International spondyloarthritis interobserver reliability exercise—the INSPIRE study: I. Assessment of spinal measures," *The Journal of Rheumatology*, vol. 34, no. 8, pp. 1733–1739, 2007.
- [46] W. J. Taylor and A. A. Harrison, "Could the bath ankylosing spondylitis disease activity index (BASDAI) be a valid measure of disease activity in patients with psoriatic arthritis?" *Arthritis Care and Research*, vol. 51, no. 3, pp. 311–315, 2004.
- [47] J. L. Fernández-Sueiro, A. Willisich, S. Pértega-Díaz et al., "Validity of the bath ankylosing spondylitis disease activity index for the evaluation of disease activity in axial psoriatic arthritis," *Arthritis Care and Research*, vol. 62, no. 1, pp. 78–85, 2010.
- [48] A. Cauli, D. D. Gladman, A. Mathieu et al., "Patient global assessment in psoriatic arthritis: a multicenter GRAPPA and OMERACT study," *The Journal of Rheumatology*, vol. 38, no. 5, pp. 898–903, 2011.
- [49] J. F. Fries, P. Spitz, R. G. Kraines, and H. R. Holman, "Measurement of patient outcome in arthritis," *Arthritis and Rheumatism*, vol. 23, no. 2, pp. 137–145, 1980.
- [50] L. H. Daltroy, M. G. Larson, W. N. Roberts, and M. H. Liang, "A modification of the health assessment questionnaire for the spondyloarthritis," *The Journal of Rheumatology*, vol. 17, no. 7, pp. 946–950, 1990.
- [51] J. A. Husted, D. D. Gladman, J. A. Long, and V. T. Farewell, "A modified version of the health assessment questionnaire (HAQ) for psoriatic arthritis," *Clinical and Experimental Rheumatology*, vol. 13, no. 4, pp. 439–443, 1995.
- [52] M. G. Blackmore, D. D. Gladman, J. Husted, J. A. Long, and V. T. Farewell, "Measuring health status in psoriatic arthritis: the health assessment questionnaire and its modification," *The Journal of Rheumatology*, vol. 22, no. 5, pp. 886–893, 1995.
- [53] J. E. Ware Jr. and C. D. Sherbourne, "The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection," *Medical Care*, vol. 30, no. 6, pp. 473–483, 1992.
- [54] J. A. Husted, D. D. Gladman, J. A. Long, V. T. Farewell, and R. J. Cook, "Validating the SF-36 health survey questionnaire in patients with psoriatic arthritis," *The Journal of Rheumatology*, vol. 24, no. 3, pp. 511–517, 1997.
- [55] S. P. McKenna, L. C. Doward, D. Whalley, A. Tennant, P. Emery, and D. J. Veale, "Development of the PsAQoL: a quality of life instrument specific to psoriatic arthritis," *Annals of the Rheumatic Diseases*, vol. 63, no. 2, pp. 162–169, 2004.
- [56] A. Y. Finlay and G. K. Khan, "Dermatology life quality index (DLQI)—a simple practical measure for routine clinical use," *Clinical and Experimental Dermatology*, vol. 19, no. 3, pp. 210–216, 1994.
- [57] P. J. Mease, "Psoriatic arthritis: update on pathophysiology, assessment and management," *Annals of the Rheumatic Diseases*, vol. 70, no. 1, pp. i77–i84, 2011.
- [58] P. J. Mease and C. E. Antoni, "Psoriatic arthritis treatment: biological response modifiers," *Annals of the Rheumatic Diseases*, vol. 64, supplement 2, pp. ii78–ii82, 2005.
- [59] P. J. Mease, "Psoriatic arthritis: pharmacotherapy update," *Current Rheumatology Reports*, vol. 12, no. 2, pp. 272–280, 2010.
- [60] The EuroQol Group, "EuroQol—a new facility for the measurement of health-related quality of life," *Health Policy*, vol. 16, no. 3, pp. 199–208, 1990.
- [61] L. B. Krupp, N. G. LaRocca, J. Muir-Nash, and A. D. Steinberg, "The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus," *Archives of Neurology*, vol. 46, no. 10, pp. 1121–1123, 1989.
- [62] D. Cella, J. S. Lai, C. H. Chang, A. Peterman, and M. Slavin, "Fatigue in cancer patients compared with fatigue in the general United States population," *Cancer*, vol. 94, no. 2, pp. 528–538, 2002.
- [63] B. L. Belza, C. J. Henke, E. H. Yelin, W. V. Epstein, and C. L. Gilliss, "Correlates of fatigue in older adults with rheumatoid arthritis," *Nursing Research*, vol. 42, no. 2, pp. 93–99, 1993.
- [64] E. M. A. Smets, B. Garssen, B. Bonke, and J. C. J. M. de Haes, "The multidimensional fatigue inventory (MFI) psychometric qualities of an instrument to assess fatigue," *Journal of Psychosomatic Research*, vol. 39, no. 3, pp. 315–325, 1995.
- [65] S. J. Bowman, D. A. Booth, R. G. Platts, and The UK Sjogren's Interest Group, "Measurement of fatigue and discomfort in

- primary Sjogren's syndrome using a new questionnaire tool," *Rheumatology*, vol. 43, no. 6, pp. 758–764, 2004.
- [66] F. Wolfe, D. J. Hawley, and K. Wilson, "The prevalence and meaning of fatigue in rheumatic disease," *The Journal of Rheumatology*, vol. 23, no. 8, pp. 1407–1417, 1996.
- [67] C. T. Schentag, J. Cichon, A. MacKinnon, D. D. Gladman, and M. B. Urowitz, "Validation and normative data for the 0-10 point scale version of the fatigue severity scale (FSS)," *Arthritis and Rheumatism*, vol. 43, article S177, 2000.
- [68] V. Chandran, S. Bhella, C. Schentag, and D. D. Gladman, "Functional assessment of chronic illness therapy-fatigue scale is valid in patients with psoriatic arthritis," *Annals of the Rheumatic Diseases*, vol. 66, no. 7, pp. 936–939, 2007.
- [69] G. Eberl, A. Studnicka-Benke, H. Hitzelhammer, F. Gschnait, and J. S. Smolen, "Development of a disease activity index for the assessment of reactive arthritis (DAREA)," *Rheumatology*, vol. 39, no. 2, pp. 148–155, 2000.
- [70] V. P. Nell-Duxneuner, T. A. Stamm, K. P. Machold, S. Pflugbeil, D. Aletaha, and J. S. Smolen, "Evaluation of the appropriateness of composite disease activity measures for assessment of psoriatic arthritis," *Annals of the Rheumatic Diseases*, vol. 69, no. 3, pp. 546–549, 2010.
- [71] M. Schoels, D. Aletaha, J. Funovits, A. Kavanaugh, D. Baker, and J. S. Smolen, "Application of the DAREA/DAPSA score for assessment of disease activity in psoriatic arthritis," *Annals of the Rheumatic Diseases*, vol. 69, no. 8, pp. 1441–1447, 2010.
- [72] D. D. Gladman, B. D. M. Tom, P. J. Mease, and V. T. Farewell, "Informing response criteria for psoriatic arthritis. I: discrimination models based on data from 3 anti-tumor necrosis factor randomized studies," *The Journal of Rheumatology*, vol. 37, no. 9, pp. 1892–1897, 2010.
- [73] D. D. Gladman, B. D. M. Tom, P. J. Mease, and V. T. Farewell, "Informing response criteria for psoriatic arthritis (PsA). II: further considerations and a proposal—the PsA joint activity index," *The Journal of Rheumatology*, vol. 37, no. 12, pp. 2559–2565, 2010.
- [74] A. Mumtaz, P. Gallagher, B. Kirby et al., "Development of a preliminary composite disease activity index in psoriatic arthritis," *Annals of the Rheumatic Diseases*, vol. 70, no. 2, pp. 272–277, 2011.
- [75] O. FitzGerald, P. Helliwell, A. Mumtaz, L. Coates, R. Pedersen, and C. Molta, "Application of composite disease activity scores in psoriatic arthritis to the PRESTA dataset," *Annals of the Rheumatic Diseases*, vol. 71, no. 3, pp. 358–362, 2012.
- [76] P. S. Helliwell, O. FitzGerald, J. Fransen et al., "The development of candidate composite disease activity and responder indices for psoriatic arthritis (GRACE project)," *Annals of the Rheumatic Diseases*. In press.
- [77] P. S. Helliwell, O. FitzGerald, P. J. Mease, and D. D. Gladman, "GRAPPA responder index project (GRACE): a report from the GRAPPA, 2011 Annual Meeting," *The Journal of Rheumatology*, vol. 39, no. 11, pp. 2196–2197, 2012.
- [78] L. C. Coates, J. Fransen, and P. S. Helliwell, "Defining minimal disease activity in psoriatic arthritis: a proposed objective target for treatment," *Annals of the Rheumatic Diseases*, vol. 69, no. 1, pp. 48–53, 2010.
- [79] L. C. Coates, R. Cook, K. A. Lee, V. Chandran, and D. D. Gladman, "Frequency, predictors, and prognosis of sustained minimal disease activity in an observational psoriatic arthritis cohort," *Arthritis Care and Research*, vol. 62, no. 7, pp. 970–976, 2010.
- [80] L. C. Coates and P. S. Helliwell, "Validation of minimal disease activity criteria for psoriatic arthritis using interventional trial data," *Arthritis Care and Research*, vol. 62, no. 7, pp. 965–969, 2010.