ABSTRACT. Objective. Patient-reported outcomes (PRO) in rheumatoid arthritis (RA) provide important information regarding disease effect. The study objective was to assess the frequency of PRO use in recent RA studies and compare results with a previous systematic review (SR) in 2005–2007.

Methods. An SR was performed in PubMed MEDLINE (January 2015). Publications were identified using these MEDical Subject Headings terms: “arthritis, rheumatoid” with a limitation to “humans,” “all adults: 19+ years,” “English,” “published in the last 2 years,” and “clinical trials.” All studies were assessed, whatever their designs. All PRO reported in publications were classified according to general domains of health by 2 authors. Statistics were descriptive.

Results. Two hundred fifty articles were analyzed. Of them, 113 (45.2%) were randomized controlled trials; 138 different PRO were reported. The most frequent PRO, similar to the 2007 SR, were function (68.0%), pain (40.0%), patient’s global assessment (49.2%), and health-related quality of life (18.4%). Fatigue (14.4%), morning stiffness (10.0%), psychological status (9.6%), productivity losses (6.4%), utility (5.2%), sleep disturbance (2.4%), and coping (2.0%) were rarely reported. Although frequent domains were reported using well-validated questionnaires, the others were reported using heterogeneous questionnaires.

Conclusion. The PRO collected and reported in RA studies are remarkably consistent with those seen in 2005–2007, and reflect the existing RA Core Set measures. Other domains of health prioritized by patients including fatigue, psychological status, productivity losses, sleep disturbance, and coping remain rarely reported. Further, heterogeneity in outcome measures used presents challenges in interpreting true disease effect and response to therapy. (First Release April 15 2016; J Rheumatol 2016;43:1300–5; doi:10.3899/jrheum.151177)

Key Indexing Terms:
- Rheumatoid Arthritis
- Patient-reported
- Systematic Review
- Outcomes Measures
only domains frequently assessed in RCT or other studies were physical function (83.4%), patient’s global assessment (PtGA; 63.3%), and pain (55.9%). Other domains assessed far less frequently were morning stiffness (26.6%), health-related quality of life (HRQOL: 19.2%), utility (16.5%), fatigue (13.7%), self-reported painful joint count (9.1%), psychological status (7.3%), coping (6.4%), productivity losses (5.5%), well-being (3.6%), sleep disturbance (1.8%), and leisure (0.9%). Over the last decade, there have been substantial advances toward awareness of importance of PRO in rheumatology. It is unknown whether these advances have translated into their increased incorporation or reporting in RCT. The objective of our study was to assess PRO reported in RA studies published in the last 2 years and to compare the frequency of questionnaires and domains reported with our prior SR of PRO in RA studies.

MATERIALS AND METHODS
To obtain all recently published articles reporting any type of PRO in RA, we conducted an SR using the PubMed MEDLINE database on January 1, 2015. To make a more direct comparison with the results from our prior SR, we did not include EMBASE or other databases for our search. Publications were identified through a search that used the following exploded MEDical Subject Headings terms: “arthritis, rheumatoid” with limitations to “humans,” “all adults: 19+ years,” “English,” “published in the last 2 years,” and “clinical trials.” Publications were limited to articles referenced in PubMed in the prior 2 years (January 1, 2013, to December 31, 2014) to obtain an overview of the status of recent research.

Inclusion criteria were articles reporting any type of clinical study that included patients with RA and reported PRO results. Articles were excluded if they did not concern RA or if they did not report any patient-based outcome measures (e.g., articles reporting only laboratory outcomes, radiographic scores, or genetic information). Reviews, editorials, and letters were excluded because we were interested in obtaining information from primary original research articles. If there was more than 1 publication related to a single RCT (e.g., extension of main clinical study, pooled, or subanalyses), we selected the study that included the most PRO.

The initial selection process by 1 author (AE) was based on titles and abstracts of the articles, followed by full text review. Data were extracted from the full-text articles. Publications were assessed using a checklist of items developed by the 2 reviewers, LK and UK. Reviewers were not blinded to the journal name or authors. Data were obtained on year of publication, study design (RCT or other studies), and number of patients. Demographic data such as percentage of women, mean age, mean disease duration, treatments under evaluation, and maximum duration of followup were recorded for each report. Patient-reported outcomes. All PRO measures were noted. Outcome measures that were not patient-reported, such as biological results (e.g., CRP, rheumatoid factor, anticyclic citrullinated peptide antibodies) or radiographs, were not assessed. If available, composite indices such as the Disease Activity Score (DAS)13, American College of Rheumatology (ACR) response criteria14, European League Against Rheumatism (EULAR) response criteria14, Simplified Disease Activity Index (SDAI)15, or Clinical Disease Activity Index (CDAI)16 were noted. These composite indices included PRO from the “core set” (e.g., PtGA for all) and domains that were not patient-reported. However, if their results were only presented as global results (e.g., ACR20 without reporting the constituent core set domains), the PRO included in the ACR criteria were not considered as reported.

Domains of health. PRO were classified by the authors according to a domain framework used previously. Questionnaires were divided into 2 parts. “Most frequently” reported questionnaires for each domain were defined as “major questionnaires.” If the frequency of the report of questionnaire was more than 5% of articles, it was defined as a “secondary questionnaire.”

RESULTS
Publications: selection process and description. Of the 479 publications identified by the literature search, 250 (52.2%) were included in our analysis. The majority of the 229 excluded publications either did not have any focus or reporting of PRO (n = 110) or were not about the selected disease (n = 74; Figure 1). Of the 250 publications remaining, 113 (45.2%) were RCT and 137 (54.8%) were other types of studies (open-label trial, prospective cohort, retrospective study, and other study designs).

The total number of patients included in the reports was 143,670, and the mean (SD) number of patients per article was 579 (1365). Mean age was 54.8 years (4.2), mean disease duration was 7.8 years (4.6), and 75.2% were women. The most commonly assessed treatments were disease-modifying antirheumatic drugs, either biological or conventional synthetic (n = 140, 63.2%).

Composite indices and response criteria reported across studies were the DAS and/or EULAR (n = 205, 82.0%), ACR (n = 83, 33.2%), SDAI (n = 40, 16.0%), and CDAI (n = 37, 14.8%), whereas none of these were reported in 38 studies (15.2%). Composite indices were more frequently used in RCT (90.3% in RCT vs 80.3% in other studies, p = 0.029), especially the ACR response (59.3% vs 11.7%, p < 0.0001) and SDAI score (21.2% vs 11.7%, p = 0.04).

PRO: overview. Across the 250 articles, 138 PRO measures were reported. The mean numbers of PRO per article was 2.7 (2.5), and the mean number of questionnaires used across the articles to report a specific domain was 9.9 (8.0). The distribution of all domains and measures are presented in Supplementary Table 1 (available from the authors on request). The 138 PRO instruments were spread across 14 domains of health, i.e., function, PtGA, pain, morning stiffness, HRQOL, utility, fatigue, self-reported painful joint count, psychological status, coping, productivity losses, well-being, sleep disturbance, and leisure. Physical function/disability was reported in 68.0% of studies, the vast majority using the Health Assessment Questionnaire (HAQ; 89.4%). PtGA was reported in 49.2%, mostly using visual analog scales (VAS) or numeric rating scales (NRS; 83.7%). Pain was reported in 40.0% of studies, predominantly also using VAS or NRS (89.0%; Table 1).

Generic HRQOL was evaluated in 46 studies (18.4%), most frequently using the Medical Outcomes Study Short Form-36 (SF-36; n = 29, 63.0%)17. HRQOL was more frequently reported in RCT than non-RCT studies (27.4% vs...
Fatigue was reported in 36 articles (14.4%). Fatigue VAS/NRS (n = 18, 50%) and the functional assessment of chronic illness therapy (FACIT)-fatigue (n = 12, 33.3%) were the main instruments used. Fatigue was more often reported in RCT than non-RCT studies (21.2% vs 8.8%, p = 0.005).

Stiffness was evaluated in 25 studies (10.0%), which was less frequently than in our previous SR (10.0% vs 26.6%) and was mostly assessed through “morning stiffness duration.”

Psychological status was reported in 24 articles (9.6%). The Hospital Anxiety and Depression Scale (n = 6, 25.0% of 24) and the Beck Depression Inventory (n = 6, 25.0% of 24) were the most frequently used questionnaires, but 23 different questionnaires were used to assess psychological status (Supplementary Table 1, available from the authors on request).

Assessments of productivity were reported in 16 studies (6.4%), primarily through employment status (n = 8, 50.0% of 16) and the Work Productivity and Activity Impairment questionnaire (n = 4, 25.0% of 16).

Utility was infrequently reported (n = 13, 5.2%) and mainly using the EQ-5D (n = 11, 84.6% of 13). Utility was reported less frequently than in our previous SR (5.2% vs 16.5%). The Short-Form Health Survey-6D was reported in only 1 study in our current analysis (7.7% of 13). Other domains such as sleep disturbance (2.4%), coping (2.0%), and leisure (0.4%) were infrequently reported.

DISCUSSION
In our present SR, the well-recognized RA core domains (function, pain, and PtGA)² were reported in a majority of RA studies, but other less-recognized domains of health were rarely reported (e.g., fatigue, sleep, productivity, HRQOL, coping, morning stiffness, and utility). Our current SR of RA...
studies from the last 2 years shows that a large gap remains between the reporting of outcomes prioritized by physicians/researchers and the reporting of the range of areas of health that have been prioritized by patients with RA. Even though there has been an increasing call for the inclusion of PRO in RA research, there has been little change in their reporting in recent studies compared with our first SR 9 years ago of studies published between 2005 and 2007, though of note, we have not analyzed publications between 2007 and 2013.

Compared with the previous SR, function, pain, and PtGA were also reported in the majority of RA studies in the 2005–2007 review. This is in keeping with the recognition of these outcomes as core outcomes. On the other hand, other outcomes including the HRQOL and fatigue were not more frequently reported in the present SR than in the previous one. This is surprising given the importance of these outcomes for the patients and their recognition by the Outcome Measures in Rheumatology (OMERACT).

In total, 138 different PRO measures were used across 250 studies. Instruments to measure the existing RA Core Set PRO domains (function, pain, and PtGA) are well established and relatively homogeneous. For example, the HAQ-Disability Index and less frequently the modified HAQ are used almost exclusively to assess physical function in patients with RA. In contrast, for domains such as fatigue, productivity, psychological status, sleeping, or coping, there is high heterogeneity among the instruments used for their measurement in the studies we reviewed, reflecting a lack of consensus on what needs to be measured in RA and the lack of optimal, appropriately validated instruments for most of these domains.

In total, 138 different PRO measures were used across 250 studies. Instruments to measure the existing RA Core Set PRO domains (function, pain, and PtGA) are well established and relatively homogeneous. For example, the HAQ-Disability Index and less frequently the modified HAQ are used almost exclusively to assess physical function in patients with RA. In contrast, for domains such as fatigue, productivity, psychological status, sleeping, or coping, there is high heterogeneity among the instruments used for their measurement in the studies we reviewed, reflecting a lack of consensus on what needs to be measured in RA and the lack of optimal, appropriately validated instruments for most of these domains.

### Table 1. Frequency of reporting of domains of health and questionnaires in the SR of 250 studies published from 2013–2014 in RA and for comparison purposes, similar results from an SR of 109 studies.

<table>
<thead>
<tr>
<th>Domain of Health</th>
<th>Questionnaires used for Domains</th>
<th>Frequency of Reporting in 250 Studies</th>
<th>Frequency of Reporting in a Previous SR of 109 Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
<td>HAQ</td>
<td>170 (68.0)</td>
<td>91 (83.4)</td>
</tr>
<tr>
<td></td>
<td>Modified HAQ</td>
<td>152 (89.4)</td>
<td>67 (73.6)</td>
</tr>
<tr>
<td></td>
<td>PtGA VAS/NRS</td>
<td>13 (7.7)</td>
<td>14 (15.4)</td>
</tr>
<tr>
<td></td>
<td>General health VAS/NRS*</td>
<td>123 (49.2)</td>
<td>69 (63.3)</td>
</tr>
<tr>
<td></td>
<td>Pain VAS/NRS</td>
<td>103 (83.7)</td>
<td>53 (76.8)</td>
</tr>
<tr>
<td></td>
<td>Morning stiffness</td>
<td>23 (18.7)</td>
<td>10 (14.5)</td>
</tr>
<tr>
<td></td>
<td>Duration</td>
<td>100 (40.0)</td>
<td>61 (55.9)</td>
</tr>
<tr>
<td></td>
<td>HRQOL</td>
<td>89 (89.0)</td>
<td>52 (85.3)</td>
</tr>
<tr>
<td></td>
<td>Fatigue</td>
<td>25 (10.0)</td>
<td>29 (26.6)</td>
</tr>
<tr>
<td></td>
<td>Fatigue VAS/NRS</td>
<td>46 (18.4)</td>
<td>21 (19.2)</td>
</tr>
<tr>
<td></td>
<td>General health VAS/NRS</td>
<td>29 (63.0)</td>
<td>16 (76.2)</td>
</tr>
<tr>
<td></td>
<td>Fatigue VAS/NRS</td>
<td>36 (14.4)</td>
<td>15 (13.7)</td>
</tr>
<tr>
<td></td>
<td>Self-reported painful joint count</td>
<td>13 (5.2)</td>
<td>18 (16.5)</td>
</tr>
<tr>
<td></td>
<td>Psychological status</td>
<td>11 (84.6)</td>
<td>11 (61.1)</td>
</tr>
<tr>
<td></td>
<td>Productivity losses</td>
<td>7 (2.8)</td>
<td>10 (9.1)</td>
</tr>
<tr>
<td></td>
<td>Utility VAS/NRS</td>
<td>24 (9.6)</td>
<td>8 (7.3)</td>
</tr>
<tr>
<td></td>
<td>EQ-5D</td>
<td>16 (6.4)</td>
<td>6 (5.5)</td>
</tr>
<tr>
<td></td>
<td>Sleep disturbance</td>
<td>5 (2.0)</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td></td>
<td>Coping</td>
<td>1 (0.4)</td>
<td>1 (0.9)</td>
</tr>
</tbody>
</table>
| * For the domains function and PtGA, there were secondary outcome measures used in more than 5% of cases. ** The n (%) given are n (%) of all studies for each domain, and n (%) of studies reporting this domain, for each questionnaire. SR: systematic review; RA: rheumatoid arthritis; HAQ: Health Assessment Questionnaire; PtGA: patient’s global assessment; VAS: visual analog scale; NRS: numeric rating scale; HRQOL: health-related quality of life; SF-36: Medical Outcomes Study Short Form-36.

---

*Personal non-commercial use only. The Journal of Rheumatology Copyright © 2016. All rights reserved.*
Data and the RA Impact of Disease score, and are being included more frequently in assessments. With evolving guidelines for PRO development and validation using advanced psychometric methods, other measurement symptoms such as the PRO Measurement Information System have potential for use in future investigations.

There were certain limitations to our SR. First, we only used PubMed as the search source, thus we may have missed abstract data (for instance, in EMBASE) or in other sources such as the US Food and Drug Administration or the European Medicines Agency reports and submissions. Second, our literature review spanned only 2 years and did not include extraction of all variables as has been recommended by groups such as the COSMIN (Consensus-based Standards for the selection of health status Measurement Instruments). However, it is important to recognize that the purpose of our review was to compare PRO collected during an earlier interval with a more recent time frame, which would be most appropriate using the same search variables. Moreover, it was not our intent to evaluate the degree of validation of individual PRO measures, but rather to report frequencies of reporting of different domains. Finally, it is important to recognize that although there has been increased recognition of including the patient perspective in RA research, there may be a considerable time lag between the design of a clinical trial, its completion, and the reporting of results. It would be interesting to compare in future studies the proportion of outcomes in studies that are laboratory clinician-reported versus patient-reported.

The existing RA Core Set PRO domains (function, pain, PtGA) are still a dominant part of RA studies and their measurement is generally reported using consistent instruments. Other domains that have been prioritized by patients including fatigue, productivity, sleep disturbance, and coping are infrequently reported, and when they are, there is tremendous heterogeneity in the specific instruments selected. Better understanding of the barriers to more comprehensive and consistent PRO collection requires additional studies, and further work by groups such as the OMERACT is needed to develop consensus on what domains should be collected and how to best measure them.

REFERENCES

20. Castrejón J, Gossec L, Carmona L. The EULAR Outcome Measures


22. Hewlett S, Dures E, Almeida C. Measures of fatigue: Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAF MDQ), Bristol Rheumatoid Arthritis Fatigue Numerical Rating Scales (BRAF NRS) for severity, effect, and coping, Chalder Fatigue Questionnaire (CFQ), Checklist Individual Strength (CIS20R and CIS8R), Fatigue Severity Scale (FSS), Functional Assessment Chronic Illness Therapy (Fatigue) (FACT-F), Multi-Dimensional Assessment of Fatigue (MAF), Multi-Dimensional Fatigue Inventory (MFI), Pediatric Quality Of Life (PedsQL) Multi-Dimensional Fatigue Scale, Profile of Fatigue (ProF), Short Form 36 Vitality Subscale (SF-36 VT), and Visual Analog Scales (VAS). Arthritis Care Res 2011;63 Suppl 11:S263-86.


