Test-retest Reliability and Correlations of 5 Global Measures Addressing At-work Productivity Loss in Patients with Rheumatic Diseases

Sarah Leggett, Antje van der Zee-Neuen, Annelies Boonen, Dorcas E. Beaton, Mihai Bojinca, Ailsa Bosworth, Sabrina Dadoun, Bruno Fautrel, Sofia Hagel, Catherine Hofstetter, Diane Lacaille, Denise Linton, Carina Mihai, Ingemar F. Petersson, Pam Rogers, Jamie C. Sergeant, Carlo Sciré, and Suzanne M.M. Verstappen, on behalf of the At-work Productivity Global Measure Working Group

ABSTRACT. *Objective.* Several global measures to assess at-work productivity loss or presenteeism in patients with rheumatic diseases have been proposed, but the comparative validity is hampered by the lack of data on test-retest reliability and comparative concurrent and construct validity. Our objective was to test-retest 5 global measures of presenteeism and to compare the association between these scales and health-related well-being.

Methods. Sixty-five participants with inflammatory arthritis or osteoarthritis in paid employment were recruited from 7 countries (UK, Canada, Netherlands, France, Sweden, Romania, and Italy). At baseline and 2 weeks later, 5 global measures of presenteeism were evaluated: the Work Productivity Scale–Rheumatoid Arthritis (WPS-RA), Work Productivity and Activity Impairment Questionnaire (WPAI), Work Ability Index (WAI), Quality and Quantity questionnaire (QQ), and the WHO Health and Performance Questionnaire (HPQ). Agreement between the 2 timepoints was assessed using single-measure intraclass correlations (ICC) and correlated between each other and with visual analog scale general well-being scores at followup by Spearman correlation.

Results. ICC between measures ranged from fair (HPQ 0.59) to excellent (WPS-RA 0.78). Spearman correlations between measures were moderate ($Q_{quality}$ vs WAI, r = 0.51) to strong (WPS-RA vs WPAI, r = 0.88). Correlations between measures and general well-being were low to moderate, ranging from $-0.44 \le r \le 0.66$.

Conclusion. Test-retest results of 4 out of 5 global measures were good, and the correlations between these were moderate. The latter probably reflect differences in the concepts, recall periods, and references used in the measures, which implies that some measures are probably not interchangeable. (J Rheumatol First Release December 1 2015; doi:10.3899/jrheum.141605)

Key Indexing Terms:
AT-WORK PRODUCTIVITY LOSS	PRESENTEEISM	GLOBAL MEASURES
TEST-RETEST	CORRELATIONS	RHEUMATIC DISEASES

Program Clinical Research Unit, St. Michael's Hospital; C. Mihai, MD, PhD, Internal Medicine and Rheumatology Clinic, Dr. Ion Cantacuzino Hospital, Carol Davila University of Medicine and Pharmacy; I.F. Petersson, MD, FRCPC, Orthopaedics, Clinical Sciences Lund, Lund University; P. Rogers, Arthritis Research Centre of Canada; J. Sergeant, PhD, Lecturer in Medical Statistics, Arthritis Research UK Centre for Epidemiology, NIHR Manchester Musculoskeletal Biomedical Research Unit; C. Sciré, MD, PhD, Epidemiology Unit, Italian Society for Rheumatology; S.M. Verstappen, PhD, Senior Research Fellow, Arthritis Research UK Centre for Epidemiology, Manchester Academic Health Science Centre, University of Manchester, and Arthritis Research UK/MRC Centre for Musculoskeletal Health and Research, University of Southampton.

Address correspondence to Dr. S.M. Verstappen, Arthritis Research UK Centre for Epidemiology, Centre for Musculoskeletal Research, Institute of Inflammation and Repair, The University of Manchester, Manchester Academic Health Science Centre, Stopford Building, Oxford Road, Manchester, M13 9PT, United Kingdom.
E-mail: suzanne.verstappen@manchester.ac.uk
Accepted for publication September 1, 2015.

Inflammatory arthritis (IA) and osteoarthritis (OA) are chronic diseases known to affect many aspects of an individual's life including social, psychological, financial, and occupational consequences[1,2,3]. Many workers with rheumatic disorders such as rheumatoid arthritis (RA), psoriatic arthritis (PsA), ankylosing spondylitis (AS), and OA may experience restrictions participating in the workforce, potentially leading to sick days from work and early work cessation[4]. Rates of work disability in rheumatic patients are well documented, varying from 16% to 39% in patients with PsA in association with longer disease duration[5]; 67% after an average disease duration of 15 years in patients with RA[6]; and somewhat lower rates in patients with AS varying from 3% to 50% after 18 and 45 years' disease duration, respectively[7]. Work disability literature for OA is limited, although it suggests much lower rates compared to inflammatory arthritis[8]. However, with advances in disease management for rheumatic diseases in the last decade, individuals with rheumatic diseases may be able to remain in employment[9]. This can be related to both earlier diagnosis and earlier and more effective pharmacologic treatment, suggesting that measures of productivity could be included as routine outcome measures in clinical trials and in observational studies[10,11,12,13]. Nonetheless, for those who remain in employment, sickness absence and at-work productivity loss may be experienced. At-work productivity loss, also known as presenteeism, comprises the level of difficulty experienced at work due to health conditions and the amount of productivity lost to such difficulties. Measuring presenteeism is relevant for patients as it reflects the difficulties and challenges patients experience in order to function at their work. A change in worker productivity for an individual can jeopardize the outcomes considered important for most workers, such as self-esteem and family esteem, social inclusion and participation, and material standards of living[2]. The importance of presenteeism as part of work-related outcomes is becoming increasingly recognized as a disease-related outcome, and subsequently, attention is shifting to presenteeism as a potential predictor of work disability, offering the opportunity to identify and support people at risk for becoming work-disabled. Thus, presenteeism measurements have become an integral part of identifying the overall effect of rheumatic disorders on work and also on worker productivity both in trials and in clinical followup.

Different measures are available to evaluate presenteeism, including global measures and multi–item measures[14]. Global measures record a general perception of presenteeism using 1 or 2 single items, while multi-item measures use a number of questions, with a focus on one or multiple aspects of presenteeism [e.g., time management, physical demands, mental interpersonal and output demands — The Work Limitations Questionnaire (WLQ-25[15])]. Global measures are particularly attractive as they are more feasible in large clinical studies with interests in many outcome domains. Despite increasing research on the clinimetric properties of these global measures for at-work productivity, several questions remain[14]. First, no data on test-retest reliability are available for the measures in an arthritis population, and comparative construct and concurrent validity is scarce[16,17]. Second, data on comparative validity are limited but needed when aiming to recommend a global measure for use in a rheumatic working population. Finally, whether the global measures are applicable to differing occupational classes is an important but unanswered question.

Our aim was to investigate test-rest reliability of 5 global measures in patients with IA and OA including the Work Productivity Scale–Rheumatoid Arthritis (WPS-RA), Work Productivity and Activity Impairment Questionnaire (WPAI), Work Ability Index (WAI), Quality and Quantity questionnaire (QQ), and the WHO Health and Performance Questionnaire (HPQ); and to compare the association between these measures with each other and with general well-being to determine whether the data between the scales are interchangeable. The 5 global measures at the center of investigation were chosen after a survey among researchers, patients, and clinicians with expertise in the field of worker productivity.

MATERIALS AND METHODS

Study population. Individuals with a physician diagnosis of IA (including RA, PsA, and AS) or OA were recruited for study via 1 outpatient rheumatology or orthopedic clinic from 7 participating countries (United Kingdom, Canada, The Netherlands, France, Sweden, Romania, and Italy). Patients had to be age ≥ 18 years, in paid full or part-time employment, and able to communicate verbally and in writing in the language of the participating country. The reliability data collected in this study were part of a larger study that applied cognitive debriefing to investigate the content validity of the measures, the results of which will be presented in a different report. Recruitment aimed at 10 participants per country, with an equal distribution of age, sex, manual/nonmanual jobs, and disease type. All patients provided written informed consent. Ethical approval was obtained from each participating center according to national ethical guidelines.

Data collection. The following 5 global measures were selected after a survey among researchers, rheumatologists, and patients[14]: WPAI[18], WPS-RA[19], WAI[20], QQ[21], and the HPQ[22]. All 5 global measures in this study use an 11-point visual analog scale (VAS) for scoring presenteeism. Characteristics of the global measures can be found in Table 1. The WPAI, WPS-RA, WAI, and HPQ presenteeism items were extracted from the large absenteeism and presenteeism questionnaires with permission from the developers. The HPQ can be summed to provide 2 scores: a ratio of presenteeism in comparison to co-workers, or a standalone single-item measure[23]. For this study, the single-item measure was used (question C of the 3 HPQ questions presented to study patients).

If available, translated global measures were used in each country, otherwise the measures were translated into the corresponding language of the participating country by a company specializing in translations in medicine (PharmaQuest Ltd.). Translations were conducted using cognitive debriefing interviews with 5 patient volunteers with rheumatic diseases for each language of interest[24], and utilized a forward and backward translation methodology.

Participants completed the set of global measures at 2 timepoints: at baseline and 2 weeks thereafter. An interval of 2 weeks has been recommended for use in test-retest reliability analyses[25], and we believed 2 weeks would reflect little or no change in work productivity. Participants were sent hard copies of the baseline questionnaires by post 2 weeks before they were scheduled to attend the hospital/research center for the cognitive interview study. Participants completed the baseline questionnaires at home, which were then returned to each participating center in a prepaid envelope. The 2-week questionnaire along with reminders were sent to participants to ensure the timepoints were separated by a 2-week interval. Participants were instructed to complete the 2-week questionnaires at home prior to the interview. The global measures were presented in a booklet to participants, with instructions for each of the individual measures. Before the start of the cognitive interview, demographic (age, sex, highest level of education), clinical (diagnosis, symptom duration, date of diagnosis), and occupational (job description, demands, incentives for working) information was collected. Occupations were categorized into manual and nonmanual occupations according to the UK Standard Occupational Classification[26]. Finally, patients were asked about their general well-being using a VAS ranging from 0 (very well) to 100 (very poorly).

Statistical analysis. Descriptive statistics were used to summarize subjects'

demographic, clinical, and occupational characteristics. For the total study population, test-retest reliability data for each of the 5 global measures at baseline and 2-week followup were generated using single-measure intraclass correlations (ICC), applying a 2-way random-effects model based on a repeated-measure ANOVA (ICC 2,1)[27]. Missing values were accounted for automatically in the ICC analysis using listwise deletion. ICC values are in general interpreted as: > 0.75 = excellent for group level analyses, 0.40–0.75 = fair to good, and < 0.40 = poor[28]. Bland-Altman plots were created to identify error across the range of score scales for each of the 5 measures. Bland-Altman plots were considered appropriate as they allow systematic differences over the range of possible values to be detected that would otherwise be difficult to detect using simple point clouds. In a sensitivity analysis, ICC were computed to determine the reliability of the measures within the groups classified as having manual and nonmanual jobs. Spearman correlations were also used to compare the level of correlation among global measures of presenteeism, and between the measures with VAS general well-being. SPSS 20 was used to generate the ICC and Spearman correlations; STATA 13 was used for descriptive analysis and to create the Bland and Altman plots.

RESULTS

Although 70 (n = 10 per country) were recruited for study, data for 5 participants were excluded from analysis because of likely misinterpretation of the direction of the scales, with the WPAI and WPS-RA ranging from good to worse and other measures ranging from worse to good. This misinterpretation became apparent during the cognitive interviews and was confirmed using scatterplots between the different global measures. Further, fewer observations were recorded for the QQ (n = 62) and the HPQ (n = 64). The total study population (useable data) therefore comprised 65 patients with RA, PsA, AS, OA, or other, e.g., connective tissue disorder. The median age was 45 years [interquartile range (IQR) 37-52], 55% were female, and the median disease duration was 13 years (IQR 6-19). The majority of participants had RA (32%) (Table 2), and 71% had a nonmanual

Table 1. Characteristics of the 5 global presenteeism measures.

Global Measure	Construct	Recall Period	Reference Frame (if any)	Attribution	Anchors
WPAI[18]	Productivity (additional phrasing: accomplished, kind of work, carefully as usual, amount of work)	7 days	NA	Disease-specific (RA, PsA, AS, OA)	Condition had no effect on my work (0) – Condition completely prevented me from working (10)
WPS-RA[19]	Productivity & interference	Last month	NA	Arthritis	No interference (0) – Complete interference (10)
WAI[20]	Work ability	Current	Work ability at its best	NA	Completely unable to work (0) – Work ability at its best (10)
QQ[21]	Quality and quantity	Today	Compared to normal	NA	Practically nothing (0) – Normal quantity (10) & Very poor quality (0) – very good quality (10)
HPQ[22]	Performance	One/two yrs (question B) & 4 wks (question C)	Colleagues' performance (question A)	NA	Worst performance (0) – Top performance (10)

WPAI: Work Productivity and Activity Impairment Questionnaire; WPS-RA: Work Productivity Scale–Rheumatoid Arthritis; WAI: Work Ability Index; QQ: Quality and Quantity questionnaire; HPQ: World Health Organization Health and Work Performance Questionnaire; NA: not applicable; RA: rheumatoid arthritis; PsA: psoriatic arthritis; AS: ankylosing spondylitis; OA: osteoarthritis.

Table 2. Demographic and clinical characteristics (n = 65).

Characteristic	Value
Age, yrs	
Median (IQR) yrs	45 (37-52)
Mean (SD)	46 (11)
Female, n (%)	36 (55)
Symptom duration, yrs	
Median (IQR)	13 (6-19)
Mean (SD)	13 (10)
VAS for well-being	
Median (IQR)	28 (13-53)
Mean (SD)	34 (25)
Diagnosis, n (%)	
Rheumatoid arthritis	21 (32)
Ankylosing spondylitis	20 (31)
Psoriatic arthritis	11 (17)
Osteoarthritis	8 (12)
Other inflammatory disease*	5 (8)

*Other inflammatory diseases include undifferentiated arthritis and undifferentiated connective tissue disease. IQR: interquartile range; SD: standard deviation; VAS: visual analog scale.

occupation. In general, participants had mild to moderate impact of disease based on the general well-being scale. The median score at baseline was 3 for the WPAI (IQR 1-6; higher score = worse score), 3 for the WPS-RA (IQR 2-6; higher score = worse score), 7 for the WAI (IQR 5-9; higher score = best score), 72 for the total QQ (IQR 49-100; higher score = best score; QQ questions 1 and 2 multiplied to give a range of 0-100), and a median of 7 for HPQ question C (IQR 6-8; higher score = best score).

Test-retest reliability. ICC for all 5 global measures at first and second assessment are shown in Table 3. Overall, the ICC in the total sample ranged from 0.59 (95% CI 0.40-0.73; HPQ question C) to 0.78 (95% CI 0.65-0.86; WPS-RA), which

Table 3. Intraclass correlations (ICC) for assessment at 2 timepoints (baseline and 2 weeks) of the 5 global measures for the total sample.

Global Measure	N	ICC
WPAI	65	0.74*
WPS-RA	65	0.78*
WAI	65	0.75*
$Q_{Quantity}$	63	0.69*
$Q_{Quality}$	62	0.73*
QQ_{Total}	62	0.76*
HPQ question C	64	0.59*

*p < 0.001. WPAI: 0 = RA had no effect on my work, 10 = RA completely prevented me from working; WPS-RA: 0 = no interference, 10 = complete interference; WAI: 0 = completely unable to work, 10 = work ability at its best; QQ: 0 = practically nothing/very poor quality, 10 = normal quantity/normal quality; HPQ question C: 0 = worst performance, 10 = top performance. WPAI: Work Productivity and Activity Impairment Questionnaire; WPS-RA: Work Productivity Scale: Rheumatoid Arthritis; WAI: Work Ability Index; QQ: Quality and Quantity questionnaire; HPQ: World Health Organization Health and Work Performance Questionnaire.

indicates fair to excellent agreement between test-retest scores at baseline and 2-week followup at group level (p < 0.001 for all tests). Bland-Altman plots illustrate the differences between the 2 timepoints (baseline and 2 weeks) and the mean scores at baseline and 2 weeks, demonstrating the degree of variation for all 5 global measures for the total sample (See Supplementary Figure 1, available online at jrheum.org). The 95% limits of agreement vary between each of the measures, but in general, the variation of the difference around the mean for most was moderate. For each measure, a few differences outside or close to the limit lines are present, demonstrating minimal levels of disagreement between the 2 timepoints (See panels, Supplementary Figure 1, available online at jrheum.org).

ICC in both the nonmanual working group (n = 46) and the manual working group (n = 19) showed good to excellent agreement, with ICC ranging from 0.67 (95% CI 0.47–0.80; HPQ question C) to 0.83 (95% CI 0.71–0.90; $Q_{quality}$) in the nonmanual group of participants, and poor to excellent agreement rate, with a range of 0.39 (95% CI –0.08 to 0.70; HPQ question C) to 0.76 (95% CI 0.47–0.90; WPS-RA) in the manual group (Table 4).

Construct validity. Spearman rank correlations between the 5 global measures yielded values ranging from strong (WPAI and WPS-RA, r = 0.88) to moderate ($Q_{quality}$ vs WAI, r = 0.51). Assuming that these measures should assess the same construct, namely presenteeism, then ICC are justified. These ranged from fair [–0.46 (95% CI –0.64 to –0.25) for WPAI and $Q_{quantity}$; negative values as a result of opposing anchors on scales] to excellent [0.85 (95% CI 0.76 to 0.90) for WPAI and WPS-RA]. Correlations between each of the individual measures and VAS general well-being were low to moderate, ranging from r = –0.44 (QQ total) to r = –0.66 (WPS-RA) (p < 0.001 for all tests; Table 5).

DISCUSSION
The results showed that 4 out of the 5 measures, the WPS-RA, WAI, QQ, and WPAI (considered acceptable as only

Table 4. Intraclass correlations (ICC) for assessments over time of the 5 global measures for the nonmanual and manual job groups.

Global Measure	Nonmanual Job		Manual Job	
	N	ICC	N	ICC
WPAI	46	0.77*	19	0.68*
WPS-RA	46	0.78*	19	0.76*
WAI	46	0.77*	19	0.67*
$Q_{Quantity}$	45	0.75*	18	0.53**
$Q_{Quality}$	44	0.83*	19	0.59**
QQ_{Total}	44	0.81*	18	0.65*
HPQ question C	45	0.67*	19	0.39*

* p ≤ 0.001; ** p < 0.01. WPAI: Work Productivity and Activity Impairment Questionnaire; WPS-RA: Work Productivity Scale–Rheumatoid Arthritis; WAI: Work Ability Index; QQ: Quality and Quantity questionnaire; HPQ: World Health Organization Health and Work Performance Questionnaire;

Table 5. Spearman correlations between 5 at-work productivity loss measures and visual analog scale (VAS) for general well-being. Note negative correlations are present when the comparison is between 2 global measures with opposing visual analog scales (e.g., WPAI vs QQ).

	WPAI	WPS-RA	WAI	$Q_{Quantity}$	$Q_{Quality}$	QQ_{Total}	HPQ question C
WPAI	–						
WPS-RA	0.88	–					
WAI	–0.69	–0.73	–				
Q Quantity	–0.60	–0.67	0.65	–			
Q Quality	–0.57	–0.65	0.51	0.81	–		
QQ Total	–0.63	–0.68	0.56	0.94	0.91	–	
HPQ question C	–0.67	–0.74	0.66	0.61	0.55	0.61	
VAS well-being	0.58	0.66	–0.61	–0.45	–0.46	–0.44	–0.53

P values all < 0.001. WPAI: 0 = RA had no effect on my work, 10 = RA completely prevented me from working; WPS-RA: 0 = no interference, 10 = complete interference; WAI: 0 = completely unable to work, 10 = work ability at its best; QQ: 0 = practically nothing/very poor quality, 10 = normal quantity/normal quality; HPQ question C: 0 = worst performance, 10 = top performance. WPAI: Work Productivity and Activity Impairment Questionnaire; WPS-RA: Work Productivity Scale–Rheumatoid Arthritis; WAI: Work Ability Index; QQ: Quality and Quantity questionnaire; HPQ: World Health Organization Health and Work Performance Questionnaire; VAS: visual analog scale.

marginally short at 0.74), met or exceeded the 0.75 threshold of acceptability for agreement between a baseline and 2-week interval, an interval we felt would reflect a stable situation with little or no change in work productivity. The correlations between the global measures suggested mostly moderate construct validity. Low to moderate convergent validity between the measures and general well-being was evident. Although the ICC of the 5 global measures have not previously been compared, individually our results are congruent to the limited reliability evidence used within different chronic conditions such as irritable bowel syndrome[29] and Crohn's disease[30], while the data from the other global measures serve to build upon the sparse evidence base.

The discrepancy between ICC in the manual and nonmanual participant groups was an interesting finding. Consistently higher agreement for the nonmanual subgroup was evident, with the most extreme difference lying within the QQ quality question (nonmanual 0.83; manual 0.59). This may be partly determined by a variance in physical exertion in manual work over time compared to sedentary work, as reflected in inconsistent scoring over the 2 timepoints. Manual work is known to be a risk factor for presenteeism in musculoskeletal disorders[31,32,33], highlighting a contextualized relationship between rheumatic diseases and work productivity. Although this would normally account for differences in the magnitude of work ability, in this study it also seemed to suggest some instability in scores (lower reliability) in jobs with more fluctuating demands. As such, it may be worthwhile to account for intensity of job demands in future analysis.

Correlations between the 5 global measures in our study were moderate. This is similar to the literature in which previous findings demonstrate moderate correlations between at-work productivity outcome measures [e.g., WALS and the Endicott Work Productivity Scale (r = 0.55) and WALS and the WLQ-25 index (r = 0.61)][9]. Although each measure at its core measures presenteeism, different concepts are included such as "productivity" (WPAI), "performance" (HPQ), and "ability" (WAI), as well as different timeframes, for example, "today" (QQ), "7 days" (WPAI), and "one month" (WPS-RA). However, the strong association between the WPAI and the WPS-RA (r = 0.88) suggests congruence between the 2 measures, potentially with regard to their conceptual foci, as both measures utilize the term "productivity."

The low to moderate results of the VAS general well-being and global measure correlations imply that productivity loss may be only partly captured by general health issues, suggesting that other factors, including contextual factors, may contribute to productivity loss. Recent research into contextual factors of work disability and absenteeism supports this notion, showing that factors such as family support towards work, work modifications, and physical job demands influence both work disability and absenteeism[34,35,36]; the same factors may also apply to presenteeism.

There were a few limitations to our study. The reversal of the VAS for the WPAI and WPS-RA accounted for the loss of 5 patients in the analysis, as a misinterpretation of the direction of the anchors on the scales was apparent. The opposing scores on the remaining measures indicated a clear intention to score similar scores on the measures with reversed VAS. The collection of general well-being VAS only at the 2-week timepoint is a limitation. However, an interval of 2 weeks for test-retest investigations is often thought to be appropriate as there is little risk for recall bias and clinical change[25,37]. Further, it is possible that an element of social desirability bias may have influenced the results; however, participants were reassured of data confidentiality to eliminate any uncertainty of employers seeing participants' responses, which we hope minimizes the possibility of social desirability bias. As well, although the questionnaires were rigorously translated using an approved translation company, some differences in translation may contribute to the moderate results; we expect the effect of this to be minimal due to procedures used during translation.

Test-retest reliability of the 5 global measures of presenteeism was mostly good, with 4 of the 5 measures meeting or exceeding the threshold of acceptable reliability. The WPS-RA performed best in the test-retest reliability analysis, and with the WPAI in the comparison between measures, suggesting interchangeability between available data using the 2 methods. Whether the remaining measures can be considered interchangeable is tentative. The findings suggest that different instruments in different diseases and discourses should perhaps be used. Based on our findings, the HPQ would not be recommended for use as a global measure of presenteeism in rheumatic working populations. The extent to which other measures provide different information likely depends on predictive ability with regard to absence and work disability. Additional work is warranted to support psychometric evidence of the 5 global measures, paying particular attention to potential contextual factors, such as manual and nonmanual occupations, that may influence study results.

ACKNOWLEDGMENT

ONLINE SUPPLEMENT
Supplementary data for this article are available online at jrheum.org.

REFERENCES

1. Waddell G, Burton AK. Is work good for your health and well-being? Norwich: The Stationery Office; 2006.
2. Black C. Working for a healthier tomorrow: Work and health in Britain. Norwich: The Stationery Office; 2008.
3. Bevan S, Quadrello T, McGee R, Mahdon M, Vavrovsky A, Barham L. Fit for work? Musculoskeletal disorders in the European workforce. London: The Work Foundation; 2009.
4. Dunlop DD, Manheim LM, Yelin EH, Song J, Chang RW. The costs of arthritis. Arthritis Rheum 2003;49:101-13.
5. Tillett W, de-Vries C, McHugh NJ. Work disability in psoriatic arthritis: A systematic review. Rheumatology 2012;51:275-83.
6. Verstappen SM, Bijlsma JW, Verkleij H, Buskens E, Blaauw AA, ter Borg EJ, et al. Overview of work disability in rheumatoid arthritis patients as observed in cross-sectional and longitudinal surveys. Arthritis Rheum 2004;51:488-97.
7. Boonen A, de Vet H, van der Heijde D, van der Linden S. Work status and its determinants among patients with ankylosing spondylitis. A systematic literature review. J Rheumatol 2001;28:1056-62.
8. Gabriel SE, Crowson CS, Campion ME, O'Fallon WM. Indirect and nonmedical costs among people with rheumatoid arthritis and osteoarthritis compared with nonarthritic controls. J Rheumatol 1997;24:43-8.
9. Beaton DE, Tang K, Gignac MA, Lacaille D, Badley EM, Anis AH, et al. Reliability, validity, and responsiveness of five at-work productivity measures in patients with rheumatoid arthritis or osteoarthritis. Arthritis Care Res 2010;62:28-37.
10. Ziegler S, Huscher D, Karberg K, Krause A, Wassenberg S, Zink A. Trends in treatment and outcomes of rheumatoid arthritis in Germany 1997-2007: Results from the National Database of the German Collaborative Arthritis Centres. Ann Rheum Dis 2010;69:1803-8.
11. Neovius M, Simard JF, Klareskog L, Askling J. Sick leave and disability pension before and after initiation of antirheumatic therapies in clinical practice. Ann Rheum Dis 2011;70:1407-14.
12. Olofsson T, Englund M, Saxne T, Joud A, Jacobsson LT, Geborek P, et al. Decrease in sick leave among patients with rheumatoid arthritis in the first 12 months after start of treatment with tumour necrosis factor antagonists: A population-based controlled cohort study. Ann Rheum Dis 2010;69:2131-6.
13. Kristensen LE, Englund M, Neovius M, Askling J, Jacobsson LT, Petersson IF. Long-term work disability in patients with psoriatic arthritis treated with anti-tumour necrosis factor: A population-based regional Swedish cohort study. Ann Rheum Dis 2013;72:1675-9.
14. Tang K, Boonen A, Verstappen SM, Escorpizo R, Luime JJ, Lacaille D, et al. Worker productivity outcome measures: OMERACT filter evidence and agenda for future research. J Rheumatol 2014; 41:165-76.
15. Lerner D, Amick BCI, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. Med Care 2001;39:72-85.
16. Zhang W, Bansback N, Boonen A, Young A, Singh A, Anis AH. Validity of the Work Productivity and Activity Impairment Questionnaire - General health version in patients with rheumatoid arthritis. Arthritis Res Ther 2010;12:R177.
17. Lambert J, Hansen BB, Arnould B, Grataloup G, Guillemin I, Hojbjerre L, et al. Linguistic validation into 20 languages and content validity of the rheumatoid arthritis-specific Work Productivity and Activity Impairment Questionnaire. Patient 2014;7:171-6.
18. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics 1993;4:353-65.
19. Osterhaus JT, Purcaru O, Richard L. Discriminant validity, responsiveness and reliability of the rheumatoid arthritis-specific Work Productivity Survey (WPS-RA). Arthritis Res Ther 2009;11:R73.
20. Tuomi K, Ilmarinen J, Jahkola A, Katajarinne L, Tulkki A. Work ability index. 2nd ed. Helsinki: Finnish Institute of Occupational Health; 1998.
21. Brouwer WB, Koopmanschap MA, Rutten FF. Productivity losses without absence: Measurement validation and empirical evidence. Health Policy 1999;48:13-27.
22. Kessler RC, Barber C, Beck A, Berglund P, Cleary PD, McKenas D, et al. The World Health Organization Health and Work Performance Questionnaire (HPQ). J Occup Environ Med 2003;45:156-74.
23. Kessler RC, Petukhova M, McInnes K. HPQ Short Form (absenteeism and presenteeism questions and scoring rules). [Internet. Accessed October 21, 2015.] Available from: http://www.hcp.med.harvard.edu/hpq/ftpdir/ absenteeism%20presenteeism%20scoring%2005050107.pdf
24. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000;25:3186-91.
25. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. J Clin Epidemiol 2003;56:730-5.
26. Office for National Statistics. Standard occupational classification 2000. Volume 1. Structure and description of unit groups. London: The Stationery Office; 1990.
27. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979;86:420-8.
28. Fleiss JL. Reliability of measurement. The design and analysis of clinical experiments. New York, Chichester: Wiley; 1986:1-32.
29. Bushnell DM, Reilly MC, Galani C, Martin ML, Ricci JF, Patrick DL, et al. Validation of electronic data capture of the Irritable Bowel Syndrome-Quality of Life Measure, the Work Productivity and

Activity Impairment Questionnaire for Irritable Bowel Syndrome and the EuroQol. Value Health 2006;9:98-105.

30. Vergara M, Montserrat A, Casellas F, Villoria A, Suarez D, Maudsley M, et al. A new validation of the Spanish Work Productivity and Activity Impairment Questionnaire-Crohn's disease version. Value Health 2011;14:859-61.

31. Agaliotis M, Fransen M, Bridgett L, Nairn L, Votrubec M, Jan S, et al. Risk factors associated with reduced work productivity among people with chronic knee pain. Osteoarthritis Cartilage 2013;21:1160-9.

32. Ng YG, Tamrin SB, Yik WM, Yusoff IS, Mori I. The prevalence of musculoskeletal disorder and association with productivity loss: A preliminary study among labour intensive manual harvesting activities in oil palm plantation. Ind Health 2014;52:78-85.

33. Geuskens GA, Hazes JM, Barendregt PJ, Burdorf A. Predictors of sick leave and reduced productivity at work among persons with early inflammatory joint conditions. Scand J Work Environ Health 2008;34:420-9.

34. Lacaille D, Sheps S, Spinelli JJ, Chalmers A, Esdaile JM. Identification of modifiable work-related factors that influence the risk of work disability in rheumatoid arthritis. Arthritis Rheum 2004;51:843-52.

35. Allaire SH, Anderson JJ, Meenan RF. Reducing work disability associated with rheumatoid arthritis: Identification of additional risk factors and persons likely to benefit from intervention. Arthritis Care Res 1996;9:349-57.

36. Boonen A, Boone C, Albert A, Mielants H. Understanding limitations in at-work productivity in patients with active ankylosing spondylitis: The role of work-related contextual factors. J Rheumatol 2015;42:93-100.

37. Streiner DL, Norman GR. Health measurement scales: A practical guide to their development and use. 4th ed. Oxford: Oxford University Press; 2008.