OMERACT Filter Evidence Supporting the Measurement of At-work Productivity Loss as an Outcome Measure in Rheumatology Research

Dorcas E. Beaton, Sarah Dyer, Annelies Boonen, Suzanne M.M. Verstappen, Reuben Escorpizo, Diane V. Lacaille, Ailsa Bosworth, Monique A.M. Gignac, Amye Leong, Oana Purcaru, Sarah Leggett, Cathy Hofstetter, Ingemar F. Peterson, Kenneth Tang, Bruno Fautrel, Claire Bombardier, and Peter S. Tugwell

ABSTRACT. ***Objective.*** Indicators of work role functioning (being at work, and being productive while at work) are important outcomes for persons with arthritis. As the worker productivity working group at OMERACT (Outcome Measures in Rheumatology), we sought to provide an evidence base for consensus on standardized instruments to measure worker productivity [both absenteeism and at-work productivity (presenteeism) as well as critical contextual factors].

Methods. Literature reviews and primary studies were done and reported to the OMERACT 12 (2014) meeting to build the OMERACT Filter 2.0 evidence for worker productivity outcome measurement instruments. Contextual factor domains that could have an effect on scores on worker productivity instruments were identified by nominal group techniques, and strength of influence was further assessed by literature review.

Results. At OMERACT 9 (2008), we identified 6 candidate measures of absenteeism, which received 94% endorsement at the plenary vote. At OMERACT 11 (2012) we received over the required minimum vote of 70% for endorsement of 2 at-work productivity loss measures. During OMERACT 12 (2014), out of 4 measures of at-work productivity loss, 3 (1 global; 2 multiitem) received support as having passed the OMERACT Filter with over 70% of the plenary vote. In addition, 3 contextual factor domains received a 95% vote to explore their validity as core contextual factors: nature of work, work accommodation, and workplace support.

Conclusion. Our current recommendations for at-work productivity loss measures are: WALS (Workplace Activity Limitations Scale), WLQ PDmod (Work Limitations Questionnaire with modified physical demands scale), WAI (Work Ability Index), WPS (Arthritis-specific Work Productivity Survey), and WPAI (Work Productivity and Activity Impairment Questionnaire). Our future research focus will shift to confirming core contextual factors to consider in the measurement of worker productivity. (First Release September 1 2015; J Rheumatol 2016;43:214–22; doi:10.3899/jrheum.141077)

Key Indexing Terms:
ARTHRITIS | AT-WORK PRODUCTIVITY LOSS | EMPLOYMENT
OUTCOME MEASURES | RHEUMATOLOGY | WORKER PRODUCTIVITY

From the Arthritis Research Centre of Canada, Vancouver, British Columbia, Canada; Arthritis Research UK Epidemiology Unit, University of Manchester; Manchester, UK; Bone and Joint Decade, the Global Alliance for Musculoskeletal Health, Santa Barbara, California, USA; Consumer Advisory Council, Canadian Arthritis Network, Canada; Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada; Department of Rehabilitation and Movement Science, The University of Vermont, Burlington, Vermont, USA; Department of Rheumatology, Pierre et Marie Curie University (UPMC - Paris 6), APHP Pitié-Salpétrière Hospital, Paris, France; Department of Rheumatology, Skåne University Hospital, Malmo; Department of Orthopedics, Clinical Sciences, Lund University, Lund, Sweden; Division of Rheumatology, Maastricht University Medical Center, and CAPHRI Research Institute, Maastricht, the Netherlands; Division of Rheumatology, University of British Columbia, Vancouver, British Columbia, Canada; Health Economics, GMAP, Immunology, UCB BioPharma SPRL, Brussells, Belgium; Institute for Work and Health; Musculoskeletal Health and Outcomes Research, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Ontario, Canada; National Rheumatoid Arthritis Society, Maidenhead, Berkshire, UK; Toronto General Research Institute at the University Health Network, Toronto, Ontario, Canada.

Supported by research grants from the Canadian Arthritis Network, a network of centers of excellence; European League Against Rheumatology; an unrestricted grant from Abbott, and OMERACT. D. Lacaille holds the Mary Pack Chair in Arthritis Research from The Arthritis Society of Canada and the University of British Columbia. C. Bombardier holds a Canada Research Chair in Knowledge Transfer for Musculoskeletal Care and a Pfizer Chair in Rheumatology, University of Toronto, Faculty of Medicine's Rheumatology Division.

D.E. Beaton, PhD, Musculoskeletal Health and Outcomes Research, Li Ka Shing Knowledge Institute, St. Michael's Hospital, and Institute for Work and Health, and Institute of Health Policy, Management and Evaluation, Rehabilitation Sciences Institute, and Department of Occupational Science and Occupational Therapy, University of Toronto; S. Dyer, MSc, Musculoskeletal Health and Outcomes Research, Li Ka Shing Knowledge Institute, St. Michael's Hospital; A. Boonen, MD, PhD, Division of Rheumatology, Maastricht University Medical Center, and CAPHRI Research Institute; S.M. Verstappen, PhD, Arthritis Research UK

Epidemiology Unit, the University of Manchester; R. Escorpizo, PT, DPT, MSc, Department of Rehabilitation and Movement Science, University of Vermont, and Department of Health Sciences and Health Policy, University of Lucerne, and Swiss Paraplegic Research and the ICF Research Branch of WHO Collaborating Centre for the Family of International Classifications in Germany; D.V. Lacaille, MD, MHSc, FRCPC, Mary Pack Chair in Arthritis Research, Division of Rheumatology, University of British Columbia, and Arthritis Research Centre of Canada; A. Bosworth, National Rheumatoid Arthritis Society; M.A. Gignac, PhD, Institute for Work and Health, and Division of Health Care and Outcomes Research, Toronto Western Research Institute at the University Health Network, and Arthritis Community Research and Evaluation Unit, and Dalla Lana School of Public Health, University of Toronto; A. Leong, MBA, Bone and Joint Decade, the Global Alliance for Musculoskeletal Health, and Healthy Motivation; O. Purcaru, PhD, Health Economics, GMAP, Immunology, UCB BioPharma SPRL; S. Leggett, MSc, Arthritis Research UK Epidemiology Unit, University of Manchester; C. Hofstetter, Consumer Advisory Council, Canadian Arthritis Network; I.F. Petersson, MD, FRCPC, Epi-centrum Skåne, and Department of Rheumatology, Skåne University Hospital, and Department of Orthopaedics, Clinical Sciences, Lund University; K. Tang, MSc(PT), MSc, Institute of Health Policy, Management and Evaluation, University of Toronto, and Musculoskeletal Health and Outcomes Research, Li Ka Shing Knowledge Institute, St. Michael's Hospital, and Institute for Work and Health; B. Fautrel, MD, PhD, Department of Rheumatology, Pierre et Marie Curie University (UPMC-Paris 6), APHP Pitié-Salpétrière Hospital; C. Bombardier, MD, FRCPC, Division of Rheumatology, Faculty of Medicine and Institute of Health Policy, Management and Evaluation, University of Toronto, and Institute for Work and Health; Toronto General Research Institute at the University Health Network, and Mount Sinai Hospital; P.S. Tugwell, MD, Department of Epidemiology and Community Medicine, University of Ottawa, and Centre for Global Health, Institute of Population Health, University of Ottawa, and Ottawa Hospital.

Address correspondence to Dr. D.E. Beaton, Muscuoskeletal Health and Outcomes Research, Li Ka Shing Knowledge Institute, St. Michael's Hospital, 30 Bond St., Toronto, Ontario M5B 1W8, Canada.
E-mail: beatond@smh.ca

Work has meaning to individuals in terms of their societal role, income, access to benefits, and social networking. For people with arthritis, the ability to maintain or regain a work role with a new treatment is an important issue in their lives. However, work-role functioning is rarely included in clinical trials. The Outcome Measures in Rheumatology (OMERACT) worker productivity group has identified available instruments and is building an OMERACT Filter 2.0 evidence base to support the measurement of this important outcome in arthritis research. Over the past 6 years[1,2,3,4] we have moved closer to our goal of standardizing the measurement of worker productivity in rheumatology. The purpose of this article is to review the accumulated material that was presented at the worker productivity special interest group (SIG) meeting at OMERACT 12 on our slate of 6 at-work productivity loss measurement instruments in terms of truth, discrimination, and feasibility concepts of the OMERACT Filter 2.0[5]; to share our emerging evidence on contextual factors of importance to the accurate measurement of worker productivity; and to share the results of plenary votes taken supporting our work at the plenary session of OMERACT 12.

Background

Difficulties in worker productivity include absence from work or a reduction in productivity or in the ease of producing while at work (at-work productivity loss, sometimes called "presenteeism"). People can transition back and forth across this threshold between not working, working but with difficulty, and working with no difficulty. The transitions might be driven by the health and abilities of a worker compared to their job's demands, or equally by shifting the job's demands to accommodate the worker's abilities. The context of the job situation always accompanies the description and rating that someone will give to their productivity. Contextual factors must be part of the accurate measurement and interpretation of worker productivity.

Indicators of absence from work were endorsed (94% in support) at a previous OMERACT meeting to include: (1) work days missed due to arthritis (sick days), (2) vacation days taken because of arthritis, (3) part days/hours missed because of arthritis, (4) change in number of hours worked per week, (5) temporary work cessation (work disability/sick leave), and (6) permanent work cessation due to arthritis[1,6].

Our attention subsequently shifted to at-work productivity loss, a concept that can be experienced in 2 important ways. First, a level of difficulty doing the tasks of work, and second the level of productivity loss (the amount of work that is not getting done because of the health limitation)[5]. To date, there is still no agreed-upon scale out of > 21 instruments now available to facilitate assessment of this part of worker productivity[7,8,9,10]. In 2008, our group led OMERACT attendees through a process of assessing the feasibility and truth (content) of the many available measures of at-work productivity loss. We were guided to narrow our work down to what are now 6 candidate instruments: WAI (Work Ability Index)[11]; QQ (Quantity and Quality Method)[12]; WPAI (Work Productivity and Activity Impairment Questionnaire)[13]; WPS [Arthritis-specific Work Productivity Survey (formerly WPS-RA, Rheumatoid arthritis-specific work productivity survey)][14], which now has evidence of use in 3 rheumatologic conditions, and is arthritis-specific[15,16]; WALS (Workplace Activity Limitations Scale)[6,17], and the WLQ-25 PDmod (Work Limitations Questionnaire)[18], with modified physical demands scale, where instruction for the physical demands subscale was reoriented to be consistent with other subscales, with the agreement of the manufacturer (personal communications with developer D. Lerner). Two of these, the WPAI and the WPS, received > 70% endorsement that they had met the OMERACT Filter at OMERACT 11[4]. The work included in the present article summarizes our ongoing work with the other instruments to complete OMERACT Filter evidence examination[2,19,20,21], and supplementing what we know about the WPS and WPAI.

Our attention has also been focused on contextual factors. Early in our work in worker productivity, it became apparent from discussions with patients that context is critically

important to the correct measurement and interpretation of worker productivity[4]. Contextual factors are factors that relate to the worker and to the environments in the workplace (physical, social, psychological). Based on the World Health Organization's International Classification of Functioning, Disability, and Health framework[22], contextual factors refer to personal factors and environmental factors. Both worker coping strategies and self-efficacy, as well as alterations in job-related demands can have an important influence on the score obtained on a worker productivity instrument. Therefore these factors need to be considered when interpreting the results of worker productivity outcome measures both in describing a state at one point in time or when evaluating change over time, where the job situation rather than personal capacity could be equally responsible for improving a level of at-work productivity[4]. During a SIG meeting at OMERACT 10, an exhaustive list of possible contextual factors was generated by experts and patient research partners, and after a "dot voting" exercise, 24 contextual factors received at least 1 vote. Of interest was that 1/24 factors received only 1 vote and that the 2 factors that received the most votes received only 13% of all votes, showing the wide diversity in the character of relevant contextual factors. Following the SIG meeting, contextual factors considered were clustered into 15 domains that described either personal or environmental contextual factors[4]. While it is undeniable that each of the factors could be relevant for the understanding of productivity for an individual person, we were also interested in the degree to which these factors could cause confounding bias in observational studies or clinical trials. At OMERACT 11, a list of criteria was presented to guide the selection of contextual factors that could confound the measurement of at-work productivity loss. Criteria included the quality of the study (low risk of bias), the strength of the association after adjustment (requiring a sufficient sample size), evidence of a temporal relation in the case of absenteeism, and sufficient strength of association to identify a possible confounding influence. In addition, guidance would be needed for deciding on level of evidence needed for each contextual factor (number of studies, consistency of findings, magnitude of results)[3].

The purpose of this article is to describe the progress of the OMERACT 12 SIG on worker productivity in both instrument selection and determination of relevant contextual factor domains.

MATERIALS AND METHODS

Taxonomy for At-work Productivity Loss Measures

In the past we published an organizing framework for instruments measuring worker productivity. First, they can be multiitem scales or single items offering 1 global rating of the concept. Second, some are focused on a time or output performance as the key concept (Are you as productive as before your arthritis, how much time do you have difficulty) while others are focused on the ability/amount of difficulty the respondent has doing the task. We therefore encountered 4 different types of measures in our work. The organization of these into a taxonomy is shown in Figure 1, which also shows that all of these sit on a background of the contextual factors that are describing the situation and circumstances in which the difficulty/productivity is being measured. Thus we acknowledge that the difficulty/productivity described is only in that context. Another context could lead to another level or type of difficulty/productivity being expressed. We sought measurement instruments for each cell in this framework and for the contextual factors of importance.

Gathering Filter Evidence

Our work has summarized and followed new measurement-related evidence for these scales in the literature, as well as conducted studies to create evidence to fill gaps in the OMERACT Filter 2.0[3,23,24]. We present both the methods used to update the literature, and the studies conducted to complete the OMERACT Filter 2.0 evidence.

Review of the Literature and Update of Evidence Tables

Every 2 years we conducted an update of our systematic literature review of psychometric evidence of worker productivity outcome instruments in arthritis or musculoskeletal populations. The most recent update was in December 2013. All studies were obtained through reviews of key references for each instrument (citation searches and database searches). Measurement studies were then sought through a selection phase carried out by a single trained observer. Relevant studies were identified and reviewed by assigned leads for each instrument and their team. Biweekly teleconferences were used to share updates and decide how the evidence should be presented in evidence tables.

New Studies Completed to Fill in Gaps in Evidence Needed According to OMERACT Filter 2.0

Two independent studies were conducted to add to this body of evidence. First, we conducted a study to complete our understanding of both the patient acceptable states (PAS) in worker productivity, and the minimal (clinically) important difference (MID), as well as boundaries of measurement error. (We call this the MID/PAS study.) Second, we conducted a multicountry cognitive debriefing study, which assessed the meaning of the responses to the candidate measures from across international patient groups. In this study we also fielded additional items to allow for international testing of construct validity and test-retest reliability of these scales (we call this the international cognitive debriefing study). These studies were conducted for the purpose of OMERACT and are integrated into the evidence synthesis below.

Testing the Preliminary Criteria to Identify "Relevant" Contextual Factors in Clinical Studies

A systematic review was performed, exploring the role of contextual factors either on presenteeism, sick leave, or work disability in patients with ankylosing spondylitis (AS), in which the proposed criteria to assess the relevance of contextual factors and to summarize evidence across studies were applied.

RESULTS AND DISCUSSION

Description of Candidate Measures

Moving into OMERACT 12 (2014), we had a mandate to move forward with 6 candidate measures, which are summarized in Table 1 along with their acronyms. Four were single-item instruments, 1 with a difficulty focus (WAI), and 3 oriented more toward a concept of level of productivity (production, efficiency) in their indicators of at-work productivity loss (WPS, WPAI, QQ). Two multiitem indices were tracked, the WLQ-25 PDmod, with the modification to the physical demands subscale to reorient it in the same direction as the other subscale, and the WALS, a more difficulty-oriented scale.

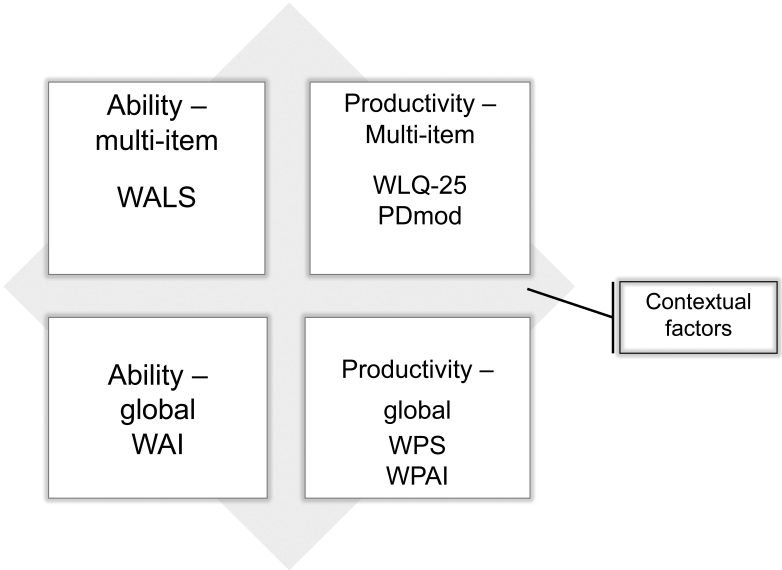

Figure 1. Organization into a taxonomy of 4 different types of work productivity measures. WALS: Workplace Activity Limitations Scale; WLQ-25 PDmod: Work Limitations Questionnaire with modified physical demands scale; WAI: Work Ability Index; WPS: Arthritis-specific Work Poductivity Survey; WPAI: Work Productivity and Activity Impairment Questionnaire.

Table 1. Characteristics of the 6 candidate measures of at-work limitations/productivity (presenteeism). Adapted from Tang, *et al*. J Rheumatol 2014;41:165-76[3].

Global Measures	Content Source	Concept	Recall Period	Disease Attribution	Comparative Referencing	Scaling
WAI	Item 1 from the Work Ability Index	Work ability	Current	None	In relation to lifetime best	0–10 (0 = completely unable to work; assume best work ability = 10)
QQ	Multiplication of 2 items	How much work performed and the quality of the work	Last workday	None (N/A)	Compared to a normal "workday"	Quantity item: 0–10 (practically nothing to normal quantity); Quality item: 0–10 (very poor to normal quality)
WPAI	Item 5 from the WPAI Questionnaire (specific health problem version)	Work productivity	Last 7 days	Can be adapted to any health condition	None	0–10 (health problem had no effect on my work to completely prevented me from working)
WPS	Item 4 from the WPS survey	Interference with work productivity	Last month	Arthritis	None	0–10 (no interference to complete interference)
Multiitem Measures	Concept		Scored Scales and Number of Items			Time Frame
WALS	Amount/level of difficulty		Single scale — summed score of 12 items			Not specified
WLQ-25 PDmod	Frequency/proportion of time having difficulty		25 items across 4 work demands: physical demands; mental-interpersonal; time management; output demands			Past 2 weeks

WPAI: Work Productivity and Activity Impairment Questionnaire; WPS: Arthritis-specific Work Productivity Survey; QQ: Quantity and Quality Method; WAI: Work Ability Index; WALS: Workplace Activity Limitations Scale; WLQ-25 PDmod: Work Limitations Questionnaire with modified physical demands scale.

OMERACT Filter Evidence

A full description of the evidence from the literature can be found in Supplementary Table 1 (use of instruments in clinical trials) and Supplementary Table 2 (accumulated filter evidence; both available online at jrheum.org); these results are summarized below by component of the OMERACT Filter.

Feasibility and Face/Content Validity (Truth)

The summary of evidence shows that the 6 scales show feasibility of use (low burden, accessible, low frequencies of missing data; Supplementary Table 2, available online at jrheum.org).

In our comparison of the content validity of response options of 5 measures including the WALS and WLQ-25 in

workers with OA or RA[24], both measures showed good results, with support for feasibility criteria of the OMERACT Filter. In response to a forced-choice question regarding which of the 5 measures participants preferred overall, the WALS was ranked first (32.6% support) and the WLQ-25 second (30% support). For the WLQ-25, the reverse direction of instructions in the PD subscale was a source of confusion, but this issue was resolved with the modified WLQ-25, now called WLQ-25 PDmod with the agreement of the scale developer (personal communications with D. Lerner).

The summary of evidence (Supplementary Table 2, available online at jrheum.org) revealed that 5 of the 6 candidate scales had strong evidence of face/content validity, and the QQ had some evidence for these criteria.

The results from the international cognitive debriefing study examining interpretation of the questionnaires by country demonstrated some differences among Canada, France, Italy, the Netherlands, Romania, Sweden, and the UK. A finding common to all countries was an initial lack of association to the word "productivity" (WPAI), as many found it difficult to rate their productivity if their job did not involve the "production of products." However, specifications ("accomplished," "kind of work," "carefully as usual," and "amount of work") found in the more detailed instructions in the WPAI clarified the term and were consistently understood across countries (Table 2). "Interference" used inthe stem and anchor of the WPS-RA caused difficulties specifically for the Romanian participants, reflecting a lack of understanding of the term. Time frames for recall of productivity loss differed across the measurement instruments. Seventy percent of patients said that a 7-day recall period (WPAI) was an accurate recall representation of how their condition affects work productivity, while 58% reported a recall period of "last workday" (QQ) to be inaccurate. The phrase "compared to normal" reference (QQ) also caused difficulty because of the ambiguous and relative nature of the word "normal." Overall, 29% of patients said the WPAI was the most relevant to them, making it the most favored measure, while the WAI was the least favored, with 12% of votes.

Construct Validity (Truth)

In one example of construct validity, Pearson correlation results from our MID/PAS study demonstrated that the global measures WPS, WPAI, and QQ were good to very good in their correlation with the multiitem measures (WALS and WLQ-25). The exception was the WAI, which was moderately correlated with the multiitem measures. Table 3 depicts the individual correlations between measures.

Additional evidence of construct validity was available on each tool either from the literature or from our own primary studies[3,23] (Supplementary Table 2, available online at jrheum.org).

Table 2A. Test-retest reliability of candidate measures of at-work limitations/productivity (presenteeism) from the first of our 2 test-retest reliability studies: MID/PAS study. Adapted from Tang, *et al*. J Rheumatol 2014;41:165-76[3].

Instrument	Content Source	Study 1: MID/PAS Study n (stable)*	Mean Baseline Score (SD)	Mean 2-week Followup	ICC2,1
Global measures					
WPS (item 4)	Item 4 from the Rheumatoid Arthritis-specific Work Productivity Survey	35	3.7 (3.0)	3.7 (3.4)	0.87
WPAI (item 5)	Item 5 from the Work Productivity and Activity Impairment Questionnaire (specific health problem version)	34	3.9 (3.0)	3.4 (2.9)	0.84
QQ method	Multiplication of 2 items (E1 and E2) from the Productivity and Disease Questionnaire (ProDISQ)	33	61.2 (33.3)	57.6 (34.2)	0.77
WAI (item 1)	Item 1 from the Work Ability Index	6‡	7.5 (2.1)	6.7 (2.5)	0.80
Multiitem measures					
WALS	Workplace Activity Limitations Scale (12 items)	37	12.5 (7.3)	12.2 (8.0)	0.93
WLQ-25 PDmod†	Work Limitations Questionnaire-25; (4 subscales: TM = time management [5 items], PD = physical demands [6 items], MI = mental interpersonal [9 items], OD = output demands [5 items]); Index score = weighted sum of subscales	37	TM = 36.5 (28.2)	TM = 34.2 (26.5)	TM = 0.93
		37	PD = 47.1 (32.9)	PD = 42.6 (32.9)	PD = 0.95
		37	MI = 25.6 (23.4)	MI = 27.0 (24.6)	MI = 0.79
		34	OD = 34.4 (28.9)	OD = 32.5 (31.6)	OD = 0.86
		34	Index = 9.6 (7.6)	Index = 9.3 (7.8)	Index = 0.93

*No change on an external anchor fielded at 2-week followup (compared to when you completed the first questionnaire package, how would you rate your ability to do your usual work activities? 0 = much worse, 5 = no change, 10 = much better); ‡: The WAI was added to the study at a later date, allowing only small accrual. Results should be taken with caution; †WLQ-25 physical demands subscale (PD) was modified in this study — instruction for the PD subscale was reoriented to be consistent with the other subscales. WPAI: Work Productivity and Activity Impairment Questionnaire; WPS: Arthritis-specific Work Productivity Survey; QQ: Quantity and Quality Method; WAI: Work Ability Index; WALS: Workplace Activity Limitations Scale; WLQ-25 PDmod: Work Limitations Questionnaire with modified physical demands scale; ICC2,1: intraclass correlation coefficient, type 2,1.

Table 2B. Test-retest reliability of candidate measures of at-work limitations/productivity (presenteeism) from the second of our 2 test-retest reliability studies: international cognitive debriefing study. Adapted from Tang, *et al*. J Rheumatol 2014;41:165-76[3].

Instrument	Content Source	Study 2: International Cognitive Debriefing n	Mean Baseline Score (SD)	Mean 2-wk Followup Score (SD)	ICC2,1
Global measures					
WPS (item 4)	Item 4 from the Arthritis-specific Work Productivity Survey (WPS)	65	3.5 (2.6)	3.5 (2.3)	0.78
WPAI (item 5)	Item 5 from the Work Productivity and Activity Impairment Questionnaire (WPAI; specific health problem version)	65	3.4 (2.6)	3.5 (2.8)	0.74
QQ method	Multiplication of 2 items (E1 and E2) from the Productivity and Disease Questionnaire	64 baseline; 63 followup	67.8 (27.4)	65.0 (27.2)	0.74
WAI (item 1)	Item 1 from the Work Ability Index (WAI)	65	7.1 (2.1)	7.2 (2.1)	0.75

QQ: Quantity and Quality Method; ICC2,1: intraclass correlation coefficient, type 2,1.

Table 3. Pearson correlations between WALS/WLQ-25 and the global measures of at-work productivity loss from our MID/PAS study.

Scale	Mean Score, n (SD)	WLQ-25 PDmod	WALS	WPS	WPAI	QQ	WAI*
WLQ-25	9.41 (7.10)	1.00, 175	0.79, 173	0.75, 148	0.80, 146	–0.74, 145	–0.49**, 51
WALS	13.16 (7.32)		1.00, 184	0.73, 152	0.81, 149	–0.77, 148	0.54, 50
WPS	3.90 (3.12)			1.00, 154	0.88, 150	–0.78, 149	–0.54, 51
WPAI	3.66 (2.85)				1.00, 151	–0.79, 146	–0.56, 50
QQ	64.07 (32.04)					1.00	0.56, 51
WAI	6.48 (2.00)						1.00, 52

*It should be noted that the WAI was introduced to the survey later than the other measures, and therefore was not completed by all participants (n = 52 vs n = 184). This lesser sample might have affected the correlation. ** $p = 0.0003$; $p \leq 0.0001$ for all other correlations. WALS: Workplace Activity Limitations Scale; WLQ-25 PDmod: Work Limitations Questionnaire with modified physical demands scale; WPS: Arthritis-specific Work Productivity Survey; WPAI: Work Productivity and Activity Impairment Questionnaire; QQ: Quantity and Quality Method; WAI: Work Ability Index.

Discrimination

The largest gap in the summary of evidence was in the area of discrimination, which encompasses 4 main properties: reliability and internal consistency, responsiveness (within-group discrimination), use in randomized clinical trials (between-group discrimination), and score interpretability. It is these gaps in the criteria that the MID/PAS study and the international cognitive debriefing study were intended to fill.

Test-retest Reliability

The published test-retest results from the MID/PAS test-retest study[3] (for all candidate at-work productivity measures) showed moderate-to-high intraclass correlation coefficients (ICC; 0.77–0.93), which indicate good-to-excellent agreement between baseline and 2-week followup. Table 2 depicts individual ICC for each measure.

The international cognitive debriefing study repeated a test-retest reliability assessment and showed a moderate range of ICC (0.74–0.78), which indicates good agreement between baseline and 2-week followup. Table 2 also shows the ICC for each measure from that study.

Within-group Discrimination (longitudinal construct validity or responsiveness)

The summary of evidence revealed that the WPAI, WPS, WAI, WALS, and WLQ-25 have passed the responsiveness criteria, while evidence for responsiveness was provided for the QQ through the MID/PAS study where change in QQ correlated moderately with change in productivity over the past 2 weeks (rs = 0.60), and ability to do usual work (rs = 0.59). Area under the curve, often used to summarize responsiveness, against 8 anchors of ability/productivity, ranged from 0.62–0.90 (Supplementary Table 2, available online at jrheum.org).

Between-group Discrimination (application in RCT or cohorts with improved and not improved groups)

The OMERACT Filter requires evidence that the instrument can discriminate between 2 arms in randomized controlled trials (RCT). In the OMERACT Filter 2.0 revisions[25], this can also be tested with a lesser degree of confidence with discrimination between 2 groups using a single arm cohort, divided into subgroups of responders and nonresponders, and

comparing change distributions in the target instrument. This can be referred to as bronze level evidence, as it is provided through results of the MID/PAS study to address the criterion of between-group discrimination in the absence of current RCT evidence.

The summary of evidence (Supplementary Table 1 for a full description of trials fielding worker productivity instruments) revealed that 3 of the 4 global measures (WPAI, WPS, and QQ) have passed the between-group criteria of discrimination.

The WAI was used in 1 RCT, showing no difference between groups[26]. This provides some evidence that the WAI does not show change where none is expected. In absence of other RCT using the WAI to date, evidence for between-group discrimination was provided for the WAI through the MID/PAS study results. In that study, change scores were compared between people who had a positive change versus those who did not, according to 8 external anchors of change. A difference in effect size was calculated [standardized response mean (SRM) improved–SRM not improved] and differences were found in SRM of 0.5 to 1.46. This indicates much more change was detected in people who improved versus people who did not.

Similarly to the WAI, the between-group discrimination criterion for the WALS and WLQ-25 PDmod is provided by the MID/PAS study, in the absence of positive RCT (showing a difference between groups) in the literature. Differences of 0.4–1.7 in SRM were found, showing much more change in people who improved relative to those who did not. There were also negative trials (showing no difference between groups) that supported the ability to *not* change in the absence of effects. No other anchors were available in these studies to describe subgroups that may have responded. There are also trials underway with the WALS.

Thresholds of Meaning (interpretability)

The MID/PAS study provided rigorous calculations for thresholds of meaning for all candidate scales against multiple anchors validated at OMERACT 11 (2012)[3]. As an example, evidence for interpretability in the WPS was provided by the MID/PAS study, where several anchors were fielded for calculating PAS, with values ranging between 3 and 7 with a median of 5. MID were calculated for improvement and deterioration, and varied depending on the anchor used, ranging from 1–3 for improvement and 1–2 for deterioration. Minimum detectable change (95% CI) was calculated (3.10). Both the MDC-95 and the MID would need to be surpassed to be confident in interpreting change. Although some suggest MID should be greater than MDC-95, we hold that the opposite could be true, as long as the change score being interpreted as improvement was greater than both MID and boundaries of error. The thresholds for all measures are summarized in Supplementary Table 2, available online at jrheum.org.

Summary of OMERACT Filter Evidence

The synthesis of the accumulated and new information is presented for each candidate measure in Table 4. This table summarizes evidence only in this context of use, that is, in persons with arthritis or with other relevant musculoskeletal disorders.

Contextual Factors

When searching, appraising, and summarizing the literature on the role of contextual factors in AS and worker outcomes, we found 20 reports addressing employment status, 6 addressing sick leave, and 3 presenteeism. For employment, there was strong evidence for the role of age; moderate evidence for personal skills/abilities (such as coping), (absence of) work accommodations, the nature of work and (absence of) workplace support, and poor evidence for the role of marital status. Evidence was insufficient for sex, education, and physical environment. For sick leave and presenteeism there were too few studies to perform a best-evidence synthesis for the role of contextual factors. These results along with those reported in our previous OMERACT work[4] were presented at the SIG for discussion. Available evidence provides a limited view because this field is new, and may need to be supplemented by participant opinion until the evidence grows with additional contextual factors being assessed in conjunction with indicators of worker productivity.

Worker Productivity SIG and Plenary Vote Results from OMERACT 2014

In our worker productivity SIG session we provided a brief overview of our completed work and presented the OMERACT Filter evidence for the 6 candidate measures in a "speed dating" type of format where participants moved around stations where they heard of the OMERACT Filter evidence for each candidate measure in a dynamic, high-energy participatory process. A presentation on the contextual factors of work was also given that included summarizing evidence and proposing this as a research agenda item for future study.

At the final plenary session, the OMERACT attendees voted on whether there was sufficient evidence of the OMERACT Filter requirements for the 4 remaining worker productivity instruments (having already received endorsement at OMERACT 11 for both the WPS and the WPAI). As summarized in Table 5, all but the QQ received > 70% agreement, thus advancing the WLQ-25 PDmod, the WALS, and the WAI as now having enough OMERACT Filter evidence. Further, the agenda supporting ongoing research into contextual factors that directly affect the responses to a worker productivity instrument was strongly supported (95%), providing us with our work for the next 2 years.

More evidence has been gathered to support the

Table 4. Summary appraisal of OMERACT Filter evidence for the 4 global and 2 multiitem measures of at-work limitations/productivity. Updated from Tang, *et al*. J Rheumatol 2014;41:165-76[3] based on updated literature review (December 2013) and results of our MID/PAS and international cognitive debriefing studies.

	OMERACT						
	Truth			Discrimination			Feasibility
Global Measures	Face/Content Validity	Construct Validity	Reliability	Responsiveness	RCT	Score Interpretability	
WPAI (item 5)	++	++	++	++	++	+	++
WPS (item 4)	++	++	++	++	++	+	++
QQ	+	+	++	+	+	+	++
WAI (item 1)	++	++	++	++	(+)*	+	++
Multiitem measures							
WALS	++	++	++	++	(+)#	+	++
WLQ-25	++	++	++	++	(+)#	+	++

Rating criteria (based on overall appraisal of all available evidence in arthritis or MSK conditions): ++: evidence of this measurement property from 2 or more studies, in the absence of conflicting evidence; +: evidence of this measurement property from at least 1 study, and overall body of evidence supporting > refuting; *: RCT shows no difference where one is not expected; not exclusively MSK population; evidence of ability to discriminate between subgroups where one group is improved and the other not (MID/PAS study), which provides some evidence of their ability to discriminate change between groups. We will continue to monitor results of ongoing trials using this tool; #: trials ongoing, or in the case of the WLQ, trials were negative, and difference in WLQ was negative. Both WALS and WLQ have evidence of their ability to discriminate between subgroups where 1 group is improved and the other not, which provides some evidence of their ability to discriminate change between groups. We will continue to monitor results of ongoing trials using these tools. WPAI: Work Productivity and Activity Impairment Questionnaire; WPS: Arthritis-specific Work Productivity Survey; QQ: Quantity and Quality Method; WAI: Work Ability Index; WALS: Workplace Activity Limitations Scale; WLQ-25 PDmod: Work Limitations Questionnaire with modified physical demands scale; MID/PAS: minimal (clinically) important difference/patient acceptable states; RCT: randomized controlled trial; MSK: musculoskeletal.

Table 5. Plenary voting results at OMERACT 12*.

Voting Questions for Candidate Measures	Yes, % (n)	No, % (n)
In your opinion is there now sufficient evidence of truth, discrimination, and feasibility for the:		
1. WLQ-25 PDmod Work Limitation Questionnaire? (multiitem scale that is productivity-oriented)	91 (71)	9 (7)
2. WALS (Workplace Activity Limitations Scale)? (multiitem scale of difficulty/ability)	86 (78)	14 (13)
3. WAI (Work Ability Index)? (single item about ability)	74 (52)	26 (18)
4. QQ (Quality and Quantity)? (2 global items Qual*Quan > Productivity)	32 (23)	68 (50)
Question for contextual factors		
5. Do you agree that these contextual factors should be prioritized on a research agenda to explore if they are core contextual factors for worker productivity: (1) Nature of work; (2) Work accommodation; (3) Workplace support?	95 (95)	5 (5)

*"Don't know" votes were not required to be included in the calculations. WLQ-25 PDmod: Work Limitations Questionnaire with modified physical demands scale; QQ: Quantity and Quality Method.

measurement of worker productivity in arthritis research. At OMERACT 12 we received support (> 70% consensus) that the WLQ-25 PDmod, WALS, and WAI had enough OMERACT Filter evidence available. They have been added to the list along with the previously endorsed WPS and WPAI. Our work allows us to recommend these 5 evidence-based measures of at-work productivity loss for studies in arthritis. This year we also got strong endorsement for 3 contextual factor domains as being important in the interpretation and measurement of worker productivity. In our research moving forward: (1) We will shift our research focus to contextual factors; (2) the QQ will continue to be monitored for improved reliability and more evidence of construct validity; we will also monitor the use of all of these measures in clinical trials; and (3) we will use our ongoing cohort study (Phase II of the cognitive debriefing study) to further verify the validity of these instruments across different cultural boundaries.

We are cognizant of the changing nature of work itself. New emerging scales may capture the dominance of knowledge and computer-based jobs over manufacturing in developed and developing countries in particular. Scales need to be re-evaluated to ensure they are still capturing the current experience of work for people with arthritis. We will continue to watch for new scales, or for revalidation of the existing ones in these new work contexts.

ACKNOWLEDGMENT

We acknowledge M. Cifaldi, RPh, MSHA, PhD, Global Lead Rheumatology, PPG, Global Health Economics and Outcomes Research, Abbott Laboratories, Abbott Park, Illinois, USA, for contribution to the data and interpretation, as part of the worker productivity working group until late 2013; Taucha Inrig, Musculoskeletal Health and Outcomes Research, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Ontario, Canada, for analysis for the MID/PAS study; and Elaine Harniman, Musculoskeletal Health and Outcomes Research, Li Ka Shing Knowledge Institute, St. Michael's Hospital, for analysis for MID/PAS study.

ONLINE SUPPLEMENT

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

REFERENCES

1. Beaton D, Bombardier C, Escorpizo R, Zhang W, Lacaille D, Boonen A, et al. Measuring worker productivity: frameworks and measures. J Rheumatol 2009;36:2100-9.
2. Escorpizo R, Bombardier C, Boonen A, Hazes JM, Lacaille D, Strand V, et al. Worker productivity outcome measures in arthritis. J Rheumatol 2007;34:1372-80.
3. Tang K, Boonen A, Verstappen S, Escorpizo R, Luime J, Lacaille D, et al. Worker productivity outcome measures: OMERACT Filter evidence and agenda for future research. J Rheumatol 2014; 41:165-76.
4. Tang K, Escorpizo R, Beaton DE, Bombardier C, Lacaille D, Zhang W, et al. Measuring the impact of arthritis on worker productivity: perspectives, methodologic issues, and contextual factors. J Rheumatol 2011;38:1776-90.
5. Boers M, Kirwan JR, Wells G, Beaton DE, Gossec L, D'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT Filter 2.0. J Clin Epidemiol 2014;67:745-53.
6. Gignac MA, Badley EM, LaCaille D, Cott CC, Adam P, Anis AH. Managing arthritis and employment: making arthritis-related work changes as a means of adaptation. Arthritis Rheum 2004;51:909-16.
7. Lerner D, Reed JI, Massaroti E, Wester LM, Burke TA. The Work Limitations Questionnaire's validity and reliability among patients with osteoarthritis. J Clin Epidemiol 2002;55:197-208.
8. Mancuso CA, Paget SA, Charlson ME. Adaptations made by rheumatoid arthritis patients to continue working: A pilot study of workplace challenges and successful adaptations. Arthritis Care Res 2000;13:89-99.
9. Reisine S, Fifield J, Walsh SJ, Feinn R. Factors associated with continued employment among patients with rheumatoid arthritis: a survival model. J Rheumatol 2001;28:2400-8.
10. Fex E, Larsson B-M, Nived K, Eberhardt K. Effect of rheumatoid arthritis on work status and social leisure time activities in patients followed 8 years from onset. J Rheumatol 1998;25:44-50.
11. Tuomi K, Ilmarinen J, Jakhola A, Katajrinne L, Tulkki A. Work Ability Index. Helsinki: Finnish Institute of Occupational Health; 1998.
12. Brouwer WB, Koopmanschap MA, Rutten FF. Productivity losses without absence: measurement validation and empirical evidence. Health Policy 1999;48:13-27.
13. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics 1993;4:353-65.
14. 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.
15. Kavanaugh A, Gladman D, van der Heijde D, Purcaru O, Mease P. Improvements in productivity at paid work and within the household, and increased participation in daily activities after 24 weeks of certolizumab pegol treatment of patients with psoriatic arthritis: results of a phase 3 double-blind randomised placebo-controlled study. Ann Rheum Dis 2015;74:44-51.
16. Osterhaus JT, Purcaru O. Discriminant validity, responsiveness and reliability of the arthritis-specific Work Productivity Survey assessing workplace and household productivity in patients with psoriatic arthritis. Arthritis Res Ther 2014;16:R140.
17. Gignac MA. Arthritis and employment: an examination of behavioral coping efforts to manage workplace activity limitations. Arthritis Rheum 2005;53:328–36.
18. Lerner D, Amick BC III, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. Med Care 2001;39:72-85.
19. Lofland JH, Pizzi L, Frick KD. A review of health-related workplace productivity loss instruments. Pharmacoeconomics 2004;22:165-84.
20. Prasad M, Wahlqvist P, Shikiar R, Shih Y-CT. A review of self-report instruments measuring health-related work productivity. A patient-reported outcomes perspective. Pharmacoeconomics 2004;22:225-44.
21. Amick BC III, Lerner D, Rogers WH, Rooney T, Katz JN. A review of health-related work outcome measures and their uses, and recommended measures. Spine 2000;25:3152-60.
22. World Health Organization. International classification of functioning, disability and health (ICF). Geneva: World Health Organization; 2001.
23. Beaton DE, Tang K, Gignace 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.
24. Tang K, Beaton DE, Lacaille D, Gignac MA, Bombardier C. Sensibility of five at-work productivity measures was endorsed by patients with osteoarthritis or rheumatoid arthritis. J Clin Epidemiol 2013;66:546-56.
25. Boers M, Kirwan JR, Tugwell PS, Beaton D, O. Bingham CO III, Conaghan PG. OMERACT Handbook. Ottawa: OMERACT; 2015. Available from: www.omeract.org/pdf/OMERACT_Handbook.pdf
26. Nurminen E, Malmivaara A, Ilmarinen J, Ylostalo P, Mutanen P, Ahonen G, et al. Effectiveness of a worksite exercise program with respect to perceived work ability and sick leaves among women with physical work. Scand J Work Environ Health 2002;28:85-93.