

Worker Productivity Outcome Measures: OMERACT Filter Evidence and Agenda for Future Research

Kenneth Tang, Annelies Boonen, Suzanne M.M. Verstappen, Reuben Escorpizo, Jolanda J. Luime, Diane Lacaille, Bruno Fautrel, Ailsa Bosworth, Mary Cifaldi, Monique A.M. Gignac, Cathy Hofstetter, Amye Leong, Pam Montie, Ingemar F. Peterson, Oana Purcaru, Claire Bombardier, Peter S. Tugwell, and Dorcas E. Beaton

ABSTRACT. The objective of the Outcome Measures in Rheumatology (OMERACT) Worker Productivity working group is to identify worker productivity outcome measures that meet the requirements of the OMERACT filter. At the OMERACT 11 Workshop, we focused on the at-work limitations/productivity component of worker productivity (i.e., presenteeism) — an area with diverse conceptualization and instrumentation approaches. Various approaches to quantify at-work limitations/productivity (e.g., single-item global and multi-item measures) were examined, and available evidence pertaining to OMERACT truth, discrimination, and feasibility were presented to conference participants. Four candidate global measures of presenteeism were put forth for a plenary vote to determine whether current evidence meets the OMERACT filter requirements. Presenteeism globals from the Work Productivity and Activity Impairment Questionnaire (72% support) and Rheumatoid Arthritis-specific Work Productivity Survey (71% support) were endorsed by conference participants; however, neither the presenteeism global item from the Health and Work Performance Questionnaire nor the Quantity and Quality method achieved the level of support required for endorsement at the present time. The plenary was also asked whether the central item from the Work Ability Index should also be considered as a candidate measure for potential endorsement in the future. Of participants at the plenary, 70% supported this presenteeism global measure. Progress was also made in other areas through discussions at individual breakout sessions. Topics examined include the merits of various multi-item measures of at-work limitations/productivity, methodological issues related to interpretability of outcome scores, and approaches to appraise and classify contextual factors of worker productivity. Feedback gathered from conference participants will inform the future research agenda of the working group. (J Rheumatol First Release Oct 15 2013; doi:10.3899/jrheum.130815)

Key Indexing Terms:

ABSENTEEISM

PRESENTEEISM

OSTEOARTHRITIS

RHEUMATOID ARTHRITIS

RELIABILITY AND VALIDITY

WORK PRODUCTIVITY

From the Institute of Health Policy, Management and Evaluation, University of Toronto; Mobility Program Clinical Research Unit, St. Michael's Hospital; and Institute for Work and Health, Toronto, Ontario, Canada; Department of Rheumatology, Maastricht University Medical Center, and CAPHRI Research Institute, Maastricht, Netherlands; Arthritis Research UK Epidemiology Unit, the University of Manchester; Manchester, UK; Department of Physical Therapy, School of Allied Health Professions, Louisiana State University, New Orleans, Louisiana, USA; Department of Health Sciences and Health Policy, University of Lucerne; Swiss Paraplegic Research, ICF Research Branch of World Health Organization (WHO) Collaborating Centre for the Family of International Classifications in Germany (DIMDI), Nottwil, Switzerland; Department of Rheumatology, Erasmus Medical Center, University Medical Center Rotterdam, Rotterdam, Netherlands; Division of Rheumatology, University of British Columbia; Arthritis Research Centre of Canada, Vancouver, British Columbia, Canada; Department of Rheumatology, Pierre et Marie Curie University (UPMC - Paris 6), APHP Pitié-Salpêtrière Hospital, Paris, France; National Rheumatoid Arthritis Society, UK; Global Lead Rheumatology, PPG, Global Health Economics and Outcomes Research, Abbott Laboratories, Abbott Park, IL, USA; Division of Health Care and Outcomes Research, Toronto Western Research Institute at the University Health Network; Arthritis Community Research and Evaluation Unit; Dalla Lana School of Public Health, University of Toronto, Institute for Work and Health, Toronto, ON, Canada; Consumer Advisory Council, Canadian Arthritis Network,

Canada; Healthy Motivation, Santa Barbara, California, USA; Consumer Advisory Board, Arthritis Research Centre of Canada, Canada; Epi-centrum Skåne; Department of Rheumatology, Skåne University Hospital; Department of Orthopaedics, Clinical Sciences, Lund University, Lund, Sweden; Health Economics, GMAP, Immunology, UCB Pharma, Bruxelles, Belgium; Division of Rheumatology, Faculty of Medicine and Institute of Health Policy, Management and Evaluation, University of Toronto; Institute for Work and Health; Toronto General Research Institute at the University Health Network; Mount Sinai Hospital, Toronto; Department of Epidemiology and Community Medicine, University of Ottawa; Centre for Global Health, Institute of Population Health, University of Ottawa; Ottawa Hospital, Ottawa; Mobility Program Clinical Research Unit, St Michael's Hospital; Institute for Work and Health; Institute of Health Policy, Management and Evaluation, Graduate Department of Rehabilitation Science and Department of Occupational Science and Occupational Therapy, University of Toronto, Toronto, ON, Canada.

K. Tang, MSc(PT), MSc, Institute of Health Policy, Management and Evaluation, University of Toronto; Mobility Program Clinical Research Unit, St. Michael's Hospital; and Institute for Work and Health; A. Boonen, MD, PhD, Assistant Professor of Rheumatology, Department of Rheumatology, Maastricht University Medical Center, and CAPHRI Research Institute; S.M.M. Verstappen, PhD, Research Fellow, Arthritis Research UK Epidemiology Unit, University of Manchester; R. Escorpizo, PT, DPT, MSc, Department of Physical Therapy, School of Allied Health

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2013. All rights reserved.

Professions, Louisiana State University; Department of Health Sciences and Health Policy, University of Lucerne; Swiss Paraplegic Research, ICF Research Branch of WHO Collaborating Centre for the Family of International Classifications in Germany; J.J. Luime, PhD, Department of Rheumatology, Erasmus Medical Center, University Medical Center Rotterdam; D. Laccaille, MD, MHSc, FRCPC, Mary Pack Chair in Arthritis Research, Division of Rheumatology, University of British Columbia; Arthritis Research Centre of Canada; B. Fautrel, MD, PhD, Department of Rheumatology, Pierre et Marie Curie University, APHP Pitié-Salpêtrière Hospital; A. Bosworth, National Rheumatoid Arthritis Society; M. Cifaldi, RPh, MSHA, PhD, Global Lead Rheumatology, PPG, Global Health Economics and Outcomes Research, Abbott Laboratories; M.A.M. Gignac, PhD, Division of Health Care and Outcomes Research, Toronto Western Research Institute at the University Health Network; Arthritis Community Research and Evaluation Unit; Dalla Lana School of Public Health, University of Toronto, Institute for Work and Health; C. Hofstetter, Consumer Advisory Council, Canadian Arthritis Network; A. Leong, MBA, Healthy Motivation; P. Montie, Consumer Advisory Board, Arthritis Research Centre of Canada; I.F. Petersson, MD, FRCPC, Epi-centrum Skåne; Department of Rheumatology, Skåne University Hospital; Department of Orthopaedics, Clinical Sciences, Lund University; O. Purcaru, PhD, Health Economics, GMAP, Immunology, UCB Pharma; C. Bombardier, MD, FRCPC, Division of Rheumatology, Faculty of Medicine and Institute of Health Policy, Management and Evaluation, University of Toronto; Institute for Work and Health; Toronto General Research Institute at the University Health Network; Mount Sinai Hospital; P.S. Tugwell, MD, Department of Epidemiology and Community Medicine, University of Ottawa; Centre for Global Health, Institute of Population Health, University of Ottawa; Ottawa Hospital; D.E. Beaton, PhD, Mobility Program Clinical Research Unit, St Michael's Hospital; Institute for Work and Health; Institute of Health Policy, Management and Evaluation, Graduate Department of Rehabilitation Science and Department of Occupational Science and Occupational Therapy, University of Toronto.

Supported by research grants from the Canadian Arthritis Network (part of the Networks of Centres of Excellence) and the European League Against Rheumatology; by an unrestricted grant from Abbott and by funding provided by OMERACT. K. Tang is supported by a Canadian Institutes of Health Research Fellowship, a Canadian Arthritis Network Graduate Award, and a Syme Fellowship from the Institute for Work and Health. D. Laccaille is supported by an Investigator Award from The Arthritis Society of Canada and is the Nancy and Peter Paul Saunders Scholar. 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, Rheumatology Division. D.E. Beaton is supported by a Canadian Institutes of Health Research New Investigators Award.

Address correspondence to Dr. Beaton, Mobility Program and Clinical Research Unit, St. Michael's Hospital, 30 Bond Street, Toronto, Ontario M5B 1W8, Canada. E-mail: beatond@smh.ca

Since 2006, the Outcome Measures of Rheumatology (OMERACT) Worker Productivity Working Group has sought to reach consensus on work-related outcomes for use in arthritis and rheumatic conditions^{1,2,3}. What makes work particularly interesting is that there are many differing but equally important perspectives to consider when evaluating outcome measures. For patients, work fulfills a valued social role and can cultivate a sense of occupational identity, thus the ability to successfully meet job demands/obligations, sustain employment, and maintain work-life balance are examples of important issues^{4,5,6}. On the other hand, quantifying the effects of productivity loss (e.g., economic evaluations) is also of high relevance and interest for those involved such as employers, insurers, health economists, and policymakers. Currently, there are diverse approaches to

quantify health-related work disability/productivity, and our group has grappled with a number of key issues, including conceptualization (e.g., work productivity vs performance vs ability), instrumentation (e.g., summative measures vs modular measures with multiple components), as well as the issue of the work “context,” recognizing that beyond worker well-being, other factors (e.g., social support, job-related factors) can also influence worker productivity, which may have important implications for research design and data interpretation³.

Much progress has been made by our working group in recent years toward the goal of identifying worker productivity outcome measures that meet the OMERACT filter of truth, discrimination, and feasibility⁷. At OMERACT 8, we reviewed available evidence on existing measures of various aspects of work absenteeism and at-work limitations/productivity loss (i.e., presenteeism)². At OMERACT 9, we examined a theoretical framework that explored the inter-relationship between work absenteeism and presenteeism⁸, discussed findings from a head-to-head comparison of 5 multi-item worker productivity measures in arthritis⁹, and through a plenary vote, endorsed an expanded definition of work absenteeism as well as a short-list of candidate measures for further considerations¹. At OMERACT 10, our working group explored additional frameworks and methodological issues related to the application of worker productivity outcome measures, and discussed relevant contextual factors of worker productivity³.

Progress Leading up to OMERACT 11

Leading up to OMERACT 11, our focus has been the presenteeism component of worker productivity and the appraisal of relevant instruments such as patient-oriented outcome measures for use in rheumatology research. With growing collaborations within our working group and a more focused set of measures under consideration for endorsement, we have decided that our process would benefit from an organization into multiple work-streams to pursue several related topics. Currently, we have 4 active work-streams (Table 1): the first is focused on evaluating and comparing global approaches to measure at-work limitations/productivity loss (e.g., single-item measures); the second is focused on multi-item approaches to measure at-work limitations/productivity loss (e.g., summative scales); the third is focused on contextual factors of worker productivity; and finally, the fourth work-stream is focused on interpretability of worker productivity outcomes, which entails examination of scale cut-scores associated with patient acceptable state (PAS)^{10,11} and score changes that would represent a minimally important difference (MID)^{12,13}. Progress specifically pertaining to each of these 4 work-stream and current research agendas will be discussed in the following sections.

In preparation for OMERACT 11, our working group

Table 1. Current work-streams of the OMERACT Worker Productivity Working Group.

Work-stream 1	Global approaches to measure at-work limitations/productivity loss (e.g., single-item measures) Objective: to appraise available measurement evidence on global measures of presenteeism against the requirements of the OMERACT filter
Work-stream 2	Multi-item approaches to measure at-work limitations/productivity loss Objective: to appraise available measurement evidence on multi-item measures of presenteeism against the requirements of the OMERACT filter
Work-stream 3	Contextual factors of worker productivity Objective: to develop criteria for identifying contextual factors of worker productivity and guidelines to consider contextual factors in the application of worker productivity outcome measures
Work-stream 4	Interpretability of worker productivity outcome measures Objective: to estimate cut-scores associated with the Patient Acceptable State (PAS) and score changes associated with the Minimally Important Difference (MID) of worker productivity outcome measures

engaged in several initiatives to help address gaps in OMERACT filter evidence among contending presenteeism measures, particularly in the area of discrimination (i.e., reliability, responsiveness). We updated our literature review of available psychometric evidence, including randomized controlled trials (RCT, pharmacological or nonpharmacological) in arthritis or musculoskeletal populations where worker productivity measures have been applied as a study outcome. Results from these trials informed whether the presenteeism component of an instrument can afford that fine level of discrimination required to evaluate change scores in 1 treatment arm over another treatment/control arm (i.e., between-group differences of within-group change). Details of these RCT are summarized in Supplementary Table 1 available online at jrheum.org. In addition, through funding support from a Discovery Advancement Project research grant from the Canadian Arthritis Network (CAN), we initiated an international Web-based survey recruiting workers with arthritis to examine the test-retest reliability (reproducibility of scores) and score interpretability (PAS/MID estimation) of candidate measures of at-work limitations/productivity. Preliminary results on the test-retest reliability of presenteeism measures revealed a moderate-to-high range of intraclass correlation coefficients ($ICC_{2,1}$)¹⁴, suggesting good to very good agreement between test-retest scores (Table 2). Tests of agreement were conducted among participants who were considered stable in their work ability over a 2-week period, based on having indicated “no change” on an 11-point external anchor (0 = much worse, 5 = no change, 10 = much improved) assessed at 2-week followup.

Objectives and Proceedings of the OMERACT 11 Workshop

At the OMERACT 11 Workshop, our overall aim was to appraise whether various candidate global measures of at-work limitations/productivity currently meet the requirements of the OMERACT filter, through a formal plenary

vote. At the 2-h workshop, we initially provided background presentations to review our working group’s progress from previous OMERACT meetings, and a current update of OMERACT filter evidence among our candidate instruments. We then invited 2 of our consumer partners to speak and shed light on the relevance and significance of employment issues from their own personal experience. This was followed by 8 individual breakout sessions led by our working group members. During the first part of all individual breakout sessions, we focused on a common topic — a discussion of the content and available psychometric evidence for the various global measures of at-work limitations/productivity — to allow participants to compare and contrast the measures in preparation of the plenary vote. The second part of the breakout sessions discussed unique topics that corresponded to our other work-streams: 3 of the individual breakout sessions discussed global measures of presenteeism in more detail, 2 of the sessions discussed the current evidence on multi-item measures and score interpretability issues, and the 3 remaining sessions discussed contextual factors of worker productivity. Finally, at the end of the individual breakout sessions, all conference participants reconvened for a plenary voting session where a recap of OMERACT filter evidence was presented, and then specific questions were posed to participants to formally evaluate whether our candidate presenteeism global measures currently meets the requirements of the OMERACT filter.

PROGRESS ON GLOBAL MEASURES OF AT-WORK LIMITATIONS/PRODUCTIVITY

Work-stream 1

Global measures are designed to ask about “on-the-job” problems or at-work productivity (loss) using a single or just a few items. Typically, these items ask fairly broad questions and are intended to provide an overall appraisal or *gestalt* of perceived limitations or productivity at work. Although some global measures can be found in larger, modular-type

Table 2. Test-retest reliability of candidate measures of at-work limitations/productivity (presenteeism).

Instrument	Content Source	n (stable)*	Mean Baseline Score (SD)	Mean 2-wk Followup Score (SD)	ICC _{2,1}
Global measures					
WPS-RA (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
HPQ (item A12) [§]	Item A12 from the Health and Work Performance Questionnaire	36	7.1 (2.0)	7.4 (2.1)	0.83
WAI (item 1)	Item 1 from the Work Ability Index	6 [‡]	7.5 (2.1)	6.7 (2.5)	0.80
Multi-item measures					
WALS	Workplace Activity Limitations Scale (12 items)	37	12.5 (7.3)	12.2 (8.0)	0.93
WLQ-25 (adapted) ^{†§}	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 37 37 34 34	TM = 36.5 (28.2); PD = 47.1 (32.9); MI = 25.6 (23.4); OD = 34.4 (28.9); Index = 9.6 (7.6)	TM = 34.2 (26.5); PD = 42.6 (32.9); MI = 27.0 (24.6); OD = 32.5 (31.6); Index = 9.3 (7.8)	TM = 0.93; PD = 0.95; MI = 0.79; OD = 0.86; Index = 0.93

* No change on an external anchor fielded at 2-wk followup (compared to when you completed the first questionnaire package, how would you rate your ability to do your usually work activities? 0 = much worse, 5 = no change, 10 = much better); † WLQ-25 was modified in this study — instruction for the physical demands subscale was reoriented to be consistent with other subscales; ‡ low sample size because measure was introduced later in the survey. § instrument has potential for use in economic evaluations. ICC: intraclass correlation coefficients.

questionnaires that also concurrently assess other aspects of work productivity (such as work absenteeism or loss of household productivity), it is important to be able to demonstrate that the presenteeism component, in and of itself, is meeting the requirements of the OMERACT filter. Moreover, this aspect of worker productivity is often

reported as a discrete outcome in studies, as many researchers are interested in understanding whether interventions can influence both work absenteeism and presenteeism. Currently, 5 global measures of at-work limitations/productivity are under consideration (Table 3), 3 of which are based on voting from OMERACT 9. These

Table 3. Characteristics of the 5 candidate global measures of at-work limitations/productivity (presenteeism).

Global Measure	Content Source	Intended Measurement Concept	Recall Period	Disease Attribution	Comparative Referencing	Scaling
WPS-RA (item 4) [§]	Item 4 from the Rheumatoid Arthritis-Specific Work Productivity Survey	Interference with work productivity	Last month	Arthritis	None	0–10 (no interference to complete interference)
WPAI (item 5) [§]	Item 5 from the Work Productivity and Activity Impairment Questionnaire (specific health problem version)	Work productivity	Last 7 days	Can be adapted to any health condition (e.g., RA)	None	0–10 (health problem had no effect on my work to completely prevented me from working)
QQ method [§]	Multiplication of 2 items (E1 and E2) from the Productivity and Disease Questionnaire (ProDISQ)	How much work performed and quality of the work	Last workday	None (not applicable)	Compared to “a normal work-day”	Quantity item: 0–10 (practically nothing to normal quantity); Quality item: 0–10 (very poor quality to normal quality)
HPQ (item A12) [§]	Item A12 from the Health and Work Performance Questionnaire	Work performance	2 versions available: last 7 or 28 days	None	Performance of other workers (using HPQ item A10)*	0–10 (worse performance to top performance)*
WAI (item 1)	Item 1 from the Work Ability Index	Work ability	Current	None	In relation to lifetime best	0–10 (0 = completely unable to work; assume work ability at its best has a value of 10)

* Can be scored with or without comparative referencing: absolute presenteeism (item A2 multiplied by 10) or relative presenteeism (ratio of A12/A10)¹⁵. § Instrument has potential for use in economic evaluations.

include the presenteeism component of the Rheumatoid Arthritis-specific Work Productivity Survey (WPS-RA)¹⁵, Work Productivity and Activity Impairment questionnaire (WPAI)¹⁶, and the World Health Organization (WHO) Health and Work Performance Questionnaire (HPQ)^{17,18}. In addition, 2 other measures were added to address the need for measures of work ability/difficulty: the Quantity and Quality Method (QQ)¹⁹, which was favored by experts; and lastly, the central (first) item of the Work Ability Index (WAI)²⁰, which has emerged in recent years as a popular work outcome measure, particularly within the European community. All 5 global presenteeism measures use a numeric rating scale (Table 3), and all except the WAI are considered to have potential for use in economic evaluations.

At individual breakout sessions at our OMERACT 11 Workshop, we reviewed the content of the global measures and highlighted 4 key aspects where these measures exhibit some important differences. The first is the intended measurement concept, which is reflected by the different terms used to describe work issues. Global items from the WPAI and WPS-RA assess work “productivity,” the HPQ evaluates work “performance,” the QQ method focuses on “how much” (i.e., quantity) and “quality” of the work performed, while the WAI asks for a rating of work “ability,” offering a perspective that is distinct from productivity. The second element is the length of the recall period, which varies from “current” to “last month” among our 5 candidate instruments. A third element is whether there is disease attribution; that is, whether the item(s) asks respondents to consider work problems insofar as they are deemed to be directly attributed to a specific or non-specific health problem. Only the WPAI (specific health problem version) and WPS-RA are disease-specific, while all other candidates are formulated without any disease/health attributions. Fourth and finally, the measures also differ in terms of comparative referencing. This concerns whether respondents are asked to provide a rating against a “reference” situation/worker. Only the QQ method and global items from the WAI and HPQ apply a type of comparative referencing. The QQ method asks respondents to compare against “a normal work-day,” while WAI asks respondents to provide a rating “in relation to lifetime best.” As per developer recommendation²¹, the presenteeism global from HPQ can be scored 2 ways: as a ratio against another item that assesses the usual performance of most workers (“relative presenteeism”), or simply as a standalone single-item (i.e., no comparative referencing).

OMERACT Filter Evidence

Summary ratings for the various aspects of the OMERACT filter for each of the 5 candidate global measures of at-work limitations/productivity in arthritis or musculoskeletal populations are provided in Table 4. These ratings reflect an up-to-date synthesis of available and relevant measurement

evidence for individual measures (full details in Supplementary Table 2 available online at jrheum.org), which were agreed upon among working group members. Overall, our current appraisal suggests that all 5 presenteeism global measures satisfy both truth and feasibility criteria, and it is in the discrimination criterion where there are some differences. Currently, only the presenteeism global measures from the WPAI and WPS-RA are considered to have fully satisfied the discrimination criterion, given the availability of supporting evidence for test-retest reliability, responsiveness, RCT (demonstrated differences between intervention groups where efficacy would be expected), and preliminary findings on PAS/MID estimates presented at OMERACT 11 (unpublished results). On the other hand, evidence for responsiveness currently remains lacking for the HPQ global and QQ method, while supporting evidence of between-group discrimination from RCT is also not yet available for presenteeism global measures from the HPQ or WAI.

Plenary Voting at the OMERACT 11 Workshop

At the plenary voting session, we asked conference participants to vote on the following question for each of the 4 presenteeism global measures from the WPAI, WPS-RA, HPQ, and the QQ method: Do you think this measure has enough evidence to pass the OMERACT filter (response options: yes, no, don't know)? Both the WPAI (yes = 72%) and WPS-RA (yes = 71%) met the minimum 70% threshold required for endorsement. Both the HPQ (yes = 8%) and QQ method (yes = 13%) were well below the level of support required for endorsement, which was anticipated considering remaining gaps in discrimination evidence from our reviews. Because the WAI was only recently identified for OMERACT consideration and because it assesses a different perspective (i.e., work ability), for this measure, we specifically asked: Do you agree that the (presenteeism global measure from) WAI should be considered as a candidate measure of ability/difficulty at work for potential OMERACT endorsement, pending demonstration of filter evidence (response options; yes, no, don't know)? Among conference participants, 70% supported this global measure as a candidate instrument; therefore it met the level of support required. The full voting results are presented in Table 5, and it should be noted that calculations of level of endorsement were based on proportions of “yes” or “no” votes and excluded “don't know” votes, as per OMERACT guidelines. Among the questions posed, we found that proportions of “don't know” votes were generally small (6–17% of all votes).

Feedback from breakout sessions provided directions for future research for this work-stream. Conference participants suggested that the availability of evidence that compared productivity-oriented presenteeism measures to objective indicators of on-the-job productivity (e.g., quanti-

Table 4. Summary appraisal of OMERACT filter evidence for the 5 global measures of at-work limitations/productivity.

Global Measures	OMERACT Truth		OMERACT Discrimination			Score Interpretability	OMERACT Feasibility
	Face/Content Validity	Construct Validity	Reliability	Responsiveness	RCT		
WPAI (item 5)	++	++	+	++	++	+	++
WPS-RA (item 4)	+	+	+	++	++	(+)	++
QQ	+	+	+	?	+	+	++
HPQ (item A12)	++	++	+	?	?	(+)	++
WAI (item 1)	++	++	+	++	?	(+)	++

Rating criteria (based on overall appraisal of all available evidence in arthritis or musculoskeletal conditions): For construct validity, reliability, and responsiveness: ++: 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; +/-: Conflicting evidence available, current evidence does not meet ++ or +; -: Only negative evidence available; ?: No evidence available. For RCT: ++: 2 or more RCT providing evidence of discrimination (required for OMERACT endorsement); +: 1 RCT providing evidence of discrimination; (+): Ongoing RCT using worker productivity outcome measure (protocol available); -: Only evidence of lack of discrimination available; ?: No RCT available. Note: Discrimination includes between-group differences of within-group change, or between-group differences of final state. For score interpretability: ++: Estimates available from 2 or more studies; +: Estimates available from 1 study; (+): Estimates expected from an ongoing study; ?: Not yet available. RCT: randomized controlled trials.

Table 5. Plenary voting results at OMERACT 11 Workshop.

Voting question	Votes, no.	Endorsement ⁺ , %
#1. Do you think the (presenteeism global item from) WPS-RA has enough evidence to pass the OMERACT filter?		71
Yes	61	
No	25	
Don't know	17	
Total	103	
#2. Do you think the (presenteeism global item from) WPAI has enough evidence to pass the OMERACT filter?		72
Yes	66	
No	26	
Don't know	6	
Total	98	
#3. Do you think the QQ method has enough evidence to pass the OMERACT filter?		13
Yes	12	
No	80	
Don't know	8	
Total	100	
#4. Do you think the (presenteeism global item from) HPQ has enough evidence to pass the OMERACT filter?		8
Yes	7	
No	83	
Don't know	6	
Total	96	
#5. Do you agree that the (global/central item from) WAI should be considered as a candidate measure of "workability/difficulty at work" for potential OMERACT endorsement, pending demonstration of filter evidence?		70
Yes	63	
No	27	
Don't know	13	
Total	103	

* Calculations of % endorsement were based on participants who offered an opinion and exclude "don't know" vote, as per OMERACT guidelines.

fying actual work output over a defined period) could provide more definitive and convincing evidence of construct validity. Some concerns were raised about appraising only the presenteeism component of full questionnaires, although participants also understood the need to validate components of worker productivity

independently if these are to be presented as discrete outcomes in studies and trials. Some participants remarked on the complexity of the concept of worker productivity, and the potential difficulties of answering broader global items without a better understanding of what is specifically implied by the terms "work productivity" or "work

Table 6. OMERACT filter evidence for candidate multi-item measures of at-work limitations/productivity.

Multi-item Measure	OMERACT Truth		OMERACT Discrimination			Score Interpretability	OMERACT Feasibility
	Face/Content Validity	Construct Validity	Reliability	Responsiveness	RCT		
WALS	++	++	++	+	?	(+)	++
WLQ-25	++	++	++	++	-	+	++

WALS: Workplace Activity Limitations Scale; WLQ-25: Work Limitations Questionnaire. Rating criteria (based on overall appraisal of all available evidence in arthritis or musculoskeletal conditions). For construct validity, reliability, and responsiveness: ++: 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; +/-: Conflicting evidence available, current evidence does not meet ++ or +; -: Only negative evidence available; ?: No evidence available. For RCT: ++: 2 or more RCT providing evidence of discrimination (required for OMERACT endorsement); +: 1 RCT providing evidence of discrimination; (+): Ongoing RCT using worker productivity outcome measure (protocol available); -: Only evidence of lack of discrimination available; ?: No RCT available. Note: Discrimination includes between-group differences of within-group change, or between-group differences of final state. For score interpretability: ++: Estimates available from 2 or more studies; +: Estimates available from 1 study; (+): Estimates expected from an ongoing study; ?: Not yet available. RCT: randomized controlled trials.

performance” or “work ability.” To this end, it was also suggested that clear definitions of such items would be helpful for the respondent. This was thought to be an important contrast against items found in summative multi-item scales, which are typically more specific and thus may be advantageous in terms of facilitating more consistent interpretations of content across respondents. Finally, the need for more head-to-head comparison of candidate global measures to gain further insights into the relative strengths and limitations of individual measures was also suggested.

Research Agenda for Work-stream 1

The research agenda for this work-stream will focus on consolidating evidence of test-retest reliability among our candidate presenteeism global measures through our ongoing project funded by the Canadian Arthritis Network. We are also poised to complete a research project funded by the European League Against Rheumatism (EULAR), in which the comparability of the 5 candidate presenteeism global measures are investigated. The first phase of the EULAR-funded project will involve cognitive interviewing with patients with arthritis to explore their interpretation and understanding of the 5 global measures, with specific focus on the aforementioned 4 key elements (e.g., intended measurement concept, recall period) where important differences exist among the tools. Patients to be recruited for this project will represent workers with different arthritic/rheumatic conditions [e.g., rheumatoid arthritis (RA), osteoarthritis (OA), ankylosing spondylitis] from various occupational sectors (e.g., manual vs non-manual jobs) and geographic regions in Europe and Canada to explore the generalizability of findings. Results from the first phase will be applied to the second phase of the project, which will investigate the ideal recall period and construct validity of the candidate global presenteeism measures.

PROGRESS ON MULTI-ITEM MEASURES OF AT-WORK LIMITATIONS/PRODUCTIVITY

Work-stream 2

Multi-item measures are designed to assess at-work limitations/productivity through a series of summative items. These are the traditional “psychometric scales” that aim to measure an underlying unobserved (or latent) construct(s). At the OMERACT 9 plenary voting session, 3 multi-item measures received support for continued consideration: the 12-item Workplace Activity Limitations Scale (WALS)⁴, 23-item Work Instability Scale for Rheumatoid Arthritis (RA-WIS)²², and the family of Work Limitations Questionnaire (WLQ-25, WLQ-16, WLQ-8, Work Role Functioning-26/27)^{23,24}, which uniquely can be applied for economic evaluations. Given our focus on instruments specifically aimed at measuring either “worker productivity” or “worker ability/difficulty,” the RA-WIS is no longer considered a candidate outcome measure because its intended concept of work instability represents a different work concept. However, given its rigorous scale development process²², strong measurement properties in both RA and OA^{9,22,25,26}, and ability to predict future adverse work outcomes²⁷, we believe the RA-WIS has a unique role as a promising prognostic tool for use in arthritis/rheumatic conditions, and may also be an appropriate outcome in studies where perceived risk of job loss (precariousness of employment) is a primary target of interest.

Multi-item presenteeism measures were a topic of discussion at 2 individual breakout sessions at the OMERACT 11 Workshop. As was the case for global measures, we similarly recognized that the discrimination criterion remains an area with some gaps, in particular, the paucity of RCT using either the WALS or the WLQ-25 that provides evidence of between-group discrimination for these instruments (Table 6). Currently, we have found 5 published or ongoing RCT that used the WLQ-25 (or a variation of the measure), and only 1 has shown a significant

difference between treatment arms²⁸. However, this study did not directly measure WLQ-25 scores; instead, scores were imputed from other outcome measures assessed in the trial. Thus, evidence from this study may have limitations in terms of informing the discrimination of the WLQ-25. To our knowledge, no published RCT applying the WALS are available to date, although other types of discrimination evidence (e.g., responsiveness) have been demonstrated from previous psychometric studies (see Appendix 1 and 2).

Direct Comparability of Multi-item Measures

The direct comparability of these measures is a topic of interest, and head-to-head studies using a controlled sample can offer a useful approach to assess the relative strengths and limitations of candidate measures. One perspective of particular interest for direct comparison is the content validity of multi-item measures from patients' viewpoints, because the WALS and WLQ-25 quantify work problems from differing perspectives. The WALS assesses the degree of workplace activity limitations, while the WLQ-25 is aimed at quantifying the proportion of time a worker experiences work limitations. In a recent study, we directly compared the "sensitivity" (includes content validity, feasibility, understandability, suitability of response options) of 5 multi-item measures (including these 2 contenders) from the perspective of workers with either OA or RA²⁹. Overall, both the WALS and WLQ-25 performed well, providing further support for truth (face/content validity) and feasibility aspects of the OMERACT filter. In the study's final appraisal, patients were asked to provide a forced-choice selection from among the 5 measures that they preferred overall. The WALS (32.6% support) and WLQ-25 (30.0% support) were ranked first and second, respectively.

Physical Demands Subscale of the WLQ-25

The WLQ-25 has 4 subscales — time management, physical demands, mental-interpersonal demands, and output demands. While the same set of response keys is provided for all 25 items (none of the time to all of the time), in the physical demands subscale the instruction uniquely asks respondents to quantify the amount of time "without difficulty," whereas the other 3 subscales ask respondents to quantify the amount of time "with difficulty." Feedback from patient respondents has indicated that they found the reversal of wording in the physical demands subscale to be difficult. Respondents missing this subtle but critical change in instruction could represent a potential source of measurement error, which may explain some of the lower-than-expected correlations between WLQ-25 subscales observed in several studies (i.e., $r < 0.4$)^{9,30,31}. This has led us to adapt this subscale so that there are consistent instructions across the 4 subscales, and preliminary results from our ongoing CAN-funded study have shown that the intersubscale correlations are behaving in a

manner thought to be more consistent with conceptual expectations ($r = 0.71-0.87$). Because the WLQ-25 is a copyrighted instrument (Glaxo Wellcome, Drs. Debra Lerner and Benjamin C. Amick III), we are currently referring to this modified measure as the "adapted WLQ-25." We are also currently engaging developers about the acceptability of our proposed adaptation for this highly promising measure.

Research Agenda for Work-stream 2

The future research agenda for the multi-item work-stream involves several objectives. First, as with the global measures, we will continue to update the literature for OMERACT filter evidence, with emphasis on emerging discrimination evidence. Second, we will consolidate evidence of the test-retest reliability for our multi-item candidate measures through our CAN-funded project. Third, we will also explore a novel approach to measure worker productivity, which involves items having dual response keys, allowing concurrent assessments of frequency of difficulty (like the WLQ-25) and amount of difficulty (like the WALS). Both perspectives have been shown to have relevance in the quantification of on-the-job problems, and a combined scale might afford increased efficiency and even better understanding of the questions being asked. Approaches to combine these perspectives will be explored as a way to assess at-work disability that could be more comprehensive than using each of these approaches individually. Testing feasibility and performance of such a dual scaling work measure in persons with arthritis has been initiated.

PROGRESS ON CONTEXTUAL FACTORS OF WORKER PRODUCTIVITY

Work-stream 3

Contextual factors have been defined by the OMERACT Filter 2.0 development team as "variables that are not the primary focus of the study but relate to the scope or setting of the study, and may include potential confounders and effect modifiers." This has significant relevance to applications of worker productivity outcomes because a complex interaction between worker- and workplace-related factors (i.e., personal and environmental factors in the WHO International Classification of Functioning, Disability, and Health framework) is an important basis of health-related work functioning³². Just as worker well-being can influence work productivity, changes in job-related factors beyond the control of the worker (e.g., job demands) can also have significant influences on worker productivity. The need to consider potential contextual factors in clinical trials and cohort studies when interpreting worker productivity outcomes should be recognized. The main objectives of this work-stream are to (1) establish criteria to help define what would constitute a contextual factor of worker productivity, (2) achieve consensus on the best approach to measure these

contextual factors, and (3) develop guidelines on how to consider and/or incorporate contextual factors in research analysis when worker productivity is a study outcome.

Some preliminary progress has been made at previous OMERACT meetings and from ongoing literature reviews³. To date, we have identified up to 70 potential contextual factors and have organized them into 2 major domains (personal or environmental factors), and also further into 1 of 15 subdomains: health, demographic attributes, economic need, personal appraisal, related skills/abilities, work-life balance, accessibility, economic climate/ labor regulations, nature of work, workplace support, non-workplace support, organizational policies and practices, physical environment, work accommodations, and others³. The large number of potential contextual factors dictates the need to more closely examine whether some are more critical (i.e., core) to consider than others during various applications of worker productivity outcome measures.

Research Agenda for Work-stream 3

Several ideas and issues related to the classification and evaluation of contextual factors were discussed among conference participants during individual breakout sessions at the OMERACT 11 Workshop, which will guide our efforts toward this work-stream's longterm objectives. These include (1) explorations of theoretical frameworks that could be applied to help organize the large breadth of potential contextual factors (e.g., International Classification of Functioning, Disability and Health), (2) approaches to adapt existing guidelines (e.g., Cochrane Collaboration, Grading of Recommendations Assessment, Development and Evaluation to help evaluate the quality of prognostic studies to grade level of evidence for specific contextual factors, and (3) addressing the need for usable guidelines to determine the relevance of specific contextual factors for different applications of worker productivity outcomes (e.g., RCT vs longitudinal observational studies).

PROGRESS ON INTERPRETABILITY OF WORKER PRODUCTIVITY OUTCOME MEASURES

Work-stream 4

The aim of this work-stream is to advance the interpretability of worker productivity outcome measures, which remains an area with limited understanding. A number of interpretability needs have been identified by our working group. For example, applications of worker productivity outcomes in clinical trials might benefit from a "response criterion" to facilitate responder analyses; interpretation of longitudinal observational studies might benefit from an improved understanding of what constitutes a "meaningful" between-group difference or within-group change beyond statistical comparisons of scores; and finally, some understanding of what constitutes a meaningful change at the individual level might also be useful for applications in

clinical settings (e.g., vocational decision making). To these ends, we are keen to advance the interpretability of worker productivity measures, by building on efforts that have been initiated by others³³. Our aim is to draw insights from past OMERACT meetings to inform approaches to estimate cut scores associated with PAS^{10,11} and magnitude of a score change that would represent the MID^{12,13}. To adapt to our concept of interest, we have introduced the new acronym of PAS-WP, which stands for patient acceptable state in worker productivity. An internal vote among working group members at a 2011 meeting in Toronto supported the pursuit of these objectives (Table 7).

Toward a Methodology for PAS/MID Estimations

Two central methodological issues related to PAS-WP and MID estimations are the choice of external anchor (to serve as "gold standard" reference comparator), and the specific cut-score estimation approach to be applied (e.g., 75% centile approach, use of receiver-operating characteristic curves). Prior to OMERACT 11, our working group engaged in a 4-round Delphi group exercise to identify specific PAS-WP and MID anchors to be used as reference comparators in our CAN-funded (discrimination) project. Candidate anchors were initially gathered in 2 ways: adaptation of existing anchors from the literature into anchors that describe a state or change in worker productivity, or generation of new PAS-WP and MID external anchors by Delphi participants. PAS-WP anchors derived from our Delphi exercise were discussed at selected individual breakout sessions. After providing a brief overview of the conceptual bases and methodological issues related to PAS-WP, we asked conference participants to assess the face validity of each of 7 anchors from our shortlist (Table 8). Discussions revealed that in general, participants favored PAS-WP anchors that made a clear distinction between acceptable and unacceptable states (e.g., PAS 5, where only 2 response options are available) and used simpler phrasing (e.g., PAS 7). Last, we presented preliminary PAS-WP estimates for various candidate presenteeism measures from our ongoing CAN-funded project, and highlighted the variability in cut-scores based on the different anchors applied.

Research Agenda for Work-stream 4

Feedback gathered from conference participants at OMERACT 11 will be used to refine our methodology (preferred external anchor, estimation method) to be applied when estimating PAS-WP and MID cutpoints for worker productivity measures as an objective of the CAN-funded study. These results will help contribute to filling the gap in evidence for the discrimination criterion of the OMERACT filter. In addition, our group will explore whether cut-scores associated with an "acceptable" productivity state can be further informed by seeking perspectives from others, such as employers or co-workers.

Table 7. Voting results among Worker Productivity working group members on pursuing research to inform interpretability of worker productivity outcome measures.

Voting Question	Votes, no.	Votes, %
1. Do you think defining “response criteria” in work productivity is an important OMERACT objective to pursue?		
Yes	14	82.4
No	1	5.9
Unsure	2	11.8
Total	17	100.0
2. Do you think determining PAS and MID estimates for at-work productivity measures (e.g., global/multi-item) are important/useful objectives toward the goal of defining “response criteria”?		
Yes	13	72.2
Yes, but only PAS is important/useful	2	11.1
Yes, but only MID is important/useful	1	5.6
No, neither the PAS nor MID should be used toward establishing response criteria	1	5.6
Establishing response criteria is not an important goal	1	5.6
Total	18	100.0

PAS: patient acceptable state; MID: minimally important difference.

Table 8. List of candidate external anchors for estimating cut-scores associated with the patient acceptable state in worker productivity (PAS-WP) explored at the OMERACT 11 Workshop.

PAS-WP anchors (acceptable state = X)
1. Were you hindered by your arthritis at your paid work over the past week?
No, not at all (X)
Yes, to a degree (X)
Yes, very much
2. People can “get the job done” even when they are not working at 100% performance. Would you say you are “getting the job done” at your current level of performance?
Yes, I can get the job done at my current level of performance (X)
No, I cannot get the job done at my current level of performance
3. Some people work, but they are sacrificing other things to do so. They may work extra hours to catch up, for example. Thinking about your current situation, would you say your current effort is acceptable to you again if it were to continue like this on an ongoing basis?
My current situation, including any ways I am balancing things outside of work to keep up with work would be acceptable to me on an ongoing basis (X)
My current situation, including the way I may be balancing things outside of work to keep up with work would NOT be acceptable to me on an ongoing basis.
4. Taking into account all you have to do during your paid work, your level of pain, and your functional abilities, do you consider that your current state is satisfactory?
No, not at all
Yes, to a degree (X)
Yes, very much (X)
5. Considering all the ways that your arthritis affect you at work, including what you have to do outside of work to balance your work, would it be acceptable to you to continue like this on an ongoing basis?
Acceptable (X)
Not acceptable
6. If you were to remain for the rest of your life as you were during the past 2 weeks, would this be acceptable to you?
No, not at all
Yes, to a degree (X)
Yes, very much (X)
7. If you were to remain at your current level of work ability/productivity for the rest of your working life, would this be acceptable or unacceptable to you?
Acceptable (X)
Not acceptable

Feedback gathered from conference participants will inform the future activities of the working group. The following summarizes progress at OMERACT 11 and the working groups agenda.

Progress at the OMERACT 11 Workshop

- Review of the content and current evidence toward the OMERACT filter (truth, discrimination, and feasibility) of candidate global and multi-item measures of at-work

- limitations/productivity (i.e., presenteeism)
- Plenary vote to determine whether candidate global measures of at-work limitations/productivity currently meet the requirements of the OMERACT filter
- Endorsement of both presenteeism global measures from the Work Productivity and Activity Impairment questionnaire (item 5) and the RA-specific Work Productivity Survey (item 4) as having met the requirements of the OMERACT filter (received $\geq 70\%$ support)
- Consideration for OMERACT endorsement of the presenteeism global measure from the Work Ability Index (item 1) as a candidate measure of work ability/difficulty at work (received $\geq 70\%$ support)

Research Agenda of the Worker Productivity Working Group

- Ongoing literature reviews to update OMERACT filter evidence for candidate measures of worker productivity, with particular focus on the discrimination criterion where some gaps remain (e.g., emerging RCT)
- Completion of ongoing EULAR- and CAN-funded projects (led by working group members) to further explore patient appraisal, psychometric properties, and score interpretability of worker productivity outcome measures through cognitive testing, qualitative studies, and cohort studies
- Work with OMERACT to refine approaches to assess the merit of outcome measures that weigh both the quantity and methodological quality of available measurement evidence
- Development of criteria to identify and classify key contextual factors of worker productivity and guidelines to apply them for various clinical/research applications
- Exploration of the need for additional work-streams focused on issues related to the economic evaluation (monetization) of productivity loss

ACKNOWLEDGMENT

The authors thank all OMERACT 11 conference participants who made contributions to the Worker Productivity Workshop, and members of the International-Canadian Arthritis Network Work Productivity Group (I-CAN-WORK) for contributions to the field of worker productivity measurement that have influenced the progress of our work. We also thank Taucha Inrig for contributing to statistical analyses on the test-retest reliability of the worker productivity outcome measures, and Patricia Nedanovski for providing invaluable ongoing administrative support to facilitate collaborations among working group members.

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, 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.
4. 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.
5. Gignac MA, Sutton D, Badley EM. Arthritis symptoms, the work environment, and the future: measuring perceived job strain among employed persons with arthritis. *Arthritis Rheum* 2007;57:738-47.
6. Lacaille D, White MA, Backman CL, Gignac MA. Problems faced at work due to inflammatory arthritis: new insights gained from understanding patients' perspective. *Arthritis Rheum* 2007; 57:1269-79.
7. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. *J Rheumatol* 1998;25:198-9.
8. Brouwer WB, Meerding WJ, Lamers LM, Severens JL. The relationship between productivity and health-related QOL: an exploration. *Pharmacoeconomics* 2005;23:209-18.
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. Tubach F, Wells GA, Ravaud P, Dougados M. Minimal clinically important difference, low disease activity state, and patient acceptable symptom state: methodological issues. *J Rheumatol* 2005;32:2025-9.
11. Tubach F, Pham T, Skomsvoll JF, Mikkelsen K, Bjorneboe O, Ravaud P, et al. Stability of the patient acceptable symptomatic state over time in outcome criteria in ankylosing spondylitis. *Arthritis Rheum* 2006;55:960-3.
12. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol* 2002; 14:109-14.
13. Kirwan JR. Minimum clinically important difference: the crock of gold at the end of the rainbow? *J Rheumatol* 2001;28:439-44.
14. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420-8.
15. 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.
16. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics* 1993;4:353-65.
17. Loeppke R, Hymel PA, Lofland JH, Pizzi LT, Konicki DL, Anstadt GW, et al. Health-related workplace productivity measurement: general and migraine-specific recommendations from the ACOEM Expert Panel. *J Occup Environ Med* 2003;45:349-59.
18. Prasad M, Wahlqvist P, Shikhar R, Shih YC. A review of self-report instruments measuring health-related work productivity: a patient-reported outcomes perspective. *Pharmacoeconomics* 2004;22:225-44.
19. Brouwer WB, Koopmanschap MA, Rutten FF. Productivity losses without absence: measurement validation and empirical evidence. *Health Policy* 1999;48:13-27.
20. Tuomi K, Ilmarinen J, Jahkola A, Katajarinne L, Tulkki A. Work ability index, 2nd revised ed. Helsinki: Finnish Institute of Occupational Health; 1998.
21. Kessler RC, Petukhova M, McInnes K. HPQ Short Form (absenteeism and presenteeism questions and scoring rules) (2007). [Internet. Accessed July 17, 2013.] Available from:

- <http://www.hcp.med.harvard.edu/hpq/ftpd/absenteeism%20presenteeism%20scoring%20050107.pdf>.
22. Gilworth G, Chamberlain MA, Harvey A, Woodhouse A, Smith J, Smyth MG, et al. Development of a work instability scale for rheumatoid arthritis. *Arthritis Rheum* 2003;49:349-54.
 23. Amick BC 3rd, 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.
 24. Lerner D, Amick BC III, Rogers WH, Malspeis S, Bungay K, Cynn D. The work limitations questionnaire. *Med Care* 2001;39:72-85.
 25. Tang K, Beaton DE, Laccaille D, Gignac MA, Zhang W, Anis AH, et al. The Work Instability Scale for Rheumatoid Arthritis (RA-WIS): Does it work in osteoarthritis? *Qual Life Res* 2010;19:1057-68.
 26. Tang K. Disease-related differential item functioning in the work instability scale for rheumatoid arthritis: converging results from three methods. *Arthritis Care Res* 2011;63:1159-69.
 27. Tang K, Beaton DE, Gignac MA, Laccaille D, Zhang W, Bombardier C. The Work Instability Scale for rheumatoid arthritis predicts arthritis-related work transitions within 12 months. *Arthritis Care Res* 2010;62:1578-87.
 28. Anis A, Zhang W, Emery P, Sun H, Singh A, Freundlich B, et al. The effect of etanercept on work productivity in patients with early active rheumatoid arthritis: results from the COMET study. *Rheumatology* 2009;48:1283-9.
 29. Tang K, Beaton DE, Laccaille 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.
 30. Roy JS, MacDaerimid JC, Amick BC III, Shannon HS, McMurtry R, Roth JH, et al. Validity and responsiveness of presenteeism scales in chronic work-related upper-extremity disorders. *Phys Ther* 2011;91:254-66.
 31. Tang K, Beaton DE, Amick III BC, Hogg-Johnson S, Cote P, Loisel P. Confirmatory factor analysis of the Work Limitations Questionnaire (WLQ-25) in workers' compensation claimants with chronic upper-limb disorders. *J Occup Rehabil* 2013;23:228-38.
 32. Sandqvist JL, Henriksson CM. Work functioning: a conceptual framework. *Work* 2004;23:147-57.
 33. Maksymowych WP, Rao S, Boonen A, Chen N, Cifaldi M. Validation of the patient acceptable work state (PAWS): establishing thresholds for major work transition in ankylosing spondylitis [abstract]. *Arthritis Rheum* 2010;62 Suppl 10:S232.