IHE Report

A systematic review of the measurement properties of self-report instruments that assess presenteeism

December 2015

INSTITUTE OF HEALTH ECONOMICS
ALBERTA CANADA

INSTITUTE OF HEALTH ECONOMICS

The Institute of Health Economics (IHE) is an independent, not-for-profit organization that performs research in health economics and synthesizes evidence in health technology assessment to assist health policy making and best medical practices.

IHE BOARD OF DIRECTORS

INSTITUTE OF HEALTH ECONOMICS
ALBERTA CANADA

IHE Report

A systematic review of the measurement properties of self-report instruments that assess presenteeism

Prepared by:

Angus Thompson, PhD
Maria Ospina, PhD
Liz Dennett, MLIS
Arianna Waye, PhD
Philip Jacobs, D.Sc.

Acknowledgements

For permissions and information regarding instruments within their purview: Nick Bansback, PhD, School of Population and Public Health, University of British Columbia, CA; Jean Ednicott, PhD, Columbia University, New York, USA; Walter "Buzz" Stewart, PhD, Center for Health Research, Geisinger Health System, Danville, PA, USA; Debra Lerner, PhD, MS, Tufts University School of Medicine, Boston, USA.

For providing information regarding published studies and instruments: Mark Attridge, PhD, Attridge Consulting, Inc., Minneapolis, USA; Monique Gignac, PhD, Institute for Work & Health, University of Toronto, CA; Raymond W. Lam, MD, University of British Columbia, Vancouver, CA; JianLi Wang, PhD, University of Calgary, CA.

For article retrieval and layout, respectively: Ms. Debra Haas and Ms. Stefanie Kletke (IHE).

Corresponding Author

Please direct any inquiries about this report to Dr. Angus Thompson, gthompson@ihe.ca, or to any of the authors.

Funding

Production of this document has been made possible by a financial contribution from the Alberta Depression Initiative. The views expressed herein are those of the authors and are not necessarily those of either the Alberta Depression Initiative or the Institute of Health Economics.

Declared Competing Interest of Authors

Competing interest is considered to be financial interest or non-financial interest, either direct or indirect, that would affect the research contained in this report or create a situation in which a person's judgement could be unduly influenced by a secondary interest, such as personal advancement.

The authors of this publication claim no competing interest.

Suggested Citation

Thompson AH, Ospina MB, Dennett L, Waye A, Jacobs P. *A systematic review of the measurement properties of self-report instruments that assess presenteeism*. Edmonton (AB): Institute of Health Economics; 2015.

Web Address

This publication is available for free download from the IHE website at http://www.ihe.ca.

Executive Summary

Background

Presenteeism, reduced productivity while working, has come into consideration as a major occupational health problem in many countries with serious consequences for both organizations and employees. Increasing evidence shows that presenteeism represents a significant source of productivity losses that can cost organizations much more than does absenteeism, and it can lead to an increase in occupational accidents, deterioration of product quality and adverse effects on healthy employees.

Challenges exist, however, in the measurement of presenteeism. That is: (1) a number of instruments that are used to measure presenteeism do not actually measure productivity, (2) testing for presenteeism requires the measurement of work outputs, but these are often inadequately or vaguely specified, (3) when self-report testing of ability is involved, as it is here, the results are often inaccurate and biased by a general tendency of humans to optimistically place themselves in a good light in comparison to others, and (4) both historical and recent research has shown that extreme self-ratings (high or low) are related to mental health issues, particularly depression, defensiveness, and optimism

Objectives

Given the variety of instruments currently available, evidence about their measurement properties and quality are essential for an informed selection of the most appropriate tool to assess presenteeism in the workplace. Systematic reviews of measurement properties are useful for such purposes. Thus, that approach was adopted here to provide an up-to-date evaluation of the measurement properties (i.e., validity, reliability, and responsiveness) of existing self-report tests, including an evaluation of the studies that were used to assess each of the instruments in question.

Methods

Comprehensive searches of electronic databases were conducted up to October 2012. Twenty-three presenteeism instruments were identified and examined. Methodological quality of each relevant study was appraised with the COSMIN (Consensus-based Standards for the Selection of Health Status Measurement Instruments) checklist. A best-evidence synthesis was used to achieve a balance between the ratings of the measurement properties of each test with the quality of the study that produced the rating in question.

Results

The titles and abstracts of 1,767 articles were screened, with 289 of these retained for a detailed, full-text review for eligibility. Forty of these studies were deemed to have actually assessed the measurement properties of one or more of the pre-selected presenteeism instruments.

The three presenteeism instruments with the strongest level of evidence on more than 1 measurement property were the Stanford Presenteeism Scale, 6-item version (content validity, internal consistency, construct validity, convergent validity, and responsiveness); the Endicott Work Productivity Scale (internal consistency, convergent validity, and responsiveness); and the Health and Work Questionnaire (HWQ; internal consistency and structural validity). Only the HWQ was assessed for criterion validity (the strongest form of validity), but the study in question was deemed to be of inadequate quality, thus providing no basis for rating the HWQ on soundness in this case.

Conclusions

Most presenteeism instruments have been examined for some form of validity, but evidence for criterion validity is virtually absent. The selection of instruments for use in primary studies thus depends on weak forms of validity. Further research should focus on the goal of a comprehensive evaluation of the psychometric properties of existing tests of presenteeism, with emphasis on criterion validity.

Abbreviations

All abbreviations that have been used in this report are listed here unless the abbreviation is well known, has been used only once, or has been used only in tables or appendices, in which case the abbreviation is defined in the figure legend or in the notes at the end of the table.

95% CI	95% confidence interval
ALWQ	Angina-related Limitations at Work Questionnaire
AQLQ	Asthma Quality of Life Questionnaire
ASQOL	Ankilosing Spondilitis Quality of Life questionnaire
BASDAI	Bath Ankilosing Spondilitis Disease Activity Index
CDAI	Crohn's Disease Activity Index
CENTRAL	Cochrane Central Register of Controlled Trials
CESD	Center for Epidemiologic Studies Depression scale
CGI	Clinical Global Impressions
COSMIN	consensus-based standards for the selection of health measurement instruments
CSI	Caregiver Strain Index
DASH	Disabilities of the Arm, Shoulder and Hand outcome
DDAI	Dimensions of Daily Activities Index
EQ5-D	five-dimensional Euro-Qol health-related quality of life visual analogue scale
ESM	experience sample methods
EWPS	Endicott Work Productivity Scale
GHP	general health perceptions
IBDQ	Inflammatory Bowel Disease Questionnaire
ICC	intraclass correlation coefficient
IQR	interquartile range
HAM-D	Hamilton rating scale for depression
HAQ	Health Assessment Questionnaire
HLQ	Health and Labour Questionnaire
HPQ	World Health Organization health and work performance questionnaire
HRPQ-D	Health Related Productivity Questionnaire Diary
HUI-3	Health Utilities Index Mark 3
HWQ	Health and Work Questionnaire
LEAPS	Lam Employment Absence and Productivity Scale

LWPS	Lost Workplace Productivity Score
MCS	mental component summary
MDHAQ	Multidimensional Health Assessment Questionnaire
MI	mental/interpersonal demands
MIDAS	Migraine Disability Assessment Questionnaire
MOS	medical outcome study
MQoLQ	Migraine Quality of Life Questionnaire
MWPLQ	Migraine Work and Productivity Loss Questionnaire
OD	output demands
OR	odds ratio
PA	productivity while doing regular daily activities
PAM	patient activation measure
PCS	physical component summary
PD	physical demands
PW	productivity while working
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Q-Q	quantity and quality
QOLRAD	Quality of Life in Reflux and Dyspepsia
RA-WIS	Rheumatoid Arthritis Work Instability Scale
SAHAPS	Stanford/American Health Association Presenteeism Scale
SCL90	Symptom Checklist 90
SD	standard deviation
SDS	Sheehan Disability Scale
SPH	specific health problems
SF-36	Short-Form 36 health survey
SPS-6	Stanford Presenteeism Scale 6-item version
SPS-13	Stanford Presenteeism Scale 13-item version
SPS-21	Stanford Presenteeism Scale 21-item version
SRM	standardized response means
TM	time management
VOLP	Valuation of Lost Productivity questionnaire
WALS	Workplace Activity Limitations Scale

WHI	Work and Health Interview
WLQ	Work Limitations Questionnaire
WPS	Work Performance Scale
WPAI	Work Productivity and Activity Impairment
WPAI:GH	Work Productivity and Activity Impairment - General Health
WPAI:GERD	Work Productivity and Activity Impairment - Gastroesophageal Reflux Disease
WPAI:SHP	Work Productivity and Activity Impairment - Specific Health Problems
WPAI:AS	Work Productivity and Activity Impairment - Allergies
WPAI:IBS	Work Productivity and Activity Impairment - Irritable Bowel Syndrome
WPAI:PsO	Work Productivity and Activity Impairment - Psoriasis
WPAI:CD	Work Productivity and Activity Impairment - Crohn's Disease
WPAI:CG	Work Productivity and Activity Impairment - Care Givers
WPAI:ChHD	Work Productivity and Activity Impairment - Chronic Hand Dermatitis
WPAI:RLS	Work Productivity and Activity Impairment - Restless Leg Syndrome
WPAI:SpA	Work Productivity and Activity Impairment - Ankylosing Spondylitis
WPSI	Work Productivity Short Inventory
WPS-RA	Work Productivity Survey-Rheumatoid Arthritis
WRFQ	Work Role Functioning Questionnaire

Table of Contents

INTRODUCTION

Presenteeism is broadly defined as "decreased productivity and below-normal work quality" when physically present at work.[1-3] Presenteeism can be studied in relation to many factors, including health. Terms such as "impaired presenteeism",[4] "sickness presenteeism"[5] or "working through illness"[6,7] describe a phenomenon in which workers turn up at work but function at less than full productivity because of illness or other health conditions that should prompt absence from work.[4,5,8,9]

Over the past two decades, presenteeism has come into consideration as a major occupational health problem in many countries with serious consequences for both organizations and employees. Increasing evidence shows that presenteeism, based on the concept of reduced productivity while working, represents a "silent" but significant source of productivity losses that can cost organizations much more than absenteeism does.[2,10,11] Presenteeism can lead to an increase in occupational accidents, deterioration of product quality[9] and adverse effects on healthy employees.[3,12] The impact for the individual is not less; employees who turn up for work when ill have their quality of life diminished; they often experiment feelings of burnout due to inadequate recovery[13] and get trapped in a vicious circle: job demands are accumulated, less energy to cope with these demands is available, resulting in more presenteeism, and so on. Similarly, by repeatedly postponing sickness leave that may effectively resolve minor illnesses, more serious illnesses may develop.

Many challenges exist when identifying an optimal or ideal approach to the measurement and valuing of presenteeism.[12] In many instances, there is a confusion between the measurement of potential causes of lost productivity while on the job (e.g., health & mental health, personality, disability,[10,14] malingering, and irresponsibility) and presenteeism proper that creates serious methodological problems. In fact, a number of instruments are available for measuring health-related difficulties with workplace tasks, work limitations, or work impairments that, although not originally developed to quantify presenteeism, they are increasingly being used for that purpose. The conceptual approach adopted here assumes that presenteeism is a factor separate from any purported causes of lost productivity at work, thus allowing independent examinations of the relationships between lost productivity and the full list of hypothesized causal variables.

It is much more difficult to measure presenteeism than absenteeism, primarily because the former requires the measurement of outputs, which are often not specified well or at all,[15] while the latter simply involves a notation of attendance which is easier to remember and is often recorded by the employer, albeit not in all cases.[16] Presenteeism is usually assessed by self-report measures that can be generic (i.e., applicable to any job) or disease-specific. Measures vary in complexity covering single items assessing the number of days in a given period in which the person attended work when unwell,[9] time-adjustments at work due to perceptions of productivity in relation to self and/or colleagues,[17] and domain-based measures that assess health-related limitations in specific job demands.[7,18] Given the variety of instruments currently available, evidence about their measurement properties and quality are essential for an informed selection of the most appropriate tool to assess presenteeism in the workplace. Systematic reviews of measurement properties are useful for selecting the best instrument for a specific purpose based on a rigorous evaluation of their measurement properties (i.e., validity, reliability, and responsiveness).[19]

Of particular importance here is the widespread use of self-report instruments to estimate productivity decrements. This practice was presumably adopted because of the higher cost of direct measurement of work performance plus, perhaps, the above-noted difficulty in specifying measurable definitions of work output. But there is reason to be suspicious of self-report testing when an ego-involved construct like ability is involved[27] as it is here, suggesting some degree of defensiveness may be induced in these work-related threat-inducing situations. Such self-reports of performance are notoriously inaccurate and are often biased by a general tendency of humans to optimistically place themselves in a good light in comparison to others (Brown 1986; Alicke and Govorun 2005).[21,22] Indeed, a large sample of Alberta workers showed this bias, with about 50% rating their own work to be more productive than the average worker, but only about 9% admitting to lower performance than their work counterparts.[23] Furthermore, those who rated themselves to be either below average or well above it, were much more likely to be depressed, a finding that appears to echo a 1993 finding that overly positive self-evaluations are indicators of mental health issues.[28] All this, plus the admonition of Alicke and Govorun[22] suggests that test scores need to be understood in the context of personality factors (i.e., both defensiveness[24] and optimism[25]) and clinical syndromes, particularly depression.[23,26]

A number of systematic reviews have summarized the measurement properties of instruments that assess productivity loss at the workplace,[14,15,20,29] work productivity combining presenteeism and absenteeism measures,[30] or work-related outcome measures in specific clinical groups (e.g., musculoskeletal disorders).[31,32] The majority of these reviews, however, have not incorporated a systematic analysis of the methods with which these instruments have been developed.[15,20,29] In some instances, non-validated approaches have been used to appraise both the quality of studies and the measurement properties of presenteeism instruments themselves.[14] Assessing the quality of studies that evaluate the measurement properties of presenteeism instruments is an essential step in a systematic review of the measurement properties of presenteeism instruments. If the quality of a study is appropriate, the results are valid and the measurement instrument can be a useful tool in practice or research. Alternatively, if study quality is inadequate, the results cannot be trusted and the quality of the measurement instrument under study remains unclear. Concurrent comparisons of the measurement properties and quality of these instruments are also needed to help reveal the relative strengths and weaknesses of the measures and to provide evidence-based guidance for the selection of outcome instruments in future studies. Finally, it is unknown how frequent is the use of presenteeism instruments in the scientific literature, in what populations have been evaluated and for which purpose.

OBJECTIVES

The objectives of this systematic review were:

1) to describe the frequency and characteristics of use of instruments measuring presenteeism in the scientific literature,
2) to summarize the measurement properties (i.e., validity, reliability, responsiveness) of instruments assessing presenteeism, and
3) to analyze the quality of studies that have evaluated the measurement properties of presenteeism instruments.

METHODOLOGICAL APPROACH

Identification of Studies

Comprehensive searches of the Medline (including in-process citations), Embase, Cochrane Central Register of Controlled Trials (CENTRAL), PsycInfo, Web of Science, CINAHL, Business Source Complete and ABI Inform electronic databases were conducted from database inception to October 2012. The search strategy was designed by an information specialist and comprised mainly of the names of potential presenteeism self-reporting instruments. The list of names was compiled during a process that included a preliminary search of the literature, extracting the names of instruments included in other presenteeism reviews, and by contacting a number of experts in the field. Because we updated the search several times during the project, the final search only included the names of those tests that we planned to include in the review. The full search strategy is available in Appendix B. In addition, reference lists of reviews and retrieved articles were checked for relevant studies. Searches were limited to citations in English language.

Inclusion and Exclusion Criteria

Based on preliminary literature searches, a list of potential instruments assessing presenteeism was assembled and their items examined by two of us (AHT, AW). Instruments measuring presenteeism were defined in this systematic review as questionnaires measuring (at least one domain of) productivity loss or reduced productivity/performance while at work.[15,20] Items assessing presenteeism must focus on at least one of the following characteristics: a) assessment of perceived productivity loss/reduced performance; b) comparative productivity loss/reduced performance (with those of others and with one's pattern); and/or c) estimation of unproductive time while at work. Based on this definition, the following 21 instruments were examined in this review (see Appendix C for a more detailed description of the instruments): Angina-Related Limitations at Work Questionnaire (ALWQ),[33] Endicott Work Productivity Scale (EWPS),[34] Health and Labour Questionnaire (HLQ),[35] Health Related Productivity Questionnaire Diary (HRPQ-D),[36,37] Health and Work Questionnaire (HWQ),[38] Lam Employment Absence and Productivity Scale (LEAPS),[39] Migraine Disability Assessment Questionnaire (MIDAS),[40] Migraine Work and Productivity Loss Questionnaire (MWPLQ),[41] Osterhaus Technique,[42] Quantity and Quality (Q-Q) method (from the Productivity and Disease Questionnaire; PRODISQ),[43-45] Stanford Presenteeism Scale 6-item version (SPS-6),[1] Stanford/American Health Association Presenteeism Scale (SAHAPS) 32-item version (SPS-32),[46] Stanford Presenteeism Scale 13-item version (SPS-13),[5] Valuation of Lost Productivity questionnaire (VOLP),[47] Work and Health Interview (WHI),[48] the Work Performance Scale (WPS) (from the Functional Status Questionnaire),[49] Work Productivity and Activity Impairment scale (WPAI),[50] World Health Organization Health and Work Performance Questionnaire (HPQ),[51] Work Productivity Short Inventory (WPSI),[52] Work Productivity Survey-Rheumatoid Arthritis (WPS-RA),[53,54] and Work Role Functioning Questionnaire (WRFQ).[55]

To be included in the review, individual studies must be full text, peer-reviewed primary studies that: a) used any of the instruments listed above to measure presenteeism as one of the study variables, and/or b) evaluated the measurement properties (i.e., validity, reliability, responsiveness) of the English version of any of the presenteeism instruments included in the list.

We excluded studies that used an instrument from the list to measure outcomes other than presenteeism (i.e., disability, health-related quality of life). No restrictions in study design were

applied; however, editorials, book chapters, review articles, conference abstracts, unpublished studies (i.e., thesis and dissertations), case studies with fewer than 30 cases, and studies enrolling only pediatric populations (age 18 years and under) were excluded. In addition, studies published in non-English languages, or studies published in English that described the cross-adaptation of instruments or the measurement properties of non-English versions of presenteeism instruments were not considered for inclusion in the review.

Two of us (AW and MO) independently screened the titles and abstracts generated from the search strategies to identify potentially relevant articles. The full text of papers deemed relevant and of those whose abstracts and titles provided insufficient information were retrieved for a closer inspection by two independent reviewers (two of AW, LD, or MO) who determined study eligibility for the review. Disagreements about inclusion and exclusion of studies were resolved through discussions among reviewers until consensus was reached. Studies that did not meet the selection criteria were excluded and the reasons for exclusion documented. A flow chart of the study selection process was prepared according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[56]

Methodological Quality Assessment

Studies that were restricted to the measurement of presenteeism as one of the study variables were not assessed for their methodological quality in this review. Rather, this aspect of the assessment included only those studies that examined the psychometric strengths and weaknesses of one or more presenteeism tests. The analysis focused on: (1) the quality of the tests in question as derived from these studies; and (2) on the methodological quality of the studies themselves.

The methodological quality of studies evaluating the measurement properties of presenteeism instruments was appraised with the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist.[57,58] Based on international consensus, the COSMIN checklist provides an overview of measurement properties of health-related patient-reported instruments. For each study on a measurement property, the methodological quality for that particular measurement property was rated by a series of items on a 4-point rating scale (poor, fair, good, excellent). An overall score for the methodological quality of a study was determined per measurement property by taking the lowest rating of any of its items.[59] In this review, the following COSMIN domains were evaluated: reliability, internal consistency, content validity, construct validity, criterion validity, and responsiveness. In addition, the interpretability and generalizability of results were described.

Generally speaking, reliability refers to the consistency with which a particular test measures something, even if we do not know what it is that is being measured. Validity, on the other hand, tells us the degree to which a test measures what it is supposed to measure when we do know what is being measured (or purportedly does so). Note that validity is limited by reliability. If a particular test shows low reliability, it cannot show high validity – at least not within the same testing context (results may vary across different populations and personal and/or health conditions).

All this means that validity is of greater value than reliability when determining the positive worth of a test, but that low reliability can tell us when a test is not valid. Thus, we can arrange the various psychometric properties in a loose hierarchy to aid us when assessing the potential value of a test. Overall, validity stands above reliability in the hierarchy, but each of these comprises a number of sub-categories that are presented in rank-order and explained in Table 1.

TABLE 1: RANKED UTILITY OF THE VARIOUS FORMS OF RELIABILITY AND VALIDITY USED IN TEST COMPARISONS

Utility Weight	Psychometric Property	Comment
Validity		
10	Criterion validity	The degree to which a test correlates with direct measures of the same construct – a "gold standard".
7	Construct validity	The extent to which the scores for a particular test relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the constructs being studied.
6	Convergent validity	The degree of correlation between a new test and previously validated tests/measures of the same construct.
5	Content validity	The relevance, and adequacy, of test items to the domains of interest.
Reliability		
4	Responsiveness to change	The extent of agreement across two administrations of a test, but with an intervention occurring between testings. That is, the ability of an instrument to detect important changes over time in the concept that it measures.
3	Structural validity	How many things does the test measure? Concordance between, e.g., the supposed and derived number of sub-scales (if any) for a particular test. Can be taken as a reliability measure since it represents consistency of test factor structure.
2	Test-retest reliability	The extent of agreement across two administrations of a test, assuming nothing happened between testings (like treatment or other change-producing event).
2	Inter-rater reliability	The extent of agreement among two or more raters at a single testing session. Introduces an additional source of unreliability (the rater) to the test unreliability found in other domains.
1	Internal consistency	The level of homogeneity of a test items (or of sub-scales, if any) at one point in time.

Source: [60]

There are many dimensions that can be used for ranking. Here, we have used a supposed level of confidence that can be taken from each of the psychometric domains that provide evidence supporting the use of the test in question. They are presented in order with a weighting applied to each (left column) which is admittedly somewhat arbitrary. In a general sense, the ratings are based on the assumption that a strong result from a particular domain reduces the need for supporting evidence from those below it. Criterion validity stands well above the rest since it involves the use of a trusted measure, often a direct measure, of the matter in question. Thus, evidence of good criterion validity is also indicative of strong functioning within the domains below, obviating the need for their measurement (one exception could be responsiveness to change which is theoretically not guaranteed by good results above it). Construct validity is also very strong since it can provide evidence in support of the validity of, in our case here, presenteeism. But it is indirect, and does not carry the weight of criterion validity.

At the lower end lies internal consistency (often represented by Cronbach's alpha [α]), a placement that may seem counter-intuitive, considering its widespread use. But it only tells us about the level of consistency on one day and, like all forms of reliability, tells us nothing about the test's ability to

measure what it purports to measure. Strong test-retest and inter-rater reliabilities are further up the hierarchy because they are necessarily based on good internal consistency and also provide broader information about test performance. Strong internal consistency may be a prerequisite for good performance in other domains, but there is no guarantee.

Note that structural validity has been placed within the reliability section, belying the term "validity" in its title. It may be that we can talk about the "validity" of statements about test structure, but this domain says nothing about the validity of the test in the sense of the term that is used here. For example, integrity of factor structure, while a very good thing, does not tell us what the test actually measures. Similarly, the measurement of responsiveness to change provides very important information. A test that cannot detect change is limited in its usefulness. But our ability to measure change does not mean that we know what it means. About as far as we can go is to say that we have reliably measured change in whatever it is that the test measures.

The COSMIN checklist is increasingly used in systematic reviews of measurement properties[61-65] and, to date, it is the only quality assessment tool of this kind that has been validated and standardized.[59] Reviewers (two of MO, AW, or LD) independently applied the COSMIN checklist to each of the studies evaluating the measurement properties of presenteeism instruments. Disagreements among reviewers were solved by consensus.

Data Extraction and Analysis

As noted above, studies included in the review were categorized according to whether they: (1) assessed presenteeism as one of the study variables; or (2) evaluated the measurement properties of a presenteeism instrument. For studies that used presenteeism as one of the study variables, details on research design, country, clinical topic and type of presenteeism instrument were collected. The studies were classified according to the primary purpose for which the presenteeism instrument was used in the study as: a) descriptive/discriminative (i.e., study described the degree of presenteeism in the population or compared differences in presenteeism among different groups), b) predictive (i.e., study used presenteeism as a variable to predict the occurrence of another outcome), and/or c) evaluative (i.e., study assessed changes in presenteeism over time or as a result of an intervention).

For studies evaluating the measurement properties of presenteeism instruments, the following information was extracted: health condition and sample size in which the instrument has been tested, data on validity (i.e., content, construct, criterion, convergent), reliability (internal consistency, test-retest, inter-rater) and responsiveness. In all cases, data from the primary studies were extracted by one reviewer (MO) and independently verified for accuracy and completeness by a second reviewer (AW). Any discrepancies in data extraction were resolved by consensus between the data extractor and the data verifier. Study selection, methodological quality assessment, and data extraction were managed with Microsoft Excel™ (Microsoft Corporation, Redmond, WA).

Characteristics of the included studies were summarized descriptively. Evidence tables (see Appendix A) were constructed to describe the presenteeism instruments and summarize the data on their measurement properties. A best-evidence synthesis approach was used to summarize the total body of evidence for the measurement properties of individual presenteeism instruments, taking into account the number of studies, their quality ratings and the consistency of their results. For each instrument, the results of the methodological quality assessment of individual studies were combined with a rating of the measurement properties. The following criteria were used: a strong level of

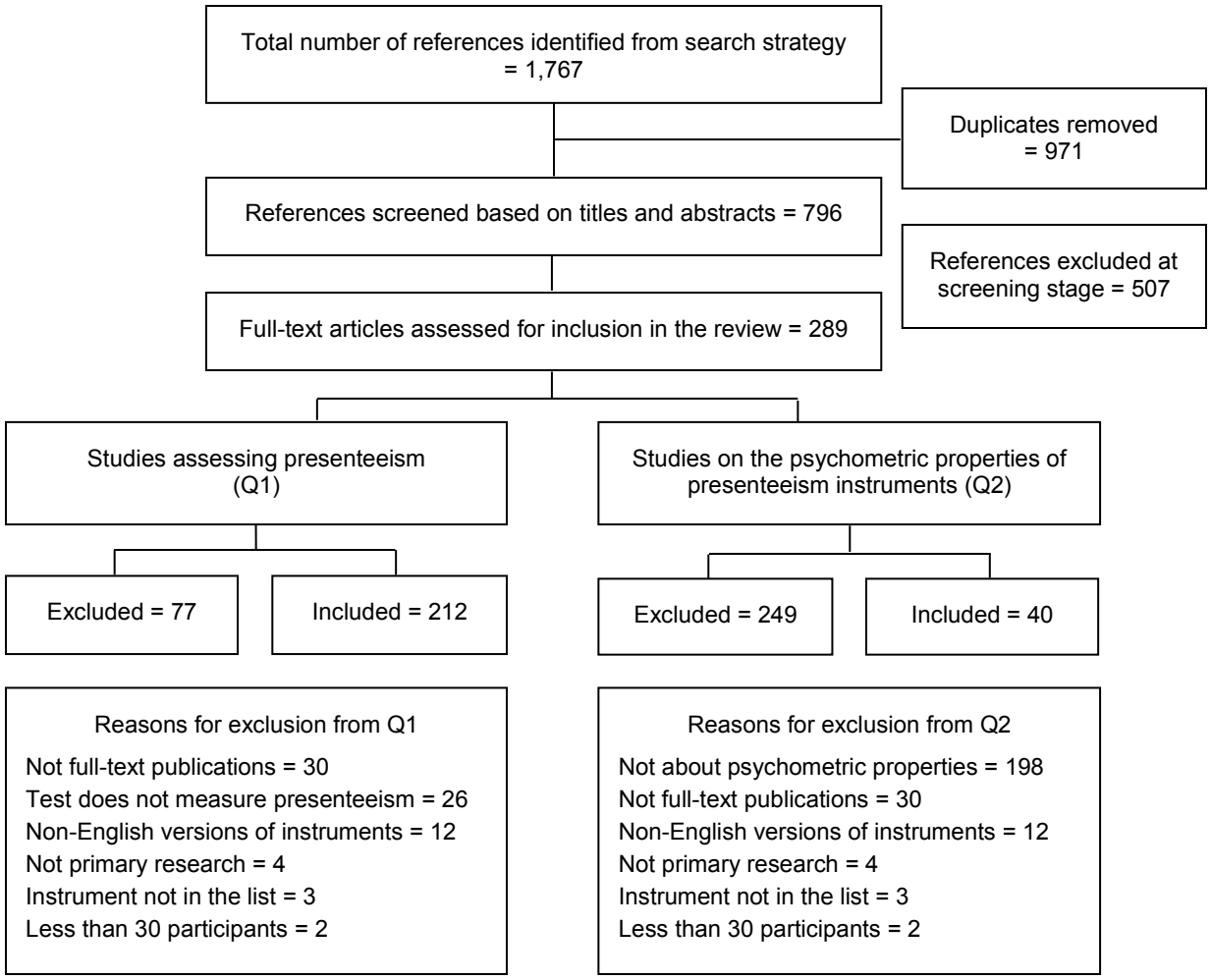
evidence (i.e., consistent findings in multiple studies of good methodological quality *or* in one study of excellent methodological quality); a moderate level of evidence (i.e., consistent findings in multiple studies of fair methodological quality *or* in one study of good methodological quality); a limited level of evidence (i.e., one study of fair methodological quality); and conflicting level of evidence (i.e., conflicting findings). When there were only studies of poor methodological quality, an unknown level of evidence was noted.

RESULTS & DISCUSSION

Search Results

Our searches identified 1,767 citations with abstracts of which 971 duplicates were removed. Titles and abstracts of the remaining 796 references were screened for relevance. This produced 289 articles judged to be potentially relevant. After applying the eligibility criteria to the full-text version of each of these studies, 212 articles that assessed presenteeism as one of the study variables were retained. Similarly, 40 studies that evaluated the measurement properties of a presenteeism instrument were identified and included in the review (Figure 1).

FIGURE 1: PRISMA FLOW DIAGRAM



Presenteeism as a Study Variable in the Scientific Literature

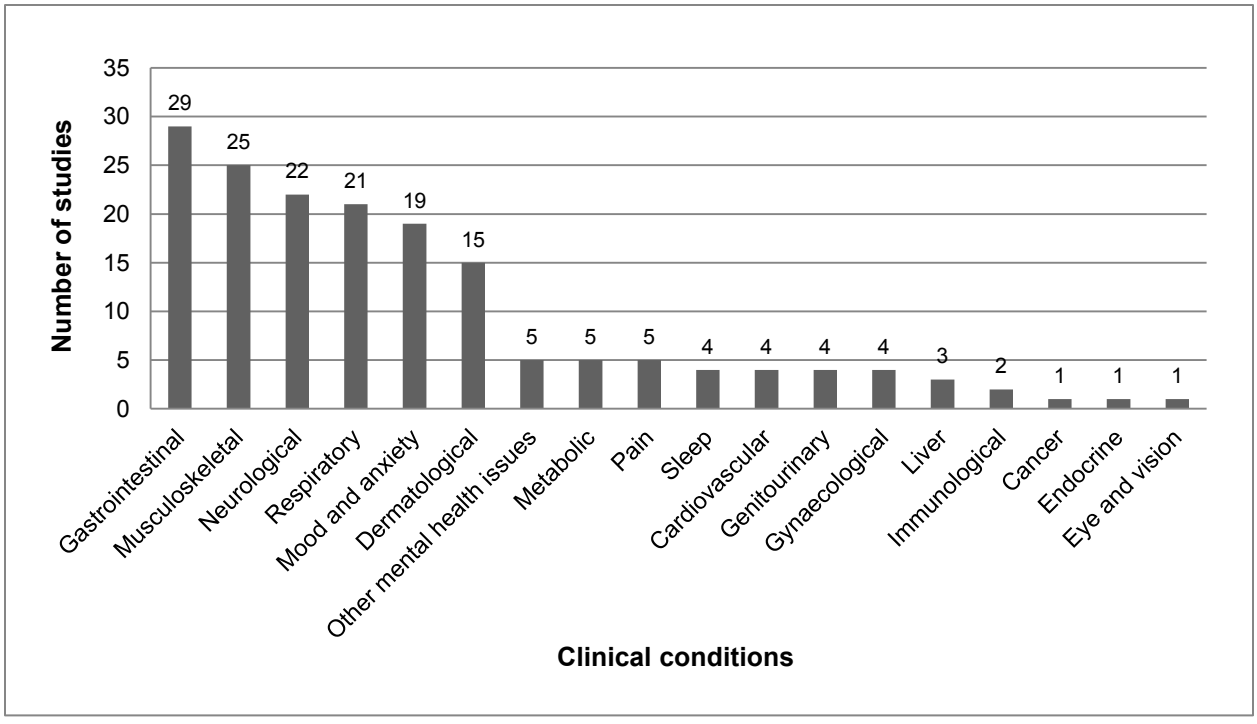
General characteristics

Two hundred and twelve studies measured presenteeism as one of the study variables using one or more instruments on our list. The majority of studies (95.2%) used only one instrument to assess presenteeism; less frequently, two or three presenteeism instruments were used per study (10 studies). The studies were published between 1996 and 2012 with a median year of publication of 2009 (interquartile range [IQR] = 2006, 2011). Studies have been mainly conducted in North America (62.7%) followed by those conducted in European studies (18.8%) and multi-country studies (10.9%). Less frequently, studies were conducted in Australasia (6.6%) or Latin-American countries (0.95%). Of the 212 studies included, 76.5% were observational studies (124 cross-sectional and 30 cohort studies) and 23.5% were intervention studies (42 randomized controlled clinical trials and eight pre-post studies).

Populations in studies measuring presenteeism

The majority of studies (66%) measured presenteeism in clinical samples, compared to studies that included healthy workers (34%). Figure 2 summarizes the distribution of studies across a variety of clinical conditions. Overall, a cluster of six clinical conditions for which presenteeism was measured proved to be the most commonly found. The conditions most studied were: (1) gastrointestinal (e.g., gastroesophageal reflux disease, Crohn's disease, irritable bowel syndrome), (2) musculoskeletal (e.g., rheumatoid arthritis), (3) neurological (e.g., migraine), (4) respiratory (e.g., asthma, chronic obstructive pulmonary disease, allergic rhinitis), (5) mental (depression, anxiety), and (6) dermatological.

FIGURE 2: CLINICAL CONDITIONS IN STUDIES MEASURING PRESENTEEISM

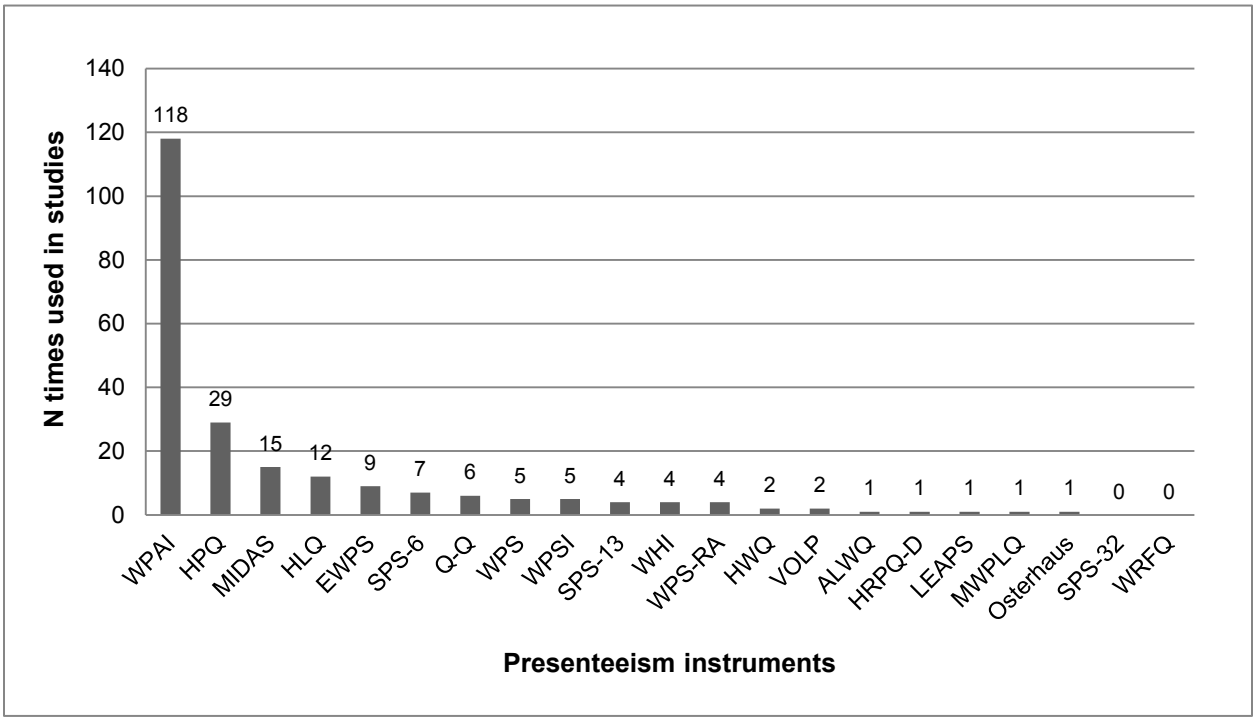


Types of presenteeism measures

In total, 227 presenteeism measurements were reported in the 212 studies that used one or more of the 21 instruments considered in this review (Figure 3). To note, five studies reported the use of two presenteeism instruments whereas another five used three instruments in their studies. Therefore, the description of the results presented below refers to the number of times the instruments were used (227) in the individual studies.

The WPAI is the most widely used instrument to measure presenteeism followed distantly by HPQ, MIDAS and HLQ. All the other 17 instruments were used infrequently to measure presenteeism in the individual studies. We did not identify studies on presenteeism that used the WRFQ or the SPS-32.

FIGURE 3: INSTRUMENTS MEASURING PRESENTEEISM IN INDIVIDUAL STUDIES



The use of a variety of versions of the WPAI has been reported in the scientific literature. By far, the general health version of the WPAI (WPAI:GH) was the most frequently used instrument in the studies (69 times), followed by the gastroesophageal reflux disease version (WPAI:GERD, 11 times). Other WPAI versions that were used in the studies were specific to: health problems (WPAI:SHP, 10 times), allergies (WPAI:AS, nine times), irritable bowel syndrome (WPAI:IBS, five times), psoriasis (WPAI:PsO, four times), Crohn's disease (WPAI:CD, three times), caregivers (WPAI:CG, two times), sleep and GERD combined (WPAI:Sleep-GERD, two times), chronic hand dermatitis (WPAI:ChHD, one time), restless leg syndrome (WPAI:RLS, one time) and ankylosing spondylitis (WPAI:SpA, one time).

Purpose of presenteeism measures

The use of presenteeism instruments to measure study variables was also classified according to its primary purpose within the studies (i.e., descriptive/discriminative, predictive and/or evaluative). In many instances, a presenteeism instrument was used for more than one purpose in a single study. Overall, presenteeism instruments were mainly used for descriptive/discriminative purposes in the individual studies (65.5% of studies). Studies using presenteeism as a variable to predict the occurrence of another outcome accounted for 43.8% of the studies. Finally, presenteeism measures were used less frequently (27.3%) for evaluative purposes, that is to say, to assess changes in presenteeism over time or as a result of an intervention.

Use of generic versus disease-specific presenteeism instruments

Presenteeism instruments can be generic or specific to certain clinical conditions. Generic presenteeism instruments include EWPS, HLQ, HPQ, HRPQ-D, HWQ, Osterhaus technique, Q-Q, SPS-32, SPS-13, SPS-6, VOLP, WHI, WPAI:GH, WPS, WPSI, and WRFQ. Alternatively, disease-specific presenteeism instruments are ALWQ, LEAPS, MIDAS, MWPLQ, WPS-RA, and successive adaptations of WPAI that were developed to assess presenteeism associated with certain clinical conditions (i.e., WPAI:AS, WPAI:CD, WPAI:CG, WPAI:ChHD, WPAI:GERD, WPAI:IBS, WPAI:PsO, WPAI:RLS, WPAI:SHP, WPAI:Sleep-GERD, and WPAI:SpA).

Overall, the use of generic presenteeism instruments in the scientific literature has more than doubled that of disease-specific presenteeism instruments (159 versus 68 times). Generic presenteeism instruments can be used in healthy samples of workers from the general population, or across a broad range of diseases and conditions. A brief description of the primary usage of each of the generic instruments in the scientific literature is as follows (see also Table A1 in Appendix A):

- EWPS: neurological, mood/anxiety and sleep disorders;
- HLQ: musculoskeletal disorders;
- HPQ: workers/general population;
- HRPQ-D: workers/general population;
- HWQ: immunological disorders and workers/general population;
- Osterhaus technique: metabolic disorders;
- Q-Q: musculoskeletal disorders;
- SPS-32: not applicable (studies using this instrument were not identified in the review);
- SPS-13: workers/general population;
- SPS-6: workers/general population;
- VOLP: musculoskeletal disorders;
- WHI: workers/general population;
- WPAI:GH: gastrointestinal disorders;
- WPS: cardiovascular disorders;
- WPSI: workers/general population;
- WRFQ: not applicable (studies using this instrument were not identified in the review).

Disease-specific presenteeism instruments can be used sometimes in populations other than that for which they were originally developed. In this review, we found that overall, disease-specific presenteeism instruments were used in all cases to assess populations for which they were intended to target. A particular example is the WPAI version developed for "specific health problems", which has been predominantly used in the scientific literature for the assessment of patients with genitourinary or musculoskeletal disorders. Table A1 in Appendix A summarizes details on the use of the 21 presenteeism instruments by population and purpose.

Evaluation of the Measurement Properties of Presenteeism Instruments

General characteristics of the studies

A total of 40 studies[1,5,7,12,33-36,38,39,47,48,50-52,54,66-89] examined the measurement properties of the 21 presenteeism instruments included in the review. The studies were published between 1993 and 2012 with a median year of publication of 2004 (IQR = 2002, 2009). The majority of these evaluations have been conducted in the United States (22 studies[1,5,33,34,36,38,48,50-52,69-72,74,76,77,79-81,83,85]). Other countries in which presenteeism measures have been evaluated include Canada,12,39,66,68,84,86 the Netherlands,[35,67,75] United Kingdom,[47,88,89] Australia,[7] Croatia,[73] and Sweden.[87] A small proportion of studies[54,78,82] have been conducted in multinational settings.

Over half of the studies[1,5,33,35,39,50-52,67,71-77,80-85,87-89] used cross-sectional research designs to evaluate the measurement properties of the presenteeism instruments. Less frequently, the instruments were examined in the context of prospective cohort studies[7,12,34,38,47,48,66] or as part of experimental studies (i.e., randomized controlled clinical trials[36,54,68,69,78,79] or single-arm clinical trials[70,86]). Sample sizes varied greatly across studies, ranging from 40 to 7,797 participants per study (median sample size = 191; IQR: 112, 354).

The measurement properties of presenteeism instruments have been examined in a broad variety of populations. The majority of these evaluations (27 studies) have been conducted among heterogeneous clinical groups, with most of them (nine studies[12,47,54,66,67,78,84,88,89]) being conducted on patients with musculoskeletal disorders. Other clinical conditions for which presenteeism instruments have been often evaluated include gastrointestinal,[68,79,80,86,87] neurological[69,81-83] and mood and anxiety disorders.[7,34,39,70] Less frequently, individuals with cardiovascular,[33,74] immunological,[36] and respiratory[71] conditions have been included. Other studies (13 studies[1,5,35,38,48,50-52,73,75-77,85]) have included samples of employees across a wide range of organizational settings (e.g., manual workforce, telecommunications, airlines, call centres) or healthy individuals responsible for the care of patients.[72]

Analysis of the measurement properties of presenteeism instruments

The evaluation of the measurement properties of presenteeism instruments was not uniform across the studies and the amount of evidence varied according to the type of measurement properties evaluated. Evidence on the content validity of presenteeism instruments was provided in eight studies.[1,33,35,38,51,52,77,89] Reliability data were reported for internal consistency in 17 studies,[1,5,7,33,34,38,39,52,66,69-71,73,74,82-84] for test-retest reliability in six studies,[34,47,50,80,82,83] and for different schedules of administration in one study.[52] Structural validity and factorial analysis were part of the analysis of four studies.[1,5,38,39] Construct validity, as determined through hypothesis testing of the relationship between the scores for presenteeism instruments and other non-presenteeism measures,

was evaluated in 28 studies.[1,5,12,33-36,38,39,47,50,54,66,69-72,75-80,84-88] Alternatively, head-to-head comparisons among different presenteeism instruments aimed at establishing their convergent validity were reported in 11 studies.[1,12,39,47,48,51,66,67,75,81,88] We identified one study[38] in which criterion validity was formally evaluated by comparing a presenteeism instrument with a "gold standard" (i.e., hours of productivity loss). Finally, eight studies[34,36,54,66,68,78,79,86] provided evidence on the responsiveness of presenteeism instruments to detect important changes in the construct over time.

Tables A2 to A4 in Appendix A describe the reported measurement properties (reliability, content validity, structural validity, construct validity, convergent validity, criterion validity and responsiveness of the presenteeism instruments). Table A5 summarizes the methodological quality of studies per measurement property and instrument. The quality assessment results will be incorporated into the analysis of the measurement properties of presenteeism instruments described below.

Content validity

Evidence on the content validity of presenteeism instruments was provided in eight studies for seven instruments: ALWQ,[33] HLQ,[35] HPQ,[51] HWQ,[38] SPS-6,[1] VOLP,[89] and WPSI.[52,77] All the studies provided a clear description of measurement aims, target populations and concepts being measured with the instrument. Differences among the studies were identified in the strategies for item selection. Only one study[51] reported on an items selection strategy based on formal reviews of other existing scales, pilot interviews, expert opinion and pretested in a target population. Another study[1] combined a strategy involving literature reviews and expert-driven item reduction. The other studies[33,35,38,52,89] pretested the instrument on target samples but failed to describe how the items were chosen up to that point. Finally, one study[77] used an item-selection strategy based on comparisons of the instrument domains with data from medical claims.

The analysis of the combined information of the methodological quality of all studies on the content validity of presenteeism instruments showed that the best level of evidence (strong) of the quality of content validity of data is available for the SPS-6 and the VOLP.

Reliability (internal consistency, test-retest, inter-rater)

Internal consistency

The internal consistency of items included in presenteeism instruments was evaluated for 10 instruments in 17 studies (three studies[39,66,70] reported internal consistency data for two different presenteeism instruments). Internal consistency data was available for ALWQ,[33] EWPS,[34,66,70] HWQ,[38] LEAPS,[39] MIDAS,[82,83] MWPLQ,[69] SPS-13,[5] SPS-6,[1,7,66,73,84] WPS,[70,71,74] and WPSI.[52] In all cases, data on the internal consistency of the total scale was reported; however, internal consistency of subscores was reported for some instruments (i.e., HQW productivity,[38] MIDAS item on presenteeism,[83] MWPLQ subscales[69]). All the studies, except one[52] used Cronbach's α to describe the internal consistency of the scales. Where Cronbach's α was reported, it always exceeded the accepted standard ($\alpha>0.70$) for scales, as recommended by Nunnally and Bernstein.[90,91]

The analysis of the combined information of the methodological quality of all studies assessing the internal consistency of presenteeism instruments showed that the best level of evidence (strong) of the quality of internal consistency data is available for EWPS, HWQ, MWPLQ, and SPS-6.

Test-retest

Test-retest reliability of presenteeism instruments was evaluated for five instruments in six studies (with one study[47] reporting on two different instruments): EWPS,[34] MIDAS,[82,83] VOLP,[47] WPAI:IBS,[80] and WPAI:GH.[47,50] Only one study[34] reported the intraclass correlation coefficient (ICC) (considered the most suitable parameter for continuous reliability measures)[91] to describe the test-retest reliability of presenteeism scores. Another study[47] correctly reported weighted Kappa (k) for nominal presenteeism data according to a pre-established cut-off point, or the mean change in scores of repeated measurements.[50] The other studies[80,82,83] chose Pearson's or Spearman's (r) correlation coefficients, which are measures often considered inadequate for reliability analysis because systematic differences are not taken into account.[90] Where ICC was reported, the value exceeded the 0.70 standard;[91] however, because the sample size included less than 50 patients, it did not receive a positive rating. Kappa values were reported as moderate (k>0.60), whereas both Pearson's and Spearman's coefficients ranged between 0.60 (moderate) to 0.92 (excellent).[91]

The analysis of the combined information of the methodological quality of all studies assessing the test-retest reliability of presenteeism instruments showed that the best level of evidence (moderate level) of the quality of test-retest reliability data is available for the MIDAS.

Inter-rater

The evaluation of different schedules of administration of a presenteeism instrument (i.e., three different recall periods) was evaluated for only one instrument (i.e., WPSI) in one study.[52] The study reported the coefficient of variation for difference in scores among three different versions of the WPSI that differed according to the length of the recall period (i.e., 12 months, 3 months, or 2 weeks). The level of evidence of the quality of inter-rater reliability data is unknown (i.e., the study was rated as of poor quality).

Structural validity

Evidence on the structural validity of presenteeism instruments was provided in four studies for four instruments: HWQ,[38] LEAPS,[39] SPS-13,[5] and SPS-6.[1] The four studies used factor analysis to determine the structure and dimensionality of the instruments. The analysis of the combined information pertaining to the methodological quality of the studies showed that the best level of evidence (strong level) of the quality of structural validity data is available for the HWQ.

Construct validity

Evidence on the construct validity of presenteeism instruments was provided in 34 studies for 21 instruments (including six different versions of the WPAI) with three studies[47,66,75] reporting on two different instruments and two studies[12,70] reporting on three instruments: ALWQ,[33] EWPS,[34,66,70] HLQ,[12,35,75] HPQ,[12,85] HRPQ-D,[36] HWQ,[38] LEAPS,[39] MWPLQ,[69] Q-Q,[75] SPS-13,[5] SPS-6,[1,66,84] VOLP,[47] WPAI (WPAI:CD,[79] WPAI:CG,[72] WPAI:GERD,[86,87] WPAI:GH,[12,47,50,70,88] WPAI:IBS,[80] WPAI:SpA,[78] WPS-RA,[54] WPS,[70,72] and WPSI.[76,77]

Presenteeism instruments were compared to other non-presenteeism measures to examine the extent to which scores correlate in a manner that is consistent with theoretically derived hypotheses concerning the relationships between the constructs being measured. A variety of non-presenteeism measures were used for the comparisons. The most frequently used measures for comparison against presenteeism instruments were health status measures (used in 17 studies) such as the Short-Form 36 Health Survey (SF-36),[5,50,54,69-71,75,78,79,86,87] the Health Assessment Questionnaire (HAQ),[12,54,66]

the Multidimensional Health Assessment Questionnaire (MDHAQ),[47,88] and the General Health Perceptions section (GHP) from the Medical Outcome Study (MOS).[50] Scales measuring disability were used in 13 studies and included the Work Limitations Questionnaire (WLQ),[5,12,66,76,80,84] Sheehan Disability Scale (SDS),[39,70] Workplace Activity Limitations Scale (WALS),[66] Disabilities of the Arm, Shoulder and Hand Outcome measure (DASH),[84] Rheumatoid Arthritis Work Instability Scale (RA-WIS),[66] Dimensions of Daily Activities Index (DDAI)[80] and measures of difficulties performing certain tasks.[66] Measures of symptoms were used as comparator in eight studies and included Clinical Global Impressions (CGI),[50] Crohn's disease Activity Index (CDAI),[79] and a variety of frequency and/or severity scores.[33,36,50,71,80,86,87] Health-related quality of life questionnaires were used in nine studies and included: Quality of Life in Reflux and Dyspepsia (QOLRAD),[86,87] the Five-dimensional Euro-Qol health-related quality of life visual analogue scale (EQ5-D),[75,79] Bath Ankilosing Spondilitis Disease Activity Index (BASDAI),[78] Ankilosing Spondilitis Quality of Life Questionnaire (ASQOL),[78] Migraine Quality of Life Questionnaire (MQoLQ),[69] Asthma Quality of Life Questionnaire (AQLQ),[71] Health Utilities Index Mark 3 (HUI-3),[78] Inflammatory Bowel Disease Questionnaire (IBDQ),[79] and Caregiver Strain Index (CSI).[72] Measures of psychopathological symptoms and mental health status were used in three studies and included Symptom Checklist 90 (SCL90),[34] Hamilton Rating Scale for Depression (HAM-D),[34] Center for Epidemiologic Studies Depression Scale (CESD),[72] and one scale assessing job stress.[1] Other measures less frequently reported were those assessing attitudes (i.e., Patient Activation Measure [PAM][85]), job satisfaction,[1,66] efficiency and performance at work,[35,38] and health resources utilization.[77]

All the studies except five that examined differences in mean scores[54,75,78-80] used Pearson's or Spearman's (r) correlation coefficients to evaluate the association between presenteeism instruments and other measures. Overall, correlations between presenteeism instruments and measures of health status, quality of life, disability, mental health, and symptoms were moderate at best, with the absolute value of correlations (i.e., both negative and positive) floating between 0.20 (little, if any correlation)[60] and 0.60 (a moderate correlation). The highest correlations were found between: (1) the WPAI:GH and the MDHAQ Global assessment of disease activity (r = 076); (2) the WPAI:GH and a measure of disability (i.e., SDS; r = 0.71); and (3) the WPAI:GERD and QOLRAD (a physical/social functioning scale; r = -0.67).

The analysis of the combined information of the methodological quality of the studies showed that the best level of evidence (strong level) of the quality of construct validity data is available for the SPS-6, WPAI:GERD, and WPSI.

Convergent validity

Head-to-head comparisons among different presenteeism instruments aimed at establishing their convergent validity were reported for ten instruments in eleven different studies (four assessing two instruments[39,47,66,75] and two studies studies[12,67] assessing three instruments): EWPS,[66] HLQ,[12,67,75] HPQ,[12,39,51] LEAPS,[39] MIDAS,[81] Q-Q,[67,75] SPS-6,[1,66] VOLP,[47] WHI,[48] and WPAI:GH.[12,47,67,88] The following head-to-head comparisons among presenteeism instruments were made: EWPS versus SPS-6,[66] HLQ versus Q-Q,[67] HLQ versus Q-Q,[75] HLQ versus WPAI:GH,[67] HLQ versus HPQ,[12] HLQ versus WPAI:GH,[12] HLQ versus WPAI:GH,[88] HPQ versus LEAPS,[39] HPQ versus WPAI:GH,[12] MIDAS versus Productivity diary,[81] Q-Q versus WPAI:GH,[67] Q-Q versus WPAI:GH,[88] SPS-6 versus SPS-32,[1] and VOLP versus WPAI:GH.[47]

Additionally, some instruments were compared to other indicators of presenteeism and work performance (i.e., HPQ versus Experience Sample Methods [ESM] evaluations and archival work

performance measures,[51] and WHI versus workplace prospective continuous performance data and diary work time measures such as time not working at work[48]).

All the studies except three used Pearson's or Spearman's (r) correlation coefficients to evaluate the association between different presenteeism instruments. The other three studies used Kappa,[75]ICC,[12] and mean differences in scores.[51] Correlations between presenteeism instruments were moderate with absolute values ranging between 0.19 to 0.89. The highest positive correlations were found for the comparison between SPS-32 and SPS-6 (r = 0.89) followed by that between EWPS and SPS-36 (r = 0.58). The highest negative correlations were for the comparison between the positively valenced HPQ global work performance versus the negatively worded LEAPS work productivity subscore (r = -0.85) and for HPQ global work performance versus LEAPS total score (r = -0.79).

The analysis of the combined information on the methodological quality of the studies showed that the best level of evidence (strong level) of the quality of convergent validity data is available for the EWPS and the SPS-6.

Criterion validity

It is immediately obvious that only one study[38] assessed the criterion validity of the HWQ against a "gold standard" (i.e., hours of productivity loss). The study reported Pearson's correlation coefficients for objectives measures of the total of hours lost while at work and HWQ scores. In all comparisons, Pearson's correlation coefficients showed negative correlations that were below 0.20, indicating a weak negative relationship.[60] The level of evidence of the quality of criterion validity, however, is to deemed "unknown" because our quality audit indicated that the study in question was not of good enough quality to allow any conclusions to be drawn about this domain for that test.

Responsiveness

Responsiveness of presenteeism instruments to important changes in the construct over time were evaluated for seven instruments in eight studies (with one study[66] reporting on two different instruments): EWPS,[34,66] HRPQ-D,[36] SPS-6,[66] WPAI:GERD,[68,86] WPAI:CD,[79] WPAI:SpA,[78] and WPS-RA.[54] All the studies but two[34,36] reported standardized response means (SRM) or Cohen's effect sizes to assess the longitudinal validity of the instruments. The other two studies used Pearson's correlation coefficients to compare changes in presenteeism instruments versus a criterion of change.

The analysis of the combined information on the methodological quality of the studies showed that the best level of evidence (strong level) of the quality of responsiveness data is available for the SPS-6.

Table 2 provides a summary of the quality of measurement properties per instrument based on a best evidence synthesis of the combined information from all studies. Note that the WPAI parallel forms have been presented separately.

TABLE 2: QUALITY ANALYSIS AND SUMMARY SCORES FOR EACH PRESENTEEISM INSTRUMENT

Instrument	Criterion validity	Construct validity	Convergent validity	Content validity	Respon-siveness	Structural "validity"	Test-retest/ inter-rater	Internal consistency	Domains rated	Traditional weighting	High Quality
Weight	10	7	6	5	4	3	2	1	(of 8)	Rating x weight	i.e. # of ratings = 3
ALWQ		1		1				1	3	13	0
EWPS		2	3		2		0	3	5	43	2
HLQ		2	2	1					3	31	0
HPQ		2	2	1					3	31	0
HRPQ-D		1			1				2	11	0
HWQ	0	1		1		3		3	5	24	2
LEAPS		1	1			1		1	4	17	0
MIDAS			1				2	2	3	12	0
MWPLQ		1						3	2	10	1
Ost-Tech									0	0	0
Q-Q		1	2						2	19	0
SPS-32									0	0	0
SPS-13		1				1		1	3	11	0
SPS-6		3	3	3	3	1		3	6	72	5
VOLP		1	1	3			1		4	30	1
WHI			1						1	6	0
WPAI:CD		2			2				2	22	0
WPAI:CG		2							1	14	0
WPAI: GERD		3			2				2	29	1
WPAI:GH		2	2				1		3	28	0
WPAI:IBS		1					0		2	7	0
WPAI:SpA		1			1				2	11	0

Instrument	Criterion validity	Construct validity	Convergent validity	Content validity	Respon-siveness	Structural "validity"	Test-retest/ inter-rater	Internal consistency	Domains rated	Traditional weighting	High Quality
WPS-RA		1			1				2	11	0
WPS		2						2	2	16	0
WPSI		3		2			0	1	4	32	1
WRFQ									0	0	0
Coverage	**4%**	**81%**	**38%**	**27%**	**27%**	**15%**	**23%**	**38%**	**2.5**	**19.2**	**0.5**

3 = strong positive evidence; 2 = moderate positive evidence; 1 = limited positive evidence; 0 = unknown, due to poor methodological quality

ALWQ: Angina-Related Limitations at Work Questionnaire; EWPS: Endicott Work Productivity Scale; HLQ: Health and Labour Questionnaire; HPQ: World Health Organization Health and Work Performance Questionnaire; HRPQ-D: Health Related Productivity Questionnaire Diary; HWQ: Health and Work Questionnaire; LEAPS: Lam Employment Absence and Productivity Scale; MIDAS: Migraine Disability Assessment Questionnaire; MWPLQ: Migraine Work and Productivity Loss Questionnaire; Ost-Tech: Osterhaus Technique; Q-Q: Quantity and Quality method; SPS-32: Stanford Presenteeism Scale (32-item version); SPS-13: Stanford Presenteeism Scale (13-item version); SPS-6: Stanford Presenteeism Scale (6-item version); VOLP: Valuation of Lost Productivity questionnaire; WHI: Work and Health Interview; WPAI:CD: Work Productivity and Activity Impairment scale-Crohn's Disease; WPAI:CG: Work Productivity and Activity Impairment scale-Caregiver; WPAI:GERD: Work Productivity and Activity Impairment scale-Gastroesophageal Reflux Disease; WPAI:GH: Work Productivity and Activity Impairment scale-General Health; WPAI:IBS: Work Productivity and Activity Impairment scale-Irritable Bowel Syndrome; WPAI:SpA: Work Productivity and Activity Impairment scale-Psoriatic Arthritis; WPS-RA: Work Productivity Survey-Rheumatoid Arthritis; WPS: Work Performance Scale from the Functional Status Questionnaire; WPSI: Work Productivity Short Inventory; WRFQ: Work Role Functioning Questionnaire

Other characteristics: Interpretability and generalizability

Almost half of the studies[7,12,33,38,47,48,50,52,54,66,69,70,74,76,79-81,84,86,89] reported the percentage of items that were missing from the analysis, with 19 of them[7,33,34,36,38,48,66,69,70,74,76,79-81,84-87,89] describing how missing items were handled in the data analysis. The reporting of the distribution of scores in the target population across the studies was variable: 12 studies[5,12,33,38,51,66,73,78,79,83-85] described the distribution of the total scores of presenteeism instruments with eleven of them[38,47,51,54,66,74,78,83-85,88] reporting the percentage of the respondents who had both the lowest and highest possible (total) scores.

The majority of the studies described the mean age, distribution of gender and sociodemographic or other important clinical characteristics of included participants. All the studies except eight[35,50-52,69,77,78,80] describe the setting in which the studies were conducted. Finally, the response rate of study participants were reported in 30 studies.[1,5,7,12,33,36,47,48,50-52,54,67,69-75,78-84,87-89] Finally, the method used to select patients across the studies was variable: ten studies[1,7,39,51,66,67,73,84,87,89] used convenience samples, whereas eleven[50,52,54,69,72,78,79,81-83,85] used random study samples. Sampling methods where not described in 19 studies.[5,12,33-36,38,47,48,68,70,71,74-77,80,86,88]

CONCLUSIONS

There are two important lessons that can be taken from this exercise. First of all, the use of self-report tests to estimate levels of presenteeism has not been comprehensively investigated. The extant reviews, including this one, have shown that there is insufficient research to inform the choice of the best measure. Secondly, our present study has indicated that even those evaluations that do exist are often not of adequate methodological quality, thus weakening our confidence in their conclusions about test validity. Note that criterion validity, arguably the most important of the attributes under study here, was examined in only the case of a single test (the HWQ) and the study in question was deemed to be poor quality, preventing us from commenting on the merits of the HWQ. Thus, there is a virtual absence of "gold-standard" studies, meaning that none of the tests have actually been shown to predict performance.

Where does this leave us? A lack of coverage of the psychometric domains means that we cannot say whether self-report tests of presenteeism are useful or not. More assessments of the psychometric properties are needed. The suggestion here is that the focus should be on criterion validity studies – we posit that the study of other domains may be wasteful in the absence of even a glimmer of knowledge that presenteeism tests have any chance of accurately estimating real-life productivity.

At this juncture, then, studies examining the equally viable hypothesis that self-report instruments cannot adequately reflect performance due to recall limitations and our human tendency to overestimate our own abilities may also be needed to supplement our knowledge of presenteeism test behaviour. Furthermore, the possibility that mental health factors, like depression, may sometimes produce a biased finding, rather than denoting test validity, add a complication that is destined to lower precision even further. It may be that presenteeism tests, as a class, have very little to add to our ability to predict changes in work performance.

APPENDICES

Appendix A: Evidence Tables

Table A1: Use of presenteeism instruments in the scientific literature by study population and purpose

Table A2: Reported measurement properties per instrument: Internal consistency, reliability, content validity and structural validity

Table A3: Reported measurement properties per instrument: Hypothesis testing

Table A4: Criterion validity and responsiveness

Table A5: Methodological quality of studies per measurement property and presenteeism instrument

Appendix B: Search Strategies for the Identification of Studies

Appendix C: Characteristics of Presenteeism Instruments Included in the Review

Appendix A: Evidence Tables

TABLE A1: USE OF PRESENTEEISM INSTRUMENTS IN THE SCIENTIFIC LITERATURE BY STUDY POPULATION AND PURPOSE

Instrument	N times used in the studies	Populations (n of times instrument was used in study samples)	Purpose of measurement within the studies**		
			Descriptive/ discriminative	Predictive	Evaluative
ALWQ	1	Cardiovascular (1)	1	1	---
EWPS	9	Neurological (2); Mood/anxiety disorders (2); Sleep disorders (2); Musculoskeletal (1); Other mental disorders (1); Workers/general population (1)	5	4	6
HLQ	12	Musculoskeletal (6); Metabolic disorders (2); Workers/general population (2); Dermatological (1); Endocrine (1)	9	3	2
HPQ	29	Workers/general population (19); Mood/anxiety disorders (7); Neurological (2); Musculoskeletal (1)	18	14	5
HRPQ-D	1	Workers/general population (1)	---	1	---
HWQ	2	Immunological (1); Workers/general population (1)	2	---	1
LEAPS	1	Mood/anxiety disorders (1)	1	1	---
MIDAS	15	Neurological (14); Mood/anxiety disorders (1)	9	4	6
MWPLQ	1	Neurological (1)	1	1	---
Osterhaus technique	1	Metabolic disorders (1)	1	---	---
Q-Q	6	Musculoskeletal (4); Workers/general population (2)	5	3	---
SPS-32	0	---	---	---	---
SPS-13	4	Workers/general population (3); Mood/anxiety disorders (1)	3	1	---
SPS-6	7	Workers/general population (3); Mood/anxiety disorders (2); Musculoskeletal (2)	6	5	1
VOLP	2	Musculoskeletal (2)	2	1	---
WHI	4	Workers/general population (3); Mood/anxiety disorders (1)	3	1	---
WPAI	118	Gastrointestinal (29); Respiratory (18);Musculoskeletal (17); Dermatological (13); Mood/anxiety disorders (6); Genitourinary (5); Pain (5); Gynaecological (4); Liver diseases (3); Metabolic disorders (3); Neurological (3); Other mental disorders (3); Workers/general population	75	56	36

		(3); Caregivers (2); Sleep disorders (2); Cancer (1); Cardiovascular (1); Eyes and vision (1); Immunological (1)			
WPAI:AS	9	Respiratory (9)	1	1	7
WPAI:CD	3	Gastrointestinal (3)	1	---	3
WPAI:CG	2	Caregivers (2)	1	1	1
WPAI:ChHD	1	Dermatological (1)	1	1	1
WPAI:GERD	11	Gastrointestinal (11)	5	5	4
WPAI:GH	69	Musculoskeletal (12); Dermatological Respiratory (8); (7); Gastrointestinal (7); Mood/anxiety disorders (6); Pain (4); Gynaecological (3); Liver diseases (3); Metabolic disorders (3); Other mental disorders (3); Workers/general population (3); Genitourinary (2); Neurological (2); Sleep disorders (2); Cancer (1); Cardiovascular (1); Immunological (1); Eyes and vision (1)	55	41	10
WPAI:IBS	5	Gastrointestinal (5)	2	2	2
WPAI:PsO	4	Dermatological (4)	1	1	3
WPAI:RLS	1	Neurological (1)	1	1	---
WPAI:SHP	10	Genitourinary (3); Musculoskeletal (2); Dermatological (1); Gastrointestinal (1); Gynaecological (1); Pain (1); Respiratory (1)	6	3	3
WPAI:Sleep-GERD	2	Gastrointestinal (2)	---	---	2
WPAI:SpA	1	Musculoskeletal (1)	1	---	---
WPS	5	Cardiovascular (2); Respiratory (2); Mood/anxiety disorders (1);	4	2	1
WPSI	5	Workers/general population (3); Other mental disorders (1); Respiratory (1)	5	2	---
WPS-RA	4	Musculoskeletal (3); Dermatological (1)	2	---	3
WRFQ	0	---	---	---	---
Total	**227**		**227**	**156**	**97**

**The total number of times an instrument was used (n=227) does not match the number of times instruments were used for a particular purpose as, in many instances, presenteeism instruments were used for more than one purpose per study.

ALWQ: Angina-Related Limitations at Work Questionnaire; EWPS: Endicott Work Productivity Scale; HLQ: Health and Labour Questionnaire; HPQ: World Health Organization Health and Work Performance Questionnaire; HRPQ-D: Health Related Productivity Questionnaire Diary; HWQ: Health and Work Questionnaire; LEAPS: Lam Employment Absence and Productivity Scale; MIDAS: Migraine Disability Assessment Questionnaire; MWPLQ: Migraine Work and Productivity Loss Questionnaire; Q-Q: Quantity and Quality method; SPS-32: Stanford/American Health Association Presenteeism Scale 32-item version; SPS-13: Stanford Presenteeism Scale 13-item version; SPS-6: Stanford Presenteeism Scale 6-item version; VOLP: Valuation of Lost Productivity questionnaire; WHI: Work and Health Interview; WPAI: Work Productivity and Activity Impairment scale; WPS: Work Performance; WPS-RA: Work Productivity Survey-Rheumatoid Arthritis; WRFQ: Work Role Functioning Questionnaire

TABLE A2: REPORTED MEASUREMENT PROPERTIES PER INSTRUMENT: INTERNAL CONSISTENCY, RELIABILITY, CONTENT VALIDITY AND STRUCTURAL VALIDITY

Instrument and study	Population	Reliability			Content validity	Structural validity
		Internal consistency	Population test-retest	Results		
ALWQ Lerner et al., 1998[33]	N=40, chronic angina pectoris	Crombach's α=0.97	---	---	Pretested in a clinical sample	
EWPS Beaton et al., 2010[66]	N=250, rheumatoid arthritis or osteoarthritis	Crombach's α=0.94	---	---	---	
Endicott et al., 1997[34]	N=42, depression and anxiety	Crombach's α=0.93	N=42, depression and anxiety	Test-retest: ICC (EWPS total score)=0.92	---	
Erickson et al., 2009[70]	N=76, anxiety disorders	Crombach's α=0.95	---	---	---	
HLQ van Roijen et al., 1996[35]	---	---	---	---	Pretested in a clinical sample	
HPQ Kessler et al., 2003[51]	---	---	---	---	Review of other existing scales, pilot interviews, development of preliminary questions by survey experts, pretesting in population	
HWQ Shikiar et al., 2004[38]	N=294, workers (airlines)	Crombach's α: HWQ total score=0.81; HWQ productivity=0.96; HWQ productivity/own assessment=0.91; Productivity/other's assessment=0.96	---	---	Pretested in a clinical sample	Factor analysis
LEAPS Lam et al., 2009[39]	N=234, major depressive disorder	Crombach's α=0.89	---	---	---	Factor analysis

MIDAS Stewart et al., 1999[82]	N=97, migraine	Crombach's α: MIDAS overall score=0.76 (USA), 0.73 (UK)	N=97, migraine	Test-retest: Pearson's correlation (r): (1) MIDAS overall score 0.80 (USA), 0.83 (UK); (2) MIDAS reduced productivity at work 0.54 (USA), 0.75 (UK); Spearman's correlation (r): (3) MIDAS overall score 0.78 (USA), 0.77 (UK); (4) MIDAS reduced productivity at work 0.65 (USA), 0.85 (UK)	---	
Stewart et al., 1999[83]	N=177, migraine	Crombach's α: MIDAS - presenteeism=0.83	N=177, migraine	Test-retest: Spearman's correlation (r): MIDAS - presenteeism test-retest 0.71; Pearson's correlation (r): MIDAS - presenteeism test-retest 0.60	---	
MWPLQ Davies et al., 1999[69]	N=164, migraine	Crombach's α: MWPLQ subscales=0.67 to 0.91	---	---	---	
SPS-13 Turpin et al., 2004[5]	N=7,797, workers (general)	Crombach's α=0.82	---	---	---	Factor analysis
SPS-6 Beaton et al., 2010[66]	N=250, rheumatoid arthritis or osteoarthritis	Crombach's α=0.70	---	---	---	
Koopman et al., 2002[1]	N=175, workers (general)	Crombach's α=0.80	---	---	Literature review, expert-driven item reduction	Factor analysis
Lalic et al., 2012[73]	N=241, workers (manual)	Crombach's α=0.326	---	---	---	
Sanderson et al., 2007[7]	N=432, depression and anxiety	Crombach's α=0.70	---	---	---	
Tang et al., 2009[84]	N=80, shoulder and elbow injuries	Crombach's α=0.76	---	---	---	

VOLP Zhang et al., 2011[47]	---	---	N=152, rheumatoid arthritis	Test-retest: Kappa (0 vs. >0)=0.63	---	
Zhang et al., 2012[89]	---	---	---	---	Focus group and pretesting	
WPAI:IBS Reilly et al., 2004[80]	---	---	N=133, irritable bowel syndrome	Test-retest: Pearson's correlation (r) test-retest for impairment at work=0.98	---	
WPAI:GH Reilly et al., 1993[50]	---	---	N=106, workers (general)	Test-retest: WPAI:GH overall work productivity mean change (SD) Time 1 - Time=-3.9 (17.9)	---	
WPS Erickson et al., 2002[71]	N=369, asthma	Crombach's α=0.79	---	---	---	
McBurney et al., 2004[74]	N=89, myocardial infarction	Crombach's α=0.73	---	---	---	
WPSI Goetzel et al., 2003[52]	N=610, workers (general)	Kendall's tau range: 0.66 to 0.74	N=610, workers (general)	Inter-rater: Coefficient of variation for difference in values: (1) 2 wk=45.9%; (2) 3 mo=92.5% (p=0.02); (3) 12 mo= 70.8% (p=0.01)	Pretested in a clinical sample	
Ozminkowski et al., 2003[77]	---	---	---	---	Pretested; comparison of WPSI domains with data from medical claims and short-term disability	

ALWQ: Angina-Related Limitations at Work Questionnaire; EWPS: Endicott Work Productivity Scale; HLQ: Health and Labour Questionnaire; HPQ: World Health Organization Health and Work Performance Questionnaire; HWQ: Health and Work Questionnaire; ICC: intraclass correlation coefficient; LEAPS: Lam Employment Absence and Productivity Scale; MIDAS: Migraine Disability Assessment Questionnaire; mo: month(s); MWPLQ: Migraine Work and Productivity Loss Questionnaire; SD: standard deviation; SPS-13: Presenteeism Scale (13-item version); SPS-6: Stanford Presenteeism Scale (6-item version); VOLP: Valuation of Lost Productivity questionnaire; wk: week(s); WPAI:IBS: Work Productivity and Activity Impairment scale-Irritable Bowel Syndrome; WPAI:GH: Work Productivity and Activity Impairment scale-General Health; WPS: Work Performance Scale from the Functional Status Questionnaire; WPSI: Work Productivity Short Inventory

TABLE A3: REPORTED MEASUREMENT PROPERTIES PER INSTRUMENT: HYPOTHESIS TESTING

Instrument and study	Hypothesis testing				
	Population	Construct validity		Concurrent validity	
		Reference standard	Results	Reference standard	Results
ALWQ Lerner et al., 1998[33]	N=40, chronic angina pectoris	Chest pain frequency score; Rose Angina score	Spearman's correlation (r): ALWQ vs (1) Chest pain frequency score -0.35 to -0.63; (2) Rose Angina score 0.13 to 0.58	---	---
EWPS Beaton et al., 2010[66]	N=250, rheumatoid arthritis or osteoarthritis	Self-rated work productivity; perceived impact of health problems at work; self-rated difficulty doing work; satisfaction with occupational performance; self-rated ability to work; intrusion of arthritis on work ability; job performance in past week; HAQ (general perceived disability); arthritis severity; pain intensity over past week; WALS; RA-WIS; WLQ (overall, TM, PD, MI, OD)	Spearman's correlation (r): EWPS vs. (1) Self-rated work productivity=-0.54; (2) Perceived impact of health problems at work=0.64; (3) Self-rated difficulty doing work=0.58; (4) Satisfaction with occupational performance=-0.62; (5) Self-rated ability to work=-0.62; (6) Intrusion of arthritis on work ability=0.55; (7) Self-rated job performance in past week= -0.62; (8) HAQ (general perceived disability)=0.36; (9) Arthritis severity=0.36; (10) Pain intensity over past week=0.40; (11) WALS=0.55; (12) RA-WIS=0.64; (13) WLQ overall=0.61; (14) WLQ-TM= 0.45; (15) WLQ-PD=0.31; (16) WLQ-MI= 0.64; (17) WLQ-OD=0.59	SPS-6	Spearman's correlation (r): EWPS vs. SPS-6=0.58
Endicott et al., 1997[34]	N=42, depression and anxiety	CGI, Severity of Illness and Global Improvement scales; HAM-D; SCL-90	Pearson's correlation (r): EWPS vs. (1) HAM-D total score (intake)=0.27; (endpoint)=0.61; (2) Global Clinical Index of Severity (intake)=0.42; (endpont)=0.46; SCL-90 total score (intake)=0.55; (endpoint)=0.50	---	---
Erickson et al., 2009[70]	N=76, anxiety disorders	SF-36 role functioning scales; SDS work performance scale	Spearman's correlation (r): EWPS vs. (1) SF-36 role physical=-0.23; (2) SF-36 role emotional=-0.63; (3) SDS=0.63	---	---
HLQ Braakman-Jansen et al., 2012[67]	N=62, rheumatoid arthritis	---	---	Q-Q; WPAI:GH	Spearman's correlation (r): HLQ vs. Q-Q=0.34; HLQ vs. WPAI:GH=0.48

Meerding et al., 2005[75]	N=570, workers (manual)	EuroQol (EQ5D) summary measure; SF-12 (PCS, MCS)	Mean scores: HLQ reported work loss (yes, no): (1) EQ5D summary measure mean=0.79, 0.89 (p<0.05); (2) PCS-12= 42.8, 51.0 (p<0.05); (3) MCS-12=49.3, 54.0 (p<0.05)	Q-Q	Kappa work loss due to health problems HLQ vs. Q-Q=0.18 (95% CI 0.11–0.25)
van Roijen et al., 1996[35]	N=726, general population	Descriptive efficiency score	Pearson's correlation (r): HLQ vs. Descriptive efficiency score=0.41	---	---
Zhang et al., 2010[12]	N=212, rheumatoid arthritis or osteoarthritis	1) HAQ (overall score, pain score, arthritis severity score); 2) WLQ	1) Spearman's correlation (r): HLQ vs. (1) HAQ overall score=0.41; (2) HAQ pain score=0.25; (3) HAQ arthritis severity score=0.23 2) ICC (ρ) (95% CI): HLQ vs. WLQ=0.22 (0.08, 0.34)	HPQ; WPAI:GH	ICC (ρ) (95% CI): HLQ vs. (1) HPQ=0.16 (0.02, 0.29); (2) WPAI:GH=0.37 (0.25, 0.48)
Zhang et al., 2010[88]	N=150, rheumatoid arthritis	---	---	WPAI:GH	Spearman's correlation (r): HLQ vs. WPAI:GH=0.39
HPQ Kessler et al., 2003[51]	N=2,350, workers (general)	---	---	ESM evaluation; archival work performance measures	HPQ global ratings with lowest 20% of archival and ESM work performance outcome measures (OR; 95% CI): (1) HPQ score - reservation agent supervisor ratings: HPQ 0-7: OR=3.2 (1.3, 7.5); HPQ 8: OR=2.4 (1.1,5.2); HPQ 9: OR=1.0 (0.4, 2.3); HPQ 10: OR=1.0 (95% CI nr); (2) HPQ score - reservation agent ESM: HPQ 0-7: OR=6.4 (1.7, 24.0); HPQ 8: OR=1.6 (0.4, 6.1); HPQ 9: OR=2.2 (0.6, 8.2); HPQ 10: OR=1.0 (95% CI nr); (3) HPQ score-customer service representative ESM: HPQ 0-7: OR=7.3 (1.6, 33.0); HPQ 8: OR=2.8 (0.6, 13.2); HPQ 9: OR=1.6 (0.3, 8.0); HPQ 10: OR=1.0 (95% CI nr); (4) HPQ score-executive leadership scores:HPQ 0-7: OR=7.0 (1.3, 37.9); HPQ 8: OR=5.4 (1.2, 24.2); HPQ 9:

					OR=2.7 (0.6, 12.6); HPQ 10: OR=1.0 (95% CI nr); (5) HPQ score-railroad engineer performance actions: HPQ 0-7: OR=12.3 (1.3, 112.3); HPQ 8: OR=1.0 (95% CI nr); HPQ 9: OR=1.0 (95% CI nr); HPQ 10: OR=1.0 (95% CI nr); HPQ global ratings with highest 20% of archival and ESM work performance outcome measures (OR; 95% CI): (6) HPQ score-reservation agent supervisor ratings: HPQ 0-7: OR=1.0 (95% CI nr); HPQ 8: OR=5.7 (1.6, 20.1); HPQ 9: OR=3.8 (1.1, 13.1); HPQ 10: OR=5.4 (1.6, 19.4); (7) HPQ score-reservation agent ESM: HPQ 0-7: OR=1.0 (95% CI nr); HPQ 8: OR=3.7 (1.3, 10.7); HPQ 9: OR=4.4 (1.5, 13.1); HPQ 10: OR=6.4 (1.8,22.2); (8) HPQ score-customer service representative ESM: HPQ 0-7: OR=1.0 (95% CI nr); HPQ 8: OR=2.5 (2.8, 10.4); HPQ 9: OR=5.5 (2.8, 10.8); HPQ 10: OR=45.8 (11.4,184.7); (9) HPQ score-executive leadership scores: HPQ 0-7: OR=1.0 (95% CI nr); HPQ 8: OR=1.0 (0.3, 3.3); HPQ 9: OR=1.4 (0.4, 4.6); HPQ 10: OR=1.0 (0.3,4.2)

Lam et al., 2009[39]	N=234, major depressive disorder	---	---	LEAPS	Spearman's correlation (r): HPQ global work performance vs. (1) LEAPS total score=-0.79; (2) LEAPS work productivity subscore-0.85; HPQ 4 productivity items vs. (3) LEAPS total score=-0.70; (4) LEAPS work productivity subscore=-0.77; HPQ % of work hours missed in the past 2 weeks (single item) vs. (5) LEAPS total score 0.41; LEAPS work productivity subscore 0.45
Terry et al., 2010[85]	N=631, workers (manual)	PAM; health risk assessment (mental component score)	Pearson's correlation (r): HPQ vs. (1) PAM score=0.18; (2) Mental component score=0.26	---	---
Zhang et al., 2010[12]	N=212, rheumatoid arthritis or osteoarthritis	1) HAQ (overall score, pain score, arthritis severity score); 2) WLQ	1) Spearman's correlation (r): HPQ vs. (1) HAQ overall score=0.41; (2) HAQ pain score=0.25; (3) HAQ arthritis severity score=0.23 2) ICC (ρ) (95% CI): HPQ vs. WLQ=0.26 (0.13, 0.38)	HLQ; WPAI:GH	ICC (ρ) (95% CI): HLQ vs. (1) HPQ=0.16 (0.02, 0.29); (2) WPAI:GH=0.61 (0.51, 0.68)
HRPQ-D Kumar et al., 2003[36]	N=42, infectious mononucleosis	Patient symptom scores	Pearson's correlation (r): HRPQ-D presenteeism (work) vs. Patient symptom scores=-0.346	---	---
HWQ Shikiar et al., 2004[38]	N=294, workers (airlines)	Total performance points (evaluation of work quality)	Pearson's correlation (r): HWQ productivity vs. (1) Total performance points=0.060; (2) Productivity (self-assessed)=0.016; (3) Productivity (assessed by others)=0.090	---	---

LEAPS Lam et al., 2009[39]	N=234, major depressive disorder	SDS scores	Pearson's correlation (r): LAM total score vs. (1) SDS work score=0.63; LAM work productivity subscore vs. (2) SDS work score=0.50	HPQ (global work performance and 4 productivity items); % of work hours missed in the past 2 weeks (single item)	Spearman's correlation (r): LEAPS total score vs. (1) HPQ global work performance=-0.79; (2) 4 productivity items=-0.70; (3) % of work hours missed in the past 2 weeks (single item)=0.41; LEAPS work productivity subscore vs. (4) HPQ global work performance=-0.85; (5) 4 productivity items=-0.77; (6) % of work hours missed in the past 2 weeks (single item)=0.45
MIDAS Stewart et al., 2000[81]	N=144, migraine	---	---	Productivity diary	Spearman's correlation (r): MIDAS summary score vs. Equivalent diary score (days productivity was reduced by half or more)=0.42
MWPLQ Davies et al., 1999[69]	N=164, migraine	SF-36 (MCS, PCS, role physical, bodily pain); MQoLQ; headache severity (at 2 and 4 hr); functional disability (at 2 and 4 hr)	Spearman's correlation (r): MWPLQ total hours of work loss vs. (1) SF-36 MCS= -0.14; (2) SF-36 PCS=-0.32; (3) SF-36 role physical=-0.34; (4) SF-36 bodily pain =-0.34; (5) MQoLQ domains=-0.43, -0.60; (6) Headache severity (2 hr)=0.47; (7) Headache severity (4 hr)=0.43; (8) Functional disability (2 hr)=0.58; (9) Functional disability (4 hr)=0.46	---	---
Q-Q Braakman-Jansen et al., 2012[67]	N=62, rheumatoid arthritis	---	---	WPAI:GH; HLQ	Spearman's correlation (r): Q-Q vs. WPAI:GH=0.61; Q-Q vs. HLQ=0.34
Meerding et al., 2005[75]	N=570, workers (manual)	EuroQol (EQ5D) summary measure; SF-12 (PCS, MCS)	Mean scores: Q-Q reported work loss (yes, no): (1) EQ5D summary measure mean = 0.64, 0.912 ($p<0.05$); (2) PCS-12 = 38.5, 51.9 ($p<0.05$); (3) MCS-2 = 51.6, 54.0 ($p<0.05$)	HLQ	Kappa work loss due to health problems Q-Q vs. HLQ=0.18 (95% CI 0.11-0.25)
Zhang et al., 2010[88]	N=150, rheumatoid arthritis	---	---	WPAI:GH	Spearman's correlation (r): Q-Q vs. WPAI:GH=0.56

SPS-13 Turpin et al., 2004[5]	N=7,797, workers (general)	WLQ (overall, OD, MI, PD, TM); SF-36 (PCS, role limitations-physical, pain, general health, vitality, social functioning, MCS)	Pearson's correlation (r): SPS-13 work impairment score (WIS) vs. (1) WLQ overall=0.50; WLQ-OD=0.49; WLQ-MI= 0.41; WLQ-PD=0.37; WLQ-TM=0.20; SPS-13 work output score (WOS) vs. (2) WLQ overall=0.40; (3) SF-36 PCS= 0.25; SF-36 role limitations-physical=0.35; SF-36 pain=0.33; SF-36 general health= 0.45; SF-36 vitality=0.61; SF-36 social functioning=0.56; SF-36 role limitations-emotional=0.51; SF-36 MCS=0.62	---	---
SPS-6 Beaton et al., 2010[66]	N=250, rheumatoid arthritis or osteoarthritis	Self-rated work productivity; perceived impact of health problems at work; self-rated difficulty doing work; satisfaction with occupational performance; self-rated ability to work; intrusion of arthritis on work ability; job performance in past week; HAQ (general perceived disability); arthritis severity; pain intensity over past week; WALS; RA-WIS; WLQ (overall, TM, PD, MI, OD)	Spearman's correlation (r): SPS-6 vs. (1) Self-rated work productivity=-0.51; (2) Perceived impact of health problems at work=0.67; (3) Self-rated difficulty doing=0.59; (4) Satisfaction with occupational performance=-0.57; (5) Self-rated ability to work=-0.59; (6) Intrusion of arthritis on work ability=0.63; (7) Self-rated job performance in past week= -0.49; (8) HAQ (general perceived disability)=0.45; (9) Arthritis severity=0.56; (10) Pain intensity over past week=0.57; (11) WALS=0.66; (12) RA-WIS=0.69; (13) WLQ overall=0.63; (14) WLQ-TM= 0.47; (15) WLQ-PD=0.41; (16) WLQ-MI= 0.59; (17) WLQ-OD=0.61	EWPS	Spearman's correlation (r): SPS-6 vs. EWPS=0.58
Koopman et al., 2002[1]	N=175, workers (general)	Job satisfaction; job stress	Spearman's correlation (r): SPS-6 vs. (1) Job satisfaction=0.15; (2) Job stress= -0.22	SPS-32	Spearman's correlation (r): SPS-6 vs. SPS-32=0.89
Tang et al., 2009[84]	N=80, shoulder and elbow injuries	DASH (work module); WLQ (overall, TM, PD, MI, OD); RA-WIS	Spearman's correlation (r): SPS-6 vs. (1) DASH-W=-0.34; (2) WLQ overall= 0.64; (3) WLQ-TM=0.44; (4) WLQ-PD= 0.54; (5) WLQ-MI=0.58; (6) WLQ-OD= 0.58; (7) RA-WIS=-0.54	---	---
VOLP Zhang et al., 2011[47]	N=152, rheumatoid arthritis	MDHAQ (function, pain, health impact, fatigue, global assessment of disease activity)	Spearman's correlation (r): VOLP vs. (1) MDHAQ function=0.39; (2) MDHAQ pain=0.25; (3) MDHAQ health impact= 0.29; (4) MDHAQ fatigue=0.34; (5) MDHAQ global assessment of disease activity=0.27	WPAI:GH	Spearman's correlation (r): VOLP vs. WPAI:GH=0.42

WHI Stewart et al., 2004[48]	N=67, workers (call center)	---	---	Workplace prospective continuous performance data (unavailable time at work); Diary work time measures (time not working at work)	Pearson's correlation (r): WHI-presenteeism vs. (1) Unavailable time at work=0.29 (SE=0.16); (2) Time not working at work=0.19 (SE=0.12); Spearman's correlation (r): WHI-presenteeism vs. (3) Unavailable time at work=0.31 (SE=0.13); (4) Time not working at work=0.33 (SE=0.13)
WPAI:CD Reilly et al., 2008[79]	N=380, Crohn's disease	CDAI; SF-36 (PCS, MCS); IBDQ; EQ-VAS	Difference worst/best WPAI:CD score vs. (1) CDAI=-9.7 ($p<0.05$); (2) IBDQ=-23.0 ($p<0.05$); (3) SF-36 PCS=-22.3 ($p<0.05$); (4) SF-36 MCS=-16.3 ($p<0.05$); (5) EQ-VAS=-15.7 ($p<0.05$)	---	---
WPAI:CG Giovannetti et al., 2009[72]	N=308, caregivers	CSI; CESD; number of hours spent caregiving in typical week	Spearman's correlation (r): WPAI work productivity loss vs. (1) CSI=0.45; (2) CESD=0.30; (3) Number of hours spent caregiving in typical week=0.32	---	---
WPAI:GERD Wahlqvist et al., 2002[87]	N=136, gastroesophageal reflux disease	QOLRAD (emotional distress, sleep disturbance, food and drinking problems, physical/social functioning, vitality); SF-36 (PCS, role-physical, bodily pain, general health, vitality, social functioning; role-emotional, MCS); heartburn (severity, frequency), regurgitation (severity, frequency)	Pearson's correlation (r): LWPS total score vs. (1) QOLRAD (emotional distress)=-0.59, QOLRAD (sleep disturbance)=-0.50; QOLRAD (food and drinking problems)=-0.43, QOLRAD (physical/social functioning)=-0.67, QOLRAD (vitality)=-0.54; (2) SF-36 PCS= -0.24, SF-36 role-physical=-0.49, SF-36 bodily pain=-0.40, SF-36 general health= -0.09, SF-36 vitality=-0.43, SF-36 social functioning=-0.50, SF-36 role-emotional= -0.36, SF-36 mental health=-0.34; (3) Heartburn severity=0.42, heartburn frequency=0.19, regurgitation severity= 0.30, regurgitation frequency=0.19; (4) LWPS-average reduced PW vs. QOLRAD (emotional distress)=-0.58, QOLRAD (sleep disturbance)=-0.53; QOLRAD (food and drinking problems)= -0.44, QOLRAD (physical/social	---	---

			functioning)=-0.67, QOLRAD (vitality)= -0.53; (5) SF-36 PCS=-0.32, SF-36 role-physical=-0.51, SF-36 bodily pain=-0.43, SF-36 general health=-0.16, SF-36 vitality =-0.46, SF-36 social functioning=-0.49, SF-36 role-emotional=-0.35, SF-36 MCS= -0.36; (6) Heartburn severity=0.39, heartburn frequency=0.14, regurgitation severity=0.28, regurgitation frequency= 0.16; LWPS-PA vs. (7) QOLRAD (emotional distress)=-0.66, QOLRAD (sleep disturbance)=-0.57; QOLRAD (food and drinking problems)=-0.59, QOLRAD (physical/social functioning)=-0.75, QOLRAD (vitality)=-0.65; (8) SF-36 PCS= -0.24, SF-36 role-physical=-0.44, SF-36 bodily pain=-0.43, SF-36 general health= -0.17, SF-36 vitality=-0.47, SF-36 social functioning=-0.56, SF-36 role-emotional= -0.35, SF-36-MCS=-0.38; (9) Heartburn severity=0.51, heartburn frequency=0.36, regurgitation severity=0.43, regurgitation frequency=0.37		
Wahlqvist et al., 2007[86]	N=130, gastroesophageal reflux disease	SF-36 (PCS, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, MCS); QOLRAD (emotional distress, sleep disturbance, food/drinking problems); symptom severity	Pearson's correlation (r): WPAI:GERD reduced productivity at work vs. (1) SF-36 PCS=0.41; SF-36 role-physical=0.54; SF-36 bodily pain=0.52; SF-36 general health=0.31; SF-36 vitality=0.42; SF-36 social functioning=0.54; SF-36 role-emotional=0.43; SF-36 mental health= 0.34; (2) QOLRAD (emotional distress)= 0.47; QOLRAD (sleep disturbance)= 0.30; QOLRAD (food/drinking problems)= 0.36; QOLRAD (physical/social functioning)=0.52; QOLRAD (vitality)= 0.52; (3) Symptom severity-stomach pain=0.30; symptom severity-heatburn= 0.27; symptom severity-belching=0.18; symptom severity-acid reflux=0.24; symptom severity-overall symptom=0.21; symptom severity-acid reflux=0.24	---	---

WPAI:GH Braakman-Jansen et al., 2012[67]	N=62, rheumatoid arthritis	---	---	Q-Q; HLQ	Spearman's correlation (r): WPAI:GH vs. (1) Q-Q=0.61; (2) HLQ =0.48
Erickson et al., 2009[70]	N=76, anxiety disorders	SF-36 role functioning scales; SDS work performance scale	Spearman's correlation (r): (1) SF-36 role-physical vs. WPAI percent missed due to health=0.04; vs. WPAI percent impairment while working=-0.36; vs. WPAI final overall impairment due to health=-0.28; vs. WPAI percent activity impairment due to health=-0.49; (2) SF-36 role-emotional vs.WPAI percent missed due to health=-0.45; vs. WPAI percent impairment while working=-0.48; vs. WPAI final overall impairment due to health=-0.57; vs. WPAI percent activity impairment due to health=-0.50; (3) SDS vs. WPAI percent missed due to health=0.46; vs. WPAI percent impairment while working=0.68; vs. WPAI final overall impairment due to health=0.71; vs. WPAI percent activity impairment due to health=0.55	---	---
Reilly et al., 1993[50]	N=106, workers (general)	GHP from the MOS; SF-36 (PCS; role-emotional; -pain); symptom severity	Pearson's correlation (r): WPAI overall score vs.(1) GHP=0.52; (2) SF-36 role-physical=0.52; (3) SF-36 role-emotional= 0.30; (4) SF-36 pain=0.20; (5) Symptom severity score=0.27; (6) Symptom interference in work=0.62; (7) Symptom interference in regular activities=0.65	---	---
Zhang et al., 2011[47]	N=152, rheumatoid arthritis	---	---	VOLP	Spearman's correlation (r): WPAI:GH vs. VOLP=0.42
Zhang et al., 2010[12]	N=212, rheumatoid arthritis or osteoarthritis	1) HAQ (overall score, pain score, arthritis severity score); 2) WLQ	1) Spearman's correlation (r): WPAI:GH vs. HAQ overall score=0.15; HAQ pain score=0.20; HAQ arthritis severity score= 0.20 2) ICC(ρ) (95% CI): WPAI:GH vs. WLQ= 0.30 (0.17, 0.41)	HLQ; HPQ	ICC (ρ) (95% CI): WPAI:GH vs. (1) HLQ=0.37 (0.25, 0.48); (2) HPQ =0.61 (0.51, 0.68)

Zhang et al., 2010[88]	N=150, rheumatoid arthritis	MDHAQ (function, pain, health impact, fatigue, global assessment of disease activity)	Spearman's correlation (r): WPAI:GH vs. MDHAQ function=0.69; MDHAQ pain= 0.75; MDHAQ health impact=0.74; MDHAQ fatigue=0.67; MDHAQ global assessment of disease activity=0.76	Q-Q; HLQ	Spearman's correlation (r): WPAI:GH vs. (1) Q-Q=0.56; (2) HLQ=0.39
WPAI:IBS Reilly et al., 2004[80]	N=133, irritable bowel syndrome	Debriefing questionnaire (symptoms); retrospective diary (symptoms); WLQ; DDAI	WPAI:IBS vs. IBS symptom severity (level): p=0.03; vs. symptom severity (VAS): p=0.0002; vs. symptom severity (distress): p<0.0001; vs. WLQ index: p<0.001; vs. DDAI index: NA	---	---
WPAI:SpA Reilly et al., 2010[78]	N=205, ankylosing spondylitis	BASDAI; ASQOL; SF-36 (PCS, MCS); HUI-3	Difference best-worse WPAI:SpA score vs. (1) BASDAI=-20.3 (p<0.001); (2) ASQOL=-22.6 (p<0.001); (3) SF-36 PCS=-27.0 (p<0.001); (4) SF-36 MCS= -11.4 (p<0.001); (5) HUI-3=-19.5 (p<0.001)	---	---
WPS-RA Osterhaus et al., 2009[54]	N=220, rheumatoid arthritis	HAQ-DI; SF-36 (PCS, MCS)	Mean WPS-RA presenteeism scores (SD) vs. (1) HAQ-DI (cut-off 0.5 and 1.5) best= 3.4 (6.88), worst=6.8 (8.96); (2) SF-36 PCS (cut-off 21.76 and 35.26): best=2.2 (4.60), worst=8.7 (10.02); (3) SF-36 MCS (cut-off 38.36 and 54.67): best=4.0 (7.83), worst=10.6 (11.07)	---	---
WPS Erickson et al., 2002[71]	N=369, asthma	AQLQ; SF-36 (role-physical , role-emotional), perceived asthma severity; symptom-derived severity	Pearson's correlation (r): WPS vs. (1) SF-36 role-physical=0.54; (2) SF-36 role-emotional=0.44; (3) AQLQ limitations=0.57; (4) Perceived asthma severity=-0.36; (5) Symptom-derived severity=-0.29	---	---
Erickson et al., 2009[70]	N=76, anxiety disorders	SF-36 (role functioning); SDS work performance	Spearman's correlation (r): WPS vs. (1) SF-36 role-physical=0.34; (2) SF-36 role-emotional=0.66; (3) SDS work performance=0.46	---	---
WPSI Ozminkowski et al., 2003[77]	N=206, workers (general)	Use of medical care; use of short-term disability programs	Pearson's correlation (r): WPSI vs. (1) Use of medical care=0.58; (2) Use of short-term disability programs=range: 0.01 to 0.23	---	---

Ozminkowski et al., 2004[76]	N=532, workers (telecom-munications)	WLQ (overall, TM, PD, MI, OD)	Spearman's correlation (r): WPSI (presenteeism loss as a % of total work days in recall period) vs. (1) WLQ-TM= 0.32; (2) WLQ-MI=0.26; (3) WLQ-OD= 0.24; (4) WLQ-PD=0.23; (5) WLQ work productivity loss percentage=0.30	---	---

95% CI: 95% confidence interval; ALWQ: Angina-Related Limitations at Work Questionnaire; AQLQ: Asthma Quality of Life Questionnaire; ASQOL: Ankilosing Spondilitis Quality of Life Questionnaire; BASDAI: Bath Ankilosing Spondilitis Disease Activity Index; CGI: Clinical Global Impressions; CESD: Center for Epidemiologic Studies Depression Scale; CDAI: Crohn's disease Activity Index; CSI: Caregiver Strain Index; DASH: Disabilities of the Arm, Shoulder and Hand Outcome measure; DDAI: Dimensions of Daily Activities Index; EQ-5D: Five-dimensional Euro-Qol health-related quality of life; EQ-VAS: Five-dimensional Euro-Qol health-related quality of life visual analogue; ESM: Experience Sample Method; EWPS: Endicott Work Productivity Scale; GHP: General Health Perceptions; h: hour(s); HAM-D: Hamilton Rating Scale for Depression; HAQ: Health Assessment Questionnaire; HAQ-DI: Health Assessment Questionnaire Disability Index; HLQ: Health and Labour Questionnaire; HPQ: World Health Organization Health and Work Performance Questionnaire; HRPQ-D: Health Related Productivity Questionnaire Diary; HUI-3: Health Utilities Index Mark 3; HWQ: Health and Work Questionnaire; IBDQ: Inflammatory Bowel Disease Questionnaire; LEAPS: Lam Employment Absence and Productivity Scale; LWPS: Lost Workplace Productivity Score; MCS: Mental Component Summary; MDHAQ: Multidimensional Health Assessment Questionnaire; MI: mental/interpersonal demands; MIDAS: Migraine Disability Assessment Questionnaire; MOS: medical outcome study; MQoLQ: Migraine Quality of Life Questionnaire; MWPLQ: Migraine Work and Productivity Loss Questionnaire; nr: not reported; OD: output demands scale; OR: odds ratio; PA: productivity while doing regular daily activities; PAM: Patient activation measure; PCS: Physical Component Summary; PD: physical demands; PW: productivity while working; QOLRAD: Quality of Life in Reflux and Dyspepsia; Q-Q: Quantity and Quality method; RA-WIS: Rheumatoid Arthritis Work Instability Scale; SCL90: Symptom Checklist 90; SDS: Sheehan Disability Scale; SF-36: Short-Form 36 Health Survey; SPS-13: Presenteeism Scale (13-item version); SPS-6: Stanford Presenteeism Scale (6-item version); TM: time management; VOLP: Valuation of Lost Productivity questionnaire; WALS: Workplace Activity Limitations Scale; WHI: Work and Health Interview; WLQ: Work Limitations Questionnaire; WPAI:CD: Work Productivity and Activity Impairment scale-Crohn's Disease; WPAI:CG: Work Productivity and Activity Impairment scale-Caregiver; WPAI:GERD: Work Productivity and Activity Impairment scale-Gastroesophageal Reflux Disease; WPAI:GH: Work Productivity and Activity Impairment scale-General Health; WPAI:IBS: Work Productivity and Activity Impairment scale-Irritable Bowel Syndrome; WPAI:GH: Work Productivity and Activity Impairment scale-General Health; WPAI:SpA: Work Productivity and Activity Impairment scale-Psoriatic Arthritis; WPS-RA: Work Productivity Survey-Rheumatoid Arthritis; WPS: Work Performance Scale from the Functional Status Questionnaire; WPSI: Work Productivity Short Inventory

TABLE A4: CRITERION VALIDITY AND RESPONSIVENESS

Instrument and study	Population	Criterion validity	Responsiveness
EWPS Beaton et al., 2010[66]	N=250, rheumatoid arthritis or osteoarthritis	---	EWPS: 1) Improved: effect size=-0.15; SRM=-0.15; 2) Deteriorated: effect size=0.45; SRM=0.49
Endicott et al., 1997[34]	N=42, depression and anxiety	---	Pearson's correlation (r): 1) Change in EWPS scores vs. change in HAM-D scores=0.29
HRPQ-D Kumar et al., 2003[36]	N=42, infectious mononucleosis	---	Pearson's correlation (r): 1) Change in symptom scores vs. change in presenteeism scores (weeks 1-8)=-0.161
HWQ Shikiar et al., 2004[38]	N=294, workers (airlines)	Pearson's correlation (r): 1) Hours lost vs. HWQ Productivity=-0.165; vs. Productivity - self-assessed=-0.195; vs. Productivity - assessed by others=-0.219	---
SPS-6 Beaton et al., 2010[66]	N=250, rheumatoid arthritis or osteoarthritis	---	Responsiveness indices of at-work productivity measures against global indicators of change: SPS-6: 1) Improved: effect size=-0.48; SRM=-0.42; 2) Deteriorated: effect size=0.32; SRM=0.38
WPAI:CD Reilly et al., 2008[79]	N=380, Crohn's disease	---	WPAI:CD SRM=-13.0 ($p<0.05$) (CDAI as criterion of remission status); -12.8 (0.05) (CDAI as criterion of response status)
WPAI:GERD Brozek et al., 2009[68]	N=217, GERD	---	WPAI:GERD percentage of work days with GERD (SRM=2.02, 95% CI: 1.84, 2.19); WPAI:GERD related to hours absent at work (SRM=0.22, 95% CI: 0.05, 0.38), WPAI:GERD reduced productivity at work (SRM=0.66, 95% CI: 0.48, 0.83); WPAI:GERD reduced productivity during other activities (SRM=0.78, 95% CI: 0.65, 0.92)
Wahlqvist et al., 2007[86]	N=130, GERD	---	WPAI:GERD percentage reduced productivity: effect size=0.64 (95% CI=0.47, 0.81)
WPAI:SpA Reilly et al., 2010[78]	N=205, ankylosing spondylitis	---	WPAI:SpA SRM=-0.5 to -0.86 (BASDAI as criterion); -0.46 to -0.89 (ASQOL as criterion)
WPS-RA Osterhaus et al., 2009[54]	N=220, rheumatoid arthritis	---	WPS-RA SRM: 0.48 to 1.12 ($p<0.05$) (ACR20 and HAQ-D as criterion of improvement) at 24 wks

95% CI: 95% confidence interval; ACR20: American College of Rheumatology 20% improvement criteria; ASQOL: Ankylosing Spondylitis Quality of Life; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; CDAI: Crohn's Disease Activity Index ; EWPS: Endicott Work Productivity Scale; GERD: Gastroesophageal Reflux Disease; HAM-D: Hamilton Depression Rating Scale; HRPQ-D: Health Related Productivity Questionnaire Diary; HWQ: Health and Work Questionnaire; SPS-6: Stanford Presenteeism Scale (6-item version); SRM: standardized response mean; VOLP: Valuation of Lost Productivity questionnaire; WPAI:CD: Work Productivity and Activity Impairment scale-Irritable Crohn's Disease; WPAI:GERD: Work Productivity and Activity Impairment scale-Gastroesophageal Reflux Disease; WPAI:SpA: Work Productivity and Activity Impairment scale-Psoriatic Arthritis WPS-RA: Work Productivity Survey-Rheumatoid Arthritis

TABLE A5: METHODOLOGICAL QUALITY OF STUDIES PER MEASUREMENT PROPERTY AND PRESENTEEISM INSTRUMENT

Instrument and study	Internal consistency	Reliability	Content validity	Structural validity	Hypothesis testing	Criterion validity	Responsiveness
ALWQ							
Lerner et al., 1998[33]	Fair	---	Fair	---	Fair	---	---
EWPS							
Beaton et al., 2010[66]	Excellent	---	---	---	Excellent	---	Excellent
Endicott et al., 1997[34]	Fair	Poor	---	---	Fair	---	Poor
Erickson et al., 2009[70]	Good	---	---	---	Fair	---	---
HLQ							
Braakman-Jansen et al., 2012[67]	---	---	---	---	Fair	---	---
Meerding et al., 2005[75]	---	---	---	---	Fair	---	---
van Roijen et al., 1996[35]	---	---	Fair	---	Fair	---	---
Zhang et al., 2010[12]	---	---	---	---	Fair	---	---
Zhang et al., 2010[88]	---	---	---	---	Fair	---	---
HPQ							
Terry et al., 2010[85]	---	---	---	---	Fair	---	---
Kessler et al., 2003[51]	---	---	Fair	---	Fair	---	---
Lam et al., 2009[39]	---	---	---	---	Fair	---	---

Zhang et al., 2010[12]	---	---	---	---	Fair	---	---
HRPQ-D							
Kumar et al., 2003[36]	---	---	---	---	Fair	---	Poor
HWQ							
Shikiar et al., 2004[38]	Excellent	---	Fair	Excellent	Fair	Poor	---
LEAPS							
Lam et al., 2009[39]	Fair	---	---	Fair	Fair	---	---
MIDAS							
Stewart et al., 2000[81]	---	---	---	---	Fair	---	---
Stewart et al., 1999[82]	Fair	Fair	---	---	---	---	---
Stewart et al., 1999[83]	Fair	Fair	---	---	---	---	---
MWPLQ							
Davies et al., 1999[69]	Excellent	---	---	---	Fair	---	---
Q-Q							
Braakman-Jansen et al., 2012[67]	---	---	---	---	Fair	---	---
Meerding et al., 2005[75]	---	---	---	---	Fair	---	---
Zhang et al., 2010[88]	---	---	---	---	Fair	---	---
SPS-13							
Turpin et al., 2004[5]	Fair	---	---	Fair	Fair	---	---

SPS-6							
Beaton et al., 2010[66]	Excellent	---	---	---	Excellent	---	Excellent
Tang et al., 2009[84]	Good	---	---	---	Good	---	---
Lalic et al., 2012[73]	Fair	---	---	---	---	---	---
Koopman et al., 2002[1]	Fair	---	Excellent	Fair	Fair	---	---
Sanderson et al., 2007[7]	Excellent	---	---	---	---	---	---
VOLP							
Zhang et al., 2012[89]	---	---	Excellent	---	---	---	---
Zhang et al., 2011[47]	---	Fair	---	---	Fair	---	---
WHI							
Stewart et al., 2004[48]	---	---	---	---	Fair	---	---
WPAI:CD							
Reilly et al., 2008[79]	---	---	---	---	Good	---	Good
WPAI:CG							
Giovannetti et al., 2009[72]	---	---	---	---	Good	---	---
WPAI:GERD							
Brozek et al., 2009[68]	---	---	---	---	---	---	Fair
Wahlqvist et al., 2007[86]	---	---	---	---	Good	---	Fair
Wahlqvist et al., 2002[87]	---	---	---	---	Good	---	---

WPAI:GH							
Braakman-Jansen et al., 2012[67]	---	---	---	---	Fair	---	---
Erickson et al., 2009[70]	---	---	---	---	Fair	---	---
Reilly et al., 1993[50]	---	Fair	---	---	Good	---	---
Zhang et al., 2011[47]	---	---	---	---	Fair	---	---
Zhang et al., 2010[12]	---	---	---	---	Fair	---	---
Zhang et al., 2010[88]	---	---	---	---	Fair	---	---
WPAI:IBS							
Reilly et al., 2004[80]	---	Poor	---	---	Fair	---	---
WPAI:SpA							
Reilly et al., 2010[78]	---	---	---	---	Fair	---	Fair
WPS-RA							
Osterhaus et al., 2009[54]	---	---	---	---	Fair	---	Fair
WPS							
Erickson et al., 2002[71]	Fair	---	---	---	Fair	---	---
Erickson et al., 2009[70]	Good	---	---	---	Fair	---	---
McBurney et al., 2004[74]	Fair	---	---	---	---	---	---
WPSI							
Goetzel et al., 2003[52]	Fair	Poor	Fair	---	---	---	---

Ozminkowski et al., 2004[76]	---	---	---	---	Good	---	---
Ozminkowski et al., 2003[77]	---	---	Fair	---	Good	---	---

ALWQ: Angina-Related Limitations at Work Questionnaire; EWPS: Endicott Work Productivity Scale; HLQ: Health and Labour Questionnaire; HPQ: World Health Organization Health and Work Performance Questionnaire; HRPQ-D: Health Related Productivity Questionnaire Diary; HWQ: Health and Work Questionnaire; LEAPS: Lam Employment Absence and Productivity Scale; MIDAS: Migraine Disability Assessment Questionnaire; MWPLQ: Migraine Work and Productivity Loss Questionnaire; Q-Q: Quantity and Quality method; SPS-13: Presenteeism Scale (13-item version); SPS-6: Stanford Presenteeism Scale (6-item version); VOLP: Valuation of Lost Productivity questionnaire; WHI: Work and Health Interview; WPAI:CD: Work Productivity and Activity Impairment scale-Crohn's Disease; WPAI:CG: Work Productivity and Activity Impairment scale-Caregiver; WPAI:GERD: Work Productivity and Activity Impairment scale-Gastroesophageal Reflux Disease; WPAI:GH: Work Productivity and Activity Impairment scale-General Health; WPAI:IBS: Work Productivity and Activity Impairment scale-Irritable Bowel Syndrome; WPAI:GH: Work Productivity and Activity Impairment scale-General Health; WPAI:SpA: Work Productivity and Activity Impairment scale-Psoriatic Arthritis; WPS-RA: Work Productivity Survey-Rheumatoid Arthritis; WPS: Work Performance Scale from the Functional Status Questionnaire; WPSI: Work Productivity Short Inventory

Appendix B: Search Strategies for the Identification of Studies

All databases were searched last on 11 October 2012.

Search strategy for the following OVID Databases

- Medline (includes in-process citations) (1946 to 11 October 2012)
- Embase (1974 to 10 October 2012)
- CENTRAL (September 2012)
- PsycINFO (1806 to week 1 of October 2012)

1. American Productivity Audit.tw.
2. "angina-related limitations at work question?aire".tw.
3. endicott work productivity scale.tw.
4. "health and labo?r question?aire".tw.
5. "health and performance question?aire".tw.
6. "health and productivity question?aire".tw.
7. "health and work question?aire".tw.
8. health-related productivity question?aire*.tw.
9. "lam employment absence and productivity scale".tw.
10. (migraine disability assessment adj2 (question?aire or survey or scale or score*)).tw.
11. MIDAS.tw. and migraine*.mp.
12. 10 or 11
13. (productiv* or presenteeism or absenteeism or work* or employ*).mp.
14. 12 and 13
15. "migraine work and productivity loss question?aire".tw.
16. (osterhaus and (work* or productivity or presenteeism)).tw.
17. (osterhaus adj3 technique).tw.
18. "productivity and disease question?aire".tw.
19. PRODISQ.tw.
20. (quantity adj2 quality adj (method or instrument)).tw.
21. (Stanford* adj5 Presenteeism Scale).tw.
22. "work and health interview".tw.
23. work performance scale.tw.
24. "work productivity and activity impairment*".tw.
25. WPAI*.tw.
26. (US National Health and Wellness Survey).tw. and (productiv* or presenteeism or absenteeism or work*).mp.
27. "health and work performance question?aire".tw.
28. work productivity short inventory.tw.
29. wellness inventory.tw.
30. work role functioning question?aire.tw.
31. (or/1-9) or (or/14-30)
32. limit 31 to english language
33. (work productivity survey or WPS-RA).tw.
34. valuation of lost productivity.tw.
35. work limitations question?aire.tw.
36. (33 or 34) not 32
37. limit 36 to english language
38. 35 not 32
39. limit 38 to english language
40. 32 or 37
41. 32 or 37 or 39

Search strategy for the ISI platform databases

Web of Science (1900 to 11 October 2012)

1 TS="angina-related limitations at work questionnaire"
2 ((TS="endicott work productivity scale")) AND Language=(English)
3 (TS=("health and labor questionnaire" OR "health and labour questionnaire")) AND Language=(English)
4 ((TS="health and productivity questionnaire")) AND Language=(English)
5 ((TS="health and work questionnaire")) AND Language=(English)
6 ((TS="health-related productivity questionnaire")) AND Language=(English)
7 ((TS="lam employment absence and productivity scale")) AND Language=(English)
8 ((TS="migraine work and productivity loss questionnaire")) AND Language=(English)
9 ((TS="Stanford Presenteeism Scale")) AND Language=(English)
10 ((TS="health and work performance questionnaire")) AND Language=(English)
11 ((TS="work and health interview")) AND Language=(English)
12 ((TS="work performance scale")) AND Language=(English)
13 ((TS="work productivity short inventory")) AND Language=(English)
14 ((TS="American Productivity Audit")) AND Language=(English)
15 ((TS=(osterhaus and (work or productivity)))) AND Language=(English)
16 TS=("work productivity and activity impairment*")
17 TS=WPAI*
18 TS="wellness inventory"
19 TS=("work productivity survey")
20 TS=("Work role functioning*" SAME limitations)
21 TS=(osterhaus SAME technique)
22 TS=("productivity and disease questionnaire")
23 (TS=(migraine disability assessment SAME (score* OR scale OR questionnaire OR survey))) AND Language=(English)
24 (TS=(MIDAS AND migraine)) AND Language=(English)
25 (TS=Osterhaus) AND Language=(English)
26 (#23 OR #24 OR #25) AND Language=(English)
27 (TS = (work* or productivity or performance or presenteeism or absenteeism or employ*)) AND Language=(English)
28 (#26 AND #27) AND Language=(English)
29 (TS=("work role functioning questionnaire")) AND Language=(English)
30 (TS="health and performance questionnaire") AND Language=(English)
31 TS=("valuation of lost productivity")
32 TS=("work limitations questionnaire")
33 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #28 OR #29 OR #30 OR #31 OR #32) AND Language=(English)

Search strategy for the following EBSCO Databases

CINAHL (1937 to 11 October 2012)
Business Source Complete (1886 to 11 October 2012)

S26 S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11 or S12 or S13 or S14 or S15 or S16 or S17 or S18 or S19 or S20 or S21 or S22 or S23 OR S24 OR S25
S25 "work limitations questionnaire"
S24 "valuation of lost productivity"
S23 "work productivity survey"
S22 "health and performance questionnaire"
S21 "productivity and disease questionnaire"
S20 osterhaus AND (technique or presenteeism or absenteeism or productivity or work* or employ*)

S19 "Work role functioning questionnaire"
S18 "wellness inventory"
S17 WPAI*
S16 "work productivity and activity impairment*"
S15 "American Productivity Audit"
S14 "work productivity short inventory"
S13 "work performance scale"
S12 "work and health interview"
S11 "health and work performance questionnaire"
S10 "Stanford Presenteeism Scale"
S9 "migraine work and productivity loss questionnaire"
S8 "lam employment absence and productivity scale"
S7 (("migraine disability assessment" n2 (score* or scale or survey or questionnaire)) OR (MIDAS AND migraine)) AND (work* OR productivity OR presenteeism OR absenteeism OR employ*)
S6 "health-related productivity questionnaire"
S5 "health and work questionnaire"
S4 "health and productivity questionnaire"
S3 ("health and labor questionnaire") OR ("health and labour questionnaire")
S2 "endicott work productivity scale"
S1 "angina-related limitations at work questionnaire"

Search strategy for the following ProQuest databases

ABI Inform (1970 to 11 October 2012)

ALL(("American Productivity Audit") OR ("angina-related limitations at work questionnaire") OR ("endicott work productivity scale") OR ("health and labor questionnaire") OR ("health and labour questionnaire") OR ("health and productivity questionnaire") OR ("health and work questionnaire") OR ("health and performance questionnaire") OR ("health-related productivity questionnaire") OR ("lam employment absence and productivity scale") OR ("migraine work and productivity loss questionnaire") OR ("Stanford Presenteeism Scale") OR ("work performance scale") OR ("work and health interview") OR ("health and work performance questionnaire") OR ("productivity and disease questionnaire") OR WPAI* OR ("work productivity and activity impairment*") OR ("work productivity short inventory") OR (“wellness inventory”) OR ("work role functioning questionnaire") OR ("valuation of lost productivity") OR ("work productivity survey") OR ("work limitations questionnaire") OR ((("migraine disability assessment questionnaire") OR (MIDAS AND migraine) OR Osterhaus) AND (productivity or presenteeism or work* or employ*)))

Appendix C: Characteristics of Presenteeism Instruments Included in the Review

TABLE C1: CHARACTERISTICS OF PRESENTEEISM INSTRUMENTS INCLUDED IN THE REVIEW

Instrument	Content and target population	Format, mode, and time of admin-istration	Domains	Number of items		Presenteeism measures				Direct translation into monetary value	Scoring and recall period
				Total	Presenteeism	Self-rated measures	Comparative measures	Time lost measure	Quality of work measures		
Angina-Related Limitations at Work Questionnaire (ALWQ)[40]	Specific survey measures employees' difficulty with work tasks due to angina condition	Paper Self	(3) work limitation work status time lost	17	2	During the past 4 weeks, how much difficulty did you have with the following because of your angina? 1) Handling the workload 2) Finishing all your work	No	Yes	No	No	Scores were derived from standardized mean of 17 questions[32] Recall period is the last 4 weeks
Endicott Work Productivity Scale (EWPS)[41,73,77]	Generic survey Intended to be used for clinical trials, especially related to mental illness across industries	Paper Self 5-10 min	(4) attendance quality of work perform-ance capacity personal factors	27	5	During the past week, how frequently did you: 1) Just do no work at times when you would be expected to be working? 2) Fail to finish assigned tasks?	During the past week, how frequently did you: 1) Notice that your productivity for the time spent is lower than expected? 2) Notice that your efficiency for the time spent was lower than expected? 3) Work more slowly or take longer to complete tasks than expected?	No	During the past week, how frequently did you: 1) Have to do a job over because you made a mistake or your supervisor told you to do a job over? 2) Have a coworker redo something you had completed?	Yes	Total score ranges from 0 (best possible score) to 100 (worst possible score) Recall period is 1 week

Health and Labour Questionnaire (HLQ)[12,42,74,82,89]	This generic survey has been used in a number of populations, and is suitable for the general population with a focus on illness or disabilities. This survey is intended to be used to measure work performance in the presence of health problems.	Paper Self and interview 10-15 min	(4) absence from work reduced productivity unpaid work loss impedi-ments to paid and unpaid labor	23	3	1) I had to put off some of my work 2) Others had to take over some of my work	No	How many extra hours would you have to work to catch up on tasks you were unable to complete in normal working hours due to health problems over the last two weeks? NOTE: Do not count the days on which your reported sick."	No	Yes	Total score ranges from 0 (best possible score) to 100 (worst possible score) Recall period is 2 weeks There is also an efficiency score ranging from 6 (least work impediments) to 24 (most work impediments)
Health-related Productivity Questionnaire Diary (HRPQ-D)[43]	This survey is intended to examine the relationship between chronic or acute disease states and work productivity	Paper Self	(4) absent-eeism present-eeism unpaid work loss education missed	9	1	For the hours you did work outside the home, how did [your illness] impact your effectiveness?	Effectiveness is rated in terms of "% of my usual"	No, but can be derived	No	Yes	Scoring is based on hours missed in different activities (presenteeism, absenteeism, housework) Current day is the recall period Diary for 4 weeks

Health and Work Questionnaire (HWQ)[45]	Global measure intended to measure workplace productivity and employee health	Paper Self ~6 min	(6) productivity supervisor relations patience concen-tration work satisfaction non-work satisfaction	24	5	1) How would you and the following people describe your EFFICIENCY this week (self, supervisor, coworker)? 2) How would you and the following people describe the OVERALL AMOUNT of work you did this week (self, supervisor, coworker)? 3) This week how frequently did you fail to finish assigned tasks?	1) Rate your highest level of efficiency this week 2) Rate your lowest level of efficiency this week	No	How would you and the following people describe the OVERALL QUALITY of your work this week (self, supervisor, coworker)?	Yes	Scale scores are calculated. Scales calculated for each of the six Domains, as well as a total score (ranges from 1-10) Recall is past week
Lam Employment Absence and Productivity Scale (LEAPS)[46]	Generic instrument intended to assess work functioning and impairment in a depressed population	Paper Self 3-5 min	(3) work productivity trouble-some symptoms absent-eeism	10	1	Over the past 2 weeks, how often at work were you bothered by any of the following problems? 1) Getting less work done"	No	No	Over the past 2 weeks, how often at work were you bothered by any of the following problems? 1) Doing poor quality work 2) Making more mistakes	No	LEAPS score is based on 7 4-point Likert scale questions. Total score ranges from 0 to 28. A productivity subscale sums over 3 presenteeism items Recall period is 2 weeks

Migraine Disability Assessment Test (MIDAS)[88]	Specific survey for migraines. This questionnaire was designed to measure the impact that migraines have on patients' lives. It is also used as a clinical diagnosis tool.	Paper Self	(5) absent-eeism present-eeism unpaid work loss leisure activity loss headache severity	5	1	How many days in the last 3 months was your productivity at work or school reduced by half or more because of your headaches?	How many days in the last 3 months was your productivity at work or school reduced by half or more because of your headaches?	No	No	No	Total score is derived from adding the number of days of activity missed based on 5 questions. This score can range from 0 to 12 weeks. However, MIDAS sets an upper limit of 21. Recall period is 12 weeks
Migraine Work and Productivity Loss Questionnaire (MWPLQ)[48,76]	Specific survey for migraines. The instrument was developed to measure the impact of headaches on work and to assess the performance impact of migraines.	Paper Self 25 min	(7) absent-eeism work quality work quantity bodily effort inter-personal demands mental effort environ-mental factors	28	2	Over the past 2 weeks, how often at work were you bothered by [illness] and getting less work done?	No	No	Think about the entire period of time your migraine headache lasted. How much difficulty did you have with the following because of your most migraine headache or migraine headache treatment (no difficulty to so much difficulty couldn't do at all)? Doing your work carefully without making mistakes?	Yes	Unclear how scored Recall period is most recent episode

Osterhaus Technique[42]	Generic survey	Paper Self	NA	12	NA	No	How many days in the last 3 months was your productivity at work or school reduced by half or more because of your headaches?	No	NA	Yes	Recall period is 4 weeks
Quantity and Quality (Q-Q) Questionnaire[50,74,82,89]	Generic survey to measure lost productivity in terms of quantity and quality of output. Can be used for any industry.	Self	(2) quantity of work quality of work	5	2	Please circle on the scale below the degree of efficiency you consider yourself to have worked with on the days you did go to work while suffering from health problems. On this scale 10 means your work was not affected and 1 means that you were hardly capable of performing your work.	Could you indicate how much work you actually performed today during regular hours as compared to normal on the scale below?	How many hours extra would you have had to work to catch up on tasks you were unable to complete in normal working hours due to health problems IN THE PAST WEEK?	Could you indicate the quality of the work you performed today as compared to normal on the scale below?	Yes	Average scale scores are calculated. These range from 0 to 10. Hours lost are calculated separately. Recall period is past day
Stanford Presenteeism Scale (SPS-13)[5]	Generic instrument, measuring presenteeism. Intended for knowledge and production based jobs	Paper Self	(3) work stress and satisfaction health/ disability status present-eeism	13	2	"Despite having my health concerns, I was able to finish hard tasks at work"	Asked to assess percentage of usual work performed in 4 week period	No	No	Yes	Work Impairment Score can be calculated (sum of scores, ranging from 0 to 100) Recall is 4 weeks

Stanford Presenteeism Scale (SPS-6)[1,73,91]	Generic instrument, intended for knowledge and production based jobs	Paper Self	(3) work stress and satisfaction cognitive, emotional, behavioral concen-tration present-eeism	6	1	Describe your work experiences in the past month (strongly disagree to strongly agree) "Despite having my health concerns, I was able to finish hard tasks at work"	No	No	No	No	Total scores can also be calculated Recall period is 4 weeks
Stanford/American Health Association Presenteeism Scale (SPS-32)[46]	Generic measure, assesses worker's ability to focus on work given health concerns	Paper Self 7-10 min	(2) present-eeism ability to focus	42	3	Describe your work experiences in the past month (strongly disagree to strongly agree) "Despite having my health concerns, I was able to finish hard tasks at work"	1) "When my (health problem) bothered me, the percentage of my time that I was as productive as usual was 0-100% 2) Compared to my usual level of productivity, when my (health problem) bothered me, the percentage of my work that I was able to accomplish was 0-100%"	No	"When my (health problem) bothered me, the percentage of my work time that I was likely to make more mistakes than usual was 0-100%"	Yes	Recall period is 4 weeks

Valuation of Lost Productivity (VOLP)[55]	The VOLP is a generic questionnaire examining lost productivity resulting from a health problem (physical/mental/emotional/symptoms of the above)	Paper Self 20 min	(6) employ-ment status job character-ristics absent-eeism job perfor-mance unpaid work work environ-ment	37	7	Think of all the work you have completed during the past 7 days. Would you complete the same work in less time if you did NOT experience any health problems (i.e., any physical, mental, or emotional problems or symptoms)? If yes, please indicate the time you took to complete all your work in the past 7 days and the time you would take to complete the same work if you did NOT experience any health problems:	Yes	Yes	The following questions ask about the way YOUR HEALTH has interfered with your work in the past 7 days (please think of any physical, mental, or emotional problems or symptoms 1) I have been more likely to make mistakes at work	Yes	Score the percentage time loss while working Recall period is 1 week
Work and Health Interview (WHI)[55]	Generic survey to measure the health related lost productive work time (for 21 different health problems)	Self, using a computer assisted interview system 15 min	(6) employ-ment status episodes of health problem job visual-ization lost productivity absent-eeism demo-graphics	17	2	1) On average, how much of the time did you just do no work when you were supposed to be working? 2) On average, how much of the time were you very tired, fell asleep at work, or just felt too exhausted to do your work?	No	1) On average, how much of the time did you work more slowly or take longer to complete tasks than usual or expected? 2) When you weren't feeling well, how long, on average, did it take you to start working after you got to work?	On average, how much of the time did you spend doing a job over because you made a mistake or your supervisor told you to do a job over?"	Yes	Scoring is a calculation of productive work hours lost Recall period is 2 weeks

Work Performance Scale (WPS) (Part of the Functional Status Questionnaire)[61,77,78]	Generic module taken from the Functional Status Questionnaire. Initially designed to assess functionality of primary care patients	Paper Self	(1) work perform-ance	6		If you were employed last month, how was your work performance done as much work as others in similar jobs?	Yes	No	If you were employed last month, how was your work performance Done your job as carefully and accurately as others with similar jobs?	No	Total score ranges from 0 to 100 (max functionality). Social/role function module can be tallied, scores range from 6-24 Recall period is 4 weeks
Work Productivity and Activity Impairment (WPAI)[74,77,79,86,87,88]	Generic measure of impact of health problems on paid and unpaid work	Paper or computer Self or interview 5-7 min	(4) missed work due to health impairment while working due to health activity impairment due to health overall work impairment	6-9	1	During the past 7 days, how much did your PROBLEM affect your productivity while you were working? Scale 0-10.	No	No	No	Yes	Total score can be calculated and are specific to the version of WPAI administered. Refer to WPAI website for more information. Recall period is 1 week

WHO Health and Work Performance Questionnaire (WHO-HPQ)[12,58]	Generic measure intended to estimate the workplace cost resulting from health problems	Paper or computer Self or interview 10-20 min	(4) present-eeism absent-eeism work related accidents/ injuries work related successes/ failures	30	9	1) How much of the time was your speed of work or productivity higher than expected? 2) How much of the time did you find yourself daydreaming and not concentrating on your work? 3) Using the same scale 0-10, how would you rate your job performance during the past 30 days? 4) How many days in the past 30 was your speed of work or productivity lower than expected? 5) During the time you were at work in the past 30 days, how often did health problems limit you in the kind or amount of work you could do compared to usual?	1) How much of the time was your speed of work or productivity lower than expected? 2) On a scale of 0-10 where 0 is the worst job performance anyone could have at your job and 10 is the performance of a top worker, what number describes your overall job performance on the days you worked during the past 30 days? 3) Using the same scale 0-10, how would you rate the usual job performance of most workers on your job?	How much of the time did you do no work at times you were supposed to be working?	1) How much time did you find yourself not working as carefully as you should? 2) How much time was the quality of your work lower than expected? 3) Did you have any special work failure, make any big mistakes, or miss a major deadline at any time during the past 30 days? 4) Did you make any big mistake at work during the past 30 days that either caused an accident or created a safety risk for yourself or others?	Yes	Global scale is available Methods are available to calibrate scores into a work performance scale Recall period is either 1 or 4 weeks

Work Productivity Short Inventory (WPSI), also known as Wellness inventory[59,84]	This survey was designed to assess the overall financial impact of 15 different health conditions that lead to lost productivity	Paper Self	(2) absent-eeism present-eeism	22	1	Yes	No	During a typical 8 hour workday, estimate the total HOURS you were UNPROD-UCTIVE because of XX related problems	No	Yes	Scoring is a calculation of productive work hours lost 3 month, 2 week and 4 week recall periods are available
Work Productivity Survey-Rheumatoid Arthritis (WPS-RA)[61,77,78,81]	Specific intended to screen and/or monitor patient's health and role functions	Paper Self	(4) absent-eeism present-eeism unpaid work loss leisure activity missed	9	2	In the last month, how much has arthritis interfered with the patient's work productivity (work outside of home) on a scale of 0-10, where 0="no interference" and 10="complete interference"?	How many days in the last month was the patient's productivity at work reduced by half or more because of arthritis? (Do not include days counted in question 2) (If none, please write 0)	Yes	No	Yes	4 week recall period
Work Role Functioning Questionnaire (WRFQ)[55]	Generic instrument measuring difficulty meeting demands given health or emotional problems	Paper Self	(5) scheduling demands physical demands mental demands social demands output demands	27	2	1) Did you finish your work on time 2) Could you handle the workload? For ease of response, 100% is placed under "all the time," 50% under "half the time" and 0% under "none of the time."	No	No	Did you do your work carefully	NA	This questionnaire is a subset of the Work Functioning Questionnaire (WFQ). Subscales of the WFQ can be calculated 4 week recall period

References

1. Koopman C, Pelletier KR, Murray JF, Sharda CE, Berger ML, Turpin RS, et al. Stanford presenteeism scale: Health status and employee productivity. *J Occup Environ Med* 2002;44(1):14-20.
2. Burton WN, Chen CY, Conti DJ, Schultz AB, Pransky G, Edington DW. The association of health risks with on-the-job productivity. *J Occup Environ Med* 2005;47(8):769-777.
3. Li X, Gignac MA, Anis AH. The indirect costs of arthritis resulting from unemployment, reduced performance, and occupational changes while at work. *Med Care* 2006;44:304-310.
4. Pauly MV, Nicholson S, Polsky D, Berger ML, Sharda C. Valuing reductions in on-the-job illness: 'Presenteeism' from managerial and economic perspectives. *Health Econ* 2008;17(4):469-485.
5. Turpin RS, Ozminkowski RJ, Sharda CE, Collins JJ, Berger ML, Billotti GM, et al. Reliability and validity of the Stanford Presenteeism Scale. *J Occup Environ Med* 2004;46(11):1123-1133.
6. McKevitt C, Morgan M, Dundas R, Holland WW. Sickness absence and 'working through' illness: A comparison of two professional groups. *J Public Health Med* 1997;19:295-300.
7. Sanderson K, Tilse E, Nicholson J, Oldenburg B, Graves N. Which presenteeism measures are more sensitive to depression and anxiety? *J Affec Disorders* 2007;101(1-3):65-74.
8. Aronsson G, Gustaffsson K. Sickness presenteeism: Prevalence, attendance-pressure factors, and an outline of a model for research. *J Occup Environ Med* 2005;47:958-966.
9. Aronsson G, Gustafsson K, Dallner M. An empirical study of sickness presenteeism. *J Epidemiol Community Health* 2000;54:502-509.
10. Hemp P. Presenteeism: At work—but out of it. *Harvard Bus Rev* 2004;82(10):49-58.
11. Stewart WF, Ricci JA, Chee E, Morganstein D. Lost productive work time costs from health conditions in the United States: Results from the American Productivity Audit. *J Occup Environ Med* 2003;45(12):1234-1246.
12. Zhang W, Gignac MA, Beaton D, Tang K, Anis AH, Canadian Arthritis Network Work Productivity Group. Productivity loss due to presenteeism among patients with arthritis: Estimates from 4 instruments. *J Rheumatol* 2010;37(9):1805-1814.
13. Demerouti E, Verbeke W, Bakker AB. Exploring the relationship between a multi-dimensional and multi-faceted burnout concept and self-rated performance. *J Management* 2005;31:186-209.
14. Schultz AB, Edington DW. Employee health and presenteeism: A systematic review. *J Occup Rehabil* 2007;17:547-579.
15. Mattke S, Balakrishnan A, Bergamo G, Newberry SJ. A review of methods to measure health-related productivity loss. *Am J Manag Care* 2007;13:211-217.
16. Revicki DA, Irwin D, Reblando J. The accuracy of self-reported disability days. *Med Care* 1994;32:401-404.

17. Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. *JAMA* 2003;289(23):3135-3144.

18. Lerner D, Amick B, III, Lee JC, Rooney T, Rogers WH, Chang H, et al. Relationship of employee-reported work limitations to work productivity. *Med Care* 2003;41:649-659.

19. Mokkink LB, Terwee CB, Stratford PW, QAlonso J, Patrick DL. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Qual Life Res* 2009;18:313-333.

20. Lofland JH, Pizzi L, Frick KD. A review of health-related workplace productivity loss instruments. *Pharmacoeconomics* 2004;22(3):165-184.

21. Brown JD. Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition* 1986;4(4):353-376.

22. Alicke MD, Govorun O. The better-than-average effect. In Alicke MD, Dunning DA, Kruger JI, editors. *The self in social judgment*. New York: Psychology Press; 2005. p. 85-106.

23. Thompson AH, Jacobs P, Dewa CS. *The Alberta survey of addictive behaviours and mental health in the workforce: 2009*. Edmonton (AB): Institute of Health Economics; 2011.

24. Ottenbreit ND, Dobson KS. Avoidance. In Dobson KS, Dozois DJA editors. *Risk factors in depression*. New York: Elsevier; 2008. p. 447-470.

25. Seligman MEP. *Learned optimism*. New York: Vintage Books; 1990.

26. Tabachnik N, Crocker J, Alloy LB. Depression, social comparison, and the false-consensus effect. *J Pers Soc Psychol* 1983;45:688-699.

27. Shedler J, Mayman M, Manis M. The illusion of mental health. *Am Psychol* 1993;48(11):1117-1131.

28. Colvin, CR, Block J, Funder DC. Overly positive self-evaluations and personality: Negative implications for mental health. *J Pers Soc Psychol* 1995;68(6):1152-1162.

29. Prasad M, Wahlqvist P, Shikiar R, Shih YC. A review of self-report instruments measuring health-related work productivity: A patient-reported outcomes perspective. *Pharmacoeconomics* 2004;22(4):225-244.

30. Beaton D, Bombardier C, Escorpizo R, Zhang W, Lacaille D, Boonen A, et al. Measuring worker productivity: Frameworks and measures. *J Rheumatol* 2009;36(9):2100-2109.

31. Williams RW, Schmuck G, Allwood S, Sanchez M, Shea R, Wark G. Psychometric evaluation of health-related work outcome measures for musculoskeletal disorders: A systematic review. *J Occup Rehabil* 2007;17:504-521.

32. Roy JS, Desmeules F, MacDermid JC. Psychometric properties of presenteeism scales for musculoskeletal disorders: A systematic review. *J Rehabil Med* 2011;43:23-31.

33. Lerner DJ, Amick BC, III, Malspeis S, Rogers WH, Gomes DR, Salem DN. The Angina-related Limitations at Work Questionnaire. *Qual Life Res* 1998;7(1):23-32.

34. Endicott J, Nee J. Endicott Work Productivity Scale (EWPS): A new measure to assess treatment effects. *Psychopharmacol Bull* 1997;33(1):13-16.

35. van Roijen L, Essink-Bot ML, Koopmanschap MA, Bonsel G, Rutten FF. Labor and health status in economic evaluation of health care. The Health and Labor Questionnaire. *Int J Technol Assess Health Care* 1996;12(3):405-415.

36. Kumar RN, Hass SL, Li JZ, Nickens DJ, Daenzer CL, Wathen LK. Validation of the Health-Related Productivity Questionnaire Diary (HRPQ-D) on a sample of patients with infectious mononucleosis: Results from a phase 1 multicenter clinical trial. *J Occup Environ Med* 2003;45(8):899-907.

37. Kumar RN, Hass SL, Li JZ. Validation of the Health Related Productivity Questionnaire Diary (HRPQ-D) on a sample of patients with infectious mononucleosis: Results from an observational study. *Value Health* 2003;6(3):270.

38. Shikiar R, Halpern MT, Rentz AM, Khan ZM. Development of the Health and Work Questionnaire (HWQ): An instrument for assessing workplace productivity in relation to worker health. *Work* 2004;22(3):219-229.

39. Lam RW, Michalak EE, Yatham LN. A new clinical rating scale for work absence and productivity: Validation in patients with major depressive disorder. *BMC Psychiatry* 2009;9:78.

40. Lipton RB, Goadsby PJ, Sawyer JPC, Blakeborough P, Stewart WF. Migraine: Diagnosis and assessment of disability. *Rev Contemp Pharmaco* 2000;11(2):63-73.

41. Lerner DJ, Amick BC, III, Malspeis S, Rogers WH, Santanello NC, Gerth WC, et al. The migraine work and productivity loss questionnaire: Concepts and design. *Qual Life Res* 1999;8(8):699-710.

42. Osterhaus JT, Gutterman DL, Plachetka JR. Healthcare resource and lost labour costs of migraine headache in the US. *Pharmacoeconomics* 1992;2:67-76.

43. Brouwer WB, Koopmanschap MA, Rutten FF. Productivity losses without absence: Measurement validation and empirical evidence. *Health Policy* 1999;48:13-27.

44. Alavinia SM, Molenaar D, Burdorf A. Productivity loss in the workforce: Associations with health, work demands, and individual characteristics. *Am J Ind Med* 2009;52(1):49-56.

45. Koopmanschap MA. PRODISQ: A modular questionnaire on productivity and disease for economic evaluation studies. *Expert Rev Pharmacoecon Outcomes Res* 2005;5(1):23-28.

46. Pelletier KR, Koopman C. Stanford/American Health Association Presenteeism Scale (SAHAPS). In: Lynch W, Mercer WM, Riedel JE, editors. *Measuring employee productivity: A guide to self-assessment tools*. New York: William M. Mercer and Institute for Health and Productivity Management; 2001.

47. Zhang W, Bansback N, Kopec J, Anis AH. Measuring time input loss among patients with rheumatoid arthritis: Validity and reliability of the Valuation of Lost Productivity questionnaire. *J Occup Environ Med* 2011;53(5):530-536.

48. Stewart WF, Ricci JA, Leotta C, Chee E. Validation of the work and health interview. *Pharmacoeconomics* 2004;22(17):1127-1140.

49. Jette AM, Davies AR, Cleary PD, Calkins DR, Rubenstein LV, Fink A, et al. The functional status questionnaire: Reliability and validity when used in primary care. *J Gen Intern Med* 1986;1(3):143-149.

50. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics* 1993;4(5):353-365.

51. Kessler RC, Barber C, Beck A, Berglund P, Cleary PD, McKenas D, et al. The World Health Organization Health and Work Performance Questionnaire (HPQ). *J Occup Environ Med* 2003;45(2):156-174.

52. Goetzel RZ, Ozminkowski RJ, Long SR. Development and reliability analysis of the Work Productivity Short Inventory (WPSI) instrument measuring employee health and productivity. *J Occup Environ Med* 2003;45(7):743-762.

53. Osterhaus J, Purcaru O, Richard L. Validity and responsiveness of the Work Productivity Survey: A novel disease-specific instrument assessing work productivity within and outside the home in subjects with rheumatoid arthritis. *Value Health* 2008;11(6):A554-A555.

54. 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(3):R73.

55. Amick BCI, Lerner D, Rogers WH, Rooney T, Katz JN. A review of health-related work outcome measures and their uses, and recommended measures. *Spine* 2000;25:3152-3160.

56. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 2009;6(7):e1000097.

57. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Qual Life Res* 2010;19:539-549.

58. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Med Res Methodol* 2010;10(22).

59. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Qual Life Res* 2012;21(4):651-657.

60. Streiner DL, Norman GR. *Health Measurement Scales: A practical guide to their development and use.* 4th ed. Oxford (UK): Oxford University Press; 2008.

61. Abma FI, van der Klink JJL, Terwee CB, Amick BCI, Bultman U. Evaluation of the measurement properties of self-reported health-related work-functioning instruments among workers with common mental disorders. *Scand J Work Environ Health* 2012;38(1):5-18.

62. Abramson CE, McBride KE, Konnyu KJ, Elliott SL. Sexual health outcome measures for individuals with a spinal cord injury: A systematic review. *Spinal Cord* 2008;46:320-324.

63. Agasi-Idenburg C, Velthuis M, Wittink H. Quality criteria and user-friendliness in self-reported questionnaires on cancer-related fatigue: A review. *J Clin Epidemiol* 2010;63:705-711.

64. Albers G, Echteld MA, de Vet HC, Onwuteaka-Philipsen BD, van der Linden MH, Deliens L. Evaluation of quality-of-life measures for use in palliative care: A systematic review. *Palliat Med* 2010;24:17-37.

65. Ahmad MA, Xypnitos FN, Giannoudis PV. Measuring hip outcomes: Common scales and checklists. *Injury* 2011;42:259-264.

66. 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(1):28-37.

67. Braakman-Jansen LM, Taal E, Kuper IH, van de Laar MA. Productivity loss due to absenteeism and presenteeism by different instruments in patients with RA and subjects without RA. *Rheumatology* 2012;51(2):354-361.

68. Brozek JL, Guyatt GH, Heels-Ansdell D, Degl'Innocenti A, Armstrong D, Fallone CA, et al. Specific HRQL instruments and symptom scores were more responsive than preference-based generic instruments in patients with GERD. *J Clin Epidemiol* 2009;62(1):102-110.

69. Davies GM, Santanello N, Gerth W, Lerner D, Block GA. Validation of a migraine work and productivity loss questionnaire for use in migraine studies. *Cephalalgia* 1999;19(5):497-502.

70. Erickson SR, Guthrie S, Vanetten-Lee M, Himle J, Hoffman J, Santos SF, et al. Severity of anxiety and work-related outcomes of patients with anxiety disorders. *Depress Anxiety* 2009;26(12):1165-1171.

71. Erickson SR, Kirking DM. A cross-sectional analysis of work-related outcomes in adults with asthma. *Ann Allergy Asthma Immunol* 2002;88(3):292-300.

72. Giovannetti ER, Wolff JL, Frick KD, Boult C. Construct validity of the Work Productivity and Activity Impairment questionnaire across informal caregivers of chronically ill older patients. *Value Health* 2009;12(6):1011-1017.

73. Lalic H, Hromin M. Presenteeism towards absenteeism: manual work versus sedentary work, private versus governmental--A Croatian review. *Collegium Antropologicum* 2012;36(1):111-116.

74. McBurney CR, Eagle KA, Kline-Rogers EM, Cooper JV, Smith DE, Erickson SR. Work-related outcomes after a myocardial infarction. *Pharmacotherapy* 2004;24(11):1515-1523.

75. Meerding WJ, IJzelenberg W, Koopmanschap MA, Severens JL, Burdorf A. Health problems lead to considerable productivity loss at work among workers with high physical load jobs. *J Clin Epidemiol* 2005;58(5):517-523.

76. Ozminkowski RJ, Goetzel RZ, Chang S, Long S. The application of two health and productivity instruments at a large employer. *J Occup Environ Med* 2004;46(7):635-648.

77. Ozminkowski RJ, Goetzel RZ, Long SR. A validity analysis of the Work Productivity Short Inventory (WPSI) instrument measuring employee health and productivity. *J Occup Environ Med* 2003;45(11):1183-1195.

78. Reilly MC, Gooch KL, Wong RL, Kupper H, van der HD. Validity, reliability and responsiveness of the Work Productivity and Activity Impairment Questionnaire in ankylosing spondylitis. *Rheumatology* 2010;49(4):812-819.

79. Reilly MC, Gerlier L, Brabant Y, Brown M. Validity, reliability, and responsiveness of the work productivity and activity impairment questionnaire in Crohn's disease. *Clin Ther* 2008;30(2):393-404.

80. Reilly MC, Bracco A, Ricci JF, Santoro J, Stevens T. The validity and accuracy of the Work Productivity and Activity Impairment questionnaire--irritable bowel syndrome version (WPAI:IBS). *Aliment Pharm Therap* 2004;20(4):459-467.

81. Stewart WF, Lipton RB, Kolodner KB, Sawyer J, Lee C, Liberman JN. Validity of the Migraine Disability Assessment (MIDAS) score in comparison to a diary-based measure in a population sample of migraine sufferers. *Pain* 2000;88(1):41-52.

82. Stewart WF, Lipton RB, Whyte J, Dowson A, Kolodner K, Liberman JN, et al. An international study to assess reliability of the Migraine Disability Assessment (MIDAS) score. *Neurology* 1999;53(5):988-994.

83. Stewart WF, Lipton RB, Kolodner K, Liberman J, Sawyer J. Reliability of the migraine disability assessment score in a population-based sample of headache sufferers. *Cephalalgia* 1999;19(2):107-114.

84. Tang K, Pitts S, Solway S, Beaton D. Comparison of the psychometric properties of four at-work disability measures in workers with shoulder or elbow disorders. *J Occup Rehabil* 2009;19(2):142-154.

85. Terry PE, Xi M. An examination of presenteeism measures: The association of three scoring methods with health, work life, and consumer activation. *Popul Health Manage* 2010;13(6):297-307.

86. Wahlqvist P, Guyatt GH, Armstrong D, Degl'Innocenti A, Heels-Ansdell D, El-Dika S, et al. The Work Productivity and Activity Impairment Questionnaire for Patients with Gastroesophageal Reflux Disease (WPAI-GERD): Responsiveness to change and English language validation. *Pharmacoeconomics* 2007;25(5):385-396.

87. Wahlqvist P, Carlsson J, Stalhammar NO, Wiklund I. Validity of a Work Productivity and Activity Impairment questionnaire for patients with symptoms of gastro-esophageal reflux disease (WPAI-GERD)--Results from a cross-sectional study. *Value Health* 2002;5(2):106-113.

88. Zhang W, Bansback N, Boonen A, Young A, Singh A, Anis AH. Validity of the work productivity and activity impairment questionnaire--General health version in patients with rheumatoid arthritis. *Arthritis Res Ther* 2010;12(5):R177.

89. Zhang W, Bansback N, Boonen A, Severens JL, Anis AH. Development of a composite questionnaire, the valuation of lost productivity, to value productivity losses: Application in rheumatoid arthritis. *Value Health* 2012;15(1):46-54.

90. Nunnally J, Bernstein I. *Psychometric Theory*. 3rd ed. New York: McGraw Hill; 1994.

91. Bedrosian RC, Striegel RH, Wang C, Schwartz S. Association of binge eating with work productivity impairment, adjusted for other health risk factors. *J Occup Environ Med* 2012;54(4):385-393.

Author Contribution Statements

While there were specific roles taken on by each of the authors, each made significant shared contributions. Philip Jacobs' major involvement came during start-up and again at the end (review and distribution strategy). The remaining four met as a group throughout the course of this project to develop the overall approach, set strategy, review progress, and co-ordinate activities - each reviewed hundreds of abstracts/summaries and dozens of complete papers. Each of us wrote sections of the report that matched our input most closely. Angus Thompson and Maria Ospina were primarily responsible for consolidating such input to produce this version, which has been reviewed and approved by all of us. A true team approach.

Angus Thompson, as a psychologist/psychiatric epidemiologist and the Principal Investigator for the project, secured funding and is responsible for the development and overall management of a longer-term project on workplace health and productivity that includes this systematic review.

Maria Ospina, as an epidemiologist/psychologist expert in the conduct of systematic reviews, took the operations lead of this project.

Liz Dennett, a librarian and information specialist, was responsible for the literature search and was a second reviewer in the full text selection and in the application of the COSMIN checklist..

Arianna Waye, a health economist, coordinated much of the activity of the larger project, and contributed to the conceptual design through a focus on the development of our definition of presenteeism, the delineation of inclusion/exclusion criteria, and acting as the second reader for the systematic review.

Philip Jacobs is a highly respected Health Economist who has published many scientific papers in the field as well as a modern textbook of health economics. He contributed to the formation of the study and served as a liaison between the study group and institutional support.

INSTITUTE OF
HEALTH ECONOMICS
ALBERTA CANADA

Institute of Health Economics
1200 – 10405 Jasper Avenue
Edmonton AB Canada T5J 3N4
Tel. 780.448.4881 Fax. 780.448.0018
info@ihe.ca

www.ihe.ca

ISBN 978-1-926929-66-8 (online)