

METHODS

A Systematic Review of Measurement Properties of Instruments Assessing Presenteeism

Maria B. Ospina, PhD; Liz Dennett, MLIS; Arianna Waye, PhD; Philip Jacobs, DPhil; and Angus H. Thompson, PhD

Presenteeism has been broadly defined as “decreased productivity and below-normal work quality” when physically present at work.¹ Presenteeism can be studied in relation to many factors, including health. Terms such as “impaired presenteeism,” “sickness presenteeism,” or “working through illness” describe a phenomenon involving less than full productivity because of illness or other health conditions among individuals who opt to come to work, even when they arguably could stay at home.³

Presenteeism is a major occupational health problem in many countries, with serious consequences for both organizations and employees. Increasing evidence shows that presenteeism represents a “silent” but significant source of productivity losses that can cost organizations much more than absenteeism does.³ Presenteeism can lead to an increase in occupational accidents, deterioration of product quality, and adverse effects on healthy employees.³ The impact for the individual is not less—employees who turn up for work when ill have their quality of life diminished. They often experience feelings of burnout due to inadequate recovery, and get trapped in a vicious circle: job demands are accumulated, they have less energy to cope with these demands, more presenteeism results, and so on. Similarly, by repeatedly postponing sickness leave that may effectively resolve minor illnesses, more serious illnesses may develop.

A number of systematic reviews have summarized the measurement properties of instruments that assess productivity loss at the workplace,^{4,7} work productivity combining presenteeism and absenteeism measures,⁸ or work-related outcome measures in specific clinical groups.^{9,10} The majority of these reviews have not incorporated a systematic analysis of the methods by which these instruments have been developed,^{4,5,7} or have employed nonvalidated approaches to appraise both the quality of studies and the measurement properties of presenteeism instruments themselves.⁶ Assessing the quality of studies that evaluate the

ABSTRACT

Background

Presenteeism (decreased productivity while at work) is reported to be a major occupational problem in many countries. Challenges exist for identifying the optimal approach to measure presenteeism. Evidence of the relative value of presenteeism instruments to support their use in primary studies is needed.

Objectives

To assess and compare the measurement properties (ie, validity, reliability, responsiveness) and the quality of the evidence of presenteeism instruments.

Study Design

Systematic review.

Methods

Comprehensive searches of electronic databases were conducted up to October 2012. Twenty-three presenteeism instruments were examined. Methodological quality was appraised with the COSMIN (COnsensus-based Standards for the selection of health status Measurement INstruments) checklist. A best-evidence synthesis approach was used in the analysis.

Results

The titles and abstracts of 1767 articles were screened, with 289 full-text articles reviewed for eligibility. Of these, 40 studies assessing the measurement properties of presenteeism instruments were identified. The 3 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, with unknown quality of the evidence.

Conclusions

Most presenteeism instruments have been examined for some form of validity; evidence for criterion validity is virtually absent. The selection of instruments for use in primary studies 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.

Am J Manag Care. 2015;21(2):e171-e185

METHODS

measurement properties of presenteeism instruments is an essential step to inform the selection of presenteeism instruments for research and practice. If the quality of a study is appropriate, the results are valid and the measurement instrument can be a useful tool. Conversely, if the study quality is inadequate, the results cannot be trusted and the quality of the measurement instrument under scrutiny remains unclear.

Similarly, evidence on the relative value of presenteeism instruments is needed. Concurrent comparisons of the measurement properties and quality of presenteeism instruments may help to reveal the relative strengths and weaknesses of the measures and provide evidence-based guidance for the selection of instruments for research and management. To fill these knowledge gaps, we conducted a systematic review to assess and compare the measurement properties (ie, validity, reliability, responsiveness) and the quality of the evidence of presenteeism instruments.

METHODS

Data Sources

We conducted comprehensive electronic searches of MEDLINE, Embase, Cochrane Central Register of Controlled Trials, PsycINFO, Web of Science, Cumulative Index to Nursing and Allied Health Literature, Business Source Complete, and ABI/INFORM from database inception to October 2012 for studies reporting psychometric properties of instruments assessing presenteeism. Two reviewers (AHT, AW) compiled a list of presenteeism instruments based on preliminary searches of the literature, contact with experts in the field, and examination of individual items. An information specialist (LD) designed the search strategy using the names of instruments as keywords. In addition, we examined the references of identified articles for additional studies. Searches were limited to citations in the English language.

Article Selection

Presenteeism instruments were defined in this systematic review as questionnaires measuring at least 1 domain of productivity loss or reduced productivity/performance while at work. Items assessing presenteeism were required to have focused on at least 1 of the following characteristics: a) perceived productivity loss/reduced performance, b) comparative productivity loss/reduced performance (compared with those of others and with one's pattern), and/or c) estimation of unproductive time while at work. Based on this definition, 21

presenteeism instruments were considered for the review. Studies included in the review were full-text, peer-reviewed primary studies that evaluated the measurement properties (ie, validity, reliability, responsiveness) of the English version of any of the presenteeism instruments listed in **Table 1**. Tests are identified herein by their acronyms, a potentially confusing array. To minimize the effects of this, the full name of each, with the acronym in parentheses, is presented on first appearance. In addition, Table 1 can serve as a glossary. Having said that, the instruments under study are serving as members of a class of tests, and as such, are like subjects in most empirical research, with no need to identify individuals when considering most of the issues under discussion.

No restrictions in study design were applied during article selection; however, editorials, book chapters, review articles, conference abstracts, unpublished studies, case studies with fewer than 30 cases, and studies enrolling only pediatric populations (subjects 18 years and younger) were excluded. Studies published in non-English languages, or studies published in English that reported 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 reviewers (AW, MO) independently screened the titles and abstracts generated from the search strategy. The full text of articles deemed relevant and those with abstracts and titles that provided insufficient information were retrieved for a closer inspection. Two independent reviewers (either AW and LD, or AW and MO) determined study eligibility for the review, with disagreements about inclusion and exclusion of studies being solved through consensus among reviewers. Reasons for exclusions were documented and a flow chart of study selection was prepared according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement.

Methodological Quality Assessment

Two reviewers (either AW and LD, or AW and MO) independently applied the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) checklist¹¹ to assess the methodological quality of the measurement property reported in the included studies, with disagreements among reviewers being solved by consensus. The COSMIN checklist was developed through an international Delphi study¹² with the specific goal of facilitating the methodological assessment of outcome measures to enable the selection of the best instrument for a specific purpose.¹¹ The checklist includes the following properties: reliability, internal

consistency, content validity, construct validity, criterion validity, and responsiveness (see definitions in **Table 2**). Briefly, the quality of each measurement property reported in a study is assessed by a series of items including design requirements and preferred statistical methods and rated on a 4-point rating scale (poor, fair, good, excellent) depending on the information reported by the authors. A total score is determined by taking the lowest rating of the items for each measurement property.¹¹ The COSMIN checklist is increasingly used in systematic reviews of measurement properties, and to date, it is the only quality assessment tool of this kind that has been validated and standardized.^{11,13}

Data Extraction and Synthesis

One reviewer (MO) extracted the following information from the studies: health condition and sample size in which the instrument was tested; validity (ie, content, construct, criterion, convergent); reliability (internal consistency, test-retest, inter-rater); and responsiveness data. A second reviewer (AW) independently verified the accuracy and completeness of data extraction, with discrepancies between the data extractor and the data verifier being resolved by consensus.

We used a best-evidence synthesis approach to summarize the evidence on the measurement properties of presenteeism instruments, taking into account the number of studies, quality ratings, and consistency across their results. For each instrument, we combined the results of the methodological quality assessment of individual studies (poor, fair, good, or excellent) with a composite rating of the level of the evidence for the measurement properties of each instrument. The resulting level of evidence for the measurement properties of each instrument was classified for each property according to the following criteria: 1) strong (ie, consistent positive findings from multiple studies of good methodological quality *or* in 1 study of excellent methodological quality); 2) moderate (ie, consistent positive findings from multiple studies of fair methodological quality *or* in 1 study of good methodological quality); 3) limited (ie, positive findings from 1 study of fair methodological quality); and 4) conflicting (ie, conflicting findings in individual studies).¹⁴ When there were only studies of poor methodological quality, an unknown level of evidence was noted.

RESULTS

Our searches identified 1767 citations, with 971 duplicates being removed. Titles and abstracts of the remain-

ing 796 references were screened for relevance, yielding 289 articles judged as potentially relevant for the review. After applying the eligibility criteria to the full-text versions of these studies, we identified 40 studies that evaluated the measurement properties of 21 presenteeism instruments (**Figure**). The complete list of excluded studies and reasons for exclusion is available upon request.

General Characteristics of the Studies

The 40 studies¹⁵⁻⁵⁴ examined the measurement properties of 18 of the 21 presenteeism instruments included in the review. We did not identify studies that evaluated the measurement properties of the Osterhaus technique, the Work Role Functioning Questionnaire, or the Stanford/American Health Association Presenteeism Scale, 32-item version. While it may seem counterintuitive to have retained tests that had no representation in the peer-reviewed literature on psychometric properties, please note that these 3 did meet the criteria for inclusion in the study (they were deemed to measure some aspect of presenteeism). As such, an inability to find studies dealing with test quality is a separate matter and an important finding. Measurement data, on the other hand, were identified for 6 parallel forms of the Work Productivity and Activity Impairment (WPAI) scale. These were treated as separate entities, thus bringing the total number of presenteeism instruments that were ultimately examined to 23.

Sample sizes varied greatly across studies, ranging from 40 to 7797 participants per study (median sample size = 191; interquartile range, 112-354). The majority of evaluations of presenteeism instruments (27 studies) included samples of heterogeneous clinical groups, with most of them conducted on patients with musculoskeletal disorders.^{15,16,32,35,45,51-54} Other clinical conditions for which presenteeism instruments have been evaluated include gastrointestinal,^{17,36,37,49,50} neurological,^{18,42-44} and mood and anxiety disorders.^{19,20,28,39} Less frequently, individuals with cardiovascular,^{29,30} immunological,²⁶ and respiratory²¹ conditions, or patients' caretakers,²² have been included. The other 13 studies^{23-25,27,31,33,34,38,40,41,46-48} have used samples of employees across a wide range of organizational settings (eg, manual workforce, telecommunications, airlines, call centers).

Measurement Properties of Presenteeism Instruments

Construct validity was the measurement property most frequently evaluated in the studies (28 studies^{15,18-22,25,26,28,29,31-38,40,45-53}) followed by reliability (17

METHODS

studies^{15,18-21,23,25,27-30,39,40,43-45,47}), convergent validity through head-to-head comparisons among different presenteeism instruments (11 studies^{15,16,24,25,28,31,41,42,51-53}), content validity (8 studies^{23-25,29,34,40,48,54}), responsiveness to detect important changes in the construct over time (8 studies^{15,17,19,26,32,35,36,49}), structural validity (4 studies^{25,28,40,47}), and finally, criterion validity, which was formally evaluated in only 1 study.⁴⁰

Table 3 summarizes the studies that reported the measurement properties of presenteeism instruments along with methodological quality ratings per measurement property for each study.

Evidence on the content validity of presenteeism instruments was provided for 7 instruments: Angina-Related Limitations at Work Questionnaire (ALWQ),²⁹ Health and Labour Questionnaire (HLQ),⁴⁸ World Health Organization Health and Work Performance Questionnaire (HPQ),²⁴ Health and Work Questionnaire (HWQ),⁴⁰ Stanford Presenteeism Scale (6-item version) (SPS-6),²⁵ Valuation of Lost Productivity questionnaire (VOLP),⁵⁴ and Work Productivity Short Inventory (WPSI).^{23,34} The combined information on the methodological quality of these studies showed that the SPS-6 and the VOLP had the best level of evidence (strong) in terms of the quality of content validity data.

The internal consistency of scale items was evaluated for 10 presenteeism instruments: ALWQ,²⁹ Endicott Work Productivity Scale (EWPS),^{15,19,20} HWQ,⁴⁰ Lam Employment Absence and Productivity Scale (LEAPS),²⁸ Migraine Disability Assessment questionnaire (MIDAS),^{43,44} Migraine Work and Productivity Loss Questionnaire (MWPLQ),¹⁸ Stanford Presenteeism Scale (13-item version) (SPS-13),⁴⁷ SPS-6,^{15,25,27,39,45} Work Performance Scale from the Functional Status Questionnaire (WPS),^{20,21,30} and WPSI.²³ In all cases, the internal consistency of the total scale was reported; however, internal consistency of subscores was assessed for some instruments (ie, HWQ productivity,⁴⁰ MIDAS item on presenteeism,⁴⁴ MWPLQ subscales¹⁸). After combining the methodological quality of all studies assessing the internal consistency of each presenteeism instruments, we found that the EWPS, HWQ, MWPLQ, and SPS-6 had the best level of evidence (strong) for the quality of internal consistency data.

Test-retest reliability was evaluated for 5 presenteeism instruments: EWPS,¹⁹ MIDAS,^{43,44} VOLP,⁵¹ Work Productivity and Activity Impairment scale: Irritable Bowel Syndrome (WPAI:IBS),³⁷ and WPAI: General Health (WPAI:GH).^{38,51} The MIDAS had the best level of evidence (moderate level) of the quality of test-retest reliability data.

Different schedules of administration of a presenteeism instrument (ie, inter-rater reliability) were evaluated

for only 1 instrument, WPSI.²³ The level of evidence of the quality of inter-rater reliability data is unknown (ie, the study was rated as of poor quality).

Evidence on the structural validity of presenteeism instruments was provided for 4 instruments: HWQ,⁴⁰ LEAPS,²⁸ SPS-13,⁴⁷ and SPS-6.²⁵ All studies used factor analysis to determine the structure and dimensionality of the instruments. The best level of evidence (strong level) of the quality of structural validity data is available for the HWQ.

Evidence on the construct validity of presenteeism instruments was provided for 21 instruments (including the 6 different versions of the WPAI): ALWQ,²⁹ EWPS,^{15,19,20} HLQ,^{31,48,53} HPQ,^{46,53} Health Related Productivity Questionnaire Diary (HRPQ-D),²⁶ HWQ,⁴⁰ LEAPS,²⁸ MWPLQ,¹⁸ Quantity and Quality method (Q-Q),³¹ SPS-13,⁴⁷ SPS-6,^{15,25,45} VOLP,⁵¹ WPSI,^{33,34} Work Productivity Survey:Rheumatoid Arthritis (WPS:RA),³² Work Productivity Survey,^{20,21} and the 6 WPAI versions (WPAI:Crohn's Disease [WPAI:CD],³⁶ WPAI:Caregiver,²² WPAI:Gastroesophageal Reflux Disease [WPAI:GERD],^{49,50} WPAI:GH,^{20,38,51-53} WPAI:IBS,³⁷ WPAI:Ankylosing Spondylitis [WPAI:SpA]³⁵).

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

Head-to-head comparisons among 10 different presenteeism instruments were aimed at establishing their convergent validity: EWPS,¹⁵ HLQ,^{16,31,53} HPQ,^{24,28,53} LEAPS,²⁸ MIDAS,⁴² Q-Q,^{16,31} SPS-6,^{15,25} VOLP,⁵¹ Work and Health Interview,⁴¹ and WPAI:GH.^{16,51-53} After combining study quality information, we found that the EWPS and the SPS-6 had the best level of evidence (strong level) of convergent validity.

One study⁴⁰ assessed the criterion validity of the HWQ against a "gold standard" (ie, hours of productivity loss). The level of evidence of the quality of criterion validity, however, is left unknown as the quality assessment indicated that the study in question was of poor quality and thus did not allow any conclusions to be drawn about this domain for that test.

Responsiveness to detect important changes in the construct over time were evaluated for 7 instruments: EWPS,^{15,19} HRPQ-D,²⁶ SPS-6,¹⁵ WPAI:GERD,^{17,49}

WPAI:CD,³⁶ WPAI:SpA,³⁵ and WPS:RA.³² The combined information on the methodological quality of all responsiveness evaluations in the studies showed that the SPS-6 and the EWPS had the best level of evidence (strong level) for responsiveness data.

Table 4 summarizes the overall level of evidence informing the use of instruments to measure presenteeism. The 3 presenteeism instruments with the strongest level of evidence on more than 1 measurement property were the SPS-6, the HWQ, and the EWPS. The SPS-6 had a strong level of evidence for the majority of measurement domains including content validity, internal consistency, construct validity, convergent validity, and responsiveness. Evaluations of the criterion validity of the SPS-6 have not been conducted. The EWPS showed a strong level of evidence for internal consistency, convergent validity, and responsiveness, but there was no evidence about the criterion validity of the instrument. Finally, the HWQ had a strong level of evidence for internal consistency and structural validity; however, the level of evidence of criterion validity is unknown. The level of evidence for the measurement properties of all the other presenteeism instruments oscillated between moderate and poor.

DISCUSSION

We systematically reviewed 40 studies on measurement properties of 21 presenteeism instruments and rated their methodological quality using the COSMIN checklist. We found that most presenteeism instruments have been assessed for at least some form of validity, but with evidence for criterion validity being virtually absent. The 3 presenteeism instruments supported by the best evidence regarding their measurement properties were the SPS-6, the HWQ, and the EWPS. Evidence for criterion validity, arguably the most important of the attributes under study here, was virtually absent across the board.

A number of important findings can be taken from this review. First, 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.⁴⁻¹⁰ Secondly, the evidence that does exist suggests that the selection of a presenteeism instrument for use in research and practice currently depends on weak forms of validity. Furthermore, our present study has indicated that our confidence in conclusions about presenteeism test validity generally is weakened even further by the finding that a meaning-

ful proportion of these evaluations are not of adequate methodological quality.

Although many of the studies were deemed to be methodologically strong, a virtual absence of “gold standard” investigations indicates that none of the presenteeism tests included in this review have actually been shown to predict productivity loss while at work.

A lack of coverage of the psychometric domains in the studies 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.

Many challenges exist when identifying an optimal or ideal approach to the measurement of presenteeism. In many instances, there is confusion between the measurement of potential causes of lost productivity while on the job (eg, physical and/or mental ill health, disability,¹ malingering, irresponsibility) and presenteeism 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, are increasingly being used for that purpose.

Furthermore, in this vein, although not being addressed specifically here, improvement in test quality will require more attention to the issue of cause-presumed and cause-neutral instruments. The majority of tests assume that presenteeism is affected by health, and ask either about the effects on productivity of general health or a specific health condition. The more disease-specific instruments provide only a partial explanation of productivity loss, at best, since the potential role of other health conditions is ignored. All ignore nonhealth measures, which could exert a considerable influence. For example, workplace culture has been shown to produce a dramatic effect on productivity, either directly or indirectly (eg, through effects on back pain). On the other hand, some tests, like the HPQ, query the level of productivity only, with no reference to illness. These, of course, can use manipulations in experimental design or statistical approaches to determine the relationship between any condition (health or otherwise) and presenteeism. The point is that these 2 paradigms lead to differences in the scope of causation that comes to be studied, and may not foster a comprehensive look at the range of determinants of presenteeism.

METHODS

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, 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.⁵ Presenteeism is usually assessed by self-report measures that can be generic (ie, 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, time-adjustments at work for perceptions of productivity in relation to self and/or colleagues, and domain-based measures that assess health-related limitations in specific job demands. 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.

Strengths and Limitations

The major strengths of this systematic review are the comprehensive literature searches and rating of the methodological quality of studies and measurement properties by 2 independent reviewers. Study limitations should be noted: we only evaluated peer-reviewed journal articles that were published in English, and we did not include gray literature; and the use of the COSMIN approach requires readers to be cognizant that it does not provide a method for addressing possible bias due to gaps in psychometric data that are common in this type of study.

CONCLUSIONS

We did not identify a presenteeism instrument that conjugates both acceptable reliability and validity. Therefore, the decision to use one presenteeism instrument over another must be driven by the usual matters of study purpose, research questions, and instrument domains, but with the recognition that presenteeism impact statements derived from such test data will be open to credible criticism. Given the large availability of self-report presenteeism instruments, the development of new instruments is discouraged; the field does not need another self-report test that has not been properly validated. Rather, we encourage movement toward the goal of a comprehensive description of the quality of presenteeism tests in the field, preferably using studies designed according to the COSMIN standards. Furthermore, noting that many studies are in existence that did not

meet our criteria for inclusion, we urge test developers to revisit data that they have presented at conferences or in the grey literature, and publish them in peer-reviewed outlets.

Acknowledgments

The following individuals and institutions are acknowledged for provision of information regarding published studies and instruments included in this systematic review: Nick Bansback, PhD, School of Population and Public Health, University of British Columbia, Vancouver, Canada; Walter "Buzz" Stewart, PhD, MPH, Center for Health Research, Geisinger Health System, Danville, PA; and Debra Lerner, PhD, MS, Tufts University School of Medicine, Boston, MA.

The following are acknowledged for provision of information regarding published studies and instruments included in this systematic review: Mark Attridge, PhD, Attridge Consulting, Inc, Minneapolis, MN; Monique A.M. Gignac, PhD, Institute for Work & Health and Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; Raymond W. Lam, MD, FRCPC, University of British Columbia, Vancouver, BC, Canada; JianLi Wang, PhD, University of Calgary, Calgary, AB, Canada.

Finally, we thank Ms Debra Haas, of the Institute of Health Economics, Edmonton, AB, Canada, for her assistance with article retrieval.

Author Affiliations: Institute of Health Economics (MBO, AW, PJ, AHT), Edmonton, AB, Canada; and University of Alberta (LD), Edmonton, AB, Canada.

Source of Funding: This study received funding from the Alberta Depression Initiative. The funding agency did not have any role in the collection of data, its analysis and interpretation, and/or in the right to approve or disapprove publication of the finished manuscript.

Author Disclosures: The authors report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

Authorship Information: Concept and design (MBO, LD, AW, PJ, AHT); acquisition of data (MBO, LD, AW); analysis and interpretation of data (MBO, LD, AW, AHT); drafting of the manuscript (MBO, LD, AW, AHT); critical revision of the manuscript for important intellectual content (MBO, LD, PJ, AHT); obtaining funding (PJ, AHT); administrative, technical, or logistic support (MBO, PJ); supervision (AHT).

Address correspondence to: Angus Thompson, PhD, Institute of Health Economics, Ste 1200, 10405 Jasper Ave NW, Edmonton, AB, Canada T5J 3N4. E-mail: gthompson@ihe.ca.

REFERENCES

1. Hemp P. Presenteeism: At work—but out of it. *Harv Bus Rev*. 2004;82(10):49-58.
2. McKeivitt 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(3):295-300.
3. Aronsson G, Gustafsson K. Sickness presenteeism: prevalence, attendance-pressure factors, and an outline of a model for research. *J Occup Environ Med*. 2005;47(9):958-966.
4. Lofland JH, Pizzi L, Frick KD. A review of health-related workplace productivity loss instruments. *Pharmacoeconomics*. 2004;22(3):165-184.
5. Mattke S, Balakrishnan A, Bergamo G, Newberry SJ. A review of methods to measure health-related productivity loss. *Am J Manag Care*. 2007;13(4):211-217.
6. Schultz AB, Edington DW. Employee health and presenteeism: a systematic review. *J Occup Rehabil*. 2007;17(3):547-579.
7. Prasad M, Wahlqvist P, Shikar 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.
8. Beaton D, Bombardier C, Escorpizo R, et al. Measuring worker productivity: frameworks and measures. *J Rheumatol*. 2009;36(9):2100-2109.

9. 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(3):504-521.
10. Roy JS, Desmeules F, MacDermid JC. Psychometric properties of presenteeism scales for musculoskeletal disorders: a systematic review. *J Rehabil Med*. 2011;43(1):23-31.
11. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HCW. 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.
12. Mokkink LB, Terwee CB, Patrick 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(4):539-549.
13. Mokkink LB, Terwee CB, Gibbons E, et al. Inter-rater agreement and reliability of the COSMIN (Consensus-based Standards for the selection of health status Measurement Instruments) checklist. *BMC Med Res Methodol*. 2010;10:82.
14. Terwee CB, Bot SD, de Boer MR, et al. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42.
15. Beaton DE, Tang K, Gignac MA, et al. Reliability, validity, and responsiveness of five at-work productivity measures in patients with rheumatoid arthritis or osteoarthritis. *Arthritis Care Res (Hoboken)*. 2010;62(1):28-37.
16. 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 (Oxford)*. 2012;51(2):354-361.
17. Brozek JL, Guyatt GH, Heels-Ansell D, 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.
18. 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.
19. Endicott J, Nee J. Endicott Work Productivity Scale (EWPS): a new measure to assess treatment effects. *Psychopharmacol Bull*. 1997;33(1):13-16.
20. Erickson SR, Guthrie S, Vanetten-Lee M, et al. Severity of anxiety and work-related outcomes of patients with anxiety disorders. *Depress Anxiety*. 2009;26(12):1165-1171.
21. 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.
22. Giovannetti ER, Wolff JL, Frick KD, Boulton 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.
23. 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.
24. Kessler RC, Barber C, Beck A, et al. The World Health Organization Health and Work Performance Questionnaire (HPQ). *J Occup Environ Med*. 2003;45(2):156-174.
25. Koopman C, Pelletier KR, Murray JF, et al. Stanford presenteeism scale: health status and employee productivity. *J Occup Environ Med*. 2002;44(1):14-20.
26. 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.
27. Lalić H, Hromin M. Presenteeism towards absenteeism: manual work versus sedentary work, private versus governmental: a Croatian review. *Coll Antropol*. 2012;36(1):111-116.
28. 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. doi:10.1186/1471-244X-9-78.
29. Lerner DJ, Amick BC 3rd, Malspeis S, Rogers WH, Gomes DR, Salem DN. The Angina-related Limitations at Work Questionnaire. *Qual Life Res*. 1998;7(1):23-32.
30. 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.
31. 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.
32. Osterhaus JT, Purcaru O, Richard L. Discriminant validity, responsiveness and reliability of the rheumatoid arthritis-specific Work Productivity Survey (WPS-RA) [published online May 20, 2009]. *Arthritis Res Ther*. 2009;11(3):R73. doi: 10.1186/ar2702.
33. 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.
34. 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.
35. Reilly MC, Gooch KL, Wong RL, Kupper H, van der Hejide D. Validity, reliability and responsiveness of the Work Productivity and Activity Impairment Questionnaire in ankylosing spondylitis. *Rheumatology (Oxford)*. 2010;49(4):812-819.
36. 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.
37. 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 Pharmacol Therap*. 2004;20(4):459-467.
38. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics*. 1993;4(5):353-365.
39. Sanderson K, Tilse E, Nicholson J, Oldenburg B, Graves N. Which presenteeism measures are more sensitive to depression and anxiety? *J Affect Disord*. 2007;101(1-3):65-74.
40. Shikier 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.
41. Stewart WF, Ricci JA, Leotta C, Chee E. Validation of the work and health interview. *Pharmacoeconomics*. 2004;22(17):1127-1140.
42. 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.
43. Stewart WF, Lipton RB, Whyte J, et al. An international study to assess reliability of the Migraine Disability Assessment (MIDAS) score. *Neurology*. 1999;53(5):988-994.
44. 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.
45. 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.
46. Terry PE, Xi M. An examination of presenteeism measures: the association of three scoring methods with health, work life, and consumer activation. *Popul Health Manag*. 2010;13(6):297-307.
47. Turpin RS, Ozminkowski RJ, Sharda CE, et al. Reliability and validity of the Stanford Presenteeism Scale. *J Occup Environ Med*. 2004;46(11):1123-1133.
48. 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.
49. Wahlqvist P, Guyatt GH, Armstrong D, 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.
50. Wahlqvist P, Carlsson J, Stålhammar 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.
51. 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.
52. 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 [published online September 22, 2010]. *Arthritis Res Ther*. 2010;12(5):R177. doi:10.1186/ar3141.

METHODS

53. 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.
54. 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.
55. Lipton RB, Goadsby PJ, Sawyer JPC, Blakeborough P, Stewart WF. Migraine: Diagnosis and assessment of disability. *Rev Contemp Pharmacol*. 2000;11(2):63-73.
56. Lerner DJ, Amick BC 3rd, Malspeis S, et al. The Migraine Work and Productivity Loss questionnaire: concepts and design. *Qual Life Res*. 1999;8(8):699-710.
57. Osterhaus JT, Gutterman DL, Plachetka JR. Healthcare resource and lost labour costs of migraine headache in the US. *Pharmacoeconomics*. 1992;2(1):67-76.
58. Brouwer WB, Koopmanschap MA, Rutten FF. Productivity losses without absence: measurement validation and empirical evidence. *Health Policy*. 1999;48(1):13-27.
59. 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.
60. Koopmanschap MA. PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies. *Expert Rev Pharmacoecon Outcomes Res*. 2005;5(1):23-28.
61. Pelletier KR, Koopman C. Stanford/American Health Association Presenteeism Scale (SAHAPS). In: Lynch W, Riedel JE, eds. *Measuring Employee Productivity: A Guide to Self-Assessment Tools*. New York, NY: William M. Mercer and Institute for Health and Productivity Management; 2001:22-24.
62. Jette AM, Davies AR, Cleary PD, et al. The Functional Status Questionnaire: reliability and validity when used in primary care. *J Gen Intern Med*. 1986;1(3):143-149.
63. 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.
64. Amick BC 3rd, Lerner D, Rogers WH, Rooney T, Katz JN. A review of health-related work outcome measures and their uses, and recommended measures. *Spine (Phila PA 1976)*. 2000;25(24):3152-3160.
65. Mokkink LB, Terwee CB, Knol DL, et al. The c for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10:22. doi:10.1186/1471-2288-10-22.
66. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 4th ed. Oxford, UK: Oxford University Press; 2008. ■

Take-Away Points

This article reviewed studies assessing the measurement properties of presenteeism instruments.

- We identified 40 studies assessing the measurement properties of 21 presenteeism instruments.
- Most presenteeism instruments have been assessed for some form of validity but evidence for criterion validity is virtually absent.
- Within these limitations, the 3 presenteeism instruments with the strongest level of evidence were the Stanford Presenteeism Scale, 6-item version (SPS-6); the Endicott Work Productivity Scale (EWPS); and the Health and Work Questionnaire (HWQ).
- The selection of a presenteeism instrument for research and management purposes currently depends on weak forms of validity.

■ **Table 1.** Characteristics of Presenteeism Instruments Evaluated in the Review

Instrument	Mode and Time of Administration	Number of Items		Types of Presenteeism Measures			
		Total	Presenteeism	Self-Rated	Comparative	Time Lost	Quality of Work
Angina-Related Limitations at Work Questionnaire ²⁹	Self-report (5-10 min)	17	2	Yes	No	Yes	No
Endicott Work Productivity Scale ¹⁹	Self-report (5-10 min)	27	5	Yes	Yes	No	Yes
Health and Labour Questionnaire ⁴⁸	Self-report and interview-administered (10-15 min)	23	3	Yes	No	Yes	No
Health and Work Performance Questionnaire ²⁴	Self-report or interview-administered (10-20 min)	30	9	Yes	Yes	Yes	Yes
Health-related Productivity Questionnaire Diary ^{26,55}	Self-report (3-5 min)	9	1	Yes	Yes	No	No
Health and Work Questionnaire ⁴⁰	Self-report (5-10 min)	24	5	Yes	Yes	No	Yes
Lam Employment Absence and Productivity Scale ²⁸	Self-report (3-5 min)	10	1	Yes	No	No	Yes
Migraine Disability Assessment Test ⁵⁵	Self-report (1-3 min)	5	1	Yes	Yes	No	No
Migraine Work and Productivity Loss Questionnaire ⁵⁶	Self-report (20-25 min)	28	2	Yes	No	No	Yes
Osterhaus Technique ⁵⁷	Self-report (5-10 min)	12	N/A	No	Yes	No	NA
Quantity and Quality Questionnaire from the Productivity and Disease Questionnaire ⁵⁸⁻⁶⁰	Self-report (3-5 min)	5	2	Yes	Yes	Yes	Yes
Stanford Presenteeism Scale (13-item version) ⁴⁷	Self-report (5-10 min)	13	2	Yes	Yes	No	No
Stanford Presenteeism Scale (6-item version) ²⁵	Self-report (3-5 min)	6	1	Yes	No	No	No
Stanford/American Health Association Presenteeism Scale ⁶¹	Self-report (7-10 min)	42	3	Yes	Yes	No	Yes
Valuation of Lost Productivity ⁵¹	Self-report (15-20 min)	37	7	Yes	Yes	Yes	Yes
Work and Health Interview ⁴¹	Self-report or interview-administered (10-15 min)	17	2	Yes	No	Yes	Yes
Work Performance Scale (part of the Functional Status Questionnaire) ⁶²	Self-report (3-5 min)	6		Yes	Yes	No	Yes
Work Productivity and Activity Impairment ³⁸	Self-report or interview-administered (5-7 min)	6-9	1	Yes	No	No	No
Work Productivity Short Inventory ²³	Self-report (20-25 min)	22	1	Yes	No	Yes	No
Work Productivity Survey-Rheumatoid Arthritis ^{32,63}	Self-report (3-5 min)	9	2	Yes	Yes	Yes	No
Work Role Functioning Questionnaire ⁶⁴	Self-report (20-25 min)	27	2	Yes	No	No	Yes

METHODS

■ **Table 2. Measurement Properties Evaluated in the Review**

Content validity	The degree to which the content of an instrument is an adequate reflection of the construct to be measured.
Structural validity	The degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured.
Criterion validity	The degree to which test scores are an adequate reflection of a "gold standard."
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.
Concurrent validity	The degree of correlation between a new test and other tests/measures of the same construct.
Internal consistency	The interrelatedness among items.
Inter-rater reliability	The consistency with which different examiners produce similar ratings in an instrument.
Test-retest reliability	The extent of agreement across 2 administrations of a test, assuming no pertinent changes occurred between administrations.
Responsiveness to change	The ability of an instrument to detect change over time in the construct to be measured.

Adapted from Terwee et al,¹⁴ Mokkink et al,⁶⁵ and Streiner et al.⁶⁶

■ **Table 3. Summary of Studies Assessing Measurement Properties of Presenteeism Instruments**

Instrument and Studies	Population	Properties Evaluated	Quality
ALWQ			
Lerner et al, 1998 ²⁹	N = 40, Chronic angina pectoris	Reliability (internal consistency)	Fair
		Content validity	Fair
		Construct validity	Fair
EWPS			
Beaton et al, 2010 ¹⁵	N = 250, Rheumatoid arthritis or osteoarthritis	Reliability (internal consistency)	Excellent
		Construct validity	Excellent
		Concurrent validity	Excellent
		Responsiveness	Excellent
Endicott et al, 1997 ¹⁹	N = 42, Depression and anxiety	Reliability (internal consistency)	Fair
		Reliability (test-retest)	Poor
		Construct validity	Fair
		Responsiveness	Poor
Erickson et al, 2009 ²⁰	N = 76, Anxiety disorders	Reliability (internal consistency)	Good
		Construct validity	Fair
HLO			
Braakman-Jansen et al, 2012 ¹⁶	N = 62, Rheumatoid arthritis	Concurrent validity	Fair
Meerding et al, 2005 ³¹	N = 570, Workers (manual)	Construct validity	Fair
		Concurrent validity	Fair
van Roijen et al, 1996 ⁴⁸	N = 726, General population	Content validity	Fair
		Construct validity	Fair
Zhang et al, 2010 ⁵³	N = 212, Rheumatoid arthritis or osteoarthritis	Construct validity	Fair
		Concurrent validity	Fair
Zhang et al, 2010 ⁵²	N = 150, Rheumatoid arthritis	Concurrent validity	Fair

(continued)

■ **Table 3.** Summary of Studies Assessing Measurement Properties of Presenteeism Instruments (*continued*)

Instrument and Studies	Population	Properties Evaluated	Quality
HPQ			
Kessler et al, 2003 ²⁴	N = 2350, Workers (general)	Content validity	Fair
		Concurrent validity	Fair
Lam et al, 2009 ²⁸	N = 234, Major depressive disorder	Concurrent validity	Fair
Terry et al, 2010 ⁴⁶	N = 631, Workers (manual)	Construct validity	Fair
Zhang et al, 2010 ⁵³	N = 212, Rheumatoid arthritis or osteoarthritis	Construct validity	Fair
		Concurrent validity	Fair
HRPQ-D			
Kumar et al, 2003 ²⁶	N = 42, Infectious mononucleosis	Construct validity	Fair
		Responsiveness	Poor
HWQ			
Shikiar et al, 2004 ⁴⁰	N = 294, Workers (airlines)	Reliability (internal consistency)	Excellent
		Content validity	Fair
		Structural validity	Excellent
		Construct validity	Fair
		Criterion validity	Poor
LEAPS			
Lam et al, 2009 ²⁸	N = 234, Major depressive disorder	Reliability (internal consistency)	Fair
		Structural validity	Fair
		Construct validity	Fair
		Concurrent validity	Fair
MIDAS			
Stewart et al, 1999 ⁴³	N = 97, Migraine	Reliability (internal consistency)	Fair
		Reliability (test-retest)	Fair
Stewart et al, 1999 ⁴⁴	N = 177, Migraine	Reliability (internal consistency)	Fair
		Reliability (test-retest)	Fair
Stewart et al, 2000 ⁴²	N = 144, Migraine	Concurrent validity	Fair
MWPLQ			
Davies et al, 1999 ¹⁸	N = 164, Migraine	Reliability (internal consistency)	Excellent
		Construct validity	Fair
Q-Q			
Braakman-Jansen et al, 2012 ¹⁶	N = 62, Rheumatoid arthritis	Concurrent validity	Fair
Meerding et al, 2005 ³¹	N = 570, Workers (manual)	Construct validity	Fair
		Concurrent validity	Fair
Zhang et al, 2010 ⁵²	N = 150, Rheumatoid arthritis	Concurrent validity	Fair
SPS-13			
Turpin et al, 2004 ⁴⁷	N = 7797, Workers (general)	Reliability (internal consistency)	Fair
		Structural validity	Fair
		Construct validity	Fair

(continued)

Table 3. Summary of Studies Assessing Measurement Properties of Presenteeism Instruments (*continued*)

Instrument and Studies	Population	Properties Evaluated	Quality
SPS-6			
Beaton et al, 2010 ¹⁵	N = 250, Rheumatoid arthritis or osteoarthritis	Reliability (internal consistency)	Excellent
		Construct validity	Excellent
		Concurrent validity	Excellent
Koopman et al, 2002 ²⁵	N = 175, Workers (general)	Responsiveness	Excellent
		Reliability (internal consistency)	Fair
		Content validity	Excellent
		Structural validity	Fair
		Construct validity	Fair
Lalic et al, 2012 ²⁷	N = 241, Workers (manual)	Concurrent validity	Fair
		Reliability (internal consistency)	Fair
Sanderson et al, 2007 ³⁹	N = 432, Depression and anxiety	Reliability (internal consistency)	Excellent
Tang et al, 2009 ⁴⁵	N = 80, Shoulder and elbow injuries	Reliability (internal consistency)	Good
		Construct validity	Good
VOLP			
Zhang et al, 2011 ⁵¹	N = 152, Rheumatoid arthritis	Reliability (test-retest)	Fair
		Construct validity	Fair
		Concurrent validity	Fair
Zhang et al, 2012 ⁵⁴	N = 212, Rheumatoid arthritis or osteoarthritis	Content validity	Excellent
WHI			
Stewart et al, 2004 ⁴¹	N = 67, Workers (call center)	Concurrent validity	Fair
WPAI:CD			
Reilly et al, 2008 ³⁶	N = 380, Crohn's disease	Construct validity	Good
		Responsiveness	Good
WPAI:CG			
Giovannetti et al, 2009 ²²	N = 308, Caregivers	Construct validity	Good
WPAI:GERD			
Brozek et al, 2009 ¹⁷	N = 217, Gastroesophageal reflux disease	Responsiveness	Fair
Wahlqvist et al, 2002 ⁵⁰	N = 136, Gastroesophageal reflux disease	Construct validity	Good
Wahlqvist et al, 2007 ⁴⁹	N = 130, Gastroesophageal reflux disease	Construct validity	Good
		Responsiveness	Fair
WPAI:GH			
Braakman-Jansen et al, 2012 ¹⁶	N = 62, Rheumatoid arthritis	Concurrent validity	Fair
Erickson et al, 2009 ¹⁶	N = 76, Anxiety disorders	Construct validity	Fair
Reilly et al, 1993 ³⁸	N = 106, Workers (general)	Reliability (test-retest)	Fair
		Construct validity	Good
Zhang et al, 2011 ⁵¹	N = 152, Rheumatoid arthritis	Concurrent validity	Fair
Zhang et al, 2010 ⁵³	N = 212, Rheumatoid arthritis or osteoarthritis	Construct validity	Fair
Zhang et al, 2010 ⁵²	N = 150, Rheumatoid arthritis	Concurrent validity	Fair
		Construct validity	Fair
		Concurrent validity	Fair

(continued)

■ **Table 3.** Summary of Studies Assessing Measurement Properties of Presenteeism Instruments (*continued*)

Instrument and Studies	Population	Properties Evaluated	Quality
WPAI:IBS			
Reilly et al, 2004 ³⁷	N = 133, Irritable bowel syndrome	Reliability (test-retest)	Poor
		Construct validity	Fair
WPAI: SpA			
Reilly et al, 2010 ³⁵	N = 205, Ankylosing spondylitis	Construct validity	Fair
		Responsiveness	Fair
WPS:RA			
Osterhaus et al, 2009 ³²	N = 220, Rheumatoid arthritis	Construct validity	Fair
		Responsiveness	Fair
WPS			
Erickson et al, 2002 ²¹	N = 369, Asthma	Reliability (internal consistency)	Good
		Construct validity	Fair
Erickson et al, 2009 ²⁰	N = 76, Anxiety disorders	Construct validity	Fair
McBurney et al, 2004 ³⁰	N = 89, Myocardial infarction	Reliability (internal consistency)	Fair
WPSI			
Goetzel et al, 2003 ²³	N = 610, Workers (general)	Reliability (internal consistency)	Fair
		Reliability (test-retest)	Poor
		Content validity	Fair
Ozminkowski et al, 2003 ³⁴	N = 206, Workers (general)	Content validity	Fair
		Construct validity	Good
Ozminkowski et al, 2004 ³³	N = 532, Workers (telecommunications)	Construct validity	Good

ALWQ indicates Angina-Related Limitations at Work Questionnaire; EWPS, Endicott Work Productivity Scale; HLO, Health and Labour Questionnaire; HPQ, World Health Organization Health and Work Performance Questionnaire; HRPO-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:SpA, Work Productivity and Activity Impairment scale: Ankylosing Spondylitis; WPS:RA, Work Productivity Survey: Rheumatoid Arthritis; WPS, Work Performance Scale from the Functional Status Questionnaire; WPSI, Work Productivity Short Inventory.

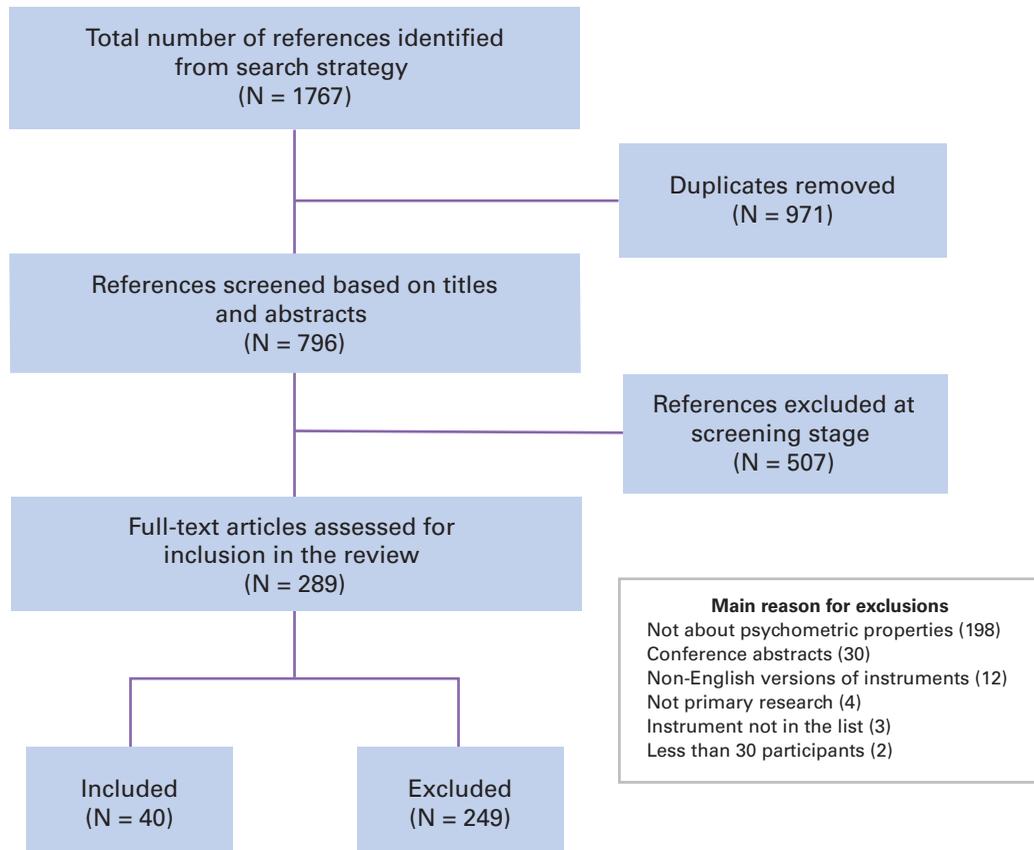
METHODS

Table 4. Level of Evidence of Quality for Presenteeism Instruments—COSMIN Appraisal

Instrument	Internal Consistency	Reliability	Content Validity	Structural Validity	Hypothesis Testing (construct and convergent validity)	Criterion Validity	Responsiveness
ALWQ	✓	—	✓	—	✓	—	—
EWPS	✓✓✓	✗	—	—	✓✓✓	—	✓✓✓
HLQ	—	—	✓	—	✓	—	—
HPQ	—	—	✓	—	✓	—	—
HRPQ-D	—	—	—	—	✓	—	✗
HWQ	✓✓✓	—	✓	✓✓✓	✓	✗	—
LEAPS	✓	—	—	✓	✓	—	—
MIDAS	✓	✓✓	—	—	✓	—	—
MWPLQ	✓✓✓	—	—	—	✓	—	—
Q-Q	—	—	—	—	✓	—	—
SPS-13	✓	—	—	✓	✓	—	—
SPS-6	✓✓✓	—	✓✓✓	✓	✓✓✓	—	✓✓✓
VOLP	—	✓	✓✓✓	—	✓	—	—
WHI	—	—	—	—	✓	—	—
WPAI:CD	—	—	—	—	✓✓	—	✓✓
WPAI:CG	—	—	—	—	✓✓	—	—
WPAI:GERD	—	—	—	—	✓✓	—	✓
WPAI:GH	—	✓	—	—	✓✓	—	—
WPAI:IBS	—	✗	—	—	✓	—	—
WPAI: SpA	—	—	—	—	✓	—	✓
WPS:RA	✓	—	—	—	✓	—	✓
WPS	✓✓	—	—	—	✓	—	—
WPSI	✓	✗	✓	—	✓✓	—	—
Level of evidence	Poor ✗	Limited ✓	Moderate ✓✓		Strong ✓✓✓		No Evidence —

ALWQ indicates Angina-Related Limitations at Work Questionnaire; COSMIN, COnsensus-based Standards for the election of health Measurement INstruments; 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: Ankylosing Spondylitis; 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.

■ **Figure.** PRISMA Flow Diagram



PRISMA indicates Preferred Reporting Items for Systematic Reviews and Meta-Analyses.