PharmacoEconomics (2015) 33:31–48
DOI 10.1007/s40273-014-0209-z

REVIEW ARTICLE

Estimating Productivity Costs in Health Economic Evaluations: A Review of Instruments and Psychometric Evidence

Kenneth Tang

Published online: 29 August 2014

Abstract Health economic evaluations (i.e. cost-effectiveness appraisal of an intervention) are useful aids for decision makers responsible for the allocation of scarce healthcare resources. The relevance of including health-related productivity costs (or benefits) in these evaluations is increasingly recognized and, as such, reliable and valid instruments to quantify productivity costs are needed. Over the years, a number of work productivity instruments have emerged in the literature, along with a growing body of psychometric evidence. The overall aim of this paper is to provide a review of available instruments with potential for estimating health-related productivity costs. This included the Health and Labor Questionnaire, Health and Work Performance Questionnaire, Health-Related Productivity Questionnaire Diary, Productivity and Disease Questionnaire, Quantity and Quality method, Stanford Presenteeism Scale 13, Valuation of Lost Productivity, Work and Health Interview, Work Limitations Questionnaire, Work Productivity and Activity Impairment Questionnaire, and Work Productivity Short Inventory. Critical discussions on the instruments' overall strengths and limitations, applicability for health economic evaluations, as well as the methodological quality of existing psychometric evidence were provided. Lastly, a set of reflective questions were proposed for users to consider when selecting an instrument for health economic evaluations.

Key Points for Decision Makers

Existing work productivity instruments differ in many respects, including breadth of content, length, approach to quantify the various sources of productivity loss (e.g. absenteeism, presenteeism, unpaid work), compatibility with different valuation approaches, and available psychometric evidence.

Users selecting an instrument for health economic evaluation should consider the five 'Ps': *purpose*, *perspective*, *practicality, population*, and *psychometrics*.

1 Introduction

Health economic evaluations (i.e. cost-effectiveness appraisal of an intervention) are useful aids for decision makers responsible for allocating scarce healthcare resources. Typically, the direct and indirect costs are distinguished. Direct costs represent expenses related to health service utilization (e.g. treatments, visits to health care providers), while indirect costs refer to other resources forgone, such as the costs of lost productivity. Productivity costs can be substantial and their inclusion in health economic evaluations is increasingly recommended [1–3]. To offer decision makers a most comprehensive view of findings, some have suggested that cost-effectiveness ratios

K. Tang (✉)
Mobility Program Clinical Research Unit,
Li Ka Shing Knowledge Institute of St. Michael's Hospital,
30 Bond Street, Toronto, ON, Canada
e-mail: ken.tang@mail.utoronto.ca

K. Tang
Institute of Health Policy, Management and Evaluation,
University of Toronto, Toronto, ON, Canada

K. Tang
Institute for Work and Health, Toronto, ON, Canada

from health economic evaluations should be presented both with and without indirect costs (i.e. a sensitivity analysis) [1]. Where a decision to include productivity costs is made, the ability to accurately estimate such costs becomes imperative. This necessitates a reliable and valid instrument. Much scholarly efforts have been dedicated to this end, as a considerable number of work productivity instruments have emerged in the literature [4, 5].

The overall objective of this paper was to provide a review of instruments with potential for estimating health-related productivity costs. The specific objectives were to (1) appraise the content comprehensiveness of existing instruments and discuss their applicability for health economic evaluations; (2) provide a critical review of existing psychometric evidence; and finally, (3) offer a series of reflective questions for users to consider when engaged in an instrument selection process.

1.1 Estimating Productivity Costs: An Overview of Key Conceptual and Methodological Issues

Estimating productivity costs requires considerations for many conceptual and methodological issues, and extensive discussions and debates on these issues are ongoing (for more detailed reviews, see Zhang et al. [6], Krol et al. [7], Krol and Brouwer [8], and Koopmanschap and Rutten [9]). To provide some background and context for the review of instruments, only a brief overview of key issues is presented here.

1.1.1 Sources of Productivity Costs (Domains)

To obtain an accurate estimate of productivity costs, the pertinent sources of productivity loss must be identified. Since employment constitutes an important societal role, productivity loss during paid work represents a significant source of productivity costs. In this context, lost productivity attributed to absenteeism or presenteeism is commonly distinguished. Absenteeism traditionally refers to *missed* work time; however, some have recently advocated for an expanded definition that would also encompass *lost* work time due to changes in employment status, such as a reduction in routine working time, temporary work cessation, job loss, or early retirement [6, 10, 11]. Presenteeism, on the other hand, refers to reduced productivity while at work due to health problems. Its relevance is increasingly recognized as the cost of presenteeism has shown to exceed that of absenteeism for various health conditions [12–17].

Beyond absenteeism and presenteeism, the relevance of compensation mechanisms and work-team dynamics has also garnered some attention. It is suggested that lost productivity can be partially compensated within a workplace if, for example, the missed work is made up at a later time by the worker (i.e. self-compensation during normal or unpaid overtime hours), some of the work tasks are taken over by colleagues, and/or new hires are made by the employer [9]. Some empirical evidence has shown that such mechanisms exist and can substantially reduce productivity loss following worker illness [18, 19]. On the other hand, the dynamics of the work team may also have important implications as illness to one worker can jeopardize the productivity output of co-workers. This is known as 'multiplier effects' [20–22]. Multiplier effects are thought to be especially relevant in certain situations; for example, if production at work is team oriented, if the ill worker is a key member of the team (i.e. difficult to replace with substitute workers), or if production cannot be easily postponed (i.e. penalty associated with not meeting an output target) [22]. As such, the need to 'correct' for both compensation mechanisms and work-team dynamics has been suggested when estimating health-related productivity costs. To date, very few attempts have been made to adjust productivity costs for these influences [18, 19, 21, 22].

In addition to paid work, poor health can also impact one's ability to perform productive activities outside the conventional labor market (i.e. unpaid work). The relevance of unpaid work is dependent on the perspective taken. From a societal perspective of health economics (i.e. where all costs incurred are relevant irrespective of the bearer), unpaid work activities has economic value and contributes to societal welfare, and therefore a cost is incurred when such activities are not performed due to health reasons [2, 3, 23, 24]. However, what constitutes as unpaid work remains of some debate. Reid's 'third person criterion' [25] is sometimes suggested as a useful definition [7, 8], which proposes that only activities replaceable by a third person should be considered unpaid work, while other activities should be considered leisure [25]. Key examples of unpaid work by this criterion would include volunteering activities, household activities (e.g. cleaning, cooking) or caregiving activities (children or elderly).

1.1.2 Measurement Issues

For paid work, direct quantification of health-related labor output loss (i.e. reduced production) is inherently difficult as many jobs do not produce tangible outputs (e.g. knowledge-based jobs). Even when possible, it may not be feasible since this can require substantial resources (e.g. cost of objective data collection) and raise sensitive ethical issues (e.g. worksite monitoring of employees) [6]. A more practical and commonly used approach is to estimate health-related labor input loss using self-report questionnaires, namely by quantifying the equivalent lost work time attributed to absenteeism and presenteeism. While health-related labor input and output losses are naturally linked, it

should be recognized that this relationship may not be directly proportional (i.e. increase or decrease in lock-step) due to compensation mechanisms, work-team dynamics, and/or some general slack in the work organization (i.e. availability of extra staff).

In terms of compensation mechanisms and work-team dynamics, few approaches to quantify these influences have been proposed to date as this remains a relatively novel research topic. One key measurement issue concerns the reliance on employees (respondents) having sufficient knowledge of how lost productivity is compensated for within their workplace, and also whether their health problems have actually impacted co-workers' productivity. If not, the involvement of employers and/or managers may be needed, which can obviously raise feasibility issues during data collection. Nonetheless, a number of recent studies have shown employee reports of these influences to be valid and feasible [18, 19, 26], which are encouraging findings on this front.

Compared with paid work, lost productivity for unpaid work is inherently more challenging to quantify since these activities are diverse and also often less structured (e.g. less formally scheduled). This raises a number of additional issues. First, recall problems become a potential concern for respondents. Second, it may be imperative to account for 'substitution' effects, that is whether unpaid work normally performed by the respondents has been postponed until a later time, or if help was received to perform such work by family, friends, or through hired help (i.e. 'replaced' help). These are also challenging to quantify with a high degree of precision. Third, on a more conceptual level, it is not always possible to clearly disentangle between unpaid work and leisure as many unpaid activities intrinsically involve elements of both (e.g. child caring) [7].

1.1.3 Valuation Approaches

Over the years, the merits of the human capital (HC) and friction cost (FC) approaches of valuating lost productivity have been a topic of much scholarly debate [27–30]. The HC approach originates from the theory of HC [31], which posits that health-related productivity costs should be represented as a function of the amount of missed work time due to health reasons and the expected wage over a given period of interest. The HC approach is commonly used, and strengths of this approach are its computational ease, intuitive plausibility, and consistency with economic theory. However, a key criticism is that productivity costs can be unreasonably high in some cases, which may reflect the *potential* rather than *actual* lost production value [32]. As a refinement to the HC approach, the FC approach argues that health-related productivity loss would only be limited to a 'friction period', which is the time needed for the previous level of production to be restored [27]. This is predicated on the assumption of the existence of involuntary unemployment, where ill/disabled workers could be replaced by healthy (unemployed) persons in society [33]. Naturally, compared with the HC approach, the FC approach should generate lower productivity cost estimates in most instances. In practice, some additional information is needed in order to apply the FC approach: (1) sufficiently detailed information of absenteeism (i.e. number of periods and duration); (2) accurate information on the friction period (which remains scarcely available at present); and also (3) an estimate of the transactional costs (e.g. for advertising, hiring, and training of new staff) during the friction period.

To assign to monetary value for lost unpaid work, a shadow price is typically assigned to the missed time dedicated for such activities. In this regard, proxy good approach and opportunity cost approach are most commonly suggested as viable alternatives in the literature [2, 34]. For the proxy good approach, the value of lost unpaid work would be determined by the value of a market substitute for the specific activity (e.g. hired housekeeper or caregiver). For the opportunity cost approach, the value of lost unpaid work would be based on the (potential) value of alternative use of time; for example, the value of such time had it been spent on performing paid work. More detailed discussions of the strengths and limitations of these valuation approaches can be found elsewhere [2, 3, 6–8, 34].

1.2 Psychometrics: What Properties Should An Instrument Demonstrate?

The psychometric soundness of an instrument should be empirically verified to ensure that it is trustworthy and functioning as intended. This implies the need for evidence of reliability and validity, which are both fundamental psychometric properties. As some instruments are designed to have broad applicability, it is often useful to (re)test an instrument in multiple settings (e.g. different occupational sectors, clinical populations) to inform the generalizability of psychometric properties.

1.2.1 Reliability

Two main types of reliability of relevance can be differentiated: reproducibility (test–retest reliability) and internal consistency. Reproducibility concerns to what extent an instrument's scores (e.g. productivity cost estimates) remain the same for tests conducted in different occasions for individuals who are unchanged. This requires an assessment of the extent of agreement in scores between the different testing occasions. Typically, intraclass

correlations (ICC, for continuous scores) [35] or kappa statistics (k, for categorical scores) [36] are most appropriate for quantifying the extent of agreement as tests of correlation are generally inadequate [37, 38]. Generally, ICC or k >0.70 have been suggested as affirmative of reproducibility [38, 39]. Internal consistency concerns the interrelatedness of items within a multi-item, summative scale. This is a relevant property for (sub)scales that intend to measure a single underlying concept based on multiple items. This is typically expressed by Cronbach's alpha (for Likert items) [40] or the Kuder–Richardson Formula 20 (for dichotomous items) [41]. For (sub)scale purported to be undimensional, Cronbach's alpha between 0.70 and 0.95 is generally considered ideal as an exceedingly high Cronbach's alpha often indicates redundancy of items [38].

1.2.2 Validity

Validity concerns the degree to which an instrument measures what it claims, or purports, to be measuring [42, 43]. With the exception of face/content validity, most other forms of validity can be empirically tested. This includes, for example, construct validity, criterion validity, and factorial (or structural) validity. Construct validity concerns whether an instrument relates to other measures (i.e. comparators) in an expected manner. Generally, a measure is expected to correlate strongly with similar measures (i.e. convergent validity) and weakly with dissimilar measures (i.e. divergent validity). This form of validity is most commonly of interest and can be assessed cross-sectionally or longitudinally [44]. In the context of estimating productivity costs, a cross-sectional assessment would be concerned with the validity of a cost estimate assessed at one point in time, whereas a longitudinal assessment would be concerned with the validity of a difference in costs gathered from two different testing occasions.[1] Criterion validity refers to the special case of construct validity testing where the comparator represents a 'gold standard' (i.e. a proven indicator), in which case a very high correlation (i.e. >0.70) with the instrument in question might be expected [38]. In the current context, objectively-measured production data might be reasonably viewed as a criterion. Lastly, factorial (or structural) validity concerns whether the subscale organization of the instrument is adequately represented. This property pertains only to multidimensional instruments.

In recent years, the methodological quality of validity testing for self-report instruments has been increasingly emphasized in the literature [38, 45]. For instance, the choice of comparators applied for validity testing (where relevant) is expected to be meaningful and justifiable. In addition, formulation of specific a priori hypotheses on the anticipated relationship is advocated, which provides the basis for the interpretation of results and, ultimately, inferences on an instrument's psychometric performance [45]. This may involve specifying the expected level of agreement/correlation between an instrument and its comparator(s).

[1] Longitudinal construct validity is sometimes considered as a type of responsiveness.

2 Review of Instruments

2.1 Inclusion Criteria and Search Strategy

To be included, this review sought instruments that (1) can be administered in the form of a questionnaire; (2) can be used to estimate productivity costs (i.e. a monetary value can be assigned to the derived metric); (3) are designed to have broad applicability (i.e. generic instruments); and (4) have undergone some formal psychometric testing (i.e. at least one study dedicated to assessments of reliability or validity could be found from the literature search).

To identify qualifying instruments and relevant psychometric evidence, the following search strategy was applied. First, multiple electronic databases including MEDLINE, EMBASE, *Psyc*INFO and Web of Science were searched for relevant articles in English from the inception of each database until 2013 (inclusive). The search strategy involved combining three sets of keywords (mapped onto Subject Heading specific to each database, where applicable) using the Boolean operator 'AND': (1) *work* or *employment* or *labor*; (2) *productivity* or *absenteeism* or *presenteeism* or *sick leave*; (3) *measure* or *instrument* or *questionnaires*; and (4) *reliability* or *validity* or *responsiveness* or *reproducibility* or *psychometrics* or *internal consistency* or *sensitivity*. After identifying relevant articles, the reference lists were hand searched to identify any additional resources that warrant further examination. Once the complete list of qualifying instruments was determined, keyword searches of each instrument name were ran to uncover any additional articles (e.g. *Health adj Labor adj Questionnaire*, or *HLQ*). Lastly, an online search was performed to identify and consult any additional sources of information provided by instrument developers. Only English versions of the instruments were considered in the review.

A total of 11 instruments qualified for the current review (listed in alphabetical order): Health and Labor Questionnaire (HLQ), Health and Work Performance Questionnaire (HPQ), Health-Related Productivity Questionnaire Diary (HRPQ-D), Productivity and Disease Questionnaire (PRODISQ), Quantity and Quality method (QQ), Stanford Presenteeism Scale 13 (SPS-13), Valuation of Lost

Productivity (VOLP), Work and Health Interview (WHI), Work Limitations Questionnaire (WLQ), Work Productivity and Activity Impairment Questionnaire (WPAI), and Work Productivity Short Inventory (WPSI).

2.2 Health and Labor Questionnaire

The HLQ was developed by researchers from the Erasmus University Institute for Medical Technology Assessment in Rotterdam, The Netherlands, to measure reduced labor performance due to illness and the associated costs of lost production [46, 47]. It uses a 2-week recall period and is divided into four main modules assessing (1) absence from paid work (absenteeism); (2) reduced productivity at paid work (presenteeism); (3) unpaid labor production; and (4) impediments to paid and unpaid labor. To assess absenteeism, respondents are asked to record whether each day of the preceding 2 weeks was a scheduled workday, and, for each scheduled workday, whether they were present or absent due to health or another reason. An additional question assessing the onset of illness is also included to facilitate an FC of valuation. Presenteeism is assessed by a direct hour-estimating method (single question), which asks respondents to estimate the number of additional hours that would be required to compensate for production losses due to health reasons on working days. This method is thought to yield a conservative estimate of productivity loss, particularly in the case where catching up on missed work is not permitted [48, 49]. To assess productivity loss during unpaid work, respondents are asked to estimate the number of hours spent (per week) for household work, shopping, odd jobs, chores around the house, and caring for children, and whether they were postponed or taken over by other household members, and/or paid workers. Completion time for the HLQ is estimated to be 10 min [46]. More recently, a short form of the HLQ has been introduced (SF-HLQ), which is a streamlined version of the original instrument [50].

The HLQ's direct hour-estimating method of assessing presenteeism has garnered the most research attention to date [46, 48, 51, 52], and these studies have generally revealed that it is moderately related to other approaches of assessing presenteeism. In van Roijen et al. [46], a moderate correlation ($r = 0.41$) was found with the Osterhaus method (i.e. self-assessment of efficiency loss [0–10] during working hours) [53]. In a study of trade firm workers, HLQ presenteeism was shown to moderately correlate with the Osterhaus method ($r = 0.38$), as well as the quantity item from the QQ method ($r = 0.40$) [48]. In a study of workers with rheumatoid arthritis (RA), HLQ presenteeism was shown to moderately correlate with the QQ method ($r = 0.34$), as well as the presenteeism item from the general health version of the WPAI ($r = 0.48$) [51]. In a notable exception, the HLQ presenteeism showed very high agreement (89 %) with the QQ method among industrial and construction workers, although the k was only 0.18 [49, 52].

HLQ strengths:

Comprehensive instrument, measures lost productivity due to absenteeism, presenteeism, and unpaid work.

Thorough assessment of unpaid work (i.e. corrects for substitution).

One of few instruments designed to have compatibility with the FC approach of valuation.

HLQ limitations:

Some potential feasibility issues as sub-optimal completion rates have been observed in previous studies [47, 49, 52].

Some workers may have cognitive difficulties with the direct hour-estimating method of assessing presenteeism [47, 52].

2.3 Health and Work Productivity Questionnaire

Developed in collaboration with the World Health Organization, the HPQ is designed to quantify the impact of health problems on productivity in the workplace [54, 55]. The instrument was initially developed through a literature review, systematic refinement of questions initially generated by experts, and additional pilot and cognitive testing. The employee (long) version of the HPQ assesses for common health problems, ongoing treatment, work performance (absenteeism and presenteeism), basic demographics, and occupational information. In addition, two clinical trial versions (recall periods of 7 days or 4 weeks), and a short version with only seven total questions on work absenteeism and presenteeism are also available. As proposed by the developers, several productivity metrics can be derived from the HPQ [56]: *absolute absenteeism*, *relative absenteeism*, *absolute presenteeism*, and *relative presenteeism*. *Absolute absenteeism* is expressed in raw hours, which is calculated as the difference between the expected and actual number of hours worked. *Relative absenteeism* is calculated by dividing *absolute absenteeism* by the number of expected work hours. *Absolute presenteeism* is calculated by multiplying 'self-rating of usual work performance (range 0–10)' by 10 (100 = best performance). Finally, *relative presenteeism* is calculated as a ratio of own work performance (item B11 in the short version) divided by the work performance of co-workers in a similar job (item B9 in the short version), and restricting the ratio to a range of 0.25–2.0 as there is no limit on the upper-bound value [56]. To derive a combined productivity metric, the developers recommend combining *relative absenteeism* with either *absolute presenteeism* or *relative*

presenteeism (for details, see Kessler et al. [56] and http://www.hcp.med.harvard.edu/hpq).

Validity of the HPQ has been examined in several studies. Kessler et al. [54] found that absenteeism metrics from the HPQ showed good concordance ($r = 0.61$–0.87) with employer payroll records (e.g. days/hours of work missed) when the instrument was applied to workers from different occupations (airline reservation agents, customer service representatives, automobile company executives, railroad engineers). In this study, HPQ presenteeism metrics were also shown to correlate with assessments of work performance based on work audits and supervisor/peer ratings [54]. In a comparison of four different approaches to quantify presenteeism among workers with arthritis, Zhang et al. [57] found that the HPQ moderately agreed with the WPAI (ICC = 0.61), but considerably less so with the WLQ Index (ICC = 0.26) and the HLQ (ICC = 0.16).

HPQ strengths:

Extensive initial instrument development process.

Metrics have been tested against meaningful productivity indicators (e.g. payroll records, supervisor ratings).

Several versions are available depending on user's objectives and feasibility considerations.

HPQ limitations:

Does not consider lost productivity for unpaid work.

Applicability of some productivity metrics is unclear (e.g. relative presenteeism).

Does not assess compensation mechanisms or work-team dynamics.

2.4 Health-Related Productivity Questionnaire Diary

The HRPQ-D is designed for assessing health-related labor force participation. It includes domains on absenteeism (missed hours), presenteeism (reduced work effectiveness), and work status changes (e.g. early retirement or reduction from full-time to part-time employment) [58]. Item development and testing were initially conducted from focus groups comprised of patients with Parkinson's disease. The initial published version of the HRPQ-D was in a condensed diary format, specifically designed for infectious mononucleosis [58]. The HRPQ-D requires data collection each day over a 1-week period. Each day, three pieces of information (number of hours planned/scheduled, number of hours missed because of health, and effectiveness during the hours worked) are collected for each of three different productivity venues—paid work (i.e. employment), housework, and educational activities (i.e. attending classes, doing homework). Metrics representing *weekly absenteeism* (total number of hours missed) and *weekly presenteeism* (productivity decreases due to reduced effectiveness for the hours spent performing paid work, housework or educational activities) can be derived for the purpose of estimating productivity costs. A combined lost productivity metric (sum of absenteeism and presenteeism) can also be derived, either for each of the three productivity venues separately or for all venues combined. Only one psychometric study on the HPRQ-D can be found to date. Kumar et al. investigated the relationship between HRPQ-D metrics and symptom scores among young adults (age range 14–32 years) suffering from infectious mononucleosis [58].

HRPQ-D strengths:

Evaluates lost productivity for both paid and unpaid work.

Short questionnaire, highly feasible.

HRPQ-D limitations:

Assessment of unpaid work does not correct for substitution.

Educational activities would not be considered as unpaid work by Reid's 'third person criterion'.

Does not assess or consider potential compensation mechanisms or work-team dynamics.

Very limited psychometric evidence is currently available.

2.5 Productivity and Disease Questionnaire

The PRODISQ is a modular questionnaire designed for estimating productivity costs for health economic evaluations [49], and initial testing used and compared aspects of the HLQ and the QQ method in specific modules. The PRODISQ consists of seven modules: (1) demographics and disease; (2) working situation and income; (3) absenteeism; (4) compensation mechanisms in case of paid work absenteeism; (5) productivity costs at work (conceptualized as 'efficiency loss'); (6) productivity costs at the organizational level; and (7) administrative and management costs. Modules 3–5 are considered most essential for estimating productivity cost, while other modules are considered as optional. Absenteeism is assessed by asking respondents two questions: 'How many working days they were absent from work during a 3-month recall period?' and 'In how many periods did this absenteeism take place?' For presenteeism, the QQ method is recommended by the developer [49], which is described in greater detail below. The module on compensation mechanisms proposed five short questions based on initial work by Severens et al. [19], asking about whether lost productivity had been compensated by self or co-workers.

PRODISQ strengths:

Fairly comprehensive, assesses lost productivity for absenteeism, presenteeism, compensation mechanisms, and work-team dynamics.

Includes additional modules for investigating employers' perspectives of productivity costs as well as administrative and management costs of worker illness.

PRODISQ limitations:

Does not assess lost productivity for unpaid work.

Questions on work-team dynamics require participation by employers or managers.

Limited psychometric evidence to date on the various modules.

2.6 Quantity and Quality Method

The QQ method was designed to assess the impact of health on the quantity and quality of work performed during the most recent workday, which serves as an indication of presenteeism [48]. The rationale for developing the QQ was the recognition that, in addition to work quantity, work quality was also important since there is a cost associated with having to repeat work of impaired quality (i.e. less is ultimately produced for a given amount of work time). The QQ method can be self-administered and consists of only two numeric rating scales (0–10). The quantity item asks respondents to indicate how much work is performed compared with normal (anchors: practically nothing–normal quantity). Similarly, the quality item asks respondents to indicate the quality of the work performed compared with normal (anchors: very poor quality–normal quality). To obtain a productivity metric (i.e. equivalent lost time), a multiplicative approach based on quantity and quality items have been proposed based on the general formula: [(10 − QQ)/10, multiplied by the number of hours worked during the most recent day] (for details, see Brouwer et al. [48]). Interestingly, moderate-to-strong correlations between the quantity and quality items have been demonstrated in several studies [48, 52], raising questions about whether respondents are able to sufficiently discriminate between the quantity and quality aspects of work [48, 49]. In some studies, only the quantity score has been considered in the QQ calculation.

Several studies have been conducted to examine the construct validity of the QQ method, including a comparison with objectively-measured production output. Among a small sample of floor layers ($n = 19$), the quantity score showed moderate correlation ($r = 0.48$) with the area of surface made during work hours according to worksite observations [52]. Compared with other self-report measures of presenteeism, a range of correlations have been observed. When applied to trade firm workers, the QQ correlated very strongly with the Osterhaus method ($r = 0.92$), and moderately with HLQ's direct hour-estimating method ($r = 0.40$) [48]. When applied to workers with RA, the QQ method correlated moderately with the WPAI ($r = 0.61$), but showed a weaker relationship with the HLQ's direct-hour estimating method ($r = 0.34$) [51].

QQ strengths:

Brevity, highly feasible for use.

Concepts of work quantity and quality are intuitive, and have broad relevance.

QQ limitations:

Assesses only presenteeism.

Some ongoing uncertainty regarding how to amalgamate the scores from both quantity and quality items.

2.7 Stanford Presenteeism Scale 13

The SPS-13 is designed to assess health-related productivity loss attributed to a single (primary) health condition, and is applicable for diverse job types [59]. It has a 4-week recall period. Other variations of this scale have also been published (e.g. SPS-6, SPS-34) [60, 61]; however, the SPS-13 is the only version with an absenteeism component. The SPS-13 begins with a list of ten major health conditions and asks respondents to declare a primary condition as the focus for the remaining questions. These conditions include allergies, arthritis or joint pain/stiffness, asthma, back or neck disorder, breathing disorder (bronchitis, emphysema), depression/anxiety or emotional disorder, diabetes, heart or circulatory problem (artery disease, high blood pressure, angina), migraines/chronic headaches, stomach or bowel disorder; and other (can be specified by the respondent). The next section consists of ten Likert-type questions that query the degree of work impairment due to the primary health condition (e.g. ability to finish tasks, focus on work goals, or work with colleagues). Responses to these questions are aggregated and then transformed to a Work Impairment Score (WIS, range 0–100), with higher scores indicating greater presenteeism. An additional single item (global question) queries the percentage of usual productivity achieved by the respondent, which generates the Work Output Score (WOS, range 0–100), with higher scores indicating less presenteeism. Between these two metrics, the WOS is considered more suitable for quantifying productivity costs [59, 62]. The final question from the SPS-13 assesses work absenteeism, which asks respondents the total number of hours missed at work over the past 4 weeks.

Only one study assessing the internal consistency and construct validity of the SPS-13 could be found to date,

which applied the instrument to workers from a large research and manufacturing corporation in the US [62]. In this study, factor analysis of the ten Likert-type questions that make up the WIS initially revealed two distinct factors—'completing work' (Cronbach's alpha = 0.97) and 'avoid distraction' (Cronbach's alpha = 0.60)—but only the former demonstrated acceptable internal consistency. Both presenteeism metrics of the SPS-13 (WIS: $r = 0.50$; WOS: $r = 0.40$) were shown to correlate moderately with the WLQ Index [62].

SPS-13 strengths:

WIS and WOS offer multiple perspectives of presenteeism.

Offers opportunity to assess lost productivity attributed only to a primary health condition.

Relatively brief questionnaire.

SPS-13 limitations:

Does not assess lost productivity for unpaid work.

Uncertain to what extent respondents are able to precisely attribute productivity loss to only one specific health condition.

Very limited psychometric evidence to date.

Does not assess compensation mechanisms or work-team dynamics.

2.8 Valuation of Lost Productivity Questionnaire

The VOLP is a self-administered, modular questionnaire developed by researchers from the Centre for Health Evaluation and Outcome Sciences in Vancouver, Canada [26]. The VOLP features six sections which assess employment status (e.g. reduced routine working time, job loss, and early retirement), job characteristics, absenteeism, work performance (presenteeism), unpaid work and dynamics of the work environment. Absenteeism is quantified by the total lost work time over the past 3 months from absent and partial workdays (where the respondents went in late or left early) due to health reasons. Presenteeism is assessed in a way similar to the HLQ's direct hour-estimating method but with some adjustments. In the VOLP, respondents are asked to indicate the number of hours they had actually spent completing their work during the past 7 days, and the estimated number of hours they would require to do the same work had they not experienced any health problems. This can be expressed as percentage time loss by taking the difference in hours and dividing it by the number of hours actually spent working. Unpaid work loss is measured as the total number of hours respondents received help on unpaid work activities (including household work, shopping, odd jobs and chores, childcare and volunteer activities) due to health reasons over the past 7 days. Lastly, the VOLP features a final module consisting of questions on work-team dynamics which assesses whether production is team-oriented, availability of replacement workers and time-sensitivity of work output. These questions are intended to facilitate the calculation of 'wage multipliers', which serve as an adjustment factor applied to initial estimates of labor input loss (for details, see Zhang et al. [26]). Additional information on the VOLP can also be found online at http://www.thevolp.com.

Psychometric properties of the VOLP have only been assessed among workers with RA to date [63]. In this study, the instrument's test–retest reliability was examined among respondents considered stable between assessment timepoints 2 weeks apart. Good agreement between test–retest scores was found for absenteeism ($k = 0.80$) and presenteeism ($k = 0.76$), although unpaid work loss appeared somewhat less reliable ($k = 0.35$). Some support for its construct validity was also demonstrated based on comparisons with corresponding metrics from the WPAI. Between these two instruments, moderate correlations were found for absenteeism ($r = 0.57$), presenteeism ($r = 0.42$), as well as unpaid work activities ($r = 0.39$) [63].

VOLP strengths:

Very comprehensive instrument, assesses lost productivity for absenteeism, presenteeism, unpaid work, compensation mechanisms, and work-team dynamics.

Module on work-team dynamics can be completed by respondents (i.e. does not require employer/manager participation).

Questionnaire enables wage multipliers to be derived as an adjustment factor, which is a promising approach to relate wage to marginal productivity.

VOLP limitations:

Can be somewhat lengthy if all modules are to be completed.

Only 'replaced' unpaid work is considered.

Reliant on employees having sufficient knowledge about work-team dynamics (to derive wage multipliers).

2.9 Work and Health Interview

The WHI was originally designed as part of the American Productivity Audit, which is a 10–15 min computer-assisted telephone survey [12, 14, 64]. It is a modular instrument that assesses employment status, usual work time, presence of 22 acute and chronic health conditions, health-related lost productive time, job characteristics, and demographics. Absenteeism is calculated as the sum of missed workdays and missed time during workdays (i.e. late start, early departure) due to health reasons over a 2-week recall period, which is converted into amount of lost productive time (in hours). Presenteeism is defined as reduced work performance during the same recall period,

and is assessed by five Likert-type questions on specific work behaviors (how often lost concentration, repeated a job, worked more slowly than usual, felt fatigued at work, and did nothing at work on days when they were at work not feeling well), plus an additional question about the average amount of time required for respondents to start working after arriving at work on days not feeling well. Response to these questions are then translated into equivalent lost productive time (for details on calculation, see Stewart et al. [64]). A combined WHI metric can be derived by summing health-related lost productive time due to absenteeism and presenteeism.

During initial development, three versions of the instrument with varying recall periods were compared and it was determined that a 2-week recall period may be best for minimizing reporting errors [65]. Construct validity of the WHI was examined in a study of 67 inbound phone call agents in northern California, where its productivity metrics were compared against workplace data (i.e. administrative data and routine continuous performance data collected in real time) and diary data collected once each hour over a 10-workday period) [66]. Results from this study found significant correlations between WHI metrics and alternative methods of assessing absenteeism ($r = 0.76$ vs. workplace data), presenteeism ($r = 0.31$ vs. workplace data; $r = 0.33$ vs. diary method), as well as total loss productive time ($r = 0.63$ with workplace data).

WHI strengths:

Reasonable completion time, despite a computer-assisted survey.

Instrument has been validated against workplace administrative data as well as diary data.

WHI limitations:

Does not quantify lost productivity for unpaid work.

Does not consider compensation mechanisms or work-team dynamics.

Limited psychometric evidence to date.

2.10 Work Limitations Questionnaire

The WLQ is a 25-item scale developed by researchers from the New England Medical Center to assess health-related limitations while performing specific job demands [67]. Its original content and format were generated from focus groups and cognitive interviews with workers with various chronic health conditions. The WLQ uses a 2-week recall period and is organized into four domains: *time management* (TM, five items), which assesses difficulties handling a job's time and scheduling demands; *physical demands* (PD, six items), which examines ability to perform job tasks that involve bodily strength, movement, endurance, coordination, and flexibility; *mental-interpersonal* (MI, nine items), which assesses problems with cognitively-demanding tasks and social interactions at work; and *output demands* (OD, five items), which examines problems meeting productivity output [67]. Item response options range from 'none of the time' to 'all of the time'. The WLQ provides four subscale scores, which is calculated by taking the mean of items within each subscale and then rescaling each mean to a 0–100 score (100 = most limitations). Alternatively, an Index score can also be generated by computing a weighted sum of the four subscale scores based on a conversion formula [68]. This formula is derived from a study of the relationship between WLQ subscale scores and objectively-measured productivity among employees from the customer service department of a large firm (assessed number of phone calls answered per payroll hour and the number of merchandise units processed per hour) [69]. The WLQ Index score is intended to represent percent productivity loss relative to healthy employees, which is a metric compatible for estimating productivity costs (due to presenteeism).

The WLQ is one of the most extensively tested work productivity instruments to date. Its internal consistency has been demonstrated in several studies on workers with arthritis. Across the four WLQ subscales, Lerner et al. [70] reported a Cronbach's alpha range of 0.93–0.97, Walker et al. [71] reported a Cronbach's alpha range of 0.83–0.88, and Beaton et al. [72] reported a Cronbach's alpha range of 0.77–0.94—all within acceptable range. In terms of construct validity, a number of studies have found low-to-moderate correlations or agreement between the WLQ and other presenteeism measures [72–74]. A recent and notable study featuring a head-to-head comparison of four instruments among workers with arthritis found that the productivity cost estimate derived from the WLQ Index only weakly agreed with the WPAI (ICC = 0.30), HPQ (ICC = 0.26) and HLQ (ICC = 0.22) [57].

WLQ strengths:

Index score formula (i.e. productivity metric) is derived from empirical relationship between WLQ summed scores and objectively-measured productivity data.

One of the most extensively tested work productivity instruments to date.

WLQ limitations:

Assesses only presenteeism.

Generalizability of the Index score formula is unclear since this was derived from workers from one particular work setting.

2.11 Work Productivity and Activity Impairment Questionnaire

The WPAI is designed to quantify the effects of health and symptoms on work productivity, as well as impairments experienced during regular (unpaid) activities, using a 7-day recall period [75]. The WPAI is available in a general health version (WPAI:GH) and can also be adapted for specific health conditions by making reference to the condition in the survey questions. As such, generic and disease-specific versions of the WPAI are actually largely identical. Detailed information on different versions of the WPAI can be found at http://www.reillyassociates.net. WPAI items were originally generated from three sources [75]: (1) review of the work productivity literature; (2) feedback from patients with allergic rhinitis; and (3) cognitive debriefing with respondents to determine final wording. The WPAI consists of six questions, assessing current employment status (Q1), number of hours missed due to health problems (Q2), number of hours missed due to other (i.e. non-health-related) reasons (Q3), hours actually worked (Q4), degree to which health affected productivity while working (Q5), and degree to which health affected regular (unpaid) activities (Q6). The last two items use a numeric rating scale (health problems had no effect [score = 0] − health problems completely prevented me from working/doing my daily activities [score = 10]). Four productivity metrics can be derived from the WPAI: (1) *absenteeism*: percent work time missed due to health, Q2/(Q2 + Q4); (2) *presenteeism*: percent impairment while working due to health, Q5/10; (3) *overall work productivity*: percent overall work impairment due to health, Q2/(Q2 + Q4) + [(1 − Q2/(Q2 + Q4)) × (Q5/10)]; and (4) *activity impairment*: percent activity impairment due to health, Q6/10. For all WPAI metrics, a higher percentage (range 0–100 %) indicates greater impact of health on productivity/activity impairment.

Like the WLQ, this is another instrument that has been tested quite extensively over the years. In terms of reliability, moderate-to-high agreements (ICC = 0.7–1.0) were shown between test–retest administrations of the WPAI (irritable bowel syndrome version) [76], and also between self- and telephone-administration (ICC = 0.5–0.9) of the WPAI (RA version) [77]. Construct validity of the various WPAI productivity metrics has been evaluated quite often [51, 57, 76, 78–87], including two studies where multiple presenteeism measures were directly compared. Among workers with RA or osteoarthritis, WPAI presenteeism moderately agreed with the HPQ (ICC = 0.61), but less so with the WLQ Index (ICC = 0.30) and the HLQ (ICC = 0.37) [57]. In another study of workers with RA, WPAI presenteeism showed moderate correlations with the QQ method ($r = 0.61$), as well as HLQ's direct hour-estimating method ($r = 0.48$) [51].

WPAI strengths:

Assesses lost productivity for both paid and unpaid work (i.e. activity impairment).

One of the most extensively tested work productivity instruments to date.

Very brief questionnaire, requires short completion time.

WPAI limitations:

Leisure activities are not distinguished from unpaid work in the assessment of 'activity impairment' (i.e. both would be similarly valued).

Does not consider compensation mechanisms or work-team dynamics.

2.12 Work Productivity Short Inventory

The WPSI is designed to quantify the impact of common health conditions on employee productivity, and three versions with varied recall period (12 months, 3 months or 2 weeks) were initially introduced [88]. The WPSI asks respondents whether they have experienced any of 15 common health conditions over the recall period (yes or no). These pre-defined health conditions were chosen through informal consultations with employers and physicians, in addition to a literature review focused on disease prevalence and associated costs to workplaces [88]. Eleven of these conditions pertain directly to the employees themselves, which include allergies, respiratory infections, arthritis, asthma, anxiety disorder, depression and bipolar disorder, stress, diabetes, hypertension, migraine/major headaches, and coronary heart disease/high cholesterol. The other four conditions pertain to caregiving provided by employees to their spouses, dependents or elders, including Alzheimer's disease, allergies, otitis media (ear ache) or respiratory infections. For every health condition that is relevant to the respondent, the WPSI assesses (1) the number of days where the condition and related symptoms were experienced; (2) the number of unproductive hours during a typical 8-h workday due to the condition and related symptoms (*presenteeism*); and (3) the overall number of absent workdays due to the condition and related symptoms (*absenteeism*). To derive productivity cost metrics, WPSI developers had specifically recommended using a cost multiplier of $34.25/h to represent compensation (salary and benefits) for company employees [13]. Given the setup of this instrument, up to 45 productivity/cost metrics can actually be generated (i.e. absenteeism, presenteeism, and total productivity loss for each of

15 specific health conditions), although such cases are probably rare.

Only few studies have been conducted to examine the psychometric properties of the WPSI to date [74, 88]. Goetzel et al. [88] applied a split-sample technique to assess and compare the reliability between three versions of the WPSI with varied recall period (12 months, 3 months and 2 weeks), and concluded that both the 12- and 3-month versions were slightly favored over the 2-week version. When tested in a sample of workers at a large telecommunications firm, WPSI presenteeism was shown to weakly correlate with metrics from the WLQ ($r = 0.23$–0.33 vs. WLQ subscale scores; $r = 0.30$ vs. Index score) [74, 88].

WPSI strengths:

Considers absenteeism and presenteeism.

Provides opportunity to assess productivity loss attributed to different health conditions.

A brief questionnaire if only few conditions are relevant for the respondent (otherwise can be lengthy).

WPSI limitations:

Does not consider lost productivity for unpaid work.

Uncertain whether respondents are able to disentangle productivity loss associated with different health conditions; logic errors in response from previous testing suggest some possibility of double counting [74, 88].

Does not consider compensation mechanisms or team-work dynamics.

Very limited psychometric evidence to date.

3 Discussion

As can be seen in Table 1, the 11 instruments reviewed differ in many important ways, including breadth of content, recall period, approach to quantify the various sources of productivity loss, and compatibility for different valuation approaches (e.g. FC approach, valuation multipliers). Accordingly, they also varied in terms of length as the instruments ranged from having very few items, such as the QQ and WPAI, to much lengthier, modular instruments, such as the PRODISQ and VOLP. The earliest of these instruments were initially introduced in the 1990s (i.e. WPAI, HLQ, QQ), while the newest of them, the VOLP, first emerged in the early 2010s. It should be mentioned that a new and promising instrument, the iMTA Productivity Cost Questionnaire (iPCQ), has been recently developed, with validation testing currently underway [8]. Overall, three main differences between instruments were perhaps most notable: (1) varied content comprehensiveness; (2) approach to measure presenteeism; and (3) extent of available psychometric evidence to date.

3.1 Content Comprehensiveness: Implications for Applicability in Health Economic Evaluations

Content comprehensiveness has important implications for the applicability of a work productivity instrument. If important content is missing, then the resulting productivity cost estimates is unlikely to be accurate. Health economic evaluations can be carried out from different perspectives, and the perspective taken has ramifications in terms of the suitability of an instrument [6, 89]. Increasingly, health economists and national guidelines are advocating a societal perspective of costing in health economic evaluation [2, 3, 23] and, in such cases, considerations of lost productivity for unpaid work becomes imperative. Therefore, instruments without this component may be inadequate for such a purpose, unless they are used in conjunction with an unpaid work section gathered from another instrument. On the other hand, these costs may be rightly excluded if cost effectiveness from an employer's perspective is sought (e.g. workplace interventions being considered as investments toward employee health) [6, 89]. In such cases, only productivity costs associated with paid work need to be considered. Among the 11 instruments reviewed, it was notable that only four considered lost productivity during unpaid work (HLQ, HRPQ-D, VOLP and WPAI). The forthcoming iPCQ is also considered sufficiently comprehensive for costing from a societal perspective [8].

Only the PRODISQ and VOLP were revealed to contain questions on compensation mechanisms and work-team dynamics. As such, these instruments offer a unique opportunity to consider these influences when estimating productivity costs. If these influences are not considered (i.e. assumed to be negligible), then potential risks for over-estimating (if compensation mechanisms are ignored) or under-estimating productivity costs (if work-team dynamics are ignored) should be recognized. In reality, it is possible that some of these influences may in fact cancel each other out, although the interaction between these influences is not well-understood [7]. Clearly, this is an area in need of further research and greater understanding. Currently, some experts are recommending that corrections for compensation mechanisms and work-team dynamics should only be conducted on the basis of a sensitivity analysis [8].

3.2 Different Approaches to Measure Presenteeism: Which is Best?

Since presenteeism can represent the most significant source of indirect health costs, accurate quantification is especially important as any measurement error can magnify differences in productivity cost estimates. In the recent literature, differences in the approach to conceptualize and measure presenteeism have garnered much discussion

Table 1 Overview of work productivity instruments reviewed

Instrument	Content	Recall period	Applicable population	Productivity domains assessed				
				Paid work				Unpaid Work
				Absenteeism	Presenteeism	Compensation mechanisms	Work-team dynamics	
HLQ	Four modules: (1) absence from paid work; (2) reduced productivity at paid work; (3) unpaid labor production; (4) impediments to paid and unpaid labor	2 weeks	Any	X	X			X
HPQ	Assesses health problems, ongoing treatment, work performance (absenteeism and presenteeism), demographics and work information	Multiple versions: 7 days or 4 weeks	Any	X	X			
HRPQ-D	Nine questions assessing (1) number of scheduled work hours; (2) number of work hours missed; (3) effectiveness on hours worked, for each of three productivity venues (paid work, housework, educational activities)	Current day (i.e. daily diary)	Any	X	X			X
PRODISQ[a]	Seven modules: (1) general information; (2) working situation and income; (3) absenteeism; (4) compensating mechanisms; (5) productivity costs at work; (6) productivity costs at the organizational level; (7) administrative and management costs	3 months for absenteeism; most recent day for presenteeism	Any	X	X	X	X	
QQ	Two questions assessing quantity and quality of work (visual analog scales)	Most recent workday	Any		X			
SPS-13	13 questions, including a 10-item Likert scale for presenteeism	4 weeks	Any	X	X			
VOLP	Six modules: (1) employment status; (2) job characteristics; (3) absenteeism; (4) work performance (presenteeism); (5) unpaid work; (6) dynamics of the work environment	3 months for absenteeism; 7 days for presenteeism	Any	X	X	X	X	X
WHI	Eight modules: (1) employment status; (2) usual work time; (3) presence of health conditions; (4) job visualization; (5) absenteeism; (6) presenteeism; (7) demographics; (8) salary information	2 weeks	22 health conditions[b]	X	X			
WLQ	25 Likert-type items organized into four domains: time management, physical demands, mental-interpersonal, output demands	2 weeks	Any		X			

Table 1 continued

Instrument	Content	Recall period	Applicable population	Productivity domains assessed				
				Paid work				Unpaid Work
				Absenteeism	Presenteeism	Compensation mechanisms	Work-team dynamics	
WPAI	Six questions assessing (1) current employment status; (2) work hours missed due to health; (3) work hours missed due to other reasons; (4) hours worked; (5) productivity while working; (6) regular (unpaid) activities	1 week	Any	X	X			X
WPSI	For each of up to 15 health conditions, assesses (1) number of days with health problems; (2) presenteeism on affected days; (3) absenteeism	Multiple versions: 2 weeks, 3 months or 12 months	15 health conditions[c]	X	X			

HLQ Health and Labor Questionnaire, *HPQ* Health and Work Productivity Questionnaire, *HRPQ-D* Health-Related Productivity Questionnaire Diary, *PRODISQ* Productivity and Disease Questionnaire, *QQ* Quantity and Quality method, *SPS-13* Stanford Presenteeism Scale, *VOLP* Valuation of Lost Productivity Questionnaire, *WHI* Work and Health Interview, *WLQ* Work Limitations Questionnaire, *WPAI* Work Productivity and Activity Impairment, *WPSI* Work Productivity Short Inventory

[a] The PRODISQ is a modular instrument and recommends QQ as its presenteeism component

[b] Includes a variety of acute conditions (e.g. common cold, influenza) and chronic conditions (e.g. diabetes, heart disease, gastrointestinal problems, depression)

[c] Includes 11 employee-specific conditions and four conditions pertaining to the care of spouses, elders or dependents (see text for details)

[90, 91]. This diversity is also readily apparent among the instruments reviewed. In this regard, perhaps two main approaches can be distinguished. Some take a 'direct' approach of asking respondents to estimate how much time it would require to make up lost production (e.g. HLQ, VOLP). In contrast, others take an 'indirect' approach that relies on transforming an initial rating/score of work difficulties into a metric compatible for valuation (i.e. equivalent lost time). Different concepts are represented by this initial rating/score; for example, work efficiency loss (i.e. QQ), work impairment (i.e. WPAI), work performance (i.e. HPQ), or work limitations (i.e. WLQ). Intuitively, although the 'indirect' approach appears less desirable (given the need for an additional conversion step), the preferred approach remains unclear. Yet, it has become quite apparent that different instruments are providing highly varied estimates of presenteeism [46, 48, 51, 52, 57, 74]. This lack of comparability is concerning, and raises other important questions. For example, when two productivity instruments do not correlate or show adequate agreement, which is 'right', and which is 'wrong'? Or, might both be 'wrong'? These are unresolved issues on the measurement of presenteeism that deserve continued research attention.

3.3 Psychometric Evidence: A Critique of Methodological Quality

The work productivity instruments reviewed also differ in the extent of psychometric testing received to date. As a whole, cross-sectional tests of validity of instruments were most frequently conducted, whereas test–retest reliability and longitudinal tests of validity have received relatively limited attention thus far. Among the 11 instruments reviewed, the WPAI and WLQ have been most frequently evaluated to date, which echoes the previous review on this topic from *Pharmacoeconomics* [4]. However, for two obvious reasons, this is not necessarily indicative of an instrument's merit. First, the quantity of testing is, in part, a function of time since an instrument was initially introduced, that is, newer measures tend to be less extensively tested since they are not yet well-known. Second, a critical factor in the strength of evidence is the methodological quality of available psychometric testing.

Among the available psychometric evidence (Table 2), only some were considered to be of high methodological quality. Tests of internal consistency—which pertained only to the SPS-13 and WLQ—were generally conducted with high methodological quality. In contrast, the methodological quality associated with assessments of test–retest reliability and validity might be considered less impressive overall. For test–retest reliability, some studies had fairly small sample sizes (i.e. $n < 50$), while for others,

Table 2 Evidence on the reliability and validity of work productivity instruments

Instrument	Reliability		Validity[a]	
	Internal consistency	Test–retest reliability[b]	Cross-sectional testing	Longitudinal testing[c]
HLQ	Not relevant	Unavailable	General population [46]; hip problems [46]; industrial and construction workers [52]**; knee problems [46]; migraine [46]; RA [51]; RA and OA [57]; spinal cord injury [46]; trade-firm workers [48]**	Unavailable
HPQ	Not relevant	Airline reservation agents [55]*	Various occupations [54, 55]; workers in education and health sectors [92]; RA and OA [57]	Various occupations [55]
HRPQ-D	Not relevant	Unavailable	Infectious mononucleosis [58]	Infectious mononucleosis [58]
PRODISQ	Not relevant	Unavailable	Various chronic health conditions [18]**; see also evidence for QQ[d]	Unavailable
QQ	Not relevant	Unavailable	Industrial and construction workers [52]**; RA [51]; trade-firm workers [48]**	Unavailable
SPS-13	Workers in a research and manufacturing firm [62]*	Unavailable	Workers in a research and manufacturing firm [62]	Unavailable
VOLP	Not relevant	RA [63]*	RA [63]	Unavailable
WHI	Not relevant	Unavailable	Inbound phone call agents [66]; population survey of US workers [65]	Unavailable
WLQ	Anxiety disorders [80]*; OA [70]*; RA [71]*; RA and OA [72]*; various chronic health conditions [67]*	Unavailable	Anxiety disorders [80]; depression and anxiety [93]; OA [70]**; RA [71, 94]; RA and OA [72]**[57]; various chronic health conditions [67]; workers in a large telecommunications firm [74]**; WRUED [73]**[95]	Anxiety disorders [80]; depression and anxiety [93]; RA and OA [72]**; WRUED [73]**
WPAI[e]	Not relevant	AS and PsA [77]; hand dermatitis [85]; IBS [76]; non-specific health problems [75]	AS [87]; asthma [79]; anxiety disorders [80]; caregivers of elderly patients [81]; Crohn's disease [82]; GERD [83, 84]; hand dermatitis [85]; IBS [76, 86]; RA [51, 78]; RA and OA [57]	AS [87]; anxiety disorders [80]; Crohn's disease [82]; hand dermatitis [85]; GERD [84, 96]
WPSI	Not relevant	Workers in a manufacturing and communications firm [88]	Workers in a large telecommunications firm [74]**; workers in a manufacturing firm [97]	Unavailable

HLQ Health and Labor Questionnaire, *HPQ* Health and Work Productivity Questionnaire, *HRPQ-D* Health-Related Productivity Questionnaire Diary, *PRODISQ* Productivity and Disease Questionnaire, *QQ* Quantity and Quality method, *SPS-13* Stanford Presenteeism Scale, *VOLP* Valuation of Lost Productivity Questionnaire, *WHI* Work and Health Interview, *WLQ* Work Limitations Questionnaire, *WPAI* Work Productivity and Activity Impairment, *WPSI* Work Productivity Short Inventory, *AS* ankylosing spondylitis, *GERD* gastroesophageal reflux disease, *OA* osteoarthritis, *RA* rheumatoid arthritis, *PsA* psoriatic arthritis, *WRUED* work-related upper-extremity disorders, *unavailable* evidence on the property has not been published to date, to the best of the author's knowledge

* Indicates reliability testing of high methodological quality (e.g. adequate sample size (i.e. $n > 50$), use of appropriate method (e.g. test of agreement), ** Indicates validity testing of high methodological quality (e.g. instrument compared with a meaningful comparator, specific a priori hypotheses provided)

[a] Includes construct, criterion, convergent, discriminant, and factorial validity

[b] Includes interrater agreement

[c] Longitudinal construct validity is sometimes considered a form of responsiveness

[d] PRODISQ is a modular instrument and recommends using QQ as its measure of presenteeism

[e] Evidence for either the generic or disease-specific versions of the WPAI is included

tests of agreement were not applied (i.e. tests of correlation are inadequate). For validity testing, a considerable number of studies exhibit one (or both) of the following limitations: (1) the lack of a meaningful comparator; and/or (2) specific a priori hypotheses were not formulated about the expected relationship between the instrument and its comparators.

Table 3 Appraising a work productivity instrument for use in health economic evaluations: the five 'Ps'

Purpose
Is the instrument compatible for estimating productivity costs (i.e. derive metrics that can be assigned a monetary value)?
Is the instrument compatible with the intended approach of valuation (e.g. human capital approach vs. frictional cost approach)?
Does the instrument assess for compensation mechanisms and/or work-team dynamics? If not, is this likely to introduce significant bias to the productivity cost estimates?
Perspective
What is the intended perspective of evaluation (e.g. worker, employer, or societal perspective)?
Given the intended perspective of evaluation, does the instrument capture all relevant sources of productivity costs (i.e. absenteeism, presenteeism, unpaid work)? If not, can it be combined with component(s) from another instrument?
Population
Is the instrument appropriate for the target population?
Is a disease-specific version of the instrument available? Would applying this version offer any advantages over its generic counterpart? (For review of available disease-specific instruments, see Prasad et al. [4] and Lofland et al. [5])
Psychometrics
Is there evidence that support the instrument's reliability and validity (for each relevant productivity component)?
Is the evidence gathered from studies of high methodological quality?
Do these studies involve a population that is reasonably similar to the intended population?
If supporting evidence does not currently exist for the instrument, would it be worthwhile to conduct an initial psychometric study?
Practicality
Is the respondent and administration burden acceptable (e.g. completion time, administration frequency, complexity of scoring)?
Is it possible to collect the same information with a shorter instrument (or version)?
Are respondents likely to be able to answer all questions from the instrument (consider level of education, complexity and applicability of questions)
Are there concerns regarding the language of the instrument (e.g. availability of translated version)?
Are there concerns regarding propriety issues and cost?

For these reasons, strong conclusions on the validity of many instruments cannot be drawn at present. Since validity is concerned with how 'well' an instrument captures the intended construct, to provide the strongest evidence, a favorable comparison with a *more established* indicator of productivity loss/cost is required. Objective data (i.e. actual production output) or workplace administrative data (e.g. payroll/attendance records) are generally most ideal. No doubt this can be a challenging proposition and, in this regard, several studies should be commended for their choice of comparators [52, 54, 66, 69]. If this is unavailable, other subjectively-measured productivity instrument(s) [e.g. self- or employer-reported] should be the next option (albeit with some limitations as discussed above). Arguably, from the standpoint of construct validation, very little is gained if one is only able to demonstrate that an instrument differs from a dissimilar measure (i.e. unrelated to productivity), since this offers little 'proof' that the instrument is measuring productivity well.

Specifying a priori hypotheses on the expected relationship between an instrument and its comparators is another potential opportunity to augment the quality of validation tests. This is important because it provides a clear and transparent basis for affirming the psychometric property in question. In addition, more specific hypotheses (along with strong rationales) are often desirable. For example, an hypothesis of a narrow range of correlation (say, a range of 0.25) with a comparator is preferable to a hypothesis based on statistical significance (e.g. probability that the correlation differs from zero). This is because if the sample size is sufficiently large, only the weakest of correlations would be non-statistically significant. Thus, if 'high' correlation is expected but only a 'low' (but also statistically significant) correlation is found, this should really be viewed as a case against an instrument's validity. In the literature, another common approach to validate an instrument is to compare productivity levels between 'known-groups' (e.g. between ill and healthy workers). While this is, overall, a sensible approach, such evidence can be potentially strengthened by providing specific hypotheses on the anticipated between-group difference in productivity (e.g. based on previous literature or expert opinion, etc.). Again, a statistically significant between-group difference (in the expected direction) may not necessarily imply a valid productivity instrument. In the case where only a 'small' difference is expected, the finding of a 'large' difference would actually suggest a subpar (i.e. imprecise) instrument.

4 Conclusions

Currently, there is no consensus on the best instrument in the field, and the abundance of tools is, in some sense, almost a 'mixed blessing'. On one hand, a variety of available options is a welcomed situation; however, making a good choice requires prudence about the strengths and limitations of the various instruments, as well as inherent assumptions (e.g. how is unpaid work defined?). A definitive recommendation may not be helpful since the ideal choice of instrument for each application can vary depending on the intended objective(s) and, in some cases, difficult decisions may be involved given competing considerations (e.g. comprehensiveness vs. length). In any case, for any given choice of instrument, it is important to be mindful of the potential direction of estimation bias (i.e. likelihood of over- or under-estimating costs) and also what type of sensitivity analysis might be most informative (if at all). To offer some general guidance on choosing an instrument for health economic evaluations, a set of reflective questions over five broad areas is proposed, summarizing key issues that users should consider (Table 3). These areas are, in no particular order: *purpose*, *perspective*, *practicality, population* and *psychometrics*—the five 'Ps'. In conclusion, it is hoped that the current review has been useful for (1) fostering an increased awareness of the strengths and limitations of available work productivity instruments; (2) encouraging more high-quality psychometric testing of work productivity instruments; and also (3) providing some general guidance (or at least a starting point) for users engaged in an instrument selection process for health economic evaluations.

Acknowledgments Kenneth Tang is recipient of a Canadian Institutes of Health Research Fellowship.

Conflict of interest None to declare.

References

1. Brouwer WB, van Exel NJ, Baltussen RM, Rutten FF. A dollar is a dollar is a dollar—or is it? Value Health. 2006;9(5):341–7.
2. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evalation of health care programmes. Oxford: Oxford University Press; 2005.
3. Gold M, Siegel J, Russell L, Weinstein M. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.
4. 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):225–44.
5. Lofland JH, Pizzi L, Frick KD. A review of health-related workplace productivity loss instruments. Pharmacoeconomics. 2004;22(3):165–84.
6. Zhang W, Bansback N, Anis AH. Measuring and valuing productivity loss due to poor health: a critical review. Soc Sci Med. 2011;72(2):185–92.
7. Krol M, Brouwer W, Rutten F. Productivity costs in economic evaluations: past, present, future. Pharmacoeconomics. 2013;31(7):537–49.
8. Krol M, Brouwer W. How to estimate productivity costs in economic evaluations. Pharmacoeconomics. 2014;32(4):335–44.
9. Koopmanschap MA, Rutten FF. Indirect costs in economic studies: confronting the confusion. Pharmacoeconomics. 1993;4(6):446–54.
10. Anis A, Zhang W, Emery P, Sun H, Singh A, Freundlich B, et al. The effect of etanercept on work productivity in patients with early active rheumatoid arthritis: results from the COMET study. Rheumatology. 2009;48(10):1283–9.
11. Escorpizo R, Bombardier C, Boonen A, Hazes JM, Lacaille D, Strand V, et al. Worker productivity outcome measures in arthritis. J Rheumatol. 2007;34(6):1372–80.
12. 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):3135–44.
13. Goetzel RZ, Guindon AM, Turshen IJ, Ozminkowski RJ. Health and productivity management: establishing key performance measures, benchmarks, and best practices. J Occup Environ Med. 2001;43(1):10–7.
14. 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):1234–46.
15. Lamb CE, Ratner PH, Johnson CE, Ambegaonkar AJ, Joshi AV, Day D, et al. Economic impact of workplace productivity losses due to allergic rhinitis compared with select medical conditions in the United States from an employer perspective. Curr Med Res Opin. 2006;22(6):1203–10.
16. 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(4):304–10.
17. 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):769–77.
18. Jacob-Tacken KH, Koopmanschap MA, Meerding WJ, Severens JL. Correcting for compensating mechanisms related to productivity costs in economic evaluations of health care programmes. Health Econ. 2005;14(5):435–43.
19. Severens JL, Laheij RJ, Jansen JB, Van der Lisdonk EH, Verbeek AL. Estimating the cost of lost productivity in dyspepsia. Aliment Pharmacol Ther. 1998;12(9):919–23.
20. Pauly MV, Nicholson S, Xu J, Polsky D, Danzon PM, Murray JF, et al. A general model of the impact of absenteeism on employers and employees. Health Econ. 2002;11(3):221–31.
21. Nicholson S, Pauly MV, Polsky D, Sharda C, Szrek H, Berger ML. Measuring the effects of work loss on productivity with team production. Health Econ. 2006;15(2):111–23.
22. 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):469–85.
23. Jonsson B. Ten arguments for a societal perspective in the economic evaluation of medical innovations. Eur J Health Econ. 2009;10(4):357–9.
24. Johannesson M. The willingness to pay for health changes, the human-capital approach and the external costs. Health Policy. 1996;36(3):231–44.
25. Reid M. Economics of household production. New York, NY: Wiley; 1934.

26. 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.
27. Koopmanschap MA, Rutten FF, van Ineveld BM, van Roijen L. The friction cost method for measuring indirect costs of disease. J Health Econ. 1995;14(2):171–89.
28. Johannesson M, Karlsson G. The friction cost method: a comment. J Health Econ. 1997;16(2):249–55; discussion 57–9.
29. Johannesson M. Avoiding double-counting in pharmacoeconomic studies. Pharmacoeconomics. 1997;11(5):385–8.
30. Brouwer WB, Koopmanschap MA, Rutten FF. Productivity costs in cost-effectiveness analysis: numerator or denominator: a further discussion. Health Econ. 1997;6(5):511–4.
31. Weisbrod BA. The valuation of human capital. J Polit Econ. 1961;69(5):425–36.
32. Drummond M. Cost-of-illness studies: a major headache? Pharmacoeconomics. 1992;2(1):1–4.
33. Koopmanschap MA, van Ineveld BM. Towards a new approach for estimating indirect costs of disease. Soc Sci Med. 1992;34(9):1005–10.
34. van den Berg B, Brouwer W, van Exel J, Koopmanschap M, van den Bos GA, Rutten F. Economic valuation of informal care: lessons from the application of the opportunity costs and proxy good methods. Soc Sci Med. 2006;62(4):835–45.
35. Fleiss JL, Cohen J, Everitt BS. Large-sample standard errors of kappa and weighted kappa. Psychol Bull. 1969;72:323–7.
36. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.
37. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. New York, NY: Oxford University Press; 1995.
38. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42.
39. Nunnally J, Bernstein I. Psychometric theory. 3rd ed. New York: McGraw-Hill; 1994.
40. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–333.
41. Kuder GF, Richardson MW. The theory of the estimation of test reliability. Psychometrika. 1937;2(3):151–60.
42. Messick S. Test validity: a matter of consequence. Soc Indic Res. 1998;45(1–3):35–44.
43. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull. 1955;52(4):281–302.
44. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res. 2003;12(4):349–62.
45. 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(4):539–49.
46. 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):405–15.
47. Hakkaart-van Roijen L, Essink-Bot ML. Manual: The Health and Labour Questionnaire. Institute for Medical Technology Assessment, Erasmus University Rotterdam; 2000. Report No.: 52. http://repub.eur.nl/pub/1313/.
48. Brouwer WB, Koopmanschap MA, Rutten FF. Productivity losses without absence: measurement validation and empirical evidence. Health Policy. 1999;48(1):13–27.
49. Koopmanschap MA. PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies. Expert Rev Pharmacoecon Outcomes Res. 2005;5(1):23–8.
50. Hakkaart-van Roijen L. Short Form – Health and Labour Questionnaire. Erasmus University Rotterdam: Institute for Medical Technology Assessment; 2010.
51. 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):354–61.
52. 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):517–23.
53. Osterhaus JT, Gutterman DL, Plachetka JR. Healthcare resource and lost labour costs of migraine headache in the US. Pharmacoeconomics. 1992;2(1):67–76.
54. 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):156–74.
55. Kessler RC, Ames M, Hymel PA, Loeppke R, McKenas DK, Richling DE, et al. Using the World Health Organization Health and Work Performance Questionnaire (HPQ) to evaluate the indirect workplace costs of illness. J Occup Environ Med. 2004;46(6 Suppl):S23–37.
56. Kessler RC, Petukhova M, Ustun TB. HPQ Short Form (absenteeism and presenteeism questions and scoring rules). 2007.
57. Zhang W, Gignac MA, Beaton D, Tang K, Anis AH. Productivity loss due to presenteeism among patients with arthritis: estimates from 4 instruments. J Rheumatol. 2010;37(9):1805–14.
58. 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.
59. Collins JJ, Baase CM, Sharda CE, Ozminkowski RJ, Nicholson S, Billotti GM, et al. The assessment of chronic health conditions on work performance, absence, and total economic impact for employers. J Occup Environ Med. 2005;47(6):547–57.
60. 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.
61. Lynch W, Riedel J. Measuring employee productivity: a guide to self-assessment tools. Scottsdale, AZ: Institute for Health & Productivity Management & William Mercer; 2001.
62. 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):1123–33.
63. 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;52(5):530–6.
64. Stewart WF, Ricci JA, Chee E, Morganstein D, Lipton R. Lost productive time and cost due to common pain conditions in the US workforce. JAMA. 2003;290(18):2443–54.
65. Stewart WF, Ricci JA, Leotta C. Health-related lost productive time (LPT): recall interval and bias in LPT estimates. J Occup Environ Med. 2004;46(6 Suppl):S12–22.
66. Stewart WF, Ricci JA, Leotta C, Chee E. Validation of the work and health interview. Pharmacoeconomics. 2004;22(17):1127–40.
67. Lerner D, Amick III BC, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. Med Care. 2001;39(1):72–85.
68. Lerner D, Rogers WH, Chang H. Technical report: scoring the work limitations questionnaire (WLQ) and the WLQ index for estimating work productivity loss. 2003.

69. Lerner D, Amick III BC, Lee JC, Rooney T, Rogers WH, Chang H, et al. Relationship of employee-reported work limitations to work productivity. Med Care. 2003;41(5):649–59.
70. Lerner D, Reed JI, Massarotti E, Wester LM, Burke TA. The Work Limitations Questionnaire's validity and reliability among patients with osteoarthritis. J Clin Epidemiol. 2002;55(2): 197–208.
71. Walker N, Michaud K, Wolfe F. Work limitations among working persons with rheumatoid arthritis: results, reliability, and validity of the work limitations questionnaire in 836 patients. J Rheumatol. 2005;32(6):1006–12.
72. 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.
73. Roy JS, MacDermid JC, Amick BC III, Shannon HS, McMurtry R, Roth JH, et al. Validity and responsiveness of presenteeism scales in chronic work-related upper-extremity disorders. Phys Ther. 2011;91(2):254–66.
74. 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):635–48.
75. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics. 1993;4(5):353–65.
76. Bushnell DM, Reilly MC, Galani C, Martin ML, Ricci JF, Patrick DL, et al. Validation of electronic data capture of the Irritable Bowel Syndrome-Quality of Life Measure, the Work Productivity and Activity Impairment Questionnaire for Irritable Bowel Syndrome and the EuroQol. Value Health. 2006;9(2):98–105.
77. Ariza-Ariza R, Hernandez-Cruz B, Navarro-Compan V, Leyva Pardo C, Juanola X, Navarro-Sarabia F. A comparison of telephone and paper self-completed questionnaires of main patient-related outcome measures in patients with ankylosing spondylitis and psoriatic arthritis. Rheumatol Int. 2013;33(11):2731–6.
78. 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.
79. Chen H, Blanc PD, Hayden ML, Bleecker ER, Chawla A, Lee JH, et al. Assessing productivity loss and activity impairment in severe or difficult-to-treat asthma. Value Health. 2008;11(2): 231–9.
80. 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): 1165–71.
81. 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):1011–7.
82. 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.
83. 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):106–13.
84. 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): 385–96.
85. Reilly MC, Lavin PT, Kahler KH, Pariser DM. Validation of the Dermatology Life Quality Index and the Work Productivity and Activity Impairment-Chronic Hand Dermatitis questionnaire in chronic hand dermatitis. J Am Acad Dermatol. 2003;48(1): 128–30.
86. 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 Ther. 2004;20(4):459–67.
87. 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):812–9.
88. Goetzel RZ, Hawkins K, Ozminkowski RJ, Wang S. The health and productivity cost burden of the "top 10" physical and mental health conditions affecting six large U.S. employers in 1999. J Occup Environ Med. 2003;45(1):5–14.
89. Berger ML, Murray JF, Xu J, Pauly M. Alternative valuations of work loss and productivity. J Occup Environ Med. 2001;43(1): 18–24.
90. Schultz AB, Edington DW. Employee health and presenteeism: a systematic review. J Occup Rehabil. 2007;17(3):547–79.
91. Brooks A, Hagen SE, Sathyanarayanan S, Schultz AB, Edington DW. Presenteeism: critical issues. J Occup Environ Med. 2010;52(11):1055–67.
92. Scuffham PA, Vecchio N, Whiteford HA. Exploring the validity of HPQ-based presenteeism measures to estimate productivity losses in the health and education sectors. Med Decis Making. 2014;34(1):127–37.
93. 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.
94. Wolfe F, Michaud K, Choi HK, Williams R. Household income and earnings losses among 6,396 persons with rheumatoid arthritis. J Rheumatol. 2005;32(10):1875–83.
95. Tang K, Beaton DE, Amick BC 3rd, Hogg-Johnson S, Cote P, Loisel P. Confirmatory factor analysis of the Work Limitations Questionnaire (WLQ-25) in workers' compensation claimants with chronic upper-limb disorders. J Occup Rehabil. 2013; 23(2):228–38.
96. 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):102–10.
97. 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):1183–95.