Systematic Literature Review

Measuring, Analyzing, and Presenting Work Productivity Loss in Randomized Controlled Trials: A Scoping Review

Wei Zhang, PhD, Paige Tocher, MPH, Jacynthe L’Heureux, MSc, Julie Sou, MSc, Huiying Sun, PhD

ABSTRACT

Objectives: This study aimed to conduct a scoping review of randomized controlled trials (RCTs) and investigate which work productivity loss outcomes were measured in these RCTs, how each outcome was measured and analyzed, and how the results for each outcome were presented.

Methods: A systematic search was conducted from January 2010 to April 2020 from 2 databases: PubMed and Cochrane Central Register of Controlled Trials. Data on country, study population, disease focus, sample size, work productivity loss outcomes measured (absenteeism, presenteeism, employment status changes), and methods used to measure, report, and analyze each work productivity loss outcome were extracted and analyzed.

Results: We found 435 studies measuring absenteeism or presenteeism, of which 155 studies (35.6%) measured both absenteeism and presenteeism and were included in our final review. Only 9 studies also measured employment status changes. The most used questionnaire was the Work Productivity and Activity Impairment Questionnaire. The analysis of absenteeism and presenteeism data was mostly done using regression models (n = 98, n = 98, respectively) for which a normal distribution was assumed (n = 77, n = 89, respectively). Absenteeism results were most often presented in time whereas presenteeism was commonly presented using a percent scale or score.

Conclusions: There is a lack of consensus on how to measure, analyze, and present work productivity loss outcomes in RCTs published in the past 10 years. The diversity of measurement, analysis, and presentation methods used in RCTs may make comparability challenging. There is a need for guidelines providing recommendations to standardize the comprehensiveness and the appropriateness of methods used to measure, analyze, and report work productivity loss in RCTs.

Keywords: absenteeism, presenteeism, randomized controlled trial, scoping review, work productivity loss.

Introduction

Health problems, such as chronic conditions, can greatly affect the ability for patients to take on and maintain work at their full capacity. The work productivity of family caregivers may also be negatively influenced because of their caregiving responsibilities. The impact of health or caregiving responsibilities on work can be assessed by measuring one's work productivity loss, which commonly includes 3 main components, absenteeism, presenteeism, and employment status changes. Absenteeism is one of the components and refers to absent work time while one is employed because of health problems for patients or caregiving responsibilities. Productivity loss may also happen while at work, which is defined as presenteeism. There have been over 20 questionnaires developed to measure this work productivity loss outcome. In summary, measuring presenteeism can be done by a direct time estimation method to gauge the extra time needed to complete usual work tasks (eg, the Health and Labor Questionnaire [HLQ] and Valuation of Lost Productivity [VOLP] questionnaires) or by asking patients and caregivers to rate their productivity loss while working using a 0 to 10 scale (eg, the Work Productivity and Activity Impairment Questionnaire [WPAI]). A third component of work productivity loss is employment status changes such as reducing routine working time and the duration of unemployment because of health problems or caregiving responsibilities. For economic evaluations or costing purposes, these components need to be first estimated in terms of work time loss and then converted into a monetary amount. Calculating total work productivity loss in terms of time requires the sum of time losses because of absenteeism, presenteeism, and employment status changes.

Increasingly, clinical trials are measuring work productivity loss as one of their outcomes because it is considered as an important patient-centered outcome and an important cost component for economic evaluations adopting a societal perspective. Randomized controlled trials (RCTs) are main valid sources for the impact of interventions and RCT-based
economic evaluations. The results of RCTs and economic evaluations are critical in informing clinical guidelines and funding decisions. Assessing current research practices equips decision makers to appraise RCTs in terms of the inclusion, analysis, and results presentation of work productivity loss outcomes. In addition, patients and caregivers might rely on RCT results to decide between treatment options and set their expectations. Therefore, it is important to understand how and which work productivity loss outcomes are measured, analyzed, and presented in literature.

In this article, we conducted a scoping review to investigate which work productivity loss outcomes were measured in RCTs published in the past 10 years, how each of these outcomes was measured and analyzed, and how their results were reported or presented.

**Methods**

According to the guidance provided by Munn et al.\(^{19}\) we chose to conduct a scoping review instead of a systematic review for the purpose of examining how research was conducted on a certain topic. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews.\(^{20}\) We limited our search and review to RCTs published between 2010 and 2020 because of the recent increasing uptake of RCTs measuring work productivity loss and their representativeness of the current research status. The review was registered on PROSPERO.\(^{21}\)

**Search Strategy**

Search strategies were developed based on a previous systematic review using Medical Subject Headings terms and keywords related to absenteeism, presenteeism, work productivity, and RCTs (see Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.015).\(^{22}\) PubMed and the Cochrane Library (Cochrane Central Register of Controlled Trials) were searched in May 2020. Articles were limited to English language and those published between January 2010 and April 2020.

**Study Selection and Selection Criteria**

At the first screening stage, 3 reviewers (J.L., J.S., and Y.H.) independently screened one-third of the studies in the literature search by title and abstract for potential inclusion. Broader inclusion criteria were applied at this stage to include studies if the population of interest was adults (regardless of patients or caregivers or employees) with at least some participants working at baseline (to whom absenteeism and presenteeism are applicable), the outcome of interest was work productivity loss (which includes at least one of absenteeism or presenteeism), and study design was an RCT, randomized trial, or an economic evaluation based on a trial. To ensure consensus among reviewers on the screening, the 3 reviewers screened 150 overlapping studies and resolved disagreement or ambiguities by consensus or via discussion with a fourth reviewer (W.Z.).

At the second screening stage, full-text articles were then screened by one reviewer (P.T.). Each full-text article was also reviewed according to the above inclusion criteria with further restriction to RCT studies (including economic evaluation based on an RCT) that measured both absenteeism and presenteeism for final review.

**Data Extraction and Result Summary**

One reviewer (P.T.) first extracted the following information from the included studies: country, population category (patient, caregiver, employee), disease focus, type of RCT, sample size, work productivity loss outcomes measured (absenteeism, presenteeism, employment status changes), work productivity loss measurement and recall period, outcome reporting, analysis method, and whether an economic evaluation was performed. Extracted information was then independently cross-checked by J.L. and J.S. Discrepancies were resolved through discussion between reviewers or via discussion with a fourth reviewer (W.Z.). Each of the 3 reviewers (P.T., J.L., and J.S.) further extracted information from one-third of the included articles such as attribution of work productivity loss (to general health or disease specific), outcome follow-up duration and assessment frequency, contextual factor consideration, costing method, and translation consideration. The fourth reviewer (W.Z.) finally checked the consistency among all the extracted information and verified the extracted information among about one-third of the included articles.

We summarized and evaluated our review results according to the key methodological areas identified by a EULAR task force who performed systematic literature reviews of studies on work participation\(^{23}\) and developed Points to Consider when designing, analyzing, and reporting work participation outcome among patients with inflammatory arthritis.\(^{24}\) The results were presented and compared by type of outcome and by study population if applicable.

**Quality Assessment**

A quality assessment of each article was completed using the Cochrane Collaboration’s tool for assessing risk bias in randomized trials.\(^{25}\) Given that this scoping review focused on presenting and summarizing common statistical analysis and result reporting and presentation methods used for work productivity loss outcomes, rather than the results of these studies themselves, we focused our assessment on the selection bias (random sequence generation and allocation concealment), performance bias (blinding of participants and personnel), and detection bias (blinding of outcome assessment) based on the information provided in the articles. Each article was rated as low, high, or unclear risk of bias for each domain. Each of 2 reviewers (H.S. and W.Z.) assessed a half of the final included studies and suggested for an independent review by the other reviewer for their uncertain assessment (39 articles).

**Results**

**Screening Process**

Figure 1 illustrates the review screening process. The literature search identified 3185 abstracts including 1200 from PubMed and 1985 from Cochrane Library. After duplicates were removed, 2262 abstracts remained. After the initial screening based on titles and abstracts, 928 full-text articles meeting the inclusion criteria were available and subsequently retrieved and reviewed. After screening the full-text articles, we identified 435 articles that met our inclusion criteria and measured absenteeism or presenteeism. Of them, 155 articles (35.6%) included both absenteeism and presenteeism results and thus were included in our final review.\(^{26-189}\) Among the final 155 articles, only 5.8% (n = 9) also measured employment status changes and approximately 18.7% (n = 29) also included an economic evaluation in their results. The extracted detailed information for each of the final 155 articles can be found in Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.015.

Figure 2 plots the 435 articles that measured absenteeism or presenteeism and the final 155 articles by their publication years.
Figure 1. PRISMA flowchart of inclusion and exclusion of articles incorporated in the review.

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

*where an article met more than one exclusion criteria, the first criterion listed was used.
The number of articles measuring absenteeism or presenteeism has been increasing over the past 10 years.

**Study Characteristics and Design**

**Study countries**

Overall, the studies included in the final review were conducted globally. For the studies that provided information on the country where the study was conducted (n = 153), 60 studies took place across multiple countries (Table 1). Outside of these, most studies came from the United States and The Netherlands (n = 29 and n = 25, respectively), with all other countries contributing < 10 studies each.

**Study populations**

The populations of interest included patients, employees, and caregivers. The study population was classified as patients if the study screened the eligibility of participants based on a particular disease focus (e.g., musculoskeletal disorders, psychiatric disorders). The study population was classified as employees if they consisted of a group of employees from a specific organization, profession, workplace, etc. Employee populations included, for example, occupations such as office workers, construction workers, hotel employees, and nurses. These studies could target employees with a specific disease (n = 9) or general employees (n = 21) regardless of whether they have a disease or not. Most articles considered patient populations only (n = 121), followed by employees (n = 30). Only one article looked specifically at unpaid or family caregivers only (n = 1), whereas 3 articles took into consideration both patients and friend, family, or other caregiver participants (n = 3).

**Disease focus**

The most common diseases of interest were musculoskeletal disorders (n = 29) (e.g., rheumatoid arthritis), psychiatric disorders (n = 18) (e.g., major depressive disorder), other risk factors/symptoms (n = 13) (e.g., back pain), gastrointestinal disorders (n = 11) (e.g., irritable bowel syndrome and Crohn’s disease), psoriasis (n = 9), and hepatitis (n = 8).

**Type of intervention**

The type of intervention included pharmacological intervention (n = 82), nonpharmacological intervention (e.g., lifestyle, occupational or workplace intervention) (n = 70), and a mix of pharmacological and nonpharmacological interventions (n = 3).

**Primary outcome**

A total of 53 articles listed a work productivity loss outcome as one of their primary outcomes, 69 studies did not, and 33 studies provided a list of outcomes without specifying which one was primary or secondary. Within the studies focusing on employees (n = 30), almost half listed a work productivity loss outcome as a primary outcome (n = 14) and half did not (n = 12) (Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.015). Studies only focusing on patients less frequently considered a work productivity loss outcome as one of their primary outcomes (37 did vs 56 did not).

**Sample size**

Only 60 articles provided information on sample size calculation or justification. Among them, 10 studies calculated or considered their sample size based on a work productivity loss outcome (n = 6 in patient populations and n = 4 in employee populations). The median total sample size of all the studies was 338 participants (range = 29-6958). Studies with patients had a little larger median total sample size at 359 (29-4233) than studies with employees at 339 (29-6958). The single caregiver study had a total sample size of 308 participants, and the 3 studies with patients and caregivers had a median of 287 participants (135-1326).
Measurement Methods

Measurement questionnaires

Absenteeism and presenteeism were most commonly measured using WPAI (n = 71 and n = 71, respectively) (Table 2). Self-report questions that were not from a preexisting validated questionnaire (n = 22) and/or administrative data (n = 16) were also 2 common sources of absenteeism data. Administrative data consisted of, for example, data provided by companies and local databases from employers. Self-report questions and 4 preexisting validated questionnaires including WLQ, the World Health Organization Health and Work Performance Questionnaire (WHO-HPQ), Work Ability Index (WAI), and Treatment Inventory of Costs in Patients with Psychiatric Disorders (TiC-P) were also commonly used to measure presenteeism (n = 14, n = 12, n = 11, n = 11, and n = 10, respectively).

Other common questionnaires that were used to measure absenteeism and/or presenteeism included the Short Form Health and Labor Questionnaire (SF-HLQ), Work Productivity Survey, and Productivity and Disease Questionnaire (PRODISQ). Among the 9 studies interested in employment status changes, productivity loss was measured using a mix of WPAI, the Economic Implications of Psoriasis Patient Questionnaire, Work Productivity Questionnaire, VOLP, and other self-report questions.

Studies with employee populations were most likely to measure absenteeism using administrative data (n = 13) and self-report question (n = 6) and to measure presenteeism using WAI (n = 9) and WHO-HPQ (n = 8) (Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.015). In contrast, studies with patient populations were most likely to measure absenteeism and presenteeism using WPAI (n = 67 and
Studies with economic evaluations were most likely to measure absenteeism using self-report questions (n = 6), TiC-P (n = 5), and administrative data (n = 4) and to measure presenteeism using self-report questions (n = 6), TiC-P (n = 5), and PRODISQ (n = 4) (Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.015).

Attribution of work productivity loss to general health
The existing questionnaires have been developed and validated to measure work productivity loss because of general health or because of a specific disease (eg, WPAI has both general health version and specific health problem version181) or because of caregiving among caregivers (eg, WPAI,182 WLQ,183 and VOLP184).

Absenteeism and presenteeism were mostly attributed to a specific disease (eg, psychiatric illness, irritable bowel syndrome, atopic dermatitis) (n = 84 and n = 78, respectively) versus attributed to general health (n = 48 and n = 38, respectively). Employment status changes were attributed to general health in 5 studies and to a specific disease in 3 studies.

Recall periods
The recall periods highly depended on the questionnaires the studies used and varied from current or 1 day to 1 year. For the studies that provided at least one recall period (n = 130 for absenteeism excluding not applicable for administrative data, n = 140 for presenteeism), the most common recall period was 7 days for both absenteeism and presenteeism (n = 77 and n = 78, respectively). The next most common recall periods were 3 months (n = 17) and 1 month (n = 13) for absenteeism and 4 weeks (n = 15) and 1 month (n = 13) for presenteeism. For the studies that reported employment status changes, 3 did not report recall periods. The other 6 studies used the duration of time since the last visit, 3 months, and 1 month as their recall periods or compared the current employment status at different time points.

### Table 2. Work productivity loss outcome measurement methods.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Category</th>
<th>Absenteeism (N = 155), n (%)</th>
<th>Presenteeism (N = 155), n (%)</th>
<th>Employment status change (N = 9), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement questionnaire*</td>
<td>WPAI</td>
<td>71 (45.8)</td>
<td>71 (45.8)</td>
<td>1 (11.1)</td>
</tr>
<tr>
<td></td>
<td>Self-report question</td>
<td>22 (14.2)</td>
<td>14 (9.0)</td>
<td>4 (44.4)</td>
</tr>
<tr>
<td></td>
<td>Administrative data</td>
<td>16 (10.3)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>TiC-P</td>
<td>10 (6.5)</td>
<td>10 (6.5)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>WHO-HPQ</td>
<td>5 (3.2)</td>
<td>11 (7.1)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>WPS</td>
<td>5 (3.2)</td>
<td>5 (3.2)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>SF-HLQ</td>
<td>5 (3.2)</td>
<td>5 (3.2)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>WLQ</td>
<td>4 (2.6)</td>
<td>12 (7.7)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>WAI</td>
<td>1 (0.6)</td>
<td>11 (7.1)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>PRODISQ</td>
<td>3 (1.9)</td>
<td>6 (3.9)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>QQ</td>
<td>0 (0.0)</td>
<td>5 (3.2)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Attribution of work productivity loss</td>
<td>General health</td>
<td>48 (31.0)</td>
<td>38 (24.5)</td>
<td>5 (55.6)</td>
</tr>
<tr>
<td></td>
<td>Disease specific</td>
<td>84 (54.2)</td>
<td>78 (50.3)</td>
<td>3 (33.3)</td>
</tr>
<tr>
<td></td>
<td>Not provided</td>
<td>23 (14.8)</td>
<td>39 (25.2)</td>
<td>1 (11.1)</td>
</tr>
<tr>
<td>Recall period*</td>
<td>7 days or 1 week</td>
<td>77 (49.7)</td>
<td>78 (50.3)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>2 weeks</td>
<td>6 (3.9)</td>
<td>12 (7.7)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>4 weeks</td>
<td>10 (6.5)</td>
<td>15 (9.7)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>1 month</td>
<td>13 (8.4)</td>
<td>13 (8.4)</td>
<td>1 (11.1)</td>
</tr>
<tr>
<td></td>
<td>3 months or 13 weeks</td>
<td>17 (11.0)</td>
<td>10 (6.5)</td>
<td>1 (11.1)</td>
</tr>
<tr>
<td>Follow-up duration</td>
<td>&lt;6 months</td>
<td>68 (43.9)</td>
<td>68 (43.9)</td>
<td>1 (11.1)</td>
</tr>
<tr>
<td></td>
<td>6-12 months</td>
<td>68 (43.9)</td>
<td>68 (43.9)</td>
<td>4 (44.4)</td>
</tr>
<tr>
<td></td>
<td>&gt;12 months</td>
<td>19 (12.3)</td>
<td>19 (12.3)</td>
<td>4 (44.4)</td>
</tr>
<tr>
<td>Assessment frequency†</td>
<td>&lt;2 weeks</td>
<td>5 (3.2)</td>
<td>4 (2.6)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>2-12 weeks</td>
<td>52 (33.5)</td>
<td>53 (34.2)</td>
<td>2 (22.2)</td>
</tr>
<tr>
<td></td>
<td>13-26 weeks</td>
<td>47 (30.3)</td>
<td>47 (30.3)</td>
<td>4 (44.4)</td>
</tr>
<tr>
<td></td>
<td>&gt;26 weeks</td>
<td>10 (6.5)</td>
<td>11 (7.1)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Wider range</td>
<td>41 (26.5)</td>
<td>40 (25.8)</td>
<td>3 (33.3)</td>
</tr>
</tbody>
</table>

Note. Self-report question refers to asking participants a question that is not part of a preexisting validated questionnaire. Administrative data include data gathered from employment records, occupational health and safety, and other databases. N indicates the number of total studies included in our final review; PRODISQ, Productivity and Disease Questionnaire; QQ, Quantity and Quality method; SF-HLQ, Short Form Health and Labor Questionnaire; TiC-P, Treatment Inventory of Costs in Patients with Psychiatric Disorders; WAI, Work Ability Index; WHO-HPQ, World Health Organization Health and Work Performance Questionnaire; WLQ, Work Limitations Questionnaire; WPAI, Work Productivity and Activity Impairment Questionnaire; WPS, Work Productivity Survey.

*Most commonly used for absenteeism or presenteeism and if using multiple questionnaires, a single study could be counted more than one time.
†The difference between 2 consecutive assessment time points.

n = 67, respectively) and self-report questions (n = 16 and n = 8, respectively).

Studies with economic evaluations were most likely to measure absenteeism using self-report questions (n = 6), TiC-P (n = 5), and administrative data (n = 4) and to measure presenteeism using self-report questions (n = 6), TiC-P (n = 5), and PRODISQ (n = 4) (Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.015).
Follow-up duration

The RCT follow-up durations varied from < 6 months (n = 65) and 6 to 12 months (n = 70) to >12 months (n = 20) with the longest follow-up of 3 years. The corresponding number of studies for absenteeism and presenteeism were n = 68, n = 68, and n = 19, respectively. Among the studies that measured employment status changes, a higher proportion of studies had a follow-up duration of >12 months (n = 4 of 9).

Outcome assessment frequency

The frequency of assessment (ie, the difference between 2 consecutive assessment time points) varied widely from 1 day to 18 months. The most common assessment frequency for absenteeism and presenteeism was 2 to 12 weeks (n = 52 and n = 53, respectively) and 13 to 26 weeks (n = 47 and n = 47, respectively). Employment status changes were assessed at a fixed frequency (ie, monthly, every 12 weeks, 13 weeks, 3 months, or 6 months) in 5 studies or at varying intervals in other 4 studies. Studies among employees were more likely to have a longer follow-up and less frequent assessment than studies among patients.

Analysis Methods

Group comparison

Given that our review focused on RCTs, most of the studies compared absenteeism and presenteeism between intervention groups (n = 139 and n = 140, respectively) and 10 studies compared absenteeism and presenteeism between clinical responders and nonresponders by pooling study participants from different intervention groups together (Table 3).

Analysed outcome

Absenteeism and presenteeism were mostly analyzed for their changes from baseline only (the difference between the outcome at a follow-up and baseline outcome) (n = 69 and n = 72, respectively). Some studies also mainly analyzed both outcomes at a certain follow-up (eg, 4-week absenteeism at the 24-week follow-up) (n = 38 for both) or the total absenteeism and presenteeism during the entire study duration (n = 21 and n = 22, respectively).

Statistical analysis method

To analyze absenteeism and presenteeism, 98 articles used a statistical method based on regression (eg, analysis of covariance [ANCOVA], ordinary least squares [OLS], mixed model for repeated measures from multiple time points and/or clusters), 33 articles were not regression based (eg, chi-square test, analysis of variance [ANOVA] or Student’s t test), and 24 articles reported descriptive statistics only (comparisons between groups were not evaluated or tested). Studies among employees were slightly more likely to analyze both outcomes using regression models than studies among patients (Appendix Table 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.015).

Among the regression-based articles, we characterized the assumed distributions in 4 categories: (1) normal; (2) binomial (logistic regression); (3) Poisson, gamma, negative binomial (NB), or ordinal; and (4) zero-inflated or 2-part model. For absenteeism, 77 studies assumed a normal distribution (41 ANCOVA and 36 linear regression model); 7 studies assumed a binomial (logistic) distribution; 10 studies assumed a Poisson, gamma, NB, or ordinal (eg, cumulative logistic regression) distribution; and 4 addressed zero-inflated distribution (2 zero-inflated NB, 1 hurdle [2-part] NB, and 1 zero-inflated Poisson). The commonly used non-regression-based methods were Student’s t test (n = 13), nonparametric test such as Mann-Whitney U and/or Wilcoxon signed-rank test (n = 9), and ANOVA (n = 4).

Similarly, for presenteeism, 89 studies assumed a normal distribution (43 ANCOVA and 46 linear regression model); 2 studies assumed a binomial (logistic) distribution; 6 studies assumed a Poisson, gamma, NB, or ordinal distribution; and 1 assumed a zero-inflated NB distribution. The commonly used nonregression-based methods were Student’s t test (n = 15), Mann-Whitney U and/or Wilcoxon signed-rank test (n = 8), and ANOVA (n = 5).

For the articles that measured employment status changes, 5 articles used a regression-based statistical method (2 using Cox regression to measure time to event, 1 linear multilevel regression, 1 logistic regression, and 1 zero-inflated NB), 3 articles reported descriptive measures only, and 1 used a non-regression-based method.

Outcome categorization

Among those studies using logistic regression and chi-squared test for absenteeism (n = 8), 4 studies defined absenteeism as yes versus no and 4 other studies used different cutoffs (6 days per 6 months, 7 days per 6 months, 9 days per year, and 10 days per year, respectively) (Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.015). For presenteeism, all relevant studies defined it as a binary based on yes versus no (with productivity loss vs without productivity loss or full productivity). In the studies using cumulative logistic regression (1 for absenteeism and 2 for presenteeism), they all analyzed the originally reported absent days or the 0 to 10 presenteeism scale without any categorization.

Contextual factors considered

Fifty-three studies considered contextual factors (other than follow-up visit time or baseline work productivity loss outcome) as covariates or confounders (Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.015). The factors that were most often adjusted for were age (n = 29), sex or gender (n = 27), race/ethnicity (n = 10), education (n = 6), and employment-related factors (n = 10) including occupation, full-time/part-time employment, type of work (manual vs administrative), contract hours, and job tenure. In addition, 20 studies considered factors via an interaction term (other than the interaction between intervention and visit time) in the regression model or subgroup analyses. The common factors being considered included age (n = 4), sex or gender (n = 4), education (n = 2), and race (n = 1).

Outcome Presentation

Absenteeism was more likely to be presented as time in hours or days (n = 44 vs n = 15) or percent of time (n = 31 vs n = 1) than presenteeism. Presenteeism was more likely to be presented as a percent score/scale (n = 18 vs n = 52) or a score/scale (n = 11 vs n = 43) than absenteeism (Table 4). Several articles also reported both outcomes in more than one format (n = 29 and n = 25, respectively). Employment status changes were reported as a percent of people (n = 5), time (n = 2), or multiple outcomes (n = 2). Studies among employees were more likely to report absenteeism in time than studies among patients (Appendix Table 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.015).

Among economic evaluation studies and studies presenting outcomes as cost, human capital approach (HCA) was more
commonly used to value the lost time using gross salary plus benefits or wage rate (n = 15 and n = 13, respectively). Although some studies did not explicitly specify that they used HCA, their costing method description suggested they applied a similar method (n = 15 and n = 19, respectively). Only a few studies conducted in The Netherlands applied friction cost approach (FCA) (n = 6 and n = 3, respectively). One study used both HCA and FCA (as a sensitivity analysis) for absenteeism. The friction period they used included 85 days, 115 days, and 23 weeks (Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.015).

Quality Assessment

Most articles (> 60%) did not provide enough information on random sequence generation and allocation concealment to enable us to clearly assess their selection bias (Appendix Table 5 and Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.015). Approximately 35% of the studies had low risk of selection bias. Given that a higher proportion of the interventions were nonpharmacologic interventions (45.2%) or mixed of pharmacological and nonpharmacologic interventions (1.9%) such as behavior changes, the blinding of the study participants was not possible. These studies had high risk of performance bias and detection bias because of the self-reported work productivity outcome assessment. Consequently, approximately 56% of articles were assessed as high risk of performance bias and detection bias, and 41% of articles were assessed as low risk.

Discussion

In this article, we reviewed RCTs published in the past 10 years and investigated how they measured, analyzed, and presented absenteeism and presenteeism. We found a large number of RCTs (n = 435) measured at least absenteeism or presenteeism in the last 10 years. Nevertheless, only 155 of them (35.6% of 435) measured both absenteeism and presenteeism and thus were included in our final review. Moreover, only 9 studies measured employment status changes in addition to absenteeism and presenteeism. These findings signal that total work productivity loss in RCTs may be underestimated given that most of the literature we reviewed did not include absenteeism, presenteeism, and employment status changes, which could lead to biased treatment effects. We found that the follow-up duration of the studies

<table>
<thead>
<tr>
<th>Topics</th>
<th>Category</th>
<th>Absenteeism (N = 155), n (%)</th>
<th>Presenteeism (N = 155), n (%)</th>
<th>Employment status change (N = 9), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group comparison</td>
<td>Yes, comparing intervention groups</td>
<td>139 (89.7)</td>
<td>140 (90.3)</td>
<td>8 (88.9)</td>
</tr>
<tr>
<td></td>
<td>Yes, comparing responders and nonresponders</td>
<td>10 (6.5)</td>
<td>10 (6.5)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>6 (3.9)</td>
<td>5 (3.2)</td>
<td>1 (11.1)</td>
</tr>
<tr>
<td>Analyzed outcome</td>
<td>Change from baseline</td>
<td>69 (44.5)</td>
<td>72 (46.5)</td>
<td>2 (22.2)</td>
</tr>
<tr>
<td></td>
<td>Point estimate at a follow-up visit</td>
<td>38 (24.5)</td>
<td>38 (24.5)</td>
<td>3 (33.3)</td>
</tr>
<tr>
<td></td>
<td>Total during the study period</td>
<td>21 (13.5)</td>
<td>22 (14.2)</td>
<td>2 (22.2)</td>
</tr>
<tr>
<td></td>
<td>Multiple*</td>
<td>26 (16.8)</td>
<td>23 (14.8)</td>
<td>2 (22.2)</td>
</tr>
<tr>
<td>Statistical analysis method</td>
<td>Non-regression-based method</td>
<td>33 (21.3)</td>
<td>33 (21.3)</td>
<td>1 (11.1)</td>
</tr>
<tr>
<td></td>
<td>Mann-Whitney U/Wilcoxon signed-rank test</td>
<td>9 (5.8)</td>
<td>8 (5.2)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Nonparametric bootstrap t test</td>
<td>3 (1.9)</td>
<td>3 (1.9)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Other nonparametric test</td>
<td>2 (1.3)</td>
<td>1 (0.6)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>t test</td>
<td>13 (8.4)</td>
<td>15 (9.7)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>ANOVA</td>
<td>4 (2.6)</td>
<td>5 (3.2)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>95% confidence interval</td>
<td>1 (0.6)</td>
<td>1 (0.6)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Chi-square</td>
<td>1 (0.6)</td>
<td>0 (0.0)</td>
<td>1 (11.1)</td>
</tr>
<tr>
<td>Regression-based method</td>
<td>Negative binomial‡</td>
<td>5 (3.2)</td>
<td>1 (0.6)</td>
<td>1 (11.1)</td>
</tr>
<tr>
<td></td>
<td>Poisson‡</td>
<td>5 (3.2)</td>
<td>1 (0.6)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>3 (1.9)</td>
<td>3 (1.9)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Cumulative logistic</td>
<td>1 (0.6)</td>
<td>2 (1.3)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Logistic</td>
<td>7 (4.5)</td>
<td>2 (1.3)</td>
<td>1 (11.1)</td>
</tr>
<tr>
<td></td>
<td>ANCOVA</td>
<td>41 (26.5)</td>
<td>43 (27.7)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Linear normal</td>
<td>36 (23.2)</td>
<td>46 (29.7)</td>
<td>1 (11.1)</td>
</tr>
<tr>
<td></td>
<td>Cox</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>2 (22.2)</td>
</tr>
<tr>
<td></td>
<td>Other: descriptive</td>
<td>24 (15.5)</td>
<td>24 (15.5)</td>
<td>3 (33.3)</td>
</tr>
</tbody>
</table>

ANCOVA indicates analysis of covariance; ANOVA, analysis of variance; N, the number of total studies included in our final review. *Analyzing more than one type of outcome listed above. ‡Including 2 zero-inflated negative binomial and 1 hurdle negative binomial for absenteeism, 1 zero-inflated negative binomial for presenteeism, and 1 zero-inflated negative binomial for employment status change, respectively. §Including 1 zero-inflated Poisson for absenteeism.
measuring employment status changes was relatively longer than other studies. Most of the studies we finally reviewed had a short follow-up duration ≤ 1 year and thus might not be able to capture employment status changes.

Most RCTs took place across multiple countries, the United States, or The Netherlands. The economic evaluation guidelines or recommendations of the United States and The Netherlands recommend adopting a societal perspective by considering work productivity loss, which might lead to more RCTs including work productivity loss as one of their outcomes. Twenty-one studies looked at general employee populations without target conditions. This is not surprising considering that work productivity loss is one of the most important outcomes for workplaces or employers.

Among the studies that had a specific disease focus, musculoskeletal and psychiatric conditions were the most common disease categories. Rheumatoid arthritis was the most common musculoskeletal condition that RCTs focused on. The psychiatric category included, but was not limited to, major depressive disorder, psychotic disorder, and anxiety disorder. Consistently, these conditions have been shown to be the top conditions having the highest absenteeism and presenteeism in previous studies. Moreover, this is important to keep developing effective interventions to help improve the work productivity of people living with these conditions.

The WPAI questionnaire was the most common measurement for both absenteeism and presenteeism. Other questionnaires were also commonly used such as WHO-HPQ, TiC-P, and SF-HLQ for both absenteeism and presenteeism. However, these questionnaires may be not the most suitable for measuring productivity loss because they require employment status changes.

The WHODAS questionnaire was the most common measurement for both absenteeism and presenteeism. Other questionnaires were also commonly used such as WHO-HPQ, TiC-P, and SF-HLQ for both absenteeism and presenteeism. However, these questionnaires may be not the most suitable for measuring productivity loss because they require employment status changes.

Table 4. Work productivity loss outcome presentation.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Category</th>
<th>Absenteeism (N = 155), n (%)</th>
<th>Presenteeism (N = 155), n (%)</th>
<th>Employment status change (N = 9), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome presentation</td>
<td>$</td>
<td>16 (10.3)</td>
<td>16 (10.3)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Time</td>
<td></td>
<td>44 (28.4)</td>
<td>15 (9.7)</td>
<td>2 (22.2)</td>
</tr>
<tr>
<td>% time</td>
<td></td>
<td>31 (20.0)</td>
<td>1 (0.6)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Score/scale</td>
<td></td>
<td>11 (7.1)</td>
<td>43 (27.7)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>% score/scale</td>
<td></td>
<td>18 (11.6)</td>
<td>52 (33.5)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>% people</td>
<td></td>
<td>6 (3.9)</td>
<td>3 (1.9)</td>
<td>5 (55.6)</td>
</tr>
<tr>
<td>Multiple*</td>
<td></td>
<td>29 (18.7)</td>
<td>25 (16.1)</td>
<td>2 (22.2)</td>
</tr>
<tr>
<td>Costing method†</td>
<td>Number of studies</td>
<td>37</td>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Human capital approach</td>
<td>15 (40.5)</td>
<td>13 (37.1)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Friction cost approach</td>
<td>6 (16.2)</td>
<td>3 (8.6)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>15 (40.5)</td>
<td>19 (54.3)</td>
<td>1 (100.0)</td>
</tr>
<tr>
<td></td>
<td>Human capital approach and friction cost approach</td>
<td>1 (2.7)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

N indicates the number of total studies included in our final review.
*Presented in more than one format.
†The denominator of % is the total number of studies that conducted economic evaluation or reported cost.

Due to different recall periods for both absenteeism and presenteeism and different measurement methods for presenteeism, work productivity loss outcomes were not comparable between RCTs. A previous study among people with arthritis has shown that questionnaires using the 3 different methods (HLQ for direct time estimation, WPAI and WHO-HPQ for 0 to 10 scale, and WLQ for multidimensional scale) gave a wide range of lost time presenteeism estimates with the highest estimates from WPAI and WHO-HPQ and the lowest estimates from HLQ. Nevertheless, it is uncertain whether the wide range of estimates from different measurement methods is consistent among other patient populations or general employee and caregiver populations. For instance, it remains unclear which questionnaire is most appropriate because the decision depends on the study purpose, either evaluating productivity loss as one patient-centered outcome or translating the loss into monetary amounts as part of an economic evaluation.

Most preexisting questionnaires have been validated in English version. In addition, the questionnaires including TiC-P, SF-HLQ, PRODISQ, Quantity and Quality method, and HLQ were originally developed and validated by researchers in The Netherlands. As the most commonly used questionnaire, WPAI has been translated and validated in many languages. Similarly, translated versions of other questionnaires such as WLQ, TiC-P, and VOP are also available. Among the studies that administered preexisting questionnaires not originally developed in a local language of the
study participants (n = 89), we found that 19 studies specifically mentioned the language version they used, described the translation, or commented on the validation of the translated version (Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.06.015). Future studies measuring work productivity loss are suggested to use one of the preexisting validated questionnaires that have been appropriately translated or to pretest their own translated version.

The most common analysis method for absenteeism and presenteeism was regression based, assuming a normal distribution. The distribution of work productivity loss is well known to be highly skewed with many zeros (ie, inflated zeros). In addition, if employment status changes are considered, a high proportion of work stoppage or unemployment might lead to a high proportion of maximum loss, that is, inflated maximum values. Different statistical models have been used to analyze this type of data, for example, 2-part models, zero-inflated models, and other mixture models.104,198–204 A recent simulation study suggested that OLS, 2-part models, or 3-part models could be appropriate for analyzing work productivity loss because of health problems in RCTs and the final model selection depends on the sample size, the proportions of zero loss and maximum loss, and the productivity loss outcome distribution between zero and maximum loss in each arm of RCTs.205 Nevertheless, in RCTs we reviewed, the most commonly used methods included ANCOVA, mixed model for repeated measures, or OLS. Using these regression models might lead to biased estimation because work productivity loss data may not meet the underlying assumptions of these models (eg, error term normally distributed).

Absenteeism was commonly presented or reported as time (days or hours) or a percent of time. Presenteeism was commonly reported using a scale/score as a percent or as other types of scores/scales. The result presentation partially depends on the questionnaires used and on the study objectives (ie, patient-centered outcome or an economic evaluation). The monetary value is required for economic evaluation purpose, and thus, work productivity losses need to be presented or converted into time losses first and then costs. To do so, studies commonly converted the percent scale/score to the time lost because of presenteeism by multiplying it by the actual work time.

To value the time lost owing to absenteeism and presenteeism, studies commonly applied the HCA approach using gross salary plus benefits or wage rate. The FCA was proposed by health economists in The Netherlands206 and their economic evaluation guideline.185 The FCA typically applies to absenteeism and assumes that productivity loss occurs only during a friction period until a replacement can be found. Nevertheless, the friction period could change over time because it depends on the level of unemployment and on the efficiency of the labor market in matching labor demand and supply in each country.205 The friction period has been updated from 14 weeks in 2000, 18 weeks in 2004, and 23 weeks in 2010 to 12 weeks (or 85 days) in 2016 by the Dutch costing manual.185,207,208 In contrast, other economic evaluation guidelines recommend using HCA instead of FCA, for example, the Second Panel on Cost-Effectiveness in Health and Medicine.19 Therefore, researchers should follow their local most recent economic evaluation guideline on how to value work productivity loss205 and, if possible, conduct a sensitivity analysis by applying both approaches.

Work and employment are important part of life. They are included as categories in the International Classification of Functioning, Disability and Health (eg, “maintaining a job” and “remunerative employment”), which is the World Health Organization framework and international standard to describe and measure health and disability.210 Work productivity loss has been considered as one of important patient-reported outcomes by patients, caregivers, and researchers211 and included as an outcome in our reviewed RCTs. Nevertheless, only 18.7% of the studies included in our review were an economic evaluation based on an RCT. This could be partially because the guidelines on economic evaluation in most countries do not require a societal perspective to include work productivity loss as one of the cost components in economic evaluations209 and partially indicate that clinical researchers might not consider cost as a clinic concept or outcome so that they did not measure work productivity loss for costing purpose or economic evaluation purpose.

A EULAR task force previously conducted 2 similar systematic reviews to our review but focused on studies in inflammatory arthritis and other chronic diseases.23 They searched 4 databases and limited their reviews to studies published within a 5-year period from January 2014 to April 2019. Their review inclusion criteria were broader than our review including both RCTs and longitudinal observational studies that assessed work status, absenteeism, or presenteeism. Nevertheless, our review still included more studies based on RCTs because of our 10-year literature search period and no disease restriction. For example, we included 30 studies among employee populations. Despite of these differences between our review and the previous reviews, we identified some similar methodological heterogeneity and issues in measuring, analyzing, and presenting work productivity outcomes. These issues included a short follow-up duration not adequate for employment status changes, sample size calculation not being reported or not based on work productivity outcomes, a lack of consideration for contextual factors and zero-inflated highly skewed data distribution in data analysis, and a lack of consensus of presenting the outcomes.

One limitation of our review is that it was limited to 2 databases, and therefore, we may have missed some studies. We also restricted our search to studies written in English and focused on the past 10 years of literature. During stage 1, 3 independent reviewers completed the abstract screening and exclusion. To minimize this limitation, 150 articles were cross-checked between 2 reviewers (agreement = 78%). Discrepancies were solved using a third reviewer. In addition, only one reviewer completed stage 2 screening, which consisted of full-text article reviews. Some data extraction was also first completed by one researcher and the extracted information was subsequently checked by 2 additional researchers (agreement = 96.8%). Considering that our purpose was to examine how research is conducted on measuring, analyzing, and presenting work productivity loss in existing RCTs, we expect the selected RCTs to be representative of the recent research practices.

Another limitation is that we only restricted our review to the studies based on RCTs. There are many observational studies (cohort studies and cross-sectional studies) that have measured, analyzed, and reported work productivity loss outcomes. In the previous systematic review of studies in inflammatory arthritis, the authors reviewed both RCTs and longitudinal observational studies and found that both types of study revealed a high methodological heterogeneity and similar methodological flaws.23 Therefore, we expect that our findings and suggestions could apply to other types of study.

Conclusions

Our study highlights that there is a lack of consensus on how to measure, analyze, and present work productivity loss outcomes, including absenteeism, presenteeism, and employment status
changes, in RCTs. The methodological diversity in RCTs may make comparability challenging. Guidelines are required to provide recommendations to standardize the comprehensiveness and the appropriateness of methods used to measure, analyze, and report work productivity loss in RCTs. It is critical to equip patients, caregivers, and decision makers to be able to effectively compare the results of RCTs when it comes to work productivity loss.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2022.06.015.

Article and Author Information

Accepted for Publication: June 8, 2022
Published Online: xxxx
doi: https://doi.org/10.1016/j.jval.2022.06.015
Author Affiliations: School of Population and Public Health, The University of British Columbia, Vancouver, BC, Canada (Zhang, L’Heureux); Centre for Health Evaluation and Outcome Sciences, Vancouver, BC, Canada (Zhang, Tocher, Sou, Sun).
Correspondence: Wei Zhang, PhD, Centre for Health Evaluation and Outcome Sciences, 588-1081 Burrard St, Vancouver, BC V6Z 1Y6, Canada. Email: wzhang@cheos.ubc.ca


Conflict of Interest Disclosures: Conflict of Interest Disclosures: The authors reported no conflicts of interest.

Funding/Sponsor: This work was supported by BC SUPPORT Unit Methods Cluster Project Award (RWCT-301). Dr Zhang acknowledges the support of the Michael Smith Foundation for Health Research Scholar Award.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgment: The authors acknowledge Yike Huang for her assistance in the first screening stage.

REFERENCES


85. Grif


