ORIGINAL RESEARCH

Methodological aspects of design, analysis and reporting of studies with work participation as an outcome domain in patients with inflammatory arthritis: results of two systematic literature reviews informing EULAR points to consider

Mary Lucy Marques,1,2 Alessia Alunno,3 Annelies Boonen,4,5 Marieke M ter Wee,6,7 Louise Falzon,8 Sofia Ramiro,1,9 Polina Putrik4,5

ABSTRACT

Objective To summarise the methodological aspects in studies with work participation (WP) as outcome domain in inflammatory arthritis (IA) and other chronic diseases.

Methods Two systematic literature reviews (SLRs) were conducted in key electronic databases (2014–2019): search 1 focused on longitudinal prospective studies in IA and search 2 on SLRs in other chronic diseases. Two reviewers independently identified eligible studies and extracted data covering pre-defined methodological areas.

Results In total, 58 studies in IA (22 randomised controlled trials, 36 longitudinal observational studies) and 24 SLRs in other chronic diseases were included. WP was the primary outcome in 26/58 (45%) studies. The methodological aspects least accounted for in IA studies were as follows (proportions of studies positively adhering to the topic are shown): aligning the studied population (16/58 (28%)) and sample size calculation (8/58 (14%)) with the work-related study objective; attribution of WP to overall health (28/58 (48%)); accounting for skewness of presenteeism/sick leave (10/52 (19%)); accounting for work-related contextual factors (25/58 (43%)); reporting attrition and its reasons (1/58 (2%)); reporting both aggregated results and proportions of individuals reaching predefined meaningful change or state (11/58 (16%)). SLRs in other chronic diseases confirmed heterogeneity and methodological flaws identified in IA studies without identifying new issues.

Conclusion High methodological heterogeneity was observed in studies with WP as outcome domain. Consensus around various methodological aspects specific to WP studies is needed to improve quality of future studies. This review informs the EULAR Points to Consider for conducting and reporting studies with WP as an outcome in IA.

Key messages

What is already known about this subject?

► Inflammatory arthritis (IA) has substantial impact on work participation (WP).

► Previous systematic literature reviews of studies with WP as an outcome documented deficiencies in the study design, analysis and reporting of results, hampering interpretation, comparison and meta-analysis.

What does this study add?

► This study provides a synthesis of the methodological choices and issues in studies with WP as an outcome domain in IA and in other chronic diseases.

► Methodological heterogeneity and flaws were identified across four key areas of potential concern: (1) study design, (2) outcome domains and measurement instruments, (3) data analysis and (4) reporting of results.

How might this impact on clinical practice?

► This study aims to inform the efforts to improve the methodological quality and homogeneity of future studies with WP as an outcome domain, and ultimately contribute to high-quality evidence on interventions to support endurable WP.

► This review informs the EULAR Points to Consider when designing, analysing and reporting studies with WP as an outcome domain in IA.

INTRODUCTION

Inflammatory arthritis (IA) encompasses a group of chronic diseases typically affecting adults in working age, and often leading to work disability with consequent loss of income.
for patients and high social expenditures for society.\(^1\) The treatment of IA aims at reaching remission or, at least, low disease activity in order to prevent structural damage and improve patients’ quality of life. Despite the proven efficacy of new therapies such as biologic (b) and targeted synthetic (ts) disease-modifying anti-rheumatic drugs (DMARDs), the burden of restricted participation in work remains high.

People living with IA have identified the ability to maintain a job and being productive while at work as a priority, ranked right after suppressing pain and improving physical function.\(^2\) Work participation (WP) is defined as an active engagement in the role of worker.\(^2\) In addition to the employment status (being employed or not), restrictions in work participation can be quantified using absenteeism (namely sick leave) and presenteeism.\(^4\) Absenteeism refers to the time missed from work due to health reasons and presenteeism refers to experienced restrictions or impaired productivity while at work due to health reasons.\(^2\) People can transition back and forth between not working, working with difficulty and working without difficulty.\(^5\)

To ensure effective interventions to support endurable WP, high-quality evidence is required. However, several systematic literature reviews in IA showed inconclusive results that could be partially attributed to methodological issues in the study design, analysis and reporting of results hampering correct interpretation, comparison and meta-analysis of studies.\(^6\)\(^,\)\(^7\)

WP is increasingly seen as an important outcome of interventions and thus as a target for improvement. During the past decade, the Outcome Measures in Rheumatology (OMERACT) Productivity Working Group focused its work on evaluating and improving the validity of outcomes and outcome measurement instruments of WP.\(^4\)\(^,\)\(^8\)\(^–\)\(^10\) Despite its continuous efforts to harmonise measurement of worker productivity loss across studies, valid instruments are not sufficient to ensure high-quality clinical studies.

The primary aim of this systematic literature review (SLR) was to inform the EULAR task force working on ‘points to consider when designing, analysing and reporting studies with WP as an outcome domain among patients with IA’. The specific objectives of the present work were (1) to summarise the methodological choices in studies with WP as an outcome domain in IA and (2) to identify the methodological issues reported in SLRs of studies with WP as an outcome domain in other chronic diseases.

**METHODS**

**Search strategy and eligibility criteria**

EULAR task force working on ‘points to consider when designing, analysing and reporting studies with WP as an outcome domain among patients with IA’ outlined the scope of the literature search and pre-identified 24 topics in seven main areas of potential concern: (1) study design, (2) outcome domains, (3) outcome measurement instruments, (4) contextual factors, (5) data analysis, (6) reporting of results and (7) estimating productivity costs. These topics were based on (a) knowledge of the literature and experience with conducting such studies and (b) potential role of the issues on bias (selection, information and statistical bias). After a careful evaluation of the seven pre-defined areas and 24 topics, and to avoid redundancy, they were grouped in four main areas (study design, work outcome domains and instruments, data analysis and reporting of results) and 16 topics (figure 1).

For topics 3 and 9, some context is needed. The follow-up time for outcome assessment should be sufficient to capture changes in the work outcome of interest (topic 3). While for presenteeism and sick leave responsiveness was demonstrated at 24 weeks of follow-up,\(^11\) for work status, a follow-up of at least 1 year is preferred. In

<table>
<thead>
<tr>
<th>A. Study design</th>
<th>B. Work outcome domains and instruments</th>
<th>C. Data analysis</th>
<th>D. Reporting of results</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) included population aligned with the work-related study objective (yes/no)</td>
<td>(5) work outcome domains defined (yes/no)</td>
<td>(10) accounting for skewness of the outcome (yes/no and how)</td>
<td>(13) work-related drop out/loss to follow-up reported (yes/no)</td>
</tr>
<tr>
<td>(2) sample size calculated for the work-related outcome (yes/no)</td>
<td>(6) validated measurement of sick leave (yes/no and which)</td>
<td>(11) accounting for interdependence of outcomes (yes/no and how)</td>
<td>(14) the size and characteristics of the (sub)groups analysed are described (yes/no)</td>
</tr>
<tr>
<td>(3) time horizon accurate for the work outcome of interest (≤ 6 months; &gt;6-12 months; &gt;12 months)</td>
<td>(7) validated measurement of presenteeism (yes/no and which)</td>
<td>(12) accounting for contextual factors (yes/no, which and how)</td>
<td>(15) both group level and patient level presented (yes/no)</td>
</tr>
<tr>
<td>(4) use of a comparator (yes/no)</td>
<td>(8) attribution of work participation to overall health vs in relation to inflammatory arthritis or no attribution</td>
<td></td>
<td>(16) provided data on volumes of production loss in cost-studies (yes/no, which)</td>
</tr>
</tbody>
</table>

Figure 1 | Representation of the 16 pre-defined topics (1 to 16) grouped by the four main methodological areas (A to D).
fact, changes in work status can only be detected over shorter follow-up periods of ≤6 months if large sample sizes are used. Work status change and, more precisely, transitions between employment and unemployment can be seen as formally the last step in a sequence of events that start with presenteeism and/or absenteeism. On the other hand, regarding the recall of the assessment instrument (topic 9), there is evidence that a recall period beyond 3 months for sick leave becomes inaccurate and that patients prefer a recall of 1 week for presenteeism (with maximal accuracy for a 4-week recall).

Two searches were conducted according to the PICOT (Population, Intervention, Comparator, Outcomes, Time of follow-up) framework—details are provided in online supplemental figure S1. Search 1 focused on studies with WP as outcome domain in IA, aiming at critically appraising methodological choices and heterogeneity across studies, and search 2 on SLRs of studies with WP as outcome domain in other chronic diseases, aiming to identify whether our pre-identified methodological issues in studies in IA were also recognised in other chronic diseases and/or new aspects were revealed.

For search 1, the following study designs were included: randomised controlled trials (RCTs), controlled clinical trials and prospective observational studies (including registries). Also, studies in IA assessing costs of changes in work participation were identified and included in order to assess whether volumes of work productivity (eg, days, hours) were reported as a separate step before converting volumes into costs. Other specific methodological aspects related to this particular type of study were considered beyond the scope for the current review. Exclusion criteria for both searches are provided in online supplemental text S1.

The search strategies were designed by an experienced librarian (LF). MEDLINE, EMBASE, CINAHL and the Cochrane Library were searched (details on search strategies in online supplemental text S2 and S3) between January 2009 and May 2019.

**Study selection and data extraction**

For both searches, references and abstracts were imported into the reference management software EndNote V.X7.0.2 and deduplicated.

As a high number of hits resulted from the initially defined broad timeframe (n=7715), it was decided to limit the review to recent studies published from January 2014 to April 2019 (n=5534). This decision was based on feasibility and with the rationale that the most recent studies would likely be of better methodological quality and better reflect current standards.

Two researchers (MLM and MMtW) independently screened all titles and abstracts. Next, full texts were reviewed to determine eligibility. Disagreements were resolved by discussion, and if necessary, the methodologists (SR and PP) were involved to make a final decision.

For both searches, study details and results of eligible studies were retrieved by two reviewers (MLM and AA) using a standardised data extraction sheet. Both reviewers (MLM and AA) retrieved data from a 20% random selection of all the included studies. Given an agreement of 89% and consensus on how to further avoid divergences in data extraction, reviewers continued to independently retrieve data of the remaining studies.

For studies in IA, general characteristics of the studies were first retrieved, such as the type of study (RCTs vs longitudinal observational studies), type of intervention (pharmacological intervention, non-pharmacological intervention and natural course of the disease), assessed WP outcome domain (work status and/or sick leave and/or presenteeism) and also if the WP outcome domain was assessed as primary or secondary outcome (online supplemental table S1). Then, the methodological choices regarding the 16 pre-defined topics (figure 1) were retrieved by area: study design (table 1), work outcome domains and instruments (table 2), data analysis (table 3) and reporting of results (table 4).

For SLRs in other chronic diseases, all the methodological issues, as reported by the authors of the SLRs, were retrieved and categorised into the 16 pre-defined topics (figure 1). The quality of the SLRs was not assessed as we were interested in reviewing which methodological flaws were reported in other chronic diseases, particularly focusing on new aspects not previously identified in IA studies. Both SLRs were registered in PROSPERO (CRD42020186798).

**RESULTS**

For the SLR in IA, the literature search yielded 7715 hits. After removing duplicates, conference abstracts and publications before 2014, 2427 articles remained for screening of titles and abstracts, leading to screening of 132 full-text articles. Twenty-three studies on costs of WP were cross-sectional or retrospective and therefore did not comply with inclusion criteria to assess general methodological choices. A total of 81 studies were included in our analysis (flowchart in online supplemental figure S2): 58 for extraction of general methodological choices, 16–73 23 for outcome reporting studies on costs of work productivity16 74–96 and one providing information on both outcomes.16

The search for SLRs in other chronic diseases yielded 10 208 hits. After excluding duplicates and studies before 2014, 3547 titles and abstracts were screened, resulting in screening of 148 full-text articles, and finally 24 were included in the analysis (flowchart in online supplemental figure S3).10–120

**General characteristics of the included studies**

The 58 IA studies appraising general methodological issues comprised 46 longitudinal observational studies.16 17 20 23 26 27 29–31 33 34 39–47 50 51 55–59 61 63 66 68 70–73 and 22 RCTs.18 19 21 22 24 25 28 32 35–38 48 49 52–54 60 62 64 67 69 The characteristics of included studies are provided in online supplemental table S1.
Table 1  Methodological choices in the area of ‘study design’

<table>
<thead>
<tr>
<th>Topics</th>
<th>Results from studies in inflammatory arthritis (n=58)</th>
<th>Aspects identified by authors of SLRs in other chronic diseases (n=24)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n/N (%)</td>
<td>Detailed information</td>
</tr>
</tbody>
</table>
| 1. The included population aligned with the work-related study objective | 16/58 (28%) | The included population specifically aligned with the work-related study objective in 16/58 (28%) studies (16/28 (57%) studies with work as primary outcome)†—n/N$:  
- 2/21 (10%) RCTs with pharmacological intervention
- 1/1 (100%) RCTs with non-pharmacological intervention
- 4/13 (31%) OBS with pharmacological intervention
- 1/2 (50%) OBS with non-pharmacological intervention
- 8/21 (38%) OBS on the natural course of the disease | 3 (12%) | Lack of clarity on the recruitment procedure
3 (12%) | Study population not representative
3 (12%) |
| 2. Sample size calculated for the work-related outcome | 8/58 (14%) | The sample size for the work-related outcome was calculated in eight studies (8/28 (29%) studies with work as primary outcome)†—n/N$:  
- 1/21 (5%) RCTs with pharmacological intervention
- 1/1 (100%) RCTs with non-pharmacological intervention
- 4/13 (31%) OBS with pharmacological intervention
- 2/2 (100%) OBS with non-pharmacological intervention
- 0/23 (0%) OBS on the natural course of the disease | 3 (12%) | No sample size calculation included studies
4 (17%) | Study population too small
21 (87%) |
| 3. Time horizon accurate for the work outcome of interest | 56/58 (97%) | The time-horizon aligned with the work outcome domain of interest in 57/58 (98%) studies—n/N$:  
- Follow-up ≤ 6 months:  
  - Work status‡: 21/21 (100%)‡1  
  - Sick leave and/or presenteeism: 30/32 (94%)‡1  
- Follow-up > 6–12 months:  
  - Work status‡: 5/21 (24%)‡1  
  - Sick leave and/or presenteeism: 28/38 (74%)‡1 | 2 (8%) | Follow-up was reported as:
3 (12%) | Highly heterogeneous across studies
4 (14%) | Too short to assess work outcomes
21 (87%) | Not done/not described
21 (87%) |
| 4. Use of a comparator | 26/58 (47%) | A comparator was used in 26/58 studies (47%)—n/N$:  
- 21/21 (100%) RCTs with pharmacological intervention
- 1/1 (100%) RCT with non-pharmacological intervention
- 2/13 (15%) OBS with pharmacological intervention
- 2/2 (100%) OBS with non-pharmacological intervention
- 0/21 (0%) OBS on the natural course of the disease | 3 (12%) | Most studies lacked a control group
1 (4%) | Unmatched control groups
21 (87%) |

*Number of systematic literature reviews reporting on the corresponding topic.†Emery et al (2016) have two different time horizons because this is a post hoc analyses of two trials: time horizon of 26 and 24 weeks for the Optimal Protocol for Treatment Initiation with Methotrexate and Adalimumab (OPTIMA) and PRevention Of Work Disability (PROWD) trials, respectively.‡The denominator may vary according to the type of intervention, work outcome of interest or type of study.

Most of the IA studies were on rheumatoid arthritis (RA) (n=33, 57%), followed by axial spondyloarthritis (axSpA) (n=16, 28%) and psoriatic arthritis (PsA) (n=6, 10%). The type of intervention and WP outcome domain for each study is presented in online supplemental table S1, and the corresponding data grouped by type of study.
Tables

Table 2: Methodological choices in the area of ‘work outcome domains and instruments’

<table>
<thead>
<tr>
<th>Topics</th>
<th>Results from studies in inflammatory arthritis (n=58)</th>
<th>Aspects identified by authors of SLRs in other chronic diseases (n=24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Work outcome domains defined</td>
<td>n/N* (%)</td>
<td>n† (%) Comments</td>
</tr>
<tr>
<td></td>
<td>Detailed information</td>
<td>High variability in the definition of a work outcome in included studies precluding data pooling/meta-analysis 103 111 113 114 117 120</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13 (54%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n/N* (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>52/58 (90%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Work status: 12/17 (71%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sick leave: 46/46 (100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presenteeism: 39/40 (98%)</td>
</tr>
<tr>
<td>6. Validated measurement of sick leave</td>
<td>42‡/46 (91%)</td>
<td>Not reported</td>
</tr>
<tr>
<td></td>
<td>Of the studies that had sick leave as work outcome domain, 42 used validated instruments to assess it, n/N*: (%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WPAI: 29/46 (93%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WPS: 5‡/46 (11%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Workdays missed due to IA (long-term sick leave assessed in registries): 9/46 (20%)</td>
</tr>
<tr>
<td>7. Validated measurement of presenteeism</td>
<td>35‡/40 (88%)</td>
<td>WPAL used only in a small number of studies 119</td>
</tr>
<tr>
<td></td>
<td>Of the studies that had presenteeism as work outcome domain, 35 used validated and OMERACT endorsed instruments, n/N*: (%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WPAI: 29/40 (73%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WLQ-25: 2/40 (5%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WPS: 5‡/40 (13%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WALS: 0/40 (0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WA: 0/40 (0%)</td>
</tr>
<tr>
<td>8. Attribution of work participation to overall health</td>
<td>29/58 (50%)</td>
<td>Not reported</td>
</tr>
<tr>
<td></td>
<td>The work outcome domain was assessed in relation to overall health (and not in relation to IA) in 46 studies, n/N*: (%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Work status: 8/17 (47%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sick leave: 23/46 (50%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presenteeism: 15/40 (38%)</td>
</tr>
</tbody>
</table>

Topics | Results from studies in inflammatory arthritis (n=58) | Aspects identified by authors of SLRs in other chronic diseases (n=24) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9. Recall period of self-reported sick leave and presenteeism</td>
<td>35/42 (82%)</td>
<td>1 (4%) WPAL used only in a small number of studies 119</td>
</tr>
<tr>
<td></td>
<td>Detailed information</td>
<td>2 (8%) Inconsistency of the recall period 103</td>
</tr>
<tr>
<td></td>
<td>The recall period of self-reported sick leave was ≤3 months in 34/37 (92%) studies (excluding registries as recall is not applicable) 16 24–26 31–39 42 43 50–56 58 60 64 67 69 71 73</td>
<td>No accounting for a possible recall bias 102 115</td>
</tr>
<tr>
<td></td>
<td>The recall period of self-reported presenteeism was of 7 days to 1 month in 34/40 (85%) studies 24–26 31–39 42 43 50–56 58 60 62 63 67 69 71 73</td>
<td></td>
</tr>
</tbody>
</table>

*The denominator may vary according to the corresponding assessed topic.
†Number of systematic literature reviews reporting on the corresponding topic.
‡Boer et al (2018) used both WPAI-RA and WPS-RA.
IA, inflammatory arthritis; n/N, number of original studies in which the methodological choice was identified/number of studies in which the topic was possible to assess; WALS, Workplace Activity Limitations Scale; WLQ-25, Work Limitations Questionnaire 25-item; WPAI, Work Productivity and Activity Impairment; WPS, Work Productivity Questionnaire.

(RCTs vs longitudinal observational studies) is shown in online supplemental table S2. Work was assessed as a primary outcome in only 26/58 (45%) of the studies, 16 17 21 23 26 27 29–31 41 42 44–47 50 56–58 61 63–64 66 68 71 rarely being the primary outcome in RCTs (n=2/22, 9%). 21 64 The time horizon for the assessment of WP outcomes varied from 24 weeks to 12 years and its distribution, as well as the frequency of assessment by work outcome domain, are both provided in online supplemental table S3.

The general characteristics of included SLRs are presented in online supplemental table S4. Most studies focused on cancer (n=15; 63%), 96 99 101 105–107 111 112 115 116 118–120 followed by stroke (n=3, 13%). 97 113 117 The most frequently

assessed work outcome was ‘return to work after a temporary absence’ (n=12, 50%).

Study design

Table 1 provides an overview of methodological choices in the area of study design. The included population was aligned with the specific work-related study objective in only 16/58 (28%) IA studies,

while the sample size calculation was performed solely in 8 (14%) studies.

Large heterogeneity was observed in the follow-up time of the IA studies, although the majority of studies assessed changes in work status within a follow-up of ≥6 months. Of the five studies assessing changes in work status over an unrealistic short follow-up period ≤6 months,

two also assessed it after 12 months (online supplemental table S3).

The frequency of assessment of sick leave in observational studies (excluding registries, n=8) was longer than 3 months in more than half of the studies (12/20 (60%)) ; however, the other 8/20 (40%) had a frequency of assessment shorter than 3 months hampering correct aggregation into cumulative sick leave.

While all RCTs had a comparator,

only 8/36 (22%) observational studies had one.

The general population, a meaningful benchmark in studies with work as an outcome, was used as a comparator solely in five observational studies.

Regarding SLRs in other chronic diseases, similar issues were reported for all the topics of study design, with the most common flaw being no mentioning of the sample size calculation for work as outcome, as reported in 21/24 (87%) SLRs.

Work outcome domains and instruments

The methodological choices regarding the work outcome domains and instruments are presented in table 2. Among studies in IA, the definition of
Table 4  Methodological choices in the area of ‘reporting of results’

<table>
<thead>
<tr>
<th>Topics</th>
<th>Results from studies in inflammatory arthritis (n=58)</th>
<th>Aspects identified by authors of SLRs in other chronic diseases (n=24)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N/N* (%)</td>
<td>N† (%) Comments</td>
</tr>
<tr>
<td>13. Work-related dropout/loss to follow-up reported</td>
<td>1/58 (2%) Reported loss to follow-up and work-related reasons for dropout</td>
<td>2 (8%) Attrition and its reasons are inconsistently reported: described as well reported in 1 SLR and inadequately reported in the other</td>
</tr>
<tr>
<td>14. The size and characteristics of the (sub)groups analysed are described</td>
<td>58/58 (100%) All studies reported the size and characteristics of the analysed (sub)groups 16-73</td>
<td>1 (4%) No subgroup analysis is performed in included studies 104</td>
</tr>
<tr>
<td>15. Group level and patient level presented</td>
<td>11/58 (19%) Presented both aggregated results (group level) and percentages according to meaningful thresholds (patient level) 19-35 35 42 51 58 62 65 66 71</td>
<td>1 (4%) Lack of patient-level data precluded meta-analysis 104</td>
</tr>
<tr>
<td>16. Volumes of production loss reported in cost studies</td>
<td>21/24‡ (88%) Of the studies reporting productivity costs, 88% provided data on natural volumes (days/hours) used to calculate costs 16 74-76 78-82 84-92</td>
<td>-</td>
</tr>
</tbody>
</table>

*The denominator may vary according to the topic assessed.
†Number of systematic literature reviews in other chronic diseases in which the authors report on the corresponding topic.
‡The denominator corresponds to the studies reporting productivity costs.

‘work status’ was described in two-thirds of studies (71%), and definitions showed large heterogeneity. Sick leave was defined in all studies assessing it, and all but one reported the definition of presenteeism.30

SLRs in other chronic diseases reported high variability in the definition of all WP outcomes in included studies precluding meta-analysis.7 In contrast, the majority of studies in IA assessed sick leave and presenteeism using validated instruments—91% and 88% of studies, respectively. The Work Productivity and Activity Impairment (WPAI) questionnaire was the outcome measurement instrument most frequently used (n=29).18–21

Overall, the work outcome domains’ attribution (to overall health, arthritis or no attribution) was heterogeneous across studies, with sick leave being the domain most frequently assessed in relation to overall health (23/46 (50%) studies).16–21

Reviews in other chronic diseases pointed out inconsistencies of the recall period (varying from 7 days to 7 years).102 103 113 On the contrary, in IA, the recall period of sick leave (excluding registriessince recall is not applicable) was accurate3,8,13 (ie, ≤3 months—figure 1) in 34/37 (92%) studies,16–21 24–26 31–39 42 43 50–56 60 62 63 67–69 71–73 and the recall of presenteeism was reliable and in line with the face validity for patients14 (ie, between 7 days and 4 weeks—figure 1) in 34/40 studies (85%).16–21 24–26 31–39 42–47 50–56 58–60 62–64 69 71–73

Data analysis

Regarding the methodological choices in the area of data analysis (table 3), only 10/53 (19%) IA studies reported skewness of sick leave and/or presenteeism and accounted for the skewness in the analyses.20 22 45–47 56 58 63 65 66 70 72

Also, only 8/22 (36%) RCTS19 25 36 37 49 52 67 69 and 17/36 (47%) observational studies16 20 23 27 44 46 50 56–58 61 63 65 66 70 72 took contextual factors into account, most frequently demographic factors, such as age and gender, while other specific work-related contextual factors (eg, nature of work and workplace support) were less frequently accounted for.27 41 50 58 63 70 72 SLRs in other chronic diseases reported that adjustment for contextual factors/confounders in the included studies, if any, was performed for very few factors.101 105 107 110 112–117 120

The majority of studies in IA (n=49/52, 94%) took interdependence between work outcomes into account acknowledging that (1) data (over time) on sick leave are less meaningful without information on the proportion of persons employed (over time) in that specific population (sick leave cannot happen if the person is not employed) and/or (2) assessing presenteeism is less meaningful if information on sick leave is not provided (eg, presenteeism cannot happen on days a person is absent due to sick leave).18–22 24–27 29 31–39 42–47 50–56 58–60 62–64 69 71–73
Reporting of results
The methodological choices in IA studies as well as the issues raised in SLRs in other chronic diseases regarding the area of reporting are described in Table 4.

The reporting loss to follow-up and the work-related reasons for drop-out were often neglected in IA studies, being reported in only one study. In other chronic diseases, this was also inconsistently reported.

All IA studies reported the size and characteristics of the (sub)groups analysed.

In IA studies, the choice on how to report study findings was heterogeneous, with only 11/58 (16%) studies presenting both aggregated results (mean/median) and percentages according to meaningful thresholds. This was also outlined by the SLRs in other chronic diseases where the lack of patient-level data was a barrier to study pooling and meta-analysis.

Data on natural volumes (days/hours) used to calculate costs was presented in the majority of the studies reporting productivity costs (21/24, 88%).

DISCUSSION
WP has been a frequently assessed endpoint in IA studies over the past 5 years; however, these studies revealed a high methodological heterogeneity and a number of important flaws. Several issues were detected in the areas of study design, work outcome definition and assessment, as well as in the analysis and reporting of the results. Review of SLRs in other chronic diseases revealed that observed methodological issues are not rheumatology specific as these are also common in studies of work outcomes in other clinical fields.

Different WP outcomes of interest apply to specific subpopulations (eg, employed/employable people) and need to be assessed in a sufficiently large group over a certain timeframe. Notwithstanding, this was often neglected, particularly when WP was not the primary outcome as occurred in the majority of RCTs. Thus, the studied population, the intermediate assessment time-points and overall follow-up time were tailored on the primary outcomes, hampering the power to detect statistically significant effects on WP outcomes and leading to follow-up times not adequate for some of the WP outcomes of interest. Moreover, even in RCTs with long-term extensions, WP outcome domains were not assessed across the extension study period as other outcomes. This pose particular challenges in studies aiming to understand the impact of an intervention on long-term employment, work disability or prolonged sick leave (eg, assessing costs of productivity loss), as having a time horizon of 6 months is not adequate. Remarkably, also studies with WP as the primary outcome had important flaws in this area, for example, the sample size calculation was often not reported.

Careful choice of which WP outcome to assess and which measurement instrument to use is of paramount importance, particularly when dealing with a comparison of interventions. As far as the definition of employment and work disability is concerned, clinical studies might want to align with definitions that are relevant for their administrative entities (eg, countries, regions, states, etc), thus likely contributing to heterogeneity in work status definitions as found in IA studies. In contrast, presenteeism and sick leave were often described in line with the frequent use of validated instruments (eg, WPAI) that include an appropriate definition for the work outcome domain. In this regard, stakeholders should strive to harmonise worldwide comparable and locally applicable definitions along with endorsing specific outcome measurement instruments, for example, as OMERACT is doing for presenteeism. Two other important methodological aspects, namely, disease attribution and recall, are relevant but not (yet) encompassed by the OMERACT framework. Regarding disease attribution, only half of the studies assessed the WP outcome domain in relation to overall health (more meaningful for benchmarking with the general population). This may be problematic since it is well established that patients have difficulties in distinguishing which restrictions can be attributable to IA, other specific health problems (eg, osteoarthritis) or overall health. Inconsistency of the recall period was often reported in studies of other chronic diseases, however less evident in IA studies. This is likely due to the widespread use of validated instruments such as WPAI (past 7 days recall) and the Work Productivity Survey (WPS; past month recall) in the field of rheumatology. WP, as any outcome, is subject to the effect of a number of variables, related to either the disease, the social environment or other aspects, which require to be considered in order to reliably assess the net change of the outcome. Contextual factors, defined by OMERACT, from a statistical viewpoint, as a “variable that is not an outcome of the study but needs to be recognized (and measured) to understand the study results”, include potential confounders and effect modifiers (https://omeract.org/handbook-resources/). The characterisation of core contextual factors (ie, when do they really matter to influence practice) remains a challenge, partially because the influence of most contextual factors tends to vary according to the setting. The International Classification of Functioning, Disability and Health (ICF) provided, in addition to the bio-psycho-social framework, also a classification distinguishing personal and environmental factors, and this was the basis for a further grouping of contextual factors relevant for WP by the OMERACT work productivity group. Lack of accounting for contextual factors was common in IA and often reported also by SLRs in other chronic disease. Work-related contextual factors such as job type, adaptations at work and more personal aspects such as ability to cope and satisfaction were often neglected. This emphasises the urgent need of action for improving and implementing feasible strategies to account for relevant work-related contextual factors.
Other methodological issues pertain to how data are analysed and reported. WP presents a continuum of subdomains which are (hierarchically) dependent on each other and/or can compete over time. The majority of studies assessing sick leave and presenteeism took interdependence between work outcomes into account, encompassing the widespread use of the WPAI, which already considers interdependence of sick leave and presenteeism (overall work impairment). SLRs in other chronic diseases reported that despite using the correct instrument (eg, WPAI), the studies frequently neglected some important subdomains. Indeed, to account for interdependence, WPAI must be comprehensively used, that is, assessing both presenteeism and sick leave plus the overall work impairment. Yet, consensus is needed on how to deal with such dependencies when instruments other than WPAI are used. It is known that distribution of presenteeism, and especially sick leave, may often be highly skewed (even zero inflated). Not accounting for this, as we observed in the majority of studies, may affect the robustness of conclusions.

Furthermore, drop-out may be related to underlying work context and thus not be at random, so the rates and reason for drop-out should be carefully considered to ensure a correct interpretation of the impact of IA on WP outcomes overtime. However, these were not reported in the majority of studies. Likewise, to enhance the insight into WP outcomes and to ensure more transparent interpretation of the differences between interventions, the mean and median values of sick leave or presenteeism and also the proportion of patients attaining a specific meaningful (change in) outcome are advisable to report. In IA studies, the choice on how to report data on work outcome domains was heterogeneous, with only 19% of studies presenting both aggregated results and percentages according to meaningful thresholds. Choice of thresholds was not uniform across studies, highlighting the needs for consensus in this respect.

This review has some limitations. Although we used a sensitive approach to identify studies with WP as an outcome domain in IA as well as SLRs in other chronic diseases, we cannot be sure that some relevant studies were missed. While retrieving data from SLRs in other chronic diseases, only the reported issues were collected, as going through the primary studies was beyond the scope. This may have resulted in missing some relevant methodological aspects not captured by the SLR authors. The exclusion of studies due to feasibility reasons implies that our summary is generalisable to issues found in recent studies.

In conclusion, a high methodological heterogeneity and important flaws were detected among the included studies in the main areas of study design, work outcome definition and assessment, analysis and reporting of results. This SLR alerts for the need of implementation of minimum quality standards around these key methodological aspects to homogenise and improve the quality of future studies in IA and likely in other chronic diseases. This review informs the EULAR Points to Consider for the conduction, analysis and reporting of studies with work as an outcome domain in IA.

**REFERENCES**


38 Olofsen T, Petersson IF, Eriksson JK, et al. Predictors of work disability in the first 3 years after diagnosis in a


63 van Lunteren M, Ez-Zaitouni Z, Fongen G, et al. Disease activity decrease is associated with improved work productivity over 1 year in early axial spondyloarthritis (SpA) patients (SpAFlyers Cohort Early cohort). Rheumatology 2017;56:2222–8.


87 Tanaka Y, Yamazaki K, Nakajima R, et al. Economic impact of adalimumab treatment in Japanese patients with rheumatoid arthritis from the adalimumab non-interventional trial for up-


