Determining the Best Discriminatory Physical Functioning Outcome Measurement Instrument for Psoriatic Arthritis Trials: A Meta-Epidemiological Study

Ying-Ying Leung, Tobias Haugegaard, Tommy Kok Annfeldt, Richard Holland, Vibeke Strand, Philip Mease, Peter Tugwell, George A. Wells, Beverley J. Shea, Ashish J Mathew, Niti Goel, Christine Lindsay, Alexis Ogdie, Ana-Maria Orbai, Laura C Coates, Dafna D Gladman, William Tillett, Jeffrey Chau, and Robin Christensen.

Affiliations (emails)

YYL: Singapore General Hospital, Singapore; Duke-NUS Medical School, Singapore (katyccc@hotmail.com).

TH: Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital; & Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark (tobias.hoerlueck@regionh.dk).

TKA: Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital. Denmark (tommy.kok.annfeldt@regionh.dk).

RH: Concord Repatriation General Hospital, Sydney, Australia

VS: Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto, California, USA

PM: Rheumatology Research, Swedish Medical Center/Providence St. Joseph Health and University of Washington School of Medicine, Seattle, Washington, USA.

PT: Division of Rheumatology, Department of Medicine, School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada

GAW: School of Epidemiology and Public Health, University of Ottawa Heart Institute, University of Ottawa, Ottawa, Canada

BJS: Ottawa Hospital Research Institute, University of Ottawa, Canada

AJM: Department of Clinical Immunology & Rheumatology, Christian Medical College, Vellore, India

NG: Patient Research Partner, Adjunct Assistant Professor, Duke University School of Medicine, Durham, North Carolina, USA

CL: Patient Research Partner, Employed by Aurinia Pharma US Inc., Prosper, Texas USA

AO: Medicine and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

AMO: Division of Rheumatology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

LC: National Institute for Health Research Research Professor, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

DDG: Medicine, University of Toronto, Senior Scientist, Schroeder Arthritis Institute, Krembil Research Institute, Co-Director, Gladman Krembil Psoriatic Arthritis Program, University Health Network, Toronto Western Hospital, Toronto, Ontario, Canada

WT: Department of Life Sciences, Centre for therapeutic Innovation, University of Bath &Royal National Hospital for Rheumatic Diseases, University of Bath, Bath, United Kingdom

JC: Patient Research Partner, Hong Kong

RC: [1]Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen & [2]Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark (Robin.Christensen@regionh.dk)

Corresponding author: Ying-Ying Leung, MD; Department of Rheumatology and Immunology, Singapore General Hospital, The Academia, Level 4, 20 College Road, Singapore 169856, Contact No.: +65 63265276, Fax no.: +65 62203321, E-mail: katyccc@hotmail.com; gmsleung@nus.edu.sg

ABSTRACT

Objectives: To empirically compare the discriminant capacities of three outcome measurement instruments for assessment of physical functioning for psoriatic arthritis (PsA): HAQ-DI, SF36-PF and SF36-PCS.

Methods: We applied a network meta-analysis technique in a sample of randomized trials (RCTs) for PsA. For randomized comparison, we calculated net effect size estimates for each outcome measurement instrument using standardized mean differences (SMDs); positive values indicated a beneficial effect of the intervention compared to the control groups. We analyzed the differences between outcome measurement instruments at the trial level by applying a multiple-treatment meta-analysis to compare the SMDs within and across randomized comparisons for each outcome measurement instrument.

Results: From 42 articles (31 RCTs), 57, 18, and 18 randomized comparisons enabled a direct comparison between HAQ-DI and SF36-PCS (difference in SMDs: 0.057, 95% confidence interval, CI: 0.003 to 0.110), SF36-PF and SF36-PCS (difference in SMDs: 0.101, 95% CI: 0.018 to 0.184); and HAQ-DI and SF36-PF (difference in SMDs: -0.059, 95% CI: -0.142 to 0.024), respectively. The network meta-analysis technique confirmed that both HAQ-DI and SF36-PF were more responsive to change than SF36-PCS, with differences between SMDs of 0.057 (95% CI: 0.003 to 0.110) and 0.109 (95% CI: 0.032 to 0.185), respectively. No difference in discriminatory capacity between HAQ-DI and SF36-PF was noted.

Conclusions: HAQ-DI and SF-36-PF were equally responsive to change and superior to SF36-PCS in PsA RCTs. We illustrated a new method for quantitative comparison of the performance of different outcome measurement instruments for a particular domain.

Systematic review registration number: Registered on PROSPERO: CRD42019129557

INTRODUCTION

When selecting an outcome measurement instrument for a randomized controlled trial (RCT), evidence is needed to ensure that the instrument will be capable of truthfully measuring change, i.e., responsiveness. More importantly, it should be able to distinguish the change in one treatment arm (assuming the treatment works) relative to the change in a comparator arm. Most recently, RCTs for psoriatic arthritis (PsA) include an active comparator (in addition to placebo) to the new intervention, often head-to-head against an approved biological (b-) disease modifying antirheumatic drug (DMARD), such as a TNF inhibitor (TNFi). These RCTs require the outcome measurement instruments to have even better discriminatory capacity such that an experimental intervention could demonstrate superiority not only against placebo comparator but also an active comparator. Building an argument that the scores of a given outcome measurement instrument can make that distinction in the target population is essential in instrument selection for trials.

Ideally, all candidate outcome measurement instruments would be studied simultaneously in RCTs evaluating the same target populations and interventions, particularly for those in the core domain set [1]. However, that is neither practical nor feasible. As a result, tradition and common practice often drive the instrument selection. Another issue is that when seeking evidence for clinical trial performance of an outcome measurement instrument, there is often no indication of what was expected in that trial for a “good” or “perfect” outcome measurement instrument evaluating a domain. If the trial’s experimental intervention(s) did not demonstrate superiority over the comparator group in a domain of interest utilizing the outcome measurement instrument, it is uncertain if the outcome measurement instrument was not sensitive enough or if it actually performed as intended to show the lack of difference in efficacy of the intervention.

PsA is a chronic inflammatory disease involving peripheral arthritis, enthesitis, dactylitis, axial involvement, psoriasis, and nail disease. It has a profound impact on the psychological, social, and physical well-being of the affected people [2]. Physical functioning is a key concern from both patients' and doctors' perspectives, and it is recognized as one of the core domains to be measured in every RCT and longitudinal observational studies [1,3]. There have been several patient-reported outcome measurement instrument that assess physical functioning in PsA [4], with the most commonly used in PsA RCTs being the Health Assessment Questionnaire Disability Index (HAQ-DI) [5] and the physical functioning domain within the Medical Outcomes Study 36-item Short Form Survey (SF36-PF) [6]. The SF-36 physical component summary score (SF36-PCS) also has been commonly reported in RCTs but does not match the physical function domain well [7]. The GRAPPA community has prioritized 6 outcome measurement instruments for physical functioning [7]. Using a qualitative synthesis based on a rigorous systematic review protocol evaluating outcome measurement instrument in PsA, we concluded that clinical trial discrimination was supported for HAQ-DI and SF36-PCS in PsA with a low risk of bias. In contrast, the SF36-PF was supported but with some concerns due to fewer available trials reporting data [6]. As part of the concerted effort of the Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) - Outcome Measures in Rheumatology (OMERACT) initiative to standardize the outcome measurement set for PsA [8], two outcome measurement instruments, namely the HAQ-DI and SF36-PF were provisionally endorsed by both GRAPPA and OMERACT organizations [9,10]. A research agenda was set to evaluate the clinical trial discrimination of these instruments quantitatively. Therefore, in this study, we aimed to quantitatively compare the sensitivity of these three outcome measurement instruments via a network meta-analysis in a representative sample of RCTs for PsA.

METHODS

Literature search and eligibility of articles

We used the data from our previous systematic review through 21 March 2019 using PubMed and Scopus [6], registered with the International Prospective Register of Systematic Reviews (PROSPERO) prior to initiation (CRD42019129557). The search aimed to identify all original full-text articles of RCTs conducted in PsA with published data in the English language. We conducted an update using the PubMed search terms on 28 March 2022 to identify new RCTs published since the last search date. The reporting of this systematic review adhered to the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) Statement [11].

Study selection

The eligibility criteria were published previously [6], and applied in the updated search. In brief, we included articles on RCTs conducted in PsA only if they reported results on any two out of the three outcome measurement instruments. If the RCT was conducted in a mixed arthritis population, we only included those with separate subgroup analyses for PsA. To reduce the possible risk of performance bias interfering with the outcome measurement instruments assessments, we excluded RCTs that did not apply a “double-blind” study design. We also limited the review to RCTs involving b-DMARDs and targeted synthetic (ts-) DMARDs. For RCTs with multiple publications, only one article from each unique RCT was included unless a publication provided new information, such as a subgroup analysis. The same researcher (YYL) conducted the initial and updated search and removed duplicates from the searches from two databases. The same researchers (YYL, RH) independently screened the titles, abstracts, and full text (if appropriate) for eligibility in all searches. Disputes were resolved by consensus.

Quality Assessment of included studies

We assessed the clinical trial discrimination of each outcome measurement instrument in each article using the OMERACT good method checklist [12,13]. Two researchers worked in pairs (YYL and RH; YYL and AJM; YYL and NG; YYL and CL) appraised each instrument in each article. The assessment form is illustrated in Supplementary Table 1. The quality of each article was rated in traffic light format as Green (good to go) or Amber (some caution, can be used) indicating good methods, Red indicating a high risk of bias and White as no data. Only studies that passed the Good Methods Check (ie. Green or Amber ratings) were included for data synthesis.

Extraction of data and data synthesis

Researchers collaborated in pairs (YYL and RH; YYL and AJM; YYL and NG; YYL and CL) to independently extract data for the three outcome measurement instruments at baseline and at the primary endpoint assessment (or end of the blinded controlled trial period) using a standardized format. Any discrepancies between researcher pairs were resolved through consensus. We only included articles that had data for at least two of the candidate outcome measurement instruments. The values measured by the outcome measurement instruments were standardized in two steps. The first step was to ensure that the direction of the scales did not differ between outcome measurement instruments. This was done by inversing the HAQ-DI scores, so a positive mean difference in HAQ-DI meant that the intervention group had better health than the comparison group, like SF36-PCS and SF36-PF. The second step was to calculate the Standardized Mean Difference (SMD, the ratio of the treatment difference to the pooled standard deviation of the mean change scores) for all outcome measurement instruments. This ensured comparability between outcome measurement instruments.

To compare the true performance of the outcome measurement instruments, a reference arm was assigned within each trial for comparison with the active intervention. For three - group - trials evaluating the efficacy of a b-DMARD against both placebo and TNFi included as comparison arms, the TNFi was also

assigned as an intervention arm, and the placebo was assigned as the reference arm. For trials comparing a combination of therapies versus a single agent [14], the monotherapy arm was assigned as the reference arm. For head-to-head studies comparing a new b- or ts-DMARD compared to TNFi with no placebo arm [15], TNFi was assigned as the reference arm. The details of the assignment of intervention and reference for each trial are shown in Supplementary Table 2. RCTs including two or more doses of the experimental intervention against one reference arm were treated as individual comparisons and presented as two (or more) rows with unique trial identifiers in a forest plot, one for each dose relative to the reference. In RCTs with two or more experimental interventions, the number of participants in the reference arm was divided by the number of active arms, treating each as an independent estimate. This approach was employed to avoid double-counting of patients and resulted in larger standard errors leading to more conservative precision estimates.

The pooled difference between intervention and comparison groups, expressed in SMDs with 95% confidence intervals (CIs), was presented in a forest plot for each outcome measurement instrument. Between-study inconsistency was assessed using the I^2 statistic, describing the percentage of total variation across trials attributable to heterogeneity rather than to chance for each outcome measurement instrument [16]. Within the framework of mixed effects models, we employed randomized comparisons as random effects by default, while trial identifiers and the name(s) of the outcome measurement instruments were applied as fixed-effect factors.

Network Meta-Analysis

To evaluate the relative responsiveness of the outcome measurement instruments, we applied a multiple-treatment meta-analyses approach, treating each outcome measurement instrument as an intervention and the difference in SMDs between outcome measurement instruments as the relative effect of one treatment over another. The analytical approach was a frequentist network

meta-analysis, with the network comprising all three outcome measurement instruments. The analysis results consisted of both direct and indirect evidence for relative responsiveness between outcome measurement instruments. The direct evidence stemmed from the available data, which was depicted in the forest plot, and the indirect evidence was estimated by mathematical combinations of the direct comparisons whenever a direct comparison of outcome measurement instruments was not available. Thus, the result of the network meta-analysis consisted of the combined direct and indirect evidence. Consistency of the network was assessed by the Tau^2 and the I^2. Consistency of the results was further assessed by presenting the direct evidence alongside the combined Network Meta-Analysis estimates. All estimates were derived from frequentist random-effects models and reported with 95% confidence intervals.

During peer review, a concern was raised about the lack of exploration of publication bias. Although this is not a standard intervention review, we explored publication bias using a funnel plot. All statistical analyses were performed using the statistical software SAS Studio and R V.3.2.3 (R Foundation for Statistical Computing) with the packages ‘meta’ and ‘metafor’.

RESULTS

We identified 77 new articles from the updated search and five from cross-reference checks of which 20 fulfilled the eligibility criteria. Together with the 33 articles from the previous search, 53 articles were included for data extraction. Eleven articles were excluded for not having data for at least two outcome measurement instruments. Finally, 42 articles from 31 RCTs were included in this analysis (Figure 1). The characteristics of the included studies and their study population are summarized in Supplementary Table 2. Quality assessments for clinical trial discrimination for each article were conducted using the OMERACT Good Method Checklist for RCT discrimination, and results are shown as Green, Amber, Red or White (Supplementary Table 2).

Figure 2 highlights the direct – head-to-head - comparisons available from the data with partially empty plots across trial comparisons indicating less frequently used or reported measures in the literature. There were 57 randomized comparisons that facilitated a direct comparison between HAQ-DI and SF36-PCS, 18 comparisons enabling a direct comparison between SF36-PF and SF36-PCS, and 18 comparisons allowing a direct comparison between HAQ-DI and SF36-PF. Figure 2 also shows the results of an individual meta-analysis for the three outcome measurement instruments, highlighting the treatment effect of each outcome measurement instrument in each study. The pooled SMDs for HAQ-DI, SF36-PCS score and SF36-PF domain were 0.444 (95% CI:0.383 to 0.504), 0.498 (95% CI: 0.437 to 0.559) and 0.393 (95% CI:0.314 to 0.473), respectively.

The results from the network meta-analysis showed that both SF36-PF and HAQ-DI were more responsive to change than SF36-PCS score, with a difference of 0.109 (95% CI: 0.032 to 0.185) and 0.057 (95% CI:0.003 to 0.110), respectively (Table 1). The findings suggested no difference in discriminatory capacity between HAQ-DI and SF36-PF, with a difference of -0.052 (95% CI: -0.128 to 0.024). The direct evidence of the network meta-analysis confirmed the robustness of the results with a difference between HAQ-DI and SF36-PCS of 0.057 (95% CI: 0.003 to 0.110) in favor of HAQ-DI, a difference between SF36-PF domain and SF36-PCS of 0.101 (95% CI: 0.018 to 0.184) in favor of SF36-PF, and no apparent difference between HAQ-DI and SF36-PF (difference: -0.059, 95% CI: -0.142 to 0.024). The findings were robust, as no inconsistency was detected between the direct and network meta-analysis estimates. The network meta-analysis confirmed that SF36-PF and HAQ-DI were more responsive to change than SF36-PCS, and no significant difference in discriminatory capacity was observed between HAQ-DI and SF36-PF. Direct evidence further supported these results, showing consistent differences between HAQ-DI, SF36-PF, and SF36-PCS.

In supplementary analyses, we present a comparison-adjusted funnel plot to assess potential publication bias or small-study effects in the network meta-

analysis. The plot displays effect estimates on the x-axis and their standard errors on the y-axis, which showed symmetrical distribution for all outcome measurement instruments (Supplementary Figure 1), providing indirect support that the summary estimates are unlikely to be distorted by small-study bias.

DISCUSSION

In this meta-epidemiological study, we employed network meta-analysis methods to evaluate the outcome measurement instruments that best distinguish interventions from reference comparators in PsA trials. We found that both HAQ-DI and SF36-PF were more responsive to change than SF36-PCS. The network meta-analysis confirmed that SF36-PF and HAQ-DI were more responsive to change than SF36-PCS, and no significant difference in discriminatory capacity was observed between HAQ-DI and SF36-PF. Direct evidence further supported these results, showing consistent differences between HAQ-DI, SF36-PF, and SF36-PCS.

Network meta-analyses have been commonly used to compare the relative efficacies of drugs or interventions across different RCTs. Nonetheless, this is the first attempt to use it to show comparative discrimination of outcome measurement instruments across RCTs. The results found from network meta-analysis were consistent with traditional meta-analysis in evaluation for responsiveness of outcome measurement instruments in a representative sample of RCTs in PsA, illustrating the robustness of these methods. The methods we used here provided a complementary and quantitative comparison across outcome measurement instruments, in addition to the current OMERACT filter 2.2 framework that gives a qualitative assessment of RCT discrimination of individual outcome measurement instruments [17]. The current study bridges that gap by providing quantitative comparisons between outcome measurement instruments for a domain. It also demonstrated a new method that allows comparison of the performance of different outcome measurement instruments in a single platform. This informs researchers on choosing an

effective outcome measurement instrument to use in clinical trials for a particular domain.

For the physical functioning domain in PsA, both HAQ-DI and SF36-PF were shown to have similar performance in RCT discrimination qualitatively [6] and were provisionally endorsed [10]. Although the current results did not show significant differences between these two outcome measurement instruments, we showed that HAQ-DI and SF36-PF were superior in responsiveness than SF36-PCS. From our previous work, both doctors and patients have concluded that SF36-PCS did not match its intended use as an instrument to measure physical functioning (lacked domain match)[10]. Unlike the SF36-PF that is one of the eight domains for HRQoL of SF36, the two summary components of SF36, namely the SF36-PCS and mental component summary scores should be more correctly conceptualized as two components of health related quality of life [18]. Therefore, SF36-PCS would be considered as part of the measurement of HRQoL rather than an outcome measurement instrument of physical functioning in PsA despite commonly reported in RCTs for PsA.

The strength of our study is the novelty of utilizing network meta-analysis methods to compare the responsiveness of outcome measurement instruments. These methods can be used for other outcome measurement instruments, not just for the physical functioning instruments utilized in PsA RCTs. This analysis paves the way for comparing the relative responsiveness of outcome measurement instruments in RCTs to facilitate researchers in choosing appropriate instruments for a domain. While we have included a representative sample of RCTs in the modern standard, we acknowledge that articles with inadequate or incomplete data for at least tw0 outcome measurement instruments could not be included in the traditional meta-analysis, and the framework of network meta-analysis has facilitated the inclusion of these studies. The results of network meta-analysis were robust with demonstrated consistency between the results of the network meta-analysis and the direct evidence. Although responsiveness of instruments could be influenced by variability in study population, treatment effects, follow-up

durations, our use of standardized differences which served as an internal comparison of the instruments' responsiveness of treatment arm over placebo arm within each study, helped to minimize the influence of heterogeneity in the comparisons. These comparisons for outcome measurement instruments for the physical functioning domain for PsA were limited to the three candidates we have chosen and not others. This study sought to compare the discriminative ability (using standardized means as surrogate) across different outcome measurement instruments for physical function, it would not provide information on comparison of other measurement properties of these instruments. Particularly, it does not provide data on proportion of patients with improvement more than a certain threshold of meaning such as minimally clinically important difference between treatment arms. Finally, the last literature search was concluded in 2022. Although we may have missed the most recent studies, the included studies span a board range of modern b-/-ts DMARDs. Importantly, our focus was on the discriminatory ability of well-established measures, which are unlikely to be materially impacted by newly published studies. Furthermore, the large number of included studies should render this analysis of outcome measurement instruments robust.

In conclusion, the outcome measurement instruments, HAQ-DI and SF36-PF, endorsed by GRAPPA and OMERCT to measure the physical functioning domain in PsA based on qualitative analysis, have shown similar performance in RCT discrimination quantitatively as well. We have illustrated a new method for quantitative comparison of the performance of different outcome measurement instruments for a particular domain.

Table 1. Comparative responsiveness of all three measures via network meta-analysis.

Outcome measurement instrument Comparisons	**Network meta-analysis**		**Direct evidence**		**Consistency**
	Difference	95% CI	Difference	95% CI	
HAQ-DI vs. SF36-PCS	0.057	0.003 to 0.110	0.057	0.003 to 0.110	Yes
HAQ-DI vs. SF36-PF	-0.052	-0.128 to 0.024	-0.059	-0.142 to 0.024	Yes
SF36-PF vs. SF36-PCS	0.109	0.032 to 0.185	0.101	0.018 to 0.184	Yes

CI: confidence Intervals; HAQ-DI: Health Assessment Questionnaire-Disability Index; SF36-PCS: Medical Outcomes Study 36-item Short Form Survey Physical Component Summary; SF36-PF: Medical Outcomes Study 36-item Short Form Survey Physical Functioning Domain.

Table 1. Comparative responsiveness of all three measures via network meta-analysis.

Figure 1. PRISMA Flow diagram outlining the process of identifying, screening, and including studies in a systematic review or meta-analysis.

Figure 2. Forest plots of the three individual meta-analyses for HAQ-DI, SF36-PF and SF36-PCS.

Supplementary Table 1. OMERACT Good Method Check List for RCT discrimination

Supplementary Table 2. Characteristics and quality assessment of 42 included articles from 31 RCTs

Supplementary Figure 1. Comparison-adjusted funnel plot for HAQ-DI, SF36-PF and SF36-PCS.

Acknowledgement

We thank the Parker Institute, Bispebjerg and Frederiksberg Hospital for the possibility of being a part of this research project.

Funding

This research received no grant from any funding agency in the public, commercial, or not-for-profit sectors. Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital is supported by a core grant from the Oak Foundation (OCAY-18-774-OFIL).

Competing interests

None of the authors have any financial competing interests. The Parker Institute is grateful for the financial support received from public and private foundations, companies, and private individuals.

References

1. Orbai AM, de Wit M, Mease P, et al. International patient and physician consensus on a psoriatic arthritis core outcome set for clinical trials. Annals of the rheumatic diseases 2017;76:673-80.
2. FitzGerald O, Ogdie A, Chandran V, et al. Psoriatic arthritis. Nature Reviews Disease Primers 2021;7:1-17.
3. Dures E, Hewlett S, Lord J, et al. Important Treatment Outcomes for Patients with Psoriatic Arthritis: A Multisite Qualitative Study. Patient 2017;10:455-62.
4. Hojgaard P, Klokker L, Orbai AM, et al. A systematic review of measurement properties of patient reported outcome measures in psoriatic arthritis: A GRAPPA-OMERACT initiative. Semin Arthritis Rheum 2018;47:654-65.
5. Mease P, Strand V, Gladman D. Functional impairment measurement in psoriatic arthritis: Importance and challenges. Semin Arthritis Rheum 2018;48:436-48.
6. Leung YY, Holland R, Mathew AJ, et al. Clinical trial discrimination of physical function instruments for psoriatic arthritis: A systematic review. Semin Arthritis Rheum 2020;50:1158-81.
7. Leung YY, Orbai AM, Ogdie A, et al. Appraisal of Candidate Instruments for Assessment of the Physical Function Domain in Patients with Psoriatic Arthritis. J Rheumatol 2021;48:58-66.
8. Tillett W, Orbai AM, Ogdie A, et al. GRAPPA-OMERACT initiative to standardise outcomes in psoriatic arthritis clinical trials and longitudinal observational studies. Ann Rheum Dis 2018;77:e23.
9. Leung YY, Orbai AM, Tillett W, et al. Instruments Measuring Physical Function for Psoriatic Arthritis Endorsed at GRAPPA 2020 Annual Meeting: Updates of the GRAPPA-OMERACT Working Group. J Rheumatol 2021.
10. Leung YY, Orbai AM, Hojgaard P, et al. OMERACT Filter 2.1 instrument selection for physical function domain in psoriatic arthritis: Provisional endorsement for HAQ-DI and SF-36 PF. Semin Arthritis Rheum 2021;51:1117-24.
11. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021;372:n160.
12. Beaton DE, Maxwell LJ, Shea BJ, et al. Instrument Selection Using the OMERACT Filter 2.1: The OMERACT Methodology. The Journal of Rheumatology 2019;46:1028-35.
13. Boers M, Kirwan JR, Tugwell P, et al. The OMERACT Handbook. [Internet. Accessed Available from: https://omeract.org/wp-content/uploads/2021/12/OMERACT-Handbook-Chapter-5_Final_June-2-2021_a.pdf.
14. Mease PJ, Gladman DD, Collier DH, et al. Etanercept and Methotrexate as Monotherapy or in Combination for Psoriatic Arthritis: Primary Results From a Randomized, Controlled Phase III Trial. Arthritis Rheumatol 2019;71:1112-24.
15. McInnes IB, Behrens F, Mease PJ, et al. Secukinumab versus adalimumab for treatment of active psoriatic arthritis (EXCEED): a double-blind, parallel-group, randomised, active-controlled, phase 3b trial. Lancet 2020;395:1496-505.
16. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539-58.
17. Maxwell LJ, Beaton DE, Boers M, et al. The evolution of instrument selection for inclusion in core outcome sets at OMERACT: Filter 2.2. Semin Arthritis Rheum 2021;51:1320-30.
18. Ware JE, Kosinski M. Interpreting SF&-36 summary health measures: A response. Quality of Life Research 2001;10:405-13; discussion 15-20.