

# **Clinical trial discrimination of physical function instruments for psoriatic arthritis: a systematic review**

Ying Ying Leung<sup>1</sup>, Richard Holland<sup>2</sup> Ashish J Mathew <sup>3,4,5</sup>, Chris Lindsay<sup>6</sup>, Niti Goel<sup>7</sup>, Alexis Odgie<sup>8</sup>, Ana-Maria Orbai<sup>9</sup>, Pil Hojgaard<sup>10</sup>, Jeffrey Chau<sup>11</sup>, Laura C Coates<sup>12</sup>, Vibeke Strand<sup>13</sup>, Dafna D Gladman<sup>14</sup>, Robin Christensen<sup>15</sup>, William Tillet<sup>16</sup>, Philip Mease<sup>17</sup>

1. Singapore General Hospital, Duke-NUS Medical School, Singapore
2. Concord Repatriation General Hospital, Sydney, Australia
3. Centre for Prognosis Studies in Rheumatic Diseases, Division of Rheumatology, Department of Medicine, University of Toronto, Toronto, Ontario Canada
4. Department of Clinical Immunology & Rheumatology, Christian Medical College, Vellore, India
5. The Copenhagen Center for Arthritis Research, Rigshospitalet Glostrup, University of Copenhagen, Copenhagen, Denmark
6. Patient research partner, Prosper, Texas USA
7. Patient Research Partner, Adjunct Assistant Professor, Duke University School of Medicine, Durham, North Carolina, USA
8. Medicine and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
9. Division of Rheumatology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
10. Copenhagen Lupus and Vasculitis Clinic, Center for Rheumatology and Spine Diseases, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark.
11. Patient Research Partner

12. National Institute for Health Research Clinician Scientist, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom
13. Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto, California, USA
14. Medicine, University of Toronto, Senior Scientist, Krembil Research Institute, Director, Psoriatic Arthritis Program, University Health Network, Toronto Western Hospital, Toronto, Ontario, Canada
15. Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital, University of Copenhagen, Copenhagen & Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark.
16. Royal National Hospital for Rheumatic Diseases, University of Bath, Bath, United Kingdom
17. Rheumatology Research, Swedish Medical Center and University of Washington School of Medicine, Seattle, Washington, USA

**Correspondence to:** 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](mailto:katyccc@hotmail.com)

**Key words:** Systematic Review, psychometric, physical function, psoriatic arthritis

## ABSTRACT

**Objectives.** Physical function (PF) is a core domain to be measured in randomized controlled trials (RCTs) of psoriatic arthritis (PsA), yet the discriminative performance of patient reported outcome measures (PROMs) for PF in RCTs has not been evaluated systematically. In this systematic review, we aimed to evaluate the clinical trial discrimination of PF-PROMs in PsA RCTs.

**Methods.** We searched PubMed and Scopus databases in English to identify all original RCTs on biological and targeted synthetic disease modifying anti-rheumatic drugs (DMARDs) conducted in PsA. We assessed quality in each article using the OMERACT good method checklist. Effect sizes (ES) for the PF-PROMs were calculated and appraised using *a priori* hypotheses. Evidence supporting clinical trial discrimination for each PF-PROM was summarized to derive recommendations.

**Results.** 35 articles from 31 RCTs were included. Four PF-PROMs had data for evaluation: HAQ-Disability Index (DI), HAQ-Spondyloarthritis (S), and Short Form 36-item Health Survey Physical Component Summary (SF-36 PCS) and Physical Functioning domain (SF-36 PF). As anticipated, higher ES values were observed for intervention groups than the control groups. Across all studies, for HAQ-DI, the median ES were -0.73 and -0.24 for intervention and control groups, respectively. Whereas for SF-36 PCS, the median ES were 0.77 and 0.23. for intervention and control groups, respectively.

**Conclusion.** Clinical trial discrimination was supported for HAQ-DI and SF-36 PCS in PsA with low risk of bias; and for SF-36 PF and HAQ-S with some caution. More studies are required for HAQ-S.

**Keywords:** Psoriatic arthritis, physical function, responsiveness, clinical trials, psychometric

## **Clinical significance**

- This is the first paper that systematically appraised the clinical trial discrimination properties for PF-PROMs in PsA.
- Data for appraisal of clinical trial discrimination were available for only four PF-PROMs (HAQ-DI, HAQ-S, SF-36 PCS and SF-36 PF).
- The HAQ-DI and SF-36 PCS demonstrated clinical trial discrimination with low risk of bias.
- Clinical trial discrimination with SF-36 PF and HAQ-S are supported with caution. More studies are needed for SF-36 PF and HAQ-S.

## 1. INTRODUCTION

Psoriatic arthritis (PsA) is a chronic inflammatory disease involving peripheral arthritis, enthesitis, dactylitis, spondylitis, psoriasis and nail disease (1). It has a profound impact on patients' physical, psychological and social well-being. The inflammation of axial and peripheral joints as well as entheses causes pain as well as joint erosion and destruction over the longer term (2). Both the acute inflammation and joint damage from PsA cause loss of physical function (PF) and disability (3). Physical function is a key concern from numerous qualitative studies among PsA patients (4, 5), and it is recognized as one of the core domains to be measured in every randomized controlled trial (RCT) and longitudinal observational study (6).

There have been several patient-reported outcome measures (PROMs) that assess PF in PsA (7), the most commonly used in RCTs of PsA being the Health Assessment Questionnaire Disability Index (HAQ-DI) (8) and the physical functioning domain within the Medical Outcomes Study 36-item Short Form Survey (SF-36 PF) (9). A few additional ones have also been validated and evaluated for use in PsA (10). The discriminative performance of these PF-PROMs in RCTs have not been evaluated systematically. In this systematic review, we aimed to evaluate the clinical trial discrimination properties of PF-PROMs in RCTs of PsA. The data derived from this study contribute to the concerted effort of the Group for Research And Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) – Outcome Measures in Rheumatology (OMERACT) initiative to standardize an outcome measurement set for PsA (11).

## 2. METHODS

The protocol of this systematic review was registered with the International Prospective Register of Systematic Reviews (PROSPERO) prior to initiation (CRD42019129557). The report of this systematic review adheres to the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA statement) (12).

### *2.2 Literature search and eligibility of articles*

We performed a database search through 21 March 2019 using PubMed and Scopus. The search aimed to identify all original full text articles of RCTs conducted in PsA with published data in the English language. The detail of search terms is summarized in Appendices (Table A.1). We included articles on RCTs conducted in PsA only. If the RCT was conducted in a mixed arthritis population, we only included those with separate subgroup analyses in PsA. We excluded RCTs that were not double blind in study design. For the purpose of deriving responsiveness data for the PF-PROMs, we limited the review to RCTs involving biological (b) disease modifying anti-rheumatic drugs (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 subgroup analyses. We also excluded articles that did not report data for PF-PROMs.

### *2.3 Selection of articles*

One researcher (YYL) removed duplicates from the searches in the two databases. Two researchers (YYL, RH) independently screened the titles, abstracts and full text (if appropriate) for eligibility. Disputes were resolved by consensus of the two researchers and a third researcher, if needed (AMO). Additional studies identified by the co-authors were

considered for inclusion. The stepwise eligibility and inclusion of articles are summarized in a flow diagram (Figure 1).

#### *2.4 Extraction of study characteristics and description of PROMs*

YYL and AO independently extracted data on the characteristics of the studies including gender, disease duration of participants, proportion of participants taking methotrexate (MTX) and previous exposure (13) to tumor necrosis factor inhibitors (TNFi), active interventions, comparators, primary outcomes and their time points. Researchers worked in pairs (YYL and RH; YYL and AJM; YYL and NG; YYL and CL) to extract data independently for PF-PROMs at baseline and the time point of assessment of the primary outcome (or end of the blinded controlled period). These assessment time points were chosen as that they represented the time to measure responsiveness of the PF-PROMs for the active interventional from the control groups, as most RCTs offered cross-over to the active intervention after the primary end point. One researcher (YYL) calculated the effect sizes (ESs) for the PF-PROMs based on published data and online supplementary materials if available. Effect sizes of PF-PROMs comparing the active interventional and control groups were evaluated using the following methods wherever data were available (3): 1) Glass's effect size ( $ES_1$ ): calculated by change scores divided by standard deviations (SDs) of baseline scores; 2) Cohen's effect size ( $ES_2$ ), calculated by change scores divided by pooled SDs; and 3) standardized response means (SRMs), calculated by change scores divided by SDs of the change scores when no comparator group was available.

For papers that provided the standard error (SE) for individual groups, SD was estimated using the formula:  $SD = SE \times \sqrt{[sample\ size, N]}$  (14). For papers that gave information on median and interquartile ranges, means (SD) were calculated using an estimation formula suggested by Wan et al (13).

## 2.5 Quality Assessment

We assessed the clinical trial discrimination of PF-PROMs in each article using the OMERACT good method checklist (15). Two researchers in groups (YYL and RH; YYL and AJM; YYL and NG; YYL and CL) appraised each instrument for the following categories using (+) yes, good methods; (+/-) some cautions; or (-) no, not achieved:

1. Was the time interval between testing stated and appropriate?
2. Was there a proportion of participants expected to change in one or both groups?
3. Were hypotheses regarding the anticipated mean differences in change scores between subgroups described *a priori*?
4. Were the statistical methods adequate for the hypotheses tested?
5. Otherwise good methods? (free of any other important flaws)

The two-researcher groups independently assessed a final decision for each article rated as 1) Green, likely low risk of bias; 2) Amber, some cautions but can be used as evidence and 3) Red, not to use as evidence. Consensus between the two researchers in the group was sought and disputes were resolved with a third member of the team, if necessary.

## 2.6 Appraisal of effect sizes using *a priori* hypotheses

The calculated ESs of each article were appraised with the following *a priori* hypotheses. These hypotheses were developed based on general knowledge of the therapeutic efficacy of various treatments:

1. At the assessment time point (primary outcome or end of blinded controlled period), patients treated with bDMARDs have significant changes in PF-PROMs, whereas



patients receiving the control intervention do not (except for alefacept [ALC] and clazakizumab [CLZ] where no significant differences were expected).

2. PF-PROM change scores among patients treated with bDMARDs are significantly better than those of the control group.
3. Within an individual trial, the ESs of PF-PROMs are higher in the bDMARD groups compared to the control groups that were treated with placebo, MTX or conventional synthetic (cs) DMARDs, but do not differ significantly from groups treated with different bDMARD doses (or groups treated with TNFi as a comparator).
4. If data for subgroup analyses are available, the ESs of change scores of the PF-PROM are higher in TNFi-naïve versus TNFi-exposed subgroups.

We assessed whether the hypotheses were satisfied using symbols of (+), when hypothesis satisfied; (+/-), if partially satisfied; or (-), not satisfied. As per OMERACT Filter 2.1 guidelines (15), we synthesized and summarized the overall evidence to support clinical trial discrimination for each PF-PROM in a Summary of Measurement Properties evidence (SOMP) table.

## *2.7 Evidence synthesis*

In the SOMP table, we presented the number of articles available for measurement properties and evidence synthesis. For each article, color coding of Green/Amber/Red indicated the quality assessment and (+), (+/-) or (-) was given to indicate the *a priori* hypotheses were satisfied, partially satisfied or not satisfied.

For each PF-PROM, a final rating for clinical trial discrimination was given as GREEN (Good to go)/AMBER (Some concern but good to go)/RED (Stop, not to use this for

evidence synthesis), or WHITE for no information available. This final rating follows the OMERACT Filter 2.1 recommendation for the quality assessment of articles, the number of good quality articles available, the consistency across articles and the PF-PROM performance in RCT discrimination (16). A PF-PROM that was supported by at least two good quality articles showing consistent findings with adequate performance in RCT discrimination would achieve GREEN. A PROM having at least two articles, but inconsistent findings; or having only one article with inadequate performance would be given a RED. In all other situations, a final rating of AMBER would be given.

### **3. RESULTS**

#### *3.1 Literature search results*

Our literature search identified a total of 676 articles. After removal of duplicates, 608 articles remained; three articles were identified through cross-referencing check. There were 439 articles excluded for the following reasons: not RCTs (344 articles), open-label trials (10 articles), not in PsA (65 articles), mixed population of arthritis without separate data reported for PsA (18 articles) and trial protocols (2 articles). We retained 172 articles for full text review as RCTs of PsA. We excluded 48 articles as trials for interventions other than bDMARDs or tsDMARDs and 70 articles as secondary analyses of RCTs in PsA. From the 45 articles of 41 unique RCTs in PsA for bDMARDs and tsDMARDs, we further excluded 10 articles without reported data on PF-PROMs. Finally, 35 relevant articles from 31 RCTs were retained for qualitative assessment (Figure 1).

The characteristics of the included articles are summarized in Table 1. There were 10 articles that did not report data on the PF domains. Among them, one article was from a phase 1b and three articles were from phase 2 trials. Among the six phase 3 RCTs that did not

include PF data, two were small pilot “proof of concept” studies with a molecular and tissue focus, three were RCTs with a primary focus on psoriasis, while one was in PsA.

Among the included 31 RCTs, PF was evaluated using four PF-PROMs, including the HAQ-DI (8), HAQ-Spondyloarthritis (HAQ-S) (17), Physical Component Summary Score of the SF-36 (SF-36 PCS) (9), and the SF-36 PF (9). New potential PF-PROMs shortlisted by GRAPPA for further evaluation (18), including the multidimensional HAQ (MDHAQ) and the Patient-Reported Outcomes Measurement Information System (PROMIS)-Short Form Physical Function 10a (PROMIS-PF10a), were not utilized in the RCTs included in this review.

### *3.2 Clinical trial discrimination for PF-PROMs*

#### 3.2.1 HAQ-DI

Results were reported for HAQ-DI in 31 articles from 30 unique RCTs. The ESs calculated for both interventional and comparator groups are shown in Table 2. Effect sizes for HAQ-DI could not be calculated from two published articles (Appendices, Table A.2) and data were not used for evidence synthesis. Quality assessment using the OMERACT good method checklist was generally affirmative (Appendices, Table A.3). Minor concerns for quality were noted in 10 articles, predominantly resulting from not clearly stating expected change scores in HAQ-DI (although most expected changes were implied), estimating ESs from median/interquartile ranges or using percentage changes that raised concern for introduction of minor errors. There were 29 articles included for evidence synthesis as shown in the SOMP table (Table 6). Results aligned with the *a priori* hypotheses that statistically significant change scores were reported in interventional but not in control groups with higher ESs for interventional compared to control groups (Table 2). The median (interquartile ranges, IQR)

ES were -0.73 (-0.84, -0.62) and -0.24 (-0.39, -0.09) for intervention and control groups, respectively. Three articles included subgroup analyses, where one demonstrated higher ESs for HAQ-DI in TNFi-naïve compared with TNFi-exposed groups (19), while two articles did not (20, 21). The working group recommended a final rating of GREEN to support HAQ-DI for clinical trial discrimination in PsA, indicating a low risk of bias and results aligned with hypotheses.

### 3.2.2 HAQ-S

There was only one article that reported results for HAQ-S (22), showing higher ESs for the ABT-122 and adalimumab groups compared to control which aligned with the *a priori* hypotheses (Table 3). Detail of the quality assessment is given in Appendices (Table A.4). The working group recommended AMBER for HAQ-S (Table 6), recognizing that more data are required for better evidence synthesis.

### 3.2.3 SF-36 PCS

The SF-36 was included in 26 out of 31 (83.9%) RCTs in PsA. All 26 RCTs reported SF-36 PCS results, but only four reported on SF-36 PF. Of the articles included for evidence synthesis for SF-36 PCS, one did not pass the OMERACT checklist for evidence synthesis (23) for two reasons. It failed to include adequate data for ES calculations. Compared to control groups, statistically significant differences in change scores were only evident with the higher dose but not the lower dose interventional groups (Table 4). For the 25 articles included for evidence synthesis for SF-36 PCS, most results aligned with the *a priori* hypotheses. The median (IQR) ES were 0.77 (0.60, 0.93) and 0.23 (0.09, 0.36) for intervention and control groups, respectively. Minor quality concerns were noted in seven

articles, that were similar to those stated for HAQ-DI (Appendix, Table A.5). The working group recommended GREEN supporting clinical trial discrimination with SF-36 PCS.

### 3.2.4 SF-36 PF

Four articles reported data for SF-36 PF. Two did not have adequate data for ES estimation and were excluded from evidence synthesis (Table 5) (Appendix, Table A.6). Of the remaining two articles, both on tofacitinib (24, 25), higher ESs in the interventional groups were shown compared with the control (Table 5). However, due to other methodologic concerns, these two articles were judged to be Amber (+) in the quality assessment (Appendix, Table A.6). Based on the limited evidence, the working group recommended AMBER for clinical trial discrimination with SF-36 PF (Table 6).

## **4. DISCUSSION**

In this systematic review, we summarized the clinical trial discrimination of the available PF-PROMs in PsA. Of the 41 unique RCTs in PsA with bDMARDs or tsDMARDs, 31 (75.6%) reported results of at least one measurement of the PF domain. This is the first paper to our knowledge that systematically appraised the clinical trial discrimination properties for PF-PROMs. Numerous instruments are available for assessing PF in PsA (7, 10). However, data for appraisal of clinical trial discrimination was available for only four PF-PROMs (HAQ-DI, HAQ-S, SF-36 PCS and SF-36 PF). Most of the studies reported data for HAQ-DI and SF-36 PCS, while only four and one reported data for SF-36 PF and HAQ-S, respectively. This systematic review supports clinical trial discrimination with HAQ-DI and SF-36 PCS with low risk of bias. Clinical trial discrimination with SF-36 PF and HAQ-S are supported with caution due to the limited studies available for data synthesis. Further research

is warranted. There are no data published to date to support clinical trial discrimination for potential PF-PROMs shortlisted by GRAPPA such as the MDHAQ and PROMIS-PF (18).

Physical function is one of the domains included in the core outcome set for reporting data in PsA RCTs and longitudinal observational studies (6). However, variations in reporting of outcomes in PsA clinical trials have been recognized and GRAPPA and OMERACT are committed to standardize the outcome measures with evidence (11). To appropriately evaluate the measurement properties of instruments using the OMERACT Filter 2.1, multiple measurement properties are considered, including domain match, feasibility, validity, test-retest reliability, longitudinal construct validity, clinical trial discrimination and threshold of meaning (15). In this study, we aimed to evaluate only the clinical trial discrimination of PF-PROMs. This represents an intermediate but important step in standardizing the outcome measurement set for PF in PsA.

The strength of the current work is the combined effort of investigators and patient research partners. The investigators are familiar with the measurement of PF in PsA with representation from 4 continents. The patient research partners have participated in a wide range of research activities, including data extraction, quality assessment of articles and appraisal of ESs. We followed the methods recommended by the OMERACT Filter 2.1 methodology in quality assessment of each article, calculating the ESs using appropriate statistics to synthesize the evidence to support clinical trial discrimination (15). This was further strengthened by setting *a priori* hypotheses on the expected magnitude of ESs of intervention groups compared with control groups.

Some limitations are recognized. We limited the evidence from RCTs of bDMARDs and tsDMARDs and excluded RCTs evaluating solely csDMARDs to best represents RCTs in the

modern era that include the appropriate core domains. We calculated the ESs of PF-PROMs from published data as access to the original dataset from the RCTs was not available. Where applicable, we used formulae to estimate the means (SD) from reported medians (IQR), which may result in variability. Nonetheless, this variability in ES estimations has been recognized and addressed in the quality assessment using the OMERACT good method checklist; and the detailed calculations are shown in the supplementary documents. Few definitions of estimated ESs (ES<sub>1</sub>, ES<sub>2</sub> or SRM) were tabulated depending on availability of published data, which may not be directly comparable. This would attribute only to minor issues as all hypothesis testing was performed intra- rather than inter- trials using the same type of ESs. We have not compared the responsiveness of one PF-PROM with another, which may be an interesting topic to pursue. Finally, we used data from published RCTs in PsA which generally enroll patients with higher disease activity (and greater limitations in physical function) than the average patient seen in clinical practice to undergo potential highly efficacious treatment, and therefore higher responsiveness for the PF-PROMs are expected. It is thus unclear if these results would be directly applicable to pragmatic trials and longitudinal observational studies.

## **5. CONCLUSION**

This systemic review supports clinical trial discrimination with HAQ-DI and SF-36 PCS with low risk of bias, while clinical trial discrimination with SF-36 PF and HAQ-S is supported with some caution. More studies are required for SF-36 PF and HAQ-S.

## **6. ACKNOWLEDGEMENTS AND FUNDING**

YYL is funded by the Clinician Scientist award of the National Medical Research Council, Singapore (NMRC/CSA-INV/0022/2017). The views expressed are those of the author(s) and not necessarily those of the NMRC. AMO is funded by the Jerome L. Greene Foundation Scholar Award, the Staurulakis Family Discovery Award, the Rheumatology Research Foundation, and the National Institutes of Health (NIH) through the Rheumatic Diseases Resource-based Core Center (P30-AR053503 Cores A and D, and P30-AR070254, Cores A and B). All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the NIH, the National Institute of Arthritis Musculoskeletal and Skin Diseases (NIAMS), or the Rheumatology Research Foundation (RRF). LCC is funded by a National Institute for Health Research Clinician Scientist award. The research was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. AO is funded by the Rheumatology Research Foundation and NIH/NIAMS R01 AR072363. The Parker Institute, Bispebjerg and Frederiksberg Hospital (PH and RC) is supported by a core grant from the Oak Foundation (OCAY-18-774-OFIL).



## 7. LIST OF ABBREVIATIONS

ABT: abatacept;  
ADA: adalimumab;  
ALC: alefacept;  
bDMARDs: biological disease modifying anti-rheumatic drugs;  
BIW: twice a week;  
BRO: brodalumab;  
CI: confidence interval;  
CLZ: clazakizumab;  
csDMARDs: conventional synthetic disease modifying anti-rheumatic drugs;  
CZP: certolizumab pegol;  
ES: effect size;  
ETN: etanercept;  
EULAR: European League Against Rheumatism;  
FIL: filgotinib;  
GOL: golimumab;  
GRAPPA: Group for Research and Assessment of Psoriasis and Psoriatic Arthritis  
GUS: guselkumab;  
HAQ-DI: Health Assessment Questionnaire – Disability Index;  
HAQ-S: Health Assessment Questionnaire – Spondyloarthritis;  
IFX: infliximab;  
IL: interleukin;  
IXE: ixekizumab;  
IV: intravenous  
IQR: interquartile range;  
LS: least squares;  
MDHAQ: multidimensional HAQ  
MTX: methotrexate;  
NA: not available;  
NS: not significant;  
OMERACT: Outcome Measures in Rheumatology  
PASI: Psoriasis Area and Severity index;  
PCB: placebo;  
PCS: physical component summary of SF-36;  
PF: physical functioning domain of SF-36;  
PROM: patient reported outcome measure;  
PROMIS: Patient-Reported Outcomes Measurement Information System;  
PROSPERO: the International Prospective Register of Systematic Reviews;  
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analysis;  
PsA: psoriatic arthritis;  
Q2W: once every 2-week;  
QW: once a week;  
RCTs: randomized controlled trials  
SD: standard deviation;  
SE: standard error;  
SEC: secukinumab;  
SF-36: Medical Outcomes Study 36-item Short Form Survey;  
TOF: tofacitinib;

TNFi: tumor necrosis factor inhibitors;  
tsDMARDs: targeted synthetic disease modifying anti-rheumatic drugs;  
UST: ustekinumab;  
vs.: versus.

## 8. TABLE AND FIGURE LEGEND

**Figure 1.** Flow diagram for article selection

**Table 1.** Characteristics of included studies

Legend:

\*early escape at Week 16;

Abbreviations:  $\Delta$ : change; ACR: American College of Rheumatology Response criteria; ABT: abatacept; ADA: adalimumab; ALC: alefacept; BIW: twice a week; BRO: brodalumab; CI: confidence interval; CLZ: clazakizumab; csDMARD: conventional synthetic disease modifying anti-rheumatic drug; CZP: certolizumab pegol; ETN: etanercept; EULAR: European League Against Rheumatism; FIL: filgotinib; GOL: golimumab; GUS: guselkumab; HAQ-DI: Health Assessment Questionnaire – Disability index; HAQ-S: Health Assessment Questionnaire – Spondyloarthritis; IFX: infliximab; IXE: ixekizumab; IV: intravenous; IL=interleukin; IR=inadequate response; MTX: methotrexate; NA: not available; NS: not significant; PASI: Psoriasis Area and Severity Index; PCB: placebo; PCS: physical component summary of SF-36; PF: physical functioning domain of SF-36; PsA=psoriatic arthritis; PsARC: Psoriatic arthritis Response Criteria; QW: once a week; Q2W: once every 2 weeks; SD: standard deviation; SE: standard error; SEC: secukinumab; SF-36: Medical Outcomes Study 36-item Short Form Survey; SPARCC: Spondyloarthritis Research Consortium of Canada enthesitis index; TOF: tofacitinib; TNFi: tumor necrosis factor inhibitor; UST: ustekinumab; vs.: versus.

**Table 2.** Effect sizes estimation for studies reporting HAQ-DI

Legend:

¶ SRM calculated using percentage change score and SD of percentage change;  $\delta$  SRM for improvement, a negative value indicate deterioration; ¥ Effect sizes estimated based on mean and SD of change scores calculated from median and IQR from original publication; \* early escape for patients with inadequate response in the PCB group to active treatment group at Week 16; \*\* option to switch TNFi at Week 24.

Abbreviations:  $\Delta$ : change; ACR: American College of Rheumatology Response criteria; ABT: abatacept; ADA: adalimumab; ALC: alefacept; BIW: twice a week; BRO: brodalumab; CI: confidence interval; CLZ: clazakizumab; CZP: certolizumab pegol; ES2: Effect size 2 (the mean difference divided by the pooled standard deviation, i.e Cohen's d); ETN: etanercept; EULAR: European League Against Rheumatism; FIL: filgotinib; GOL: golimumab; GUS: guselkumab; HAQ-DI: Health Assessment Questionnaire – Disability Index; IFX: infliximab; IXE: ixekizumab; IV: intravenous; IQR: interquartile range; MTX: methotrexate; NA: not available; NS: not significant; PASI: Psoriasis Area and Severity Index; PCB: placebo; PsARC: Psoriatic arthritis Response Criteria; QW: once a week; Q2W: once every 2-week; SD: standard deviation; SE: standard error; SEC: secukinumab; SPARCC: Spondyloarthritis Research Consortium of Canada enthesitis index; SRM: Standardized response mean (mean difference divided by the standard deviation of the differences between baseline and assessment end point); TOF: tofacitinib; TNFi: tumor necrosis factor inhibitor; UST: ustekinumab; vs.: versus.

**Table 3.** Effect sizes estimation for studies reported HAQ-S

Legend:

Abbreviations: ACR: American College of Rheumatology Response criteria; ADA: adalimumab; ES1: Effect size 1 (the mean difference divided by standard deviation of baseline score); HAQ-S: Health Assessment Questionnaire – Spondyloarthritis; PCB: placebo; vs.: versus.

**Table 4.** Effect sizes estimation for studies reported SF-36 PCS

Legend:

¶ SRM calculated using percentage change score and SD of percentage change; δ SRM for improvement, a negative value indicate deterioration; ¥ Effect sizes estimated based on mean and SD of change scores calculated from median and IQR from original publication; \* early escape for patients with inadequate response in the PCB group to active treatment group at Week 16; \*\* option to switch TNFi at Week 24.

Abbreviations: Δ: change; ACR: American College of Rheumatology Response criteria; ABT: abatacept; ADA: adalimumab; ALC: alefacept; BIW: twice a week; BRO: brodalumab; CI: confidence interval; CLZ: clazakizumab; CZP: certolizumab pegol; ES2: Effect size 2 (the mean difference divided by the pooled standard deviation, i.e Cohen's d); ETN: etanercept; EULAR: European League Against Rheumatic Diseases; GOL: golimumab; GUS: guselkumab; IFX: infliximab; IXE: ixekizumab; IV: intravenous; IQR: interquartile range; MTX: methotrexate; NA: not available; PASI: psoriasis area and severity index; PCB: placebo; PCS: physical component summary of SF-36; PF: physical functioning domain of SF-36; QW: once a week; Q2W: once every 2-week; SD: standard deviation; SE: standard error; SEC: secukinumab; SF-36: Medical Outcome Survey Short Form 36 items; SPARCC: Spondyloarthritis Research Consortium of Canada enthesitis index; SRM: Standardized response mean (mean difference divided by the standard deviation of the differences between baseline and assessment end point); TNFi: tumor necrosis factor inhibitor; UST: ustekinumab; vs.: versus.

**Table 5.** Effect sizes estimation for studies reporting SF-36 PF.

Legend:

¶ SRM calculated using percentage change score and SD of percentage change; δ SRM for improvement, a negative value indicate deterioration; ¥ Effect sizes estimated based on mean and SD of change scores calculated from median and IQR from original publication; \* early escape for patients with inadequate response in the control group to active treatment group at week 16; \*\* option to switch TNFi at Week 24.

Abbreviations: Δ: change; ACR: American College of Rheumatology Response criteria; ADA: adalimumab; ES2: Effect size 2 (the mean difference divided by the pooled standard deviation, i.e Cohen's d); FIL: filgotinib; HAQ-DI: Health Assessment Questionnaire – Disability Index; IFX: infliximab; IQR: interquartile range; PCB: placebo; PF: physical functioning domain of SF-36; SD: standard deviation; SF-36: Medical Outcomes Study 36-item Short Form Survey; SRM: Standardized response mean (mean difference divided by the standard deviation of the differences between baseline and assessment end point); TOF: tofacitinib; vs.: versus.

**Table 6.** Summary of Measurement Properties Table for clinical trial discrimination

Legend:

Color code in each box indicate study quality assessed by OMERACT good methods.

GREEN means “yes, likely low risk of bias”; AMBER means “some cautions but can be used as evidence” and RED means “No, don’t use as evidence”. WHITE (empty boxes), indicates absence of information on that property from that study. (+) indicates findings of the study had adequate performance of the instrument; (+/-) indicates equivocal performance; (-) indicates poor performance (less than adequate).

## 9. DECLARATION OF INTEREST

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

YYL is funded by the Clinician Scientist award of the National Medical Research Council, Singapore (NMRC/CSA-INV/0022/2017). AMO is funded by the Jerome L. Greene Foundation Scholar Award, the Staurulakis Family Discovery Award, the Rheumatology Research Foundation, and the National Institutes of Health (NIH) through the Rheumatic Diseases Resource-based Core Center (P30-AR053503 Cores A and D, and P30-AR070254, Cores A and B). PH and RC (The Parker Institute, Bispebjerg and Frederiksberg Hospital) is supported by a core grant from the Oak Foundation (OCAY-18-774-OFIL). LCC is funded by a National Institute for Health Research Clinician Scientist award, and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). AO is funded by the Rheumatology Research Foundation and NIH/NIAMS R01 AR072363. All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the funding agencies.

## 10. AUTHOR CONTRIBUTIONS

**Ying Ying Leung:** Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing original draft; review & editing. **Richard Holland:** Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Validation; Visualization; Writing - review & editing. **Ashish J Mathew:** Data curation; Formal analysis; Investigation; Methodology; Resources; Validation; Writing - review & editing. **Chris Lindsay:** Data curation; Formal analysis;

Investigation; Methodology; Validation; Writing - review & editing. **Niti Goel:** Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing - review & editing. **Alexis Oddie:** Conceptualization; Data curation; Validation; Writing - review & editing. **Ana-Maria Orbai:** Conceptualization; Data curation; Methodology; Validation; Writing - review & editing. **Pil Hoejgaard:** Methodology; Writing - review & editing. **Jeffrey Chau:** Methodology; Writing - review & editing. **Laura C Coates:** Methodology; Supervision; Writing - review & editing. **Vibeke Strand:** Methodology; Supervision; Writing - review & editing. **Dafna D Gladman:** Conceptualization; Methodology; Supervision; Writing - review & editing. **Robin Christensen:** Formal analysis; Methodology; Resources; Supervision; Validation; Writing - review & editing. **William Tillett:** Methodology; Supervision; Validation; Visualization; Writing - review & editing. **Philip Mease:** Methodology; Supervision; Validation; Visualization; Writing - review & editing.

## 11. REFERENCES

1. Gladman DD, Antoni C, Mease P, Clegg DO, Nash P. Psoriatic arthritis: epidemiology, clinical features, course, and outcome. *Annals of the rheumatic diseases*. 2005;64 Suppl 2:ii14-7.
2. Gladman DD, Farewell VT, Nadeau C. Clinical indicators of progression in psoriatic arthritis: multivariate relative risk model. *The Journal of rheumatology*. 1995;22(4):675-9.
3. Husted JA, Tom BD, Farewell VT, Schentag CT, Gladman DD. Description and prediction of physical functional disability in psoriatic arthritis: a longitudinal analysis using a Markov model approach. *Arthritis and rheumatism*. 2005;53(3):404-9.
4. Orbai AM, de Wit M, Mease PJ, Callis Duffin K, Elmamoun M, Tillett W, et al. Updating the Psoriatic Arthritis (PsA) Core Domain Set: A Report from the PsA Workshop at OMERACT 2016. *The Journal of rheumatology*. 2017;44(10):1522-8.
5. Dures E, Hewlett S, Lord J, Bowen C, McHugh N, Group PS, et al. Important Treatment Outcomes for Patients with Psoriatic Arthritis: A Multisite Qualitative Study. *Patient*. 2017;10(4):455-62.
6. Orbai AM, de Wit M, Mease P, Shea JA, Gossec L, Leung YY, et al. International patient and physician consensus on a psoriatic arthritis core outcome set for clinical trials. *Annals of the rheumatic diseases*. 2017;76(4):673-80.
7. Mease P, Strand V, Gladman D. Functional impairment measurement in psoriatic arthritis: Importance and challenges. *Seminars in arthritis and rheumatism*. 2018;48(3):436-48.
8. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis and rheumatism*. 1980;23(2):137-45.
9. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30(6):473-83.
10. Hojgaard P, Klokke L, Orbai AM, Holmsted K, Bartels EM, Leung YY, et al. A systematic review of measurement properties of patient reported outcome measures in psoriatic arthritis: A GRAPPA-OMERACT initiative. *Seminars in arthritis and rheumatism*. 2018;47(5):654-65.
11. Tillett W, Orbai AM, Ogdie A, Leung YY, Strand V, Gladman DD, et al. GRAPPA-OMERACT initiative to standardise outcomes in psoriatic arthritis clinical trials and longitudinal observational studies. *Ann Rheum Dis*. 2018;77(5):e23.
12. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ (Clinical research ed)*. 2009;339:b2700.
13. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*. 2014;14:135.
14. Collaboration TC. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. In: Higgins JPT, Green S, editors. 2011.
15. Beaton DE, Maxwell LJ, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Instrument Selection Using the OMERACT Filter 2.1: The OMERACT Methodology. *The Journal of rheumatology*. 2019;46(8):1028-35.
16. Boers M, Kirwan JR, Tugwell P, Beaton D, Bingham III CO, Conaghan PG, et al. *The OMERACT Handbook*. 2017.
17. Daltroy LH, Larson MG, Roberts NW, Liang MH. A modification of the Health Assessment Questionnaire for the spondyloarthropathies. *The Journal of rheumatology*. 1990;17(7):946-50.
18. Leung YY, Orbai AM, Ogdie A, Hojgaard P, Holland R, Goel N, et al. Appraisal of candidate instruments for assessment of the physical function domain in patients with psoriatic arthritis. *The Journal of rheumatology*. 2020.
19. Kavanaugh A, McInnes IB, Mease PJ, Hall S, Chinoy H, Kivitz AJ, et al. Efficacy of Subcutaneous Secukinumab in Patients with Active Psoriatic Arthritis Stratified by Prior Tumor Necrosis Factor Inhibitor Use: Results from the Randomized Placebo-controlled FUTURE 2 Study. *The Journal of rheumatology*. 2016;43(9):1713-7.



20. Ritchlin C, Rahman P, Kavanaugh A, McInnes IB, Puig L, Li S, et al. Efficacy and safety of the anti-IL-12/23 p40 monoclonal antibody, ustekinumab, in patients with active psoriatic arthritis despite conventional non-biological and biological anti-tumour necrosis factor therapy: 6-month and 1-year results of the phase 3, multicentre, double-blind, placebo-controlled, randomised PSUMMIT 2 trial. *Annals of the rheumatic diseases*. 2014;73(6):990-9.
21. Mease PJ, Gottlieb AB, van der Heijde D, FitzGerald O, Johnsen A, Nys M, et al. Efficacy and safety of abatacept, a T-cell modulator, in a randomised, double-blind, placebo-controlled, phase III study in psoriatic arthritis. *Annals of the rheumatic diseases*. 2017;76(9):1550-8.
22. Mease PJ, Genovese MC, Weinblatt ME, Peloso PM, Chen K, Othman AA, et al. Phase II Study of ABT-122, a Tumor Necrosis Factor– and Interleukin-17A–Targeted Dual Variable Domain Immunoglobulin, in Patients With Psoriatic Arthritis With an Inadequate Response to Methotrexate. *Arthritis and Rheumatology*. 2018;70(11):1778-89.
23. Mease PJ, Genovese MC, Greenwald MW, Ritchlin CT, Beaulieu AD, Deodhar A, et al. Brodalumab, an anti-IL17RA monoclonal antibody, in psoriatic arthritis. *The New England journal of medicine*. 2014;370(24):2295-306.
24. Mease P, Hall S, FitzGerald O, van der Heijde D, Merola JF, Avila-Zapata F, et al. Tofacitinib or Adalimumab versus Placebo for Psoriatic Arthritis. *The New England journal of medicine*. 2017;377(16):1537-50.
25. Gladman D, Rigby W, Azevedo VF, Behrens F, Blanco R, Kaszuba A, et al. Tofacitinib for Psoriatic Arthritis in Patients with an Inadequate Response to TNF Inhibitors. *The New England journal of medicine*. 2017;377(16):1525-36.
26. Antoni CE, Kavanaugh A, Kirkham B, Tutuncu Z, Burmester GR, Schneider U, et al. Sustained benefits of infliximab therapy for dermatologic and articular manifestations of psoriatic arthritis: results from the infliximab multinational psoriatic arthritis controlled trial (IMPACT). *Arthritis and rheumatism*. 2005;52(4):1227-36.
27. Antoni C, Krueger GG, de Vlam K, Birbara C, Beutler A, Guzzo C, et al. Infliximab improves signs and symptoms of psoriatic arthritis: results of the IMPACT 2 trial. *Annals of the rheumatic diseases*. 2005;64(8):1150-7.
28. Kavanaugh A, Antoni C, Krueger GG, Yan S, Bala M, Dooley LT, et al. Infliximab improves health related quality of life and physical function in patients with psoriatic arthritis. *Annals of the rheumatic diseases*. 2006;65(4):471-7.
29. Mease PJ, Ory P, Sharp JT, Ritchlin CT, Van den Bosch F, Wellborne F, et al. Adalimumab for long-term treatment of psoriatic arthritis: 2-year data from the Adalimumab Effectiveness in Psoriatic Arthritis Trial (ADEPT). *Annals of the rheumatic diseases*. 2009;68(5):702-9.
30. Genovese MC, Mease PJ, Thomson GT, Kivitz AJ, Perdok RJ, Weinberg MA, et al. Safety and efficacy of adalimumab in treatment of patients with psoriatic arthritis who had failed disease modifying antirheumatic drug therapy. *The Journal of rheumatology*. 2007;34(5):1040-50.
31. Mease PJ, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomised trial. *Lancet (London, England)*. 2000;356(9227):385-90.
32. Mease PJ, Woolley JM, Singh A, Tsuji W, Dunn M, Chiou CF. Patient-reported outcomes in a randomized trial of etanercept in psoriatic arthritis. *The Journal of rheumatology*. 2010;37(6):1221-7.
33. Gniadecki R, Robertson D, Molta CT, Freundlich B, Pedersen R, Li W, et al. Self-reported health outcomes in patients with psoriasis and psoriatic arthritis randomized to two etanercept regimens. *Journal of the European Academy of Dermatology and Venereology : JEADV*. 2012;26(11):1436-43.
34. Mease PJ, Gladman DD, Collier DH, Ritchlin CT, Helliwell PS, Liu L, et al. Etanercept and Methotrexate as Monotherapy or in Combination for Psoriatic Arthritis: Primary Results From a Randomized, Controlled Phase 3 Trial. *Arthritis & rheumatology (Hoboken, NJ)*. 2019.
35. Kavanaugh A, McInnes I, Mease P, Krueger GG, Gladman D, Gomez-Reino J, et al. Golumab, a new human tumor necrosis factor alpha antibody, administered every four weeks as a

- subcutaneous injection in psoriatic arthritis: Twenty-four-week efficacy and safety results of a randomized, placebo-controlled study. *Arthritis and rheumatism*. 2009;60(4):976-86.
36. Kavanaugh A, Husni ME, Harrison DD, Kim L, Lo KH, Leu JH, et al. Safety and Efficacy of Intravenous Golimumab in Patients With Active Psoriatic Arthritis: Results Through Week Twenty-Four of the GO-VIBRANT Study. *Arthritis & rheumatology (Hoboken, NJ)*. 2017;69(11):2151-61.
  37. Gladman D, Fleischmann R, Coteur G, Woltering F, Mease PJ. Effect of certolizumab pegol on multiple facets of psoriatic arthritis as reported by patients: 24-week patient-reported outcome results of a phase III, multicenter study. *Arthritis care & research*. 2014;66(7):1085-92.
  38. McInnes IB, Sieper J, Braun J, Emery P, van der Heijde D, Isaacs JD, et al. Efficacy and safety of secukinumab, a fully human anti-interleukin-17A monoclonal antibody, in patients with moderate-to-severe psoriatic arthritis: a 24-week, randomised, double-blind, placebo-controlled, phase II proof-of-concept trial. *Annals of the rheumatic diseases*. 2014;73(2):349-56.
  39. Mease PJ, McInnes IB, Kirkham B, Kavanaugh A, Rahman P, van der Heijde D, et al. Secukinumab Inhibition of Interleukin-17A in Patients with Psoriatic Arthritis. *The New England journal of medicine*. 2015;373(14):1329-39.
  40. McInnes IB, Mease PJ, Kirkham B, Kavanaugh A, Ritchlin CT, Rahman P, et al. Secukinumab, a human anti-interleukin-17A monoclonal antibody, in patients with psoriatic arthritis (FUTURE 2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet (London, England)*. 2015;386(9999):1137-46.
  41. Nash P, Mease PJ, McInnes IB, Rahman P, Ritchlin CT, Blanco R, et al. Efficacy and safety of secukinumab administration by autoinjector in patients with psoriatic arthritis: Results from a randomized, placebo-controlled trial (FUTURE 3). *Arthritis Research and Therapy*. 2018;20(1).
  42. Mease PJ, van der Heijde D, Ritchlin CT, Okada M, Cuchacovich RS, Shuler CL, et al. Ixekizumab, an interleukin-17A specific monoclonal antibody, for the treatment of biologic-naïve patients with active psoriatic arthritis: results from the 24-week randomised, double-blind, placebo-controlled and active (adalimumab)-controlled period of the phase III trial SPIRIT-P1. *Annals of the rheumatic diseases*. 2017;76(1):79-87.
  43. Nash P, Kirkham B, Okada M, Rahman P, Combe B, Burmester GR, et al. Ixekizumab for the treatment of patients with active psoriatic arthritis and an inadequate response to tumour necrosis factor inhibitors: results from the 24-week randomised, double-blind, placebo-controlled period of the SPIRIT-P2 phase 3 trial. *Lancet (London, England)*. 2017;389(10086):2317-27.
  44. Gottlieb A, Menter A, Mendelsohn A, Shen YK, Li S, Guzzo C, et al. Ustekinumab, a human interleukin 12/23 monoclonal antibody, for psoriatic arthritis: randomised, double-blind, placebo-controlled, crossover trial. *Lancet (London, England)*. 2009;373(9664):633-40.
  45. McInnes IB, Kavanaugh A, Gottlieb AB, Puig L, Rahman P, Ritchlin C, et al. Efficacy and safety of ustekinumab in patients with active psoriatic arthritis: 1 year results of the phase 3, multicentre, double-blind, placebo-controlled PSUMMIT 1 trial. *Lancet (London, England)*. 2013;382(9894):780-9.
  46. Araujo EG, Englbrecht M, Hoepken S, Finzel S, Kampylafka E, Kleyer A, et al. Effects of ustekinumab versus tumor necrosis factor inhibition on enthesitis: Results from the enthesial clearance in psoriatic arthritis (ECLIPSA) study. *Seminars in arthritis and rheumatism*. 2019;48(4):632-7.
  47. Deodhar A, Gottlieb AB, Boehncke WH, Dong B, Wang Y, Zhuang Y, et al. Efficacy and safety of guselkumab in patients with active psoriatic arthritis: a randomised, double-blind, placebo-controlled, phase 2 study. *Lancet (London, England)*. 2018;391(10136):2213-24.
  48. Mease P, Genovese MC, Gladstein G, Kivitz AJ, Ritchlin C, Tak PP, et al. Abatacept in the treatment of patients with psoriatic arthritis: results of a six-month, multicenter, randomized, double-blind, placebo-controlled, phase II trial. *Arthritis and rheumatism*. 2011;63(4):939-48.
  49. Mease PJ, Gladman DD, Keystone EC. Alefacept in combination with methotrexate for the treatment of psoriatic arthritis: results of a randomized, double-blind, placebo-controlled study. *Arthritis and rheumatism*. 2006;54(5):1638-45.

50. Strand V, de Vlam K, Covarrubias-Cobos JA, Mease PJ, Gladman DD, Graham D, et al. Tofacitinib or adalimumab versus placebo: patient-reported outcomes from OPAL Broaden-a phase III study of active psoriatic arthritis in patients with an inadequate response to conventional synthetic disease-modifying antirheumatic drugs. *RMD Open*. 2019;5(1):e000806.
51. Strand V, de Vlam K, Covarrubias-Cobos JA, Mease PJ, Gladman DD, Chen L, et al. Effect of tofacitinib on patient-reported outcomes in patients with active psoriatic arthritis and an inadequate response to tumour necrosis factor inhibitors in the phase III, randomised controlled trial: OPAL Beyond. *RMD Open*. 2019;5(1):e000808.
52. Mease P, Coates LC, Helliwell PS, Stanislavchuk M, Rychlewska-Hanczewska A, Dudek A, et al. Efficacy and safety of filgotinib, a selective Janus kinase 1 inhibitor, in patients with active psoriatic arthritis (EQUATOR): results from a randomised, placebo-controlled, phase 2 trial. *Lancet* (London, England). 2018;392(10162):2367-77.
53. Mease PJ, Gottlieb AB, Berman A, Drescher E, Xing J, Wong R, et al. The Efficacy and Safety of Clazakizumab, an Anti-Interleukin-6 Monoclonal Antibody, in a Phase IIb Study of Adults With Active Psoriatic Arthritis. *Arthritis & rheumatology* (Hoboken, NJ). 2016;68(9):2163-73.
54. Mease PJ, Gladman DD, Ritchlin CT, Ruderman EM, Steinfeld SD, Choy EH, et al. Adalimumab for the treatment of patients with moderately to severely active psoriatic arthritis: results of a double-blind, randomized, placebo-controlled trial. *Arthritis and rheumatism*. 2005;52(10):3279-89.

**Table 1.** Characteristics of included studies

Author / year/ (study acronyms)	Intervention/ Comparator	Sample size (% women)	PsA duration (years)	TNFi IR	MTX use	Primary outcome/ time point	Baseline HAQ-DI (mean)	Reported data for physical function -PROM			
								HAQ- DI	HAQ -S	SF-36 PCS	SF-36 PF
TNF inhibitors											
Antoni, et al. 2005 (IMPACT) (26)	IFX 5mg/kg vs. PCB	N=104 (42.3)	11.4	0%	NA	ACR20/ Week 16	1.2	Yes	No	No	No
Antoni, et al. 2005 (IMPACT2) (27)	IFX 5mg/kg vs. PCB	N=200 (39.0)	8	0%	46%	ACR20/ Week 14	1.1	Yes	No	-	No
Kavanaugh, et al. 2006 (IMPACT2) (28)								-	-	Yes	-
Mease, et al. 2005 (ADEPT) (29)	ADA 40mg Q2W vs. PCB	N=313 (44.4)	9.5	0%	51%	ACR20/ Week 12  Δ in modified Total Sharp Score/ Week 24	1.0	Yes	No	Yes	No
Genovese, et al. 2007 (30)	ADA 40mg Q2W vs. PCB	N=100 (46.0)	7.4	0%	47% (cs-DAMRD 66%)	ACR20/ Week 12	1.0	Yes	No	Yes	No
Mease, et al. 2000 (31)	ETN 25mg BIW vs. PCB	N=60 (43.0)	9.3	0%	47%	PsARC/ Week 12	1.3	Yes	No	No	No
Mease, et al. 2010 (32)	ETN 25mg BIW vs. PCB	N=205 (49.0)	9.1	0%	41.5%	ACR20/ Week 12	1.1	Yes	No	Yes	No
Gniadecki, et al. 2012 (PRESTA) (33)	ETN 50mg BIW/QW vs. ETN 50mg QW/QW	N=752 (37.0)	7.0	NA	NA	Psoriasis clear or almost clear/ Week 12	0.92	Yes	No	No	No
Mease, et al. 2019 (SEAM-PsA) (34)	ETN 50mg QW vs. ETN 50mg QW plus MTX vs. MTX alone	N=851 (50.8)	3.2	0%	100% (prior MTX use 0%)	ACR20/ Week 24	1.2	Yes	No	Yes	No
Kavanaugh, et al. 2009	GOL (2 doses) vs. PCB	N=405 (39.8)	7.5	0%	48%	ACR20/ Week 14	1.3	Yes	No	Yes	No

(GO-REVEAL) (35)											
Kavanaugh, et al. 2017 (GO-VIBRANT) (36)	GOL IV 2mg/kg vs. PCB	N=480 (48.0)	5.8	0%	70%	ACR20/ Week 14	1.3	Yes	No	Yes	No
Gladman, et al. 2014 (RAPID-PsA) (37)	CZP (2 doses) vs. PCB	N=409 (55.3)	8.6	19.1%	64.1%;	ACR20, EULAR response/ Week 12	1.3	Yes	No	Yes	No
<b>IL17 inhibitors</b>											
McInnes, 2014 (Phase II) (38)	SEC (10mg/kg) vs. PCB (N=42)	N=42 (64.0)	NA	35%	49% (any csDMARD 51%)	ACR20/ Week 6	1.5	Yes	No	Yes	No
Mease, et al. 2015 (FUTURE I) (39)	SEC (2 doses) vs. PCB (N=606)	N=606 (54.5)	NA	29.4%	60.7%;	ACR20/ Week 24*	1.2	Yes	No	Yes	No
McInnes, et al. 2015 (FUTURE II) (40)	SEC (3 doses) vs. PCB	N=397 (51.6)	NA	35%	47%;	ACR20/ Week 24*	1.2	Yes	No	Yes	No
Kavanaugh, et al. 2016. (FUTURE II) (19) -subgroup analysis	TNFi-naïve vs. TNFi-exposed										
Nash, et al. 2018 (FUTURE III) (41)	SEC (2 doses) vs. PCB	N=414 (54.8)	7.5	32%	47.6%	ACR20/ Week 24	1.2	Yes	No	Yes	No
Mease, et al. 2017 (SPIRIT-P1) (42)	IXE (2 doses) vs. PCB vs. ADA	N=417 (54.0)	6.7	0%	54.2% (any csDMARD 64%)	ACR20/ Week 24*	1.2	Yes	No	Yes	No
Nash, et al. 2017 (SPIRIT-P2) (43)	IXE (2 doses) vs. PCB	N=363 (53.4)	10.0	100%	41%	ACR20/ Week 24*	1.2	Yes	No	Yes	No
Mease, et al. 2014 Phase II (23)	BRO (2 doses) vs. PCB	N=168 (64.0)	8.7	51%	50%	ACR20/ Week 12	1.3	Yes	No	Yes	No
<b>IL12/23 inhibitors</b>											
Gottlieb, et al. 2009 (44)	UST (2 doses) vs. PCB	N=146 (43.8)	5.6	27.4%	20.5%	ACR20/ Week 12	0.9	Yes	No	No	No
McInnes, et al. 2013 (PSUMMIT I) (45)	UST (2 doses) vs. PCB	N=615 (46.3)	4.0	0%	48%	ACR20, EULAR response	1.3	Yes	No	Yes	No

						PASI75/ Week 24*					
Ritchlin, et al. 2014 (PSUMMIT II) (20)	UST (2 doses) vs. PCB	N=312 (52.6)	5.1	57.7%	49.7%;	ACR20, EULAR response PASI75/ Week 24*	1.0	Yes	No	Yes	No
Araugo, et al. 2019 (ECLIPSA) (46)	UST (45mg or 100mg if body weight >100kg) vs. TNFi	N=47 (40.4)	2.5	0%	91.5%	Enthesitis resolution (SPARCC=0)/ Week 24	1.0	Yes	No	Yes	No
<b>IL23 inhibitors</b>											
Deodhar, et al. 2018 (47)	GUS 100mg vs. PCB	N=149 (49.0)	7.0	8.7%	44.3%	ACR20/ Week24*	1.4	Yes	No	Yes	No
<b>T cell inhibition</b>											
Mease, et al. 2011 Phase II (48)	ABT (3 doses) vs. PCB	N=170 (44.2)	8.2	37.4%	58.3%	ACR20/ Day 169	1.2	Yes	No	Yes	No
Mease, et al. 2017 (ASTRAEA) (21)	ABT vs. PCB	N=424 (45.0)	8.5	61%	60%	ACR20/ Week 24*	1.3	Yes	No	Yes	No
Mease, et al. 2006 (49)	ALC/MTX vs. PCB/MTX	N=185 (61.1)	5	0%	100%	ACR20/ Week 12	1.1	Yes	No	No	No
<b>JAK inhibitors</b>											
Mease, et al. 2017 (OPAL Broaden) (24) Strand, et al. 2019 (for SF-36 PCS) (50)	TOF (2 doses) vs. PCB vs. ADA	N=422 (53.0)	6.1	0%	83.9%	ACR20/ 3 months	1.1	Yes	No	Yes	No
Gladman, et al. 2017 (OPAL Beyond) (25) Strand, et al. 2019 (for SF-36 PCS) (51)	TOF (2 doses) vs. PCB	N=395 (55.0)	9.4	100%	73.6%	ACR20, Δ in HAQ-DI/ 3 months	1.3	Yes	No	No	Yes
Mease, et al. 2018 (EQUATOR) (52) Phase II	FIL 200mg vs. PCB	N=131 (50.4)	7.0	0%	54.2% (any csDMARD 74%)	ACR20/ Week 16	1.4	Yes	No	No	Yes

<b>IL6 inhibitors</b>											
Mease, et al. 2016 Phase II (53)	CLZ (3 doses) vs. PCB	N=165 (59.4)	7.1	0%	69.1%	ACR20/ Week 16	1.4	Yes	No	Yes	No
<b>Others</b>											
Mease, et al. 2018 Phase II (22)	ABT-122 (2 doses) vs. PCB vs. ADA	N=240 (49.6)	7.3	0%	100%	ACR20/ Week 12	1.3	No	Yes	No	No

\*early escape at Week 16;

Abbreviations: Δ: change; ACR: American College of Rheumatology Response criteria; ABT: abatacept; ADA: adalimumab; ALC: alefacept; BIW: twice a week; BRO: brodalumab; CI: confidence interval; CLZ: clazakizumab; csDMARD: conventional synthetic disease modifying anti-rheumatic drug; CZP: certolizumab pegol; ETN: etanercept; EULAR: European League Against Rheumatism; FIL: filgotinib; GOL: golimumab; GUS: guselkumab; HAQ-DI: Health Assessment Questionnaire – Disability index; HAQ-S: Health Assessment Questionnaire – Spondyloarthritis; IFX: infliximab; IXE: ixekizumab; IV: intravenous; IL=interleukin; IR=inadequate response; MTX: methotrexate; NA: not available; NS: not significant; PASI: Psoriasis Area and Severity Index; PCB: placebo; PCS: physical component summary of SF-36; PF: physical functioning domain of SF-36; PsA: psoriatic arthritis; PsARC: Psoriatic arthritis Response Criteria; QW: once a week; Q2W: once every 2 weeks; SD: standard deviation; SE: standard error; SEC: secukinumab; SF-36: Medical Outcomes Study 36-item Short Form Survey; SPARCC: Spondyloarthritis Research Consortium of Canada enthesitis index; TOF: tofacitinib; TNFi: tumor necrosis factor inhibitor; UST: ustekinumab; vs.: versus.

**Table 2.** Effect size estimation for studies reporting HAQ-DI

Author / year/ (study acronyms)	Intervention/ comparator (sample size, N)	Primary outcome/ time point	Effect sizes at primary endpoint (unless specified)	<i>a priori</i> hypothesis satisfied
Antoni, et al. 2005 (IMPACT) (26)	IFX 5mg/kg vs. PCB (N=104)	ACR20/ Week 16	SRM <sup>†</sup> (for improvement) <sup>§</sup> at Week 16: IFX = 6.07 PCB = -0.19	1, 2, 3
Antoni, et al. 2005 (IMPACT2) (27)	IFX 5mg/kg vs. PCB (N=200)	ACR20/ Week 14	SRM <sup>†</sup> (for improvement) <sup>§</sup> at Week 14: IFX = 1.08 PCB = -0.19	1, 2, 3
Mease, et al. 2005 (ADEPT) (54)	ADA 40mg Q2W vs. PCB (N=313)	ACR20/ Week 12; Δ in modified Total Sharp Score/ Week 24	SRM at Week 12: ADA = -0.8 PCB = 0.2	1, 2, 3
Genovese, et al. 2007 (30)	ADA 40mg Q2W vs. PCB (N=100)	ACR20/ Week 12	SRM at Week 12: ADA = -0.6 PCB = -0.33	1, 2, 3
Mease, et al. 2000 (31)	ETN 25mg BIW vs. PCB (N=60)	PsARC/ Week 12	ES <sub>2</sub> <sup>‡</sup> at Week 12: ETN = -0.547 PCB = -0.237	1, 2, 3
Mease, et al. 2010 (32)	ETN 25mg BIW vs. PCB (N=205)	ACR20/ Week 12	Effect size for Week 12: NA  ES <sub>2</sub> at Week 24 (end of double-blind phase): ETN = -0.597 PCB = -0.098	1, 2, 3
Gniadecki, et al. 2012 (PRESTA) (33)	ETN 50mg BIW/QW vs. ETN 50mg QW/QW (N=752)	Psoriasis clear or almost clear/ Week 12	ES <sub>2</sub> at Week 12: ETN 50mg BIW/QW: -0.74 ETN 50mg QW/QW: -0.69	1, 3
Mease, et al. 2019 (SEAM-PsA) (34)	ETN 50mg QW vs. ETN 50mg QW plus MTX vs. MTX alone (N=851)	ACR20/ Week 24	SRM at Week 24: ETN = -0.733 ETN plus MTX = -0.685 MTX alone = -0.646	1, 3
Kavanaugh, et al. 2009 (GO-REVEAL) (35)	GOL (2 doses) vs. PCB (N=405)	ACR20/ Week 14	SRM at Week 14: GOL 100mg = -0.75 GOL 50mg = -0.62 PCB = -0.09	1, 2, 3
Kavanaugh, et al. 2017	GOL IV 2mg/kg	ACR20/	SRM at Week 14:	1, 2, 3



(GO-VIBRANT) (36)	vs. PCB (N=480)	Week 14	GOL IV = -1.13 PCB = -0.26	
Gladman, et al. 2014 (RAPID-PsA) (37)	CZP (2 doses) vs. PCB (N=409)	ACR20, EULAR response/ Week 12	SRM at Week 12: CZP 400mg Q4W = -0.83 CZP 200mg Q2W = -0.80 PCB = -0.44	1, 2, 3
McInnes, 2014 (Phase II) (38)	SEC (10mg/kg) vs. PCB (N=42)	ACR20/ Week 6	SRM at Week 6: SEC = -0.680 PCB = 0.018	1, 2, 3
Mease, et al. 2015 (FUTURE I) (39)	SEC (2 doses with loading) vs. PCB (N=606)	ACR20/ Week 24*	SRM at Week 24: SEC 150mg = -0.703 SEC 75mg = -0.721 PCB = 0.239	1, 2, 3
McInnes, et al. 2015 FUTURE II) (40)	SEC (3 doses) vs. PCB (N=397)	ACR20/ Week 24*	SRM at Week 24: SEC 300mg = -1.12 SEC 150mg = -0.96 SEC 75mg = -0.644 PCB = -0.522	1, 2, 3
Kavanaugh, et al. 2016 (FUTURE II) (19) -subgroup analysis	TNFi-naïve vs. TNFi-exposed	ACR20/ Week 24*	SRM at Week 24 (TNFi-naïve vs. -exposed): SEC 300mg: -1.20 vs. -1.02 SEC 150mg: -1.15 vs. -0.71 SEC 75mg: -0.77 vs. -0.44 PCB: -0.63 vs -0.35	4
Nash, et al. 2018 (FUTURE III) (41)	SEC (2 doses) vs. PCB (N=414)	ACR20/ Week 24*	SRM at Week 24: SEC 300mg = -0.81 SEC 150mg = -0.57 PCB = -0.24	1, 2, 3
Mease, et al. 2017 (SPIRIT-P1) (42)	IXE (2 doses) vs. PCB vs. ADA (N=417)	ACR20/ Week 24*	SRM at Week 24: IXE Q2W = -0.98 IXE Q4W = -0.85 PCB = -0.35 ADA = -0.74	1, 2, 3
Nash, et al. 2017 (SPIRIT-P2) (43)	IXE (2 doses) vs. PCB (N=363)	ACR20/ Week 24*	SRM at Week 24: IXE Q2W = -0.36 IXE Q4W = -0.55 PCB = -0.18	1, 2, 3

Mease, et al. 2014 Phase II (23)	BRO (2 doses) vs. PCB (N=168)	ACR20/ Week 12	SRM <sup>‡</sup> at Week 12: BRO 280mg = -0.60 BRO 140mg = -0.38 PCB = -0.21	1, 2, 3
Gottlieb, et al. 2009 (44)	UST (2 doses) vs. PCB (N=146)	ACR20/ Week 12	SRM <sup>‡</sup> at Week 12: UST = -0.66 PCB = -0.14	1, 2, 3
McInnes, et al. 2013 (PSUMMIT I) (45)	UST (2 doses) vs. PCB (N=615)	ACR20, EULAR response PASI75/ Week 24*	SRM <sup>‡</sup> at Week 24: UST 90mg = -0.59 UST 45mg = -0.62 PCB = -0.21	1, 2, 3
Ritchlin, et al. 2014 (PSUMMIT II) (20)	UST (2 doses) vs. PCB (N=312)	ACR20, EULAR response PASI75/ Week 24*	SRM <sup>‡</sup> at Week 24: UST 90mg = -0.66 UST 45mg = -0.59 PCB = 0.00  SRM <sup>‡</sup> at Week 24 Subgroups analysis (TNFi-naïve vs. -exposed): UST 90mg = -0.66 vs. -0.66 UST 45mg = -0.66 vs. -0.59 PCB = 0 vs. 0	1, 2, 3
Araugo, et al. 2019 (ECLIPSA) (46)	UST (45mg or 100mg if body weight >100kg) vs. TNFi (N=47)	Enthesitis resolution (SPARCC = 0) at Week 24	ES <sub>2</sub> <sup>‡</sup> at Week 24: UST = -1.81 TNFi = -1.74	1, 2, 3
Deodhar, et al. 2018 (47)	GUS 100mg vs. PCB (N=149)	ACR20/ Week 24*	SRM at Week 24: GUS = -0.82 PCB = -0.11	1, 2, 3
Mease, et al. 2011 Phase II (48)	ABT (3 doses) vs. PCB (N=170)	Day 169	Insufficient data for effect size calculation	none
Mease, et al. 2017 (ASTRAEA) (21)	ABT vs. PCB (N=424)	ACR20/ Week 24*	SRM <sup>‡</sup> at Week 24: ABT = -0.69 PCB = -0.40  SRM <sup>‡</sup> at Week 24: Subgroup analysis: (TNFi-naïve vs. -exposed) ABT = -0.62 vs. -0.64 PCB = -0.39 vs. -0.34	1, 3
Mease, et al. 2006 (49)	ALC/MTX	ACR20/	% Δ in HAQ-DI:	none

	vs. PCB/MTX (N=185)	Week 12	ALC/MTX = -24.5% PCB/MTX = -7.7% (NS) Inadequate data for ES calculations	
Mease, et al. 2017 (OPAL Broaden) (24)	TOF (2 doses) vs. PCB vs. ADA (N= 422)	ACR20/ 3 months	SRM at 3 months: TOF 10mg = -0.78 TOF 5mg = -0.69 PCB = -0.36 ADA = -0.76	1, 2, 3
Gladman, et al. 2017 (OPAL Beyond) (25)	TOF (2 doses) vs. PCB (N=395)	ACR20, Δ in HAQ-DI/ 3 months	SRM at 3 months: TOF 10mg = -0.64 TOF = -0.70 PCB = -0.26	1, 2, 3
Mease, et al. 2018 (EQUATOR) (52) Phase II	FIL 200mg vs. PCB (N=131)	ACR20/ Week 16	SRM at Week 16: FIL = -1.14 PCB = -0.56  ES <sub>2</sub> at Week 16: FIL = -1.04 PCB = -0.47	1, 2, 3
Mease, et al. 2016 Phase II (53)	CLZ (3 doses) vs. PCB (N=165)	ACR20/ Week 16	SRM <sup>‡</sup> at Week 16: CLZ 200mg = -0.51 CLZ 100mg = -0.77 CLZ 50mg = -0.83 PCB = -0.52	1, 2, 3

<sup>‡</sup> SRM calculated using percentage change score and SD of percentage change; <sup>§</sup> SRM for improvement, a negative value indicate deterioration; <sup>¥</sup> Effect sizes estimated based on mean and SD of change scores calculated from median and IQR from original publication; \* early escape for patients with inadequate response in the PCB group to active treatment group at Week 16; \*\* option to switch TNFi at Week 24.

Abbreviations: Δ: change; ACR: American College of Rheumatology Response criteria; ABT: abatacept; ADA: adalimumab; ALC: alefacept; BIW: twice a week; BRO: brodalumab; CI: confidence interval; CLZ: clazakizumab; CZP: certolizumab pegol; ES<sub>2</sub>: Effect size 2 (the mean difference divided by the pooled standard deviation, i.e Cohen's *d*); ETN: etanercept; EULAR: European League Against Rheumatism; FIL: filgotinib; GOL: golimumab; GUS: guselkumab; HAQ-DI: Health Assessment Questionnaire – Disability Index; IFX: infliximab; IXE: ixekizumab; IV: intravenous; IQR: interquartile range; MTX: methotrexate; NA: not available; NS: not significant; PASI: Psoriasis Area and Severity Index; PCB: placebo; PsARC: Psoriatic arthritis Response Criteria; QW: once a week; Q2W: once every 2-week; SD: standard deviation; SE: standard error; SEC: secukinumab; SPARCC: Spondyloarthritis Research Consortium of Canada enthesitis index; SRM: Standardized response mean (mean difference divided by the standard deviation of the differences between baseline and assessment end point); TOF: tofacitinib; TNFi: tumor necrosis factor inhibitor; UST: ustekinumab; vs.: versus.

**Table 3.** Effect size estimation for studies reported HAQ-S

Author / year/ (study acronyms)	Intervention/ comparator (sample size, N)	Primary outcome/ time point	PROM	Effect sizes at primary end point (unless specified)	<i>a priori</i> hypothesis satisfied
Mease, et al. 2018 Phase II (22)	ABT-122 (2 doses) vs. PCB vs. ADA (N=240)	ACR20/ Week 12	HAQ-S	ES <sub>1</sub> at Week 12: ABT-122 240mg: -0.93 ABT-122 120mg: -0.92 PCB: -0.47 ADA: -0.97	3

Abbreviations: ACR: American College of Rheumatology Response criteria; ADA: adalimumab; ES<sub>1</sub>: Effect size 1 (the mean difference divided by standard deviation of baseline score); HAQ-S: Health Assessment Questionnaire – Spondyloarthritis; PCB: placebo; vs.: versus.

**Table 4.** Effect size estimation for studies reporting SF-36 PCS

Author / year/ (study acronyms)	Intervention/ comparator (sample size, N)	Primary Outcome/ Time point	Effect sizes at primary end point (unless specified)	Fulfilling of <i>a priori</i> hypothesis
Antoni, et al. 2005 (IMPACT2) (27)	IFX 5mg/kg vs. PCB (N=200)	ACR20/ Week 14	SRM at Week 14: IFX = 0.98 PCB = 0.13	1, 2, 3
Mease, et al. 2005 (ADEPT) (54)	ADA 40mg Q2W vs. PCB (N=313)	ACR20/ Week 12; $\Delta$ in modified Total Sharp Score/ Week 24	SRM at Week 12: ADA = 0.93 PCB = 0.16	1, 2, 3
Genovese, et al. 2007 (30)	ADA 40mg Q2W vs. PCB (N=100)	ACR20/ Week 12	SRM at Week 12: ADA = 0.67 PCB = 0.39	1, 2, 3
Mease, et al. 2010 (32)	ETN 25mg BIW vs. PCB (N=205)	ACR20/ Week 12	SRM at Week 12: NA  ES <sub>2</sub> at Week 24 (end of double-blind phase): ETN = 0.880 PCB = 0.073	1, 2, 3
Mease, et al. 2019 (SEAM-PsA) (34)	ETN 50mg QW vs. ETN 50mg QW plus MTX vs. MTX alone (N=851)	ACR20/ Week 24	SRM at Week 24: ETN = 0.832 ETN plus MTX = 0.813 MTX alone = 0.629	1, 2, 3
Kavanaugh, et al. 2009 (GO-REVEAL) (35)	GOL (2 doses) vs. PCB (N=405)	ACR20/ Week 14	SRM at Week 14: GOL 100mg = 0.82 GOL 50mg = 0.74 PCB = 0.08	1, 2, 3
Kavanaugh, et al. 2017 (GO-VIBRANT) (36)	GOL IV 2mg/kg vs. PCB (N=480)	ACR20/ Week 14	SRM at Week 14: GOL IV = 1.14 PCB = 0.46	1, 2, 3
Gladman, et al. 2014 (RAPID-PsA) (37)	CZP (2 doses) vs. PCB (N=409)	ACR20, EULAR response/ Week 12	SRM at Week 12: CZP 400mg Q4W = 0.87 CZP 200mg Q2W = 0.82 PCB = 0.30	1, 2, 3

McInnes IB, 2014 (Phase II) (38)	SEC (10mg/kg) vs. PCB (N=42)	ACR20/ Week 6	SRM at Week 6: SEC = 0.541 PCB = -0.017	1, 2, 3
Mease, et al. 2015 (FUTURE I) (39)	SEC (2 doses with loading) vs. PCB (N=606)	ACR20/ Week 24*	SRM at Week 24: SEC 150mg = 0.785 SEC 75mg = 0.732 PCB = 0.178	1, 2, 3
McInnes, et al. 2015 FUTURE II) (40)	SEC (3 doses) vs. PCB (N=397)	ACR20/ Week 24*	SRM at Week 24: SEC 300mg = 0.98 SEC 150mg = 0.875 SEC 75mg = 0.587 PCB = 0.203	1, 2, 3
Kavanaugh, et al. 2016. (FUTURE II) (19) -subgroup analysis	TNFi-naïve vs. TNFi-exposed	ACR20/ Week 24*	SRM at Week 24: (TNFi-naïve vs. -exposed): SEC 300mg: 1.07 vs. 0.95 SEC 150mg: 1.07 vs. 0.60 SEC 75mg: 0.71 vs. 0.45 PCB: 0.22 vs 0.27	4
Nash, et al. 2018 (FUTURE III) (41)	SEC (2 doses) vs. PCB (N=414)	ACR20/ Week 24*	SRM at Week 24: SEC 300mg = 0.93 SEC 150mg = 0.49 PCB = 0.30	1, 2, 3
Mease, et al. 2017 (SPIRIT-P1) (42)	IXE (2 doses) vs. PCB vs. ADA (N=417)	ACR20/ Week 24*	SRM at Week 24: IXE Q2W = 1.79 IXE Q4W = 1.33 PCB = 0.55 ADA = 1.78	1, 2, 3
Nash, et al. 2017 (SPIRIT-P2) (43)	IXE (2 doses) vs. PCB (N=363)	ACR20/ Week 24*	SRM at Week 24: IXE Q2W = 0.30 IXE Q4W = 0.27 PCB = 0.06	1, 2, 3
Mease, et al. 2014 Phase II (23)	BRO (2 doses) vs. PCB (N=168)	ACR20/ Week 12	Difference from PCB (95% CI): BRO 280mg: 2.4 (0.1 to 4.6) BRO 140mg: 1.4 (-0.8 to 3.6) Insufficient data for effect size calculation	none
McInnes, et al. 2013 (PSUMMIT I) (45)	UST (2 doses) vs. PCB (N=615)	ACR20, EULAR response, PASI75/ Week 24*	SRM <sup>y</sup> at Week 24: UST 90mg = 0.75 UST 45mg = 0.48	1, 2, 3

			PCB = 0.24	
Ritchlin, et al. 2014 (PSUMMIT II) (20)	UST (2 doses) vs. PCB (N=312)	ACR20, EULAR response, PASI75/ Week 24*	SRM <sup>¶</sup> at Week 24: UST 90mg = 0.60 UST 45mg = 0.50 PCB = 0.30	1, 2, 3
Araugo, et al. 2019 (ECLIPSA) (46)	UST (45mg or 100mg if body weight >100kg) vs. TNFi (N=47)	Enthesitis resolution (SPARCC = 0)/ Week 24	ES <sub>2</sub> <sup>¶</sup> at Week 24: UST = 2.95 TNFi = 1.56	1, 3
Deodhar, et al. 2018 (47)	GUS 100mg vs. PCB (N=149)	ACR20/ Week 24*	SRM at Week 24: GUS = 0.88 PCB = 0.06	1, 2, 3
Mease, et al. 2011 Phase II (48)	ABT (3 doses) vs. PCB (N=170)	ACR20/ Day 169	SRM at Day 169: ABT 30/10 mg/kg: 0.59 ABT 10mg/kg: 0.77 ABT 3 mg/kg: 0.53 PCB: 0.02	1, 2, 3
Mease, et al. 2017 (ASTRAEA) (21)	ABT vs. PCB (N=424)	ACR20/ Week 24*	SRM <sup>¶</sup> at Week 24: ABT: 0.72 PCB: 0.53	1, 2, 3
Strand, et al. 2019 (OPAL Broaden) (50)	TOF (2 doses) vs. PCB vs. ADA (N= 422)	ACR20/ 3 months	SRM at 3 months: TOF 10mg = 0.75 TOF 5mg = 0.73 PCB = 0.33 ADA = 0.81	1, 2, 3
Strand, et al. 2019 (OPAL Beyond) (51)	TOF (2 doses) vs. PCB (N=395)	ACR20, Δ in HAQ-DI/ 3 months	SRM at 3 month: TOF 10mg = 0.67 TOF 5mg = 0.67 PCB = 0.22	1, 2, 3
Mease, et al. 2016 Phase II (53)	CLZ (3 doses) vs. PCB (N=165)	ACR20/ Week 16	SRM <sup>¶</sup> at Week 16: CLZ 200mg = 0.52 CLZ 100mg = 0.59 CLZ 50mg = 0.82 PCB = 0.57	1, 2, 3

<sup>¶</sup> SRM calculated using percentage change score and SD of percentage change; <sup>§</sup> SRM for improvement, a negative value indicate deterioration; <sup>¶</sup> Effect sizes estimated based on mean and SD of change scores calculated from median and IQR from original publication; \* early escape for patients with inadequate response in the PCB group to active treatment group at Week 16; \*\* option to switch TNFi at Week 24.

Abbreviations:  $\Delta$ : change. ACR: American College of Rheumatology Response criteria; ABT: abatacept; ADA: adalimumab; ALC: alefacept; BIW: twice a week; BRO: brodalumab; CI: confidence interval; CLZ: clazakizumab; CZP: certolizumab pegol; ES<sub>2</sub>: Effect size 2 (the mean difference divided by the pooled standard deviation, i.e. Cohen's *d*); ETN: etanercept; EULAR: European League Against Rheumatic Diseases; GOL: golimumab; GUS: guselkumab; IFX: infliximab; IXE: ixekizumab; IV: intravenous; IQR: interquartile range; MTX: methotrexate; NA: not available; PASI: psoriasis area and severity index; PCB: placebo; PCS: physical component summary of SF-36; PF: physical functioning domain of SF-36; QW: once a week; Q2W: once every 2-week; SD: standard deviation; SE: standard error; SEC: secukinumab; SF-36: Medical Outcome Survey Short Form 36 items; SPARCC: Spondyloarthritis Research Consortium of Canada enthesitis index; SRM: Standardized response mean (mean difference divided by the standard deviation of the differences between baseline and assessment end point); TNFi: tumor necrosis factor inhibitor; UST: ustekinumab; vs.: versus.



**Table 5.** Effect size estimation for studies reporting SF-36 PF.

Author / year/ (study acronyms)	Intervention/ comparator (sample size, N)	Primary outcome/ time point	Effect sizes at primary endpoint (unless specified)	Fulfilling of <i>a priori</i> hypothesis
Antoni, et al. 2007 (IMPACT2) (27)	IFX 5mg/kg vs. PCB (N=200)	ACR20/ Week 14	Insufficient data for effect size calculation	1, 2
Mease, et al. 2017 (OPAL Broaden)	TOF (2 doses) vs. PCB vs. ADA (N= 422)	ACR20/ 3 months	SRM at 3 months: TOF 10mg = 0.64 TOF 5mg = 0.64 PCB = 0.23 ADA = 0.58	1, 3
Gladman, et al. 2017 (OPAL Beyond)	TOF (2 doses) vs. PCB (N=395)	ACR20, $\Delta$ in HAQ-DI/ 3 months	SRM at 3 months: TOF 10mg = 0.53 TOF 5mg = 0.64 PCB = 0.22	1, 3
Mease, et al. 2018 (EQUATOR) Phase II	FIL 200mg vs. PCB (N=131)	ACR20/ Week 16	Insufficient data for effect size calculation	1, 2

<sup>†</sup> SRM calculated using percentage change score and SD of percentage change; <sup>§</sup> SRM for improvement, a negative value indicate deterioration; <sup>‡</sup> Effect sizes estimated based on mean and SD of change scores calculated from median and IQR from original publication; \* early escape for patients with inadequate response in the control group to active treatment group at week 16; \*\* option to switch TNFi at Week 24.

Abbreviations:  $\Delta$ : change; ACR: American College of Rheumatology Response criteria; ADA: adalimumab; ES<sub>2</sub>: Effect size 2 (the mean difference divided by the pooled standard deviation, i.e Cohen's *d*); FIL: filgotinib; HAQ-DI: Health Assessment Questionnaire – Disability Index; IFX: infliximab; IQR: interquartile range; PCB: placebo; PF: physical functioning domain of SF-36; SD: standard deviation; SF-36: Medical Outcomes Study 36-item Short Form Survey; SRM: Standardized response mean (mean difference divided by the standard deviation of the differences between baseline and assessment end point); TOF: tofacitinib; vs.: versus.

**Table 6.** Summary of Measurement Properties Table for clinical trial discrimination

1 <sup>st</sup> Author / year/ (study acronyms/ drug)	HAQ-DI	HAQ-S	SF-36 PCS	SF-36 PF
Antoni, 2005 (IMPACT)	+			
Antoni, 2005 (IMPACT2)	+		+	
Kavanaugh, 2006 (IMPACT2)				+
Mease, 2005 (ADEPT)	+		+	
Genovese, 2007 (ADA)	+		+	
Mease, 2000 (ETN)	+		+	
Mease. 2010 (ETN)	+		+	
Gniadecki, 2012 (PRESTA)	+			
Mease. 2019 (SEAM-PsA)	+/-		+	
Kavanaugh, 2009 (GO-REVEAL)	+		+	
Kavanaugh, 2017 (GO-VIBRANT)	+		+	
Gladman, 2014 (RAPID-PsA)	+		+	
McInnes, 2014 (SEC)	+		+	
Mease, 2015 (FUTURE I)	+		+	
McInnes, 2015 FUTURE II)	+		+	
Kavanaugh, 2016 (FUTURE II) -subgroup analysis	+		+	
Nash P, 2018 (FUTURE III)	+		+	
Mease, 2017 (SPIRIT-P1)	+		+	
Nash, 2017 (SPIRIT-P2)	+		+	
Mease, 2014 (BRO)	+		+	
Gottlieb, 2009 (UST)	+			
McInnes, 2013 (PSUMMIT I)	+		+	
Ritchlin, 2014 (PSUMMIT II)	+		+	
Araugo, 2019 (ECLIPSA)	+		+	
Deodhar, 2018 (GUS)	+		+	
Mease, 2011 (ABT)	+/-		+	
Mease, 2017 (ASTRAEA)	+		+	
Mease, 2006 (ALC)	+/-			
Mease, 2017 (OPAL Broaden)	+			+
Strand, 2019 (OPAL Broaden)			+	
Gladman, 2017 (OPAL Beyond)	+			+
Strand, 2019 (OPAL Beyond)			+	
Mease, 2018 (EQUATOR)	+			+
Mease, 2016 (CLZ)	+		+	
Mease, 2018 (ABT-122)		+		
Total available articles	31	1	26	4
Total articles for evidence synthesis	29	1	25	2
Final rating	GREEN	AMBER	GREEN	AMBER

Color code in each box indicate study quality assessed by OMERACT good methods. GREEN means “yes, likely low risk of bias”; AMBER means “some cautions but can be used as evidence” and RED means “No,

---

don't use as evidence". WHITE (empty boxes), indicates absence of information on that property from that study. (+) indicates findings of the study had adequate performance of the instrument; (+/-) indicates equivocal performance; (-) indicates poor performance (less than adequate).

## Appendices

Table A.1. Detailed search terms of the systematic review

Pubmed: "Arthritis, Psoriatic"[Mesh] AND "randomized controlled trial"[publication type] AND English[lang]
Scopus: ( TITLE-ABS-KEY ( <i>psoriatic</i> AND <i>arthritis</i> ) AND TITLE-ABS-KEY ( <i>randomized</i> AND <i>controlled</i> AND <i>trial</i> ) ) AND DOCTYPE ( <i>ar</i> ) AND ( LIMIT-TO ( PUBSTAGE , " <i>final</i> " ) ) AND ( LIMIT-TO ( LANGUAGE , " <i>English</i> " ) )

**Table A.2. Effect sizes estimation of all studies.**

Author / year/ (study acronyms)	Intervention/ comparator (sample size, N)	Primary Outcome/ Time point	PROM	Baseline PROM Mean (SD) (unless specified)	Endpoint PROM/ $\Delta$ score Mean (SD) (unless specified)	P value of $\Delta$ scores of interventions Vs. PCB group	Effect sizes at primary endpoint (unless specified)
<b>TNF inhibitors</b>							
Antoni, et al. 2005 (IMPACT)	IFX 5mg/kg vs. PCB (N=104)	ACR20/ Week 16	HAQ-DI	IFX: 1.2 (0.7) PCB: 1.2 (0.7)	<u>% improvement in HAQ-DI</u> IFX: 49.8 (8.2) PCB: -1.6 (8.3)	<0.001	SRM <sup>¶</sup> (for improvement) <sup>δ</sup> at Week 16: IFX = 6.07 PCB = -0.19
			SF-36 PF or PCS	NA	NA	NA	NA
Antoni, et al. 2005 (IMPACT2)	IFX 5mg/kg vs. PCB (N=200)	ACR20/ Week 14	HAQ-DI	IFX: 1.1 (0.6) PCB: 1.1 (0.6)	<u>% <math>\Delta</math> in HAQ-DI</u> IFX: 48.6 (43.3) PCB: -18.4 (90.5)	<0.001	SRM <sup>¶</sup> (for improvement) <sup>δ</sup> at Week 14: IFX = 1.08 PCB = -0.19
			PCS	IFX: 33.0 (9.4) PCB: 31.0 (9.0)	<u><math>\Delta</math> score</u> IFX: 9.1 (9.3) PCB: 1.1 (8.4)	< 0.001	SRM at Week 14: IFX = 0.98 PCB = 0.13
			SF-36 PF	NA	<u><math>\Delta</math> score</u> IFX: 17.7 PCB: -0.2	<0.001	Insufficient data for effect size calculation
Mease, et al. 2005 (ADEPT)	ADA 40mg Q2W vs. PCB (N=313)	ACR20/ Week 12; $\Delta$ in modified total sharp score Week 24	HAQ-DI	ADA: 1.0 (0.6) PCB: 1.0 (0.7)	<u><math>\Delta</math> score</u> Ada: -0.4 (0.5) PCB: -0.1 (0.5)	<0.001	SRM at Week 12: ADA = -0.8 PCB = 0.2
			SF-36 PCS	ADA: 33.2 (9.9) PCB: 33.3 (9.8)	<u><math>\Delta</math> score</u> ADA: 9.3 (10.0) PCB: 1.4 (8.7)	<0.001	SRM at Week 12: ADA = 0.93 PCB = 0.16
Genovese, et al. 2007	ADA40mg Q2W vs. PCB (N=100)	ACR20/ Week 12	HAQ-DI	ADA: 0.9 (0.5) PCB: 1.0 (0.7)	<u><math>\Delta</math> score</u> ADA: -0.3 (0.5) PCB: -0.1 (0.3)	0.010	SRM at Week 12: ADA = -0.6 PCB = -0.33
			SF-36 PCS	ADA: 34.9 (9.2) PCB: 32.7 (11.3)	<u><math>\Delta</math> score</u> ADA: 5.7 (8.5) PCB: 2.8 (7.1)	0.082	SRM at Week 12: ADA = 0.67 PCB = 0.39
Mease, et al. 2000	ETN 25mg BIW vs. PCB (N=60)	PsARC/ Week 12	HAQ-DI	Median (IQR) ETN: 1.3 (0.9, 1.6) PCB: 1.2 (0.8, 1.6)	<u>Endpoint Median (IQR)</u> ETN: 0.1 (0.0, 1.0) PCB: 1.1 (0.5, 1.5) <u>Calculated <math>\Delta</math> scores (pooled SD)</u>	<0.0001	ES <sub>2</sub> <sup>¶</sup> at Week 12: ETN = -0.547 PCB = -0.237

					ETN: -0.367 (0.671) PCB: -0.167 (0.705)		
			SF-36 PF or PCS	NA	NA	NA	NA
Mease, et al. 2010	ETN 25mg BIW vs. PCB (N=205)	ACR20/ Week 12  Week 24 is the end of double- blind phase	HAQ- DI	Mean [SE] ETN: 1.1 [0.1] PCB: 1.1 [0.1]	Week 4: ETN: 35.1% Δ from BL PCB: 8.0% Δ from BL  Week 12: ETN: 53.5% Δ from BL PCB: 6.3% Δ from BL  <u>Endpoint score at Week 24</u> Mean [SE] ETN: 0.5 [0.1] PCB: 1.0 [0.1]  <u>Calculated Δ sores (pooled SD) at Week 24</u> ETN: -0.6 (1.005) PCB: -0.1 (1.02)	Week 4: <0.001 for % Δ from BL  Week 12: <0.001 for % Δ from BL  Week 24: <0.001	Effect size for Week 12: Similar to that at Week 24 as per graph in paper  ES <sub>2</sub> at Week 24 (end of double-blind phase): Etanercept = -0.597 PCB = -0.098
			SF-36 PCS	Mean [SE] ETN: 35.8 [1.0] PCB: 35.7 [0.9]	Week 4: Δ from BL ETN: 5.8 PCB: 0.5  Week 12: similar to Week 24 as per graph  <u>Endpoint score at Week 24</u> Mean [SE] ETN: 45.1 [1.1] PCB: 36.4 [1.0]  <u>Calculated Δ sores (pooled SD) at Week 24</u> ETN: 9.3 (10.564) PCB: 0.7 (9.178)	Week 24 <0.001	SRM at Week 12: NA  ES <sub>2</sub> at Week 24 (end of double-blind phase): ETN = 0.880 PCB = 0.073
Gniadecki, et al. 2012 (PRESTA)	ETN 50mg BIW/QW vs.	Psoriasis clear or almost clear/	HAQ- DI	ETN BIW/QW: 0.90 (0.69) ETN QW/QW: 0.93 (0.70)	<u>Endpoint score at Week 24</u> ETN BIW/QW: 0.48 (0.60) ETN QW/QW: 0.49 (0.57)	0.792	ES <sub>2</sub> at Week 12: ETN 50mg BIW/QW: -0.74 ETN 50mg QW/QW: -0.69

	ETN 50mg QW/QW (N=752)	week 12			<u>Calculated <math>\Delta</math> sores (pooled SD)</u> ETN BIW/QW: -0.42 (0.65) ETN QW/QW: -0.44 (0.64)		
			SF-36 PCS or PF	NA	NA	NA	NA
Mease, et al. 2019 (SEAM-PsA)	ETN 50mg QW plus MTX vs. ETN 50mg QW vs. MTX alone (N=851)	ACR20/ Week 24	HAQ-DI	Mean [SE] (Calculated SD) ETN/MTX: 1.15 [0.04] (0.672) ETN: 1.15 [0.04] (0.674) MTX: 1.27 [0.04] (0.673)	<u><math>\Delta</math> score</u> Mean [SE] (Calculated SD) ETN plus MTX: -0.47 [0.04] (0.641) ETN: -0.44 [0.04] (0.642) MTX: -0.41 [0.04] (0.635)	0.67 for ETN  0.34 for ETN/MTX	SRM at Week 24: ETN = -0.733 ETN plus MTX = -0.685 MTX alone = -0.646
			SF36 PCS	Mean [SE] (Calculated SD) ETN/MTX: 37.4 [0.6] (10.058) ETN: 37.8 [0.5] (8.426) MTX: 35.6 [0.5] (8.396)	<u><math>\Delta</math> score</u> Mean [SE] (Calculated SD) ETN/MTX: 8.0 [0.6] (9.619) ETN: 7.8 [0.6] (9.60) MTX: 6.0 [0.6] (9.544)	0.033 for ETN  0.015 for ETN/MTX	SRM at Week 24: ETN = 0.832 ETN plus MTX = 0.813 MTX alone = 0.629
Kavanaugh, et al. 2009 (GO-REVEAL)	GOL (2 doses) vs. PCB (N=405)	ACR20/ Week 14	HAQ-DI	GOL 100mg: 1.05 (0.62) GOL 50mg: 0.98 (0.65) PCB: 1.03 (0.55)	<u><math>\Delta</math> score at Week 14:</u> GOL 100mg: -0.38 (0.51) GOL 50mg: -0.31 (0.50) PCB: -0.04 (0.44)  <u><math>\Delta</math> score at Week 24<sup>†</sup></u> GOL 100mg: -0.39 (0.50) GOL 50mg: -0.33 (0.55) PCB: 0.01 (0.49)	<0.001	SRM at Week 14: GOL 100mg = -0.75 GOL 50mg = -0.62 PCB = -0.09  SRM at Week 24*: GOL 100mg = -0.78 GOL 50mg = -0.60 PCB = 0.02
			SF36 PCS	GOL 100mg: 32.79 (8.90) GOL 50mg: 33.03 (10.68) PCB: 31.93 (9.25)	<u><math>\Delta</math> score</u> GOL 100mg: 7.85 (9.55) GOL 50mg: 6.53 (8.88) PCB: 0.63 (7.68)	<0.001	SRM at Week 14: GOL 100mg = 0.82 GOL 50mg = 0.74 PCB = 0.08
Kavanaugh, et al. 2017 (GO-VIBRANT)	GOL ivi 2mg/kg vs. PCB (N=480)	ACR20/ Week 14	HAQ-DI	GOL: 1.3 (0.6) PCB: 1.3 (0.6)	<u><math>\Delta</math> score</u> GOL: -0.60 (0.53) PCB: -0.12 (0.47)	<0.001	SRM at Week 14: GOL ivi = -1.13 PCB = -0.26
			SF-36 PCS	GOL: 33.1 (6.9) PCB: 34.0 (7.2)	<u><math>\Delta</math> score</u> GOL: 8.7 (7.6) PCB: 2.7 (5.9)	<0.001	SRM at Week 14: GOL ivi = 1.14 PCB = 0.46
Gladman, et al. 2014 (RAPID-PsA)	CZP (2 doses) vs. PCB (N=409)	ACR20, EULAR	HAQ-DI	CZP 400 Q4W: 1.3 (0.6) CZP 200 Q2W: 1.3 (0.7) PCB: 1.3 (0.7)	<u><math>\Delta</math> score</u> CZP 400 Q4W: -0.39 (0.47) CZP 200 Q2W: -0.45 (0.56)	<0.001	SRM at Week 12: CZP 400mg Q4W = -0.83 CZP 200mg Q2W = -0.80

		response/ Week 12			PCB: -0.16 (0.36)		PCB = -0.44
			SF-36 PCS	CZP 400 Q4W: 33.2 (7.5) CZP 200 Q2W: 33.1 (7.7) PCB: 33.8 (7.9)	<u>Δ score</u> CZP 400 Q4W: 6.7 (7.7) CZP 200 Q2W: 7.5 (9.1) PCB: 1.8 (6.1)	<0.001	SRM at Week 12: CZP 400mg Q4W = 0.87 CZP 200mg Q2W = 0.82 PCB = 0.30
<b>IL17 inhibitors</b>							
Author / year/ study acronyms	Intervention/ comparator (sample size, N)	Primary Outcome/ Time point	PROM	Baseline PROM Mean (SD) (unless specified)	Endpoint PROM/ Δ score Mean (SD) (unless specified)	P value of Δ scores of interventions Vs. PCB group	Effect sizes at primary endpoint (unless specified)
McInnes, 2014 (Phase II)	SEC (10mg/kg) vs. PCB (N=42)	ACR20/ Week 6	HAQ- DI	SEC: 1.64 (0.66) PCB: 1.20 (0.71)	<u>% Δ score</u> SEC: -20.6 (30.3) PCB: 0.25 (14.0)	0.002	SRM <sup>†</sup> at Week 6: SEC = -0.680 PCB = 0.018
			SF-36 PCS	SEC: 30.8 (10.0) PCB: 36.1 (9.0)	<u>Δ score</u> SEC: 15.2 (28.1) PCB: -0.51 (29.6)	0.030	SRM at Week 6: SEC = 0.541 PCB = -0.017
Mease, et al. 2015 (FUTURE I)	SEC (2 doses with loading) vs. PCB (N=606)	ACR20/ Week 24*	HAQ- DI	SEC 150mg: 1.2 (0.7) SEC 75mg: 1.3 (0.7) PCB: 1.2 (0.6)	<u>Δ score</u> Mean [SE] (Calculated SD)  SEC 150mg: -0.40 [0.04] (0.569) SEC 75mg: -0.41 [0.04] (0.569) PCB: -0.17 [0.05] (0.711)	<0.001	SRM at Week 24: SEC 150mg = -0.703 SEC 75mg = -0.721 PCB = 0.239
			SF-36 PCS	SEC 150mg: 36.2 (8.1) SEC 75mg: 36.9 98.1) PCB: 36.8 (8.0)	<u>Δ score</u> LS Mean [SE] (Calculated SD)  SEC 150mg: 5.91 [0.53] (7.533) SEC 75mg: 5.41 [0.52] (7.391) PCB: 1.82 [0.72] (10.233)	<0.001	SRM at Week 24: SEC 150mg = 0.785 SEC 75mg = 0.732 PCB = 0.178
McInnes, et al. 2015 FUTURE II)	SEC (3 doses) vs. PCB (N=397)	ACR20/ Week 24*	HAQ- DI	SEC 300mg: 1.3 (0.6) SEC 150mg: 1.2 (0.6) SEC 75mg: 1.2 (0.6) PCB: 1.2 (0.7)	<u>Δ score</u> LS Mean [SE] (calculated SD)  SEC 300mg: -0.56 [0.05] (0.5) SEC 150mg: -0.48 [0.05] (0.5) SEC 75mg: -0.32 [0.05] (0.497) PCB: -0.31 [0.06] (0.594)	0.004 for SEC 300mg  0.056 for SEC 150mg  0.92 for SEC 75mg	SRM at Week 24: SEC 300mg = -1.12 SEC 150mg = -0.96 SEC 75mg = -0.644 PCB = -0.522
			SF-36 PCS	SEC 300mg: 36.9 (8.0) SEC 150mg: 36.2 (8.1) SEC 75mg: 36.2 (8.1)	<u>Δ score</u> LS Mean [SE] (calculated SD)	0.0013 for SEC 300mg	SRM at Week 24: SEC 300mg = 0.98 SEC 150mg = 0.875



				PCB: 37.4 (8.8)	SEC300mg: 7.25 [0.74] (7.40) SEC150mg: 6.39 [0.73] (7.30) SEC75mg: 4.38 [0.75] (7.462) PCB: 1.95 [0.97] (9.603)	0.0057 for SEC 150mg  0.6421 for SEC 75mg	SEC 75mg = 0.587 PCB = 0.203
Kavanaugh, et al. 2016. (FUTURE II) -subgroup analysis	TNFi naïve vs. TNFi exposed	ACR20/ Week 24*	HAQ-DI	<u>TNFi naïve:</u> Pooled SEC: 1.2 (0.6) PCB: 1.2 (0.7)  <u>TNFi exposed:</u> Pooled SEC: 1.3 (0.6) PCB: 1.1 (0.7)	<u>Δ score</u> LS Mean [SE] (calculated SD)  <u>TNFi naïve:</u> SEC 300mg: -0.59 [0.06] (0.49) SEC 150mg: -0.55 [0.06] (0.48) SEC 75mg: -0.37 [0.06] (0.48) PCB: -0.35 [0.07] (0.56)  <u>TNFi exposed:</u> SEC 300mg: -0.53 [0.09] (0.52) SEC 150mg: -0.35 [0.08] (0.49) SEC 75mg: -0.23 [0.09] (0.52) PCB: -0.23 [0.11] (0.65)	<u>TNFi naïve:</u> <0.05 for 300mg and 150mg  <u>TNFi exposed:</u> <0.05 for 300mg NS for 150mg	SRM at Week 24 (TNFi naïve vs. exposed): SEC 300mg: -1.20 vs. -1.02 SEC 150mg: -1.15 vs. -0.71 SEC 75mg: -0.77 vs. -0.44 PCB: -0.63 vs -0.35
			SF-36 PCS		<u>Δ score</u> LS Mean [SE] (calculated SD)  <u>TNFi naïve:</u> SEC 300mg: 8.05 [0.92] (7.53) SEC 150mg: 7.91 [0.93] (7.38) SEC 75mg: 5.37 [0.94] (7.58) PCB: 2.08 [1.20] (9.52)  <u>TNFi exposed:</u> SEC 300mg: 6.56 [1.20] (6.89) SEC 150mg: 4.21 [1.15] (6.995) SEC 75mg: 3.15 [1.20] (7.00) PCB: 2.65 [1.66] (9.82)	<u>TNFi naïve:</u> <0.0001 for 300mg <0.001 for 150mg <0.05 for 75mg  <u>TNFi exposed:</u> NS for all doses	SRM at Week 24: (TNFi naïve Vs. exposed): SEC 300mg: 1.07 Vs. 0.95 SEC 150mg: 1.07 Vs. 0.60 SEC 75mg: 0.71 Vs. 0.45 PCB: 0.22 Vs. 0.27
Nash, et al. 2018 (FUTURE III)	SEC (2 doses) vs. PCB (N=414)	ACR20/ Week 24*	HAQ-DI	SEC 300mg: 1.1 (0.7) SEC 150mg: 1.2 (0.6) PCB: 1.2 (0.6)	<u>Δ score</u> Mean [SE] (calculated SD)  SEC 300mg: -0.38 [0.04] (0.47) SEC 150mg: -0.27 [0.04] (0.47) PCB: -0.17 [0.06] (0.70)	<0.01	SRM at week 24: SEC 300mg = -0.81 SEC 150mg = -0.57 PCB = -0.24
			SF-36 PCS	SEC 300mg: 39.2 (8.4) SEC 150mg: 37.9 (7.6)	<u>Δ score</u> Mean [SE] (calculated SD)	<0.01	SRM at week 24: SEC 300mg = 0.93

				PCB: 37.4 (8.5)	SEC 300mg: 6.46 [0.59] (6.96) SEC 150mg: 3.42 [0.60] (7.05) PCB: 2.94 [0.83] (9.71)		SEC 150mg = 0.49 PCB = 0.30
Mease, et al. 2017 (SPIRIT-P1)	IXE (2 doses) vs. PCB vs. ADA (N=417)	ACR20/ Week 24*	HAQ-DI	IXE Q2W: 1.2 (0.57) IXE Q4W: 1.2 (0.54) PCB: 1.2 (0.60) ADA: 1.1 (0.59)	<u>Δ score at Week 12</u> LS Mean [SE] (calculated SD)  IXE Q2W: -0.47 [0.05] (0.51) IXE Q4W: -0.37 [0.05] (0.52) PCB: -0.13 [0.05] (0.51) ADA: -0.35 [0.05] (0.50)	Week 12 < 0.001 for IXE Q4W/Q2W  <0.01 for ADA	SRM at Week 12: IXE Q2W = -0.92 IXE Q4W = -0.71 PCB = -0.25 ADA = -0.70
			SF-36 PCS	IXE Q2W: 34.2 (8.7) IXE Q4W: 32.4 (10.1) PCB: 34.0 (8.3) ADA: 33.9 (8.8)	<u>Δ score at Week 12</u> LS Mean [SE] (calculated SD)  IXE Q2W: 7.6 [0.8] (4.08) IXE Q4W: 5.8 [0.8] (5.00) PCB: 2.3 [0.8] (4.23) ADA: 5.7 [0.8] (3.39)	Week 12 < 0.001 for IXE Q4W/Q2W  < 0.01 for ADA	SRM at Week 12: IXE Q2W = 1.86 IXE Q4W = 1.16 PCB = 0.54 ADA = 1.68
					<u>Δ score at Week 24</u> <u>(primary endpoint)</u> LS Mean [SE] (calculated SD)  IXE Q2W: -0.50 [0.05] (0.51) IXE Q4W: -0.44 [0.05] (0.52) PCB: -0.18 [0.05] (0.51) ADA: -0.37 [0.05] (0.50)	Week 24 <0.001 for IXE Q4W/Q2W  < 0.01 for ADA	SRM at Week 24: (primary end-point) IXE Q2W = -0.98 IXE Q4W = -0.85 PCB = -0.35 ADA = -0.74
					<u>Δ score at Week 24</u> <u>(primary endpoint)</u> LS Mean [SE] (calculated SD)  IXE Q2W: 8.2 [0.9] (4.59) IXE Q4W: 7.5 [0.9] (5.62) PCB: 2.9 [1.0] (5.29) ADA: 6.8 [0.9] (3.82)	Week 24 IXE Q4W/Q2W: <0.001  ADA: <0.01	SRM at Week 24: (primary endpoint) IXE Q2W = 1.79 IXE Q4W = 1.33 PCB = 0.55 ADA = 1.78
Nash, et al. 2017 (SPIRIT-P2)	IXE (2 doses) vs. PCB (N=363)	ACR20/ Week 24*	HAQ-DI	IXE Q2W: 1.2 (0.6) IXE Q4W: 1.2 (0.6) PCB: 1.2 (0.7)	<u>Δ score</u> LS Mean [SE] (calculated SD)  IXE Q2W: -0.4 [0.1] (1.11) IXE Q4W: -0.6 [0.1] (1.10) PCB: -0.2 [0.1] (1.09)	0.0002 for IXE Q2W  <0.0001 for IXE Q4W	SRM at Week 24: IXE Q2W = -0.36 IXE Q4W = -0.55 PCB = -0.18

			SF-36 PCS	IXE Q2W: 34.3 (9.1) IXE Q4W: 34.8 (8.8) PCB: 33.9 (9.0)	<u>Δ score</u> LS Mean [SE] (calculated SD)  IXE Q2W: 4.0 [1.2] (13.31) IXE Q4W: 3.6 [1.2] (13.25) PCB: 0.9 [1.3] (14.12)	0.009 for IXE Q2W  0.02 for IXE Q4W	SRM at Week 24: IXE Q2W = 0.30 IXE Q4W = 0.27 PCB = 0.06
Mease, et al. 2014 Phase II	BRO (2 doses) vs. PCB (N=168)	ACR20/ Week 12	HAQ-DI	BRO 280mg: 1.4 (0.6) BRO 140mg: 1.2 (0.7) PCB: 1.3 (0.6)	% Δ Median [IQR] BRO 280mg: -13.3 [-36.4, 0.0] BRO 140mg -6.5 [-45.0, 6.9] PCB: -7.7 [-23.1, 13.3]  Calculated mean (SD) BRO 280mg: -16.57 (27.70) BRO 140mg -14.87 (39.47) PCB: -5.83 (27.71)	NS for all doses	SRM <sup>‡</sup> at Week 12: BRO 280mg = -0.60 BRO 140mg = -0.38 PCB = -0.21
			SF-36 PCS	NA	<u>Δ score</u> BRO 280mg: 3.0 BRO 140mg: 2.0 PCB: 0.7	<0.05 for 280mg NS for 140mg	Difference from PCB (95% CI): BRO 280mg: 2.4 (0.1 to 4.6) BRO 140mg: 1.4 (-0.8 to 3.6)  Insufficient data for effect size calculation
IL12 inhibitors							
Author / year/ study acronyms	Intervention/ comparison (sample size)	Primary Outcome/ Time point	PROM	Baseline PROM Mean (SD) (unless specified)	Endpoint PROM/ Δ score Mean (SD) (unless specified)	P value of Δ scores of interventions Vs. PCB group	Effect sizes at primary endpoint (unless specified)
Gottlieb, et al. 2009	UST (2 doses) vs. PCB (N=146)	ACR20/ Week 12	HAQ-DI	<u>0-12 Week</u> Median [IQR]  UST: 0.9 [0.5, 1.4] PCB: 0.8 [0.3, 1.1]	<u>Δ score</u> Median [IQR] UST: -0.25 [-0.5, 0.0] PCB: 0.0 [-0.25, 0.13]  Calculated mean (SD) UST: -0.25 (0.38) PCB: -0.04 (0.29)	0.0005	SRM <sup>‡</sup> at Week 12: UST = -0.66 PCB = -0.14
			SF-36 PCS or PF	NA	NA	NA	NA
McInnes, et al. 2013 (PSUMMIT I)	UST (2 doses) vs. PCB (N=615)	ACR20, EULAR response PASI75	HAQ-DI	Median [IQR] UST 90mg: 1.3 [0.8, 1.6] UST 45mg: 1.3 [0.8, 1.8] PCB: 1.3 [0.8, 1.8]	<u>Δ score</u> Median [IQR] UST 90mg: -0.25 [-0.75, 0.00] UST 45mg: -0.25 [-0.63, 0.00]	<0.0001 for both UST doses	SRM <sup>‡</sup> at Week 24: UST 90mg = -0.59 UST 45mg = -0.62 PCB = -0.21

		/Week 24*			PCB: 0.00 [-0.38, 0.13]  Calculated mean (SD) UST 90mg: -0.33 (0.56) UST 45mg: -0.29 (0.47) PCB: -0.08 (0.38)		
			SF-36 PCS	Median [IQR] UST 90mg: 36.5 [30.2, 40.1] UST 45mg: 35.5 [30.6, 40.1] PCB: 35.8 [31.8, 40.1]	<u>Δ score</u> Median [IQR] UST 90mg: 5.8 [0.6, 10.9] UST 45mg: 3.9 [-1.3, 10.7] PCB: 1.2 [-2.3, 5.2]  Calculated mean (SD) UST 90mg: 5.77 (7.69) UST 45mg: 4.33 (8.96) PCB: 1.37 (5.60)	<0.0001 for both UST doses:	SRM <sup>¥</sup> at Week 24: UST 90mg = 0.75 UST 45mg = 0.48 PCB = 0.24
Ritchlin, et al. 2014 (PSUMMIT II)	UST (2 doses) vs. PCB (N=312)	ACR20, EULAR response PASI75/ Week 24*	HAQ-DI	Median [IQR] UST 90mg: 1.3 [0.8, 1.9] UST 45mg: 1.4 [0.8, 1.9] PCB: 1.3 [0.8, 1.8]	<u>Δ score</u> Median [IQR] UST 90mg: -0.25 [-0.50, 0.00] UST 45mg: -0.13 [-0.38, 0.00] PCB: 0.00 [-0.13, 0.13]  Calculated mean (SD) UST 90mg: -0.25 (0.38) UST 45mg: -0.17 (0.29) PCB: 0.00 (0.38)  <u>TNFi naïve</u> Median [IQR] UST 90mg: -0.25 [-0.50, 0.00] UST 45mg: -0.25 [-0.50, 0.00] PCB: 0.00 [-0.25, 0.25]  Calculated mean (SD) UST 90mg: -0.25 (0.38) UST 45mg: -0.25 (0.38) PCB: 0.00 (0.38)  <u>TNFi exposed</u> Median [IQR] UST 90mg: -0.19 [-0.38, 0.00] UST 45mg: -0.13 [-0.38, 0.00]	<0.001 for UST 90mg  <0.01 for UST 45mg	SRM <sup>¥</sup> at Week 24: UST 90mg = -0.66 UST 45mg = -0.59 PCB = 0.00  Subgroups analysis: SRM <sup>¥</sup> at Week 24 (TNFi naïve vs. exposed): UST 90mg = -0.66 vs. -0.66 UST 45mg = -0.66 vs. -0.59 PCB = 0 vs. 0

					PCB: 0.00 [-0.13, 0.13]  Calculated mean (SD) UST 90mg: -0.19 (0.29) UST 45mg: -0.17 (0.29) PCB: 0.00 (0.20)		
			SF-36 PCS	Median [IQR] UST 90mg: 28.2 [21.8, 33.6] UST 45mg: 28.0 [22.6, 34.0] PCB: 29.4 [23.3, 36.2]	$\Delta$ score Median [IQR] UST 90mg: 3.5 [-0.2, 10.1] UST 45mg: 2.7 [-0.7, 9.1] PCB: 0.00 [-0.8, 4.0]  Calculated mean (SD) UST 90mg: 4.67 (7.75) UST 45mg: 3.7 (7.37) PCB: 1.07 (3.61)	<0.01 for both UST doses	SRM <sup>¥</sup> at Week 24: UST 90mg = 0.60 UST 45mg = 0.50 PCB = 0.30
Araugo, et al. 2019 (ECLIPSA)	UST (45mg or 100mg if body weight >100kg) vs. TNFi (N=47)	Enthesitis clearance (SPARCC = 0) at Week 24	HAQ-DI	UST: 0.87 (0.63) TNFi: 1.17 (0.62)	<u>Score at Week 24</u> Median [IQR] UST: 0 [0.25] TNFi: 0.3 [0.35]  <u>Calculated <math>\Delta</math> score (pooled SD)</u> UST: -0.87 (0.48) TNFi: -0.87 (0.50)	NA	ES <sub>2</sub> <sup>¥</sup> at Week 24: UST = -1.81 TNFi = -1.74
			SF-36 PCS	UST: 29.1 (9.6) TNFi: 29.5 (9.5)	<u>Score at Week 24</u> Median [IQR] UST: 52.8 [6.1] TNFi: 46 [11.6]  <u>Calculated <math>\Delta</math> score (pooled SD)</u> UST: 23.7 (8.04) TNFi: 16.5 (10.60)	NA	ES <sub>2</sub> <sup>¥</sup> at Week 24: UST = 2.95 TNFi = 1.56
IL23 inhibitors							
Author / year/ study acronyms	Intervention/ comparison (sample size)	Primary Outcome/ Time point	PROM	Baseline PROM Mean (SD) (unless specified)	End point PROM/ $\Delta$ score Mean (SD) (unless specified)	P value of $\Delta$ scores of intervention vs. PCB group	Effect sizes at primary end point (unless specified)
Deodhar, et al. 2018	GUS 100mg vs. PCB	ACR20/ Week24*	HAQ-DI	GUS: 1.4 (0.6) PCB: 1.3 (0.5)	$\Delta$ score GUS: -0.42 (0.51)	0.00025	SRM at Week 24: GUS = -0.82

	(N=149)				PCB: -0.06 (0.53)		PCB = -0.11
			SF-36 PCS	GUS: 33.5 (7.1) PCB: 34.4 (8.0)	<u>Δ score</u> GUS: 6.59 (7.47) PCB: 0.36 (6.51)	<0.0001	SRM at Week 24: GUS = 0.88 PCB = 0.06
<b>T cell inhibition</b>							
Author / year/ study acronyms	Intervention/ comparison (sample size)	Primary Outcome/ Time point	PROM	Baseline PROM Mean (SD) (unless specified)	End point PROM/ Δ score Mean (SD) (unless specified)	P value of Δ scores of intervention vs. PCB group	Effect sizes at primary endpoint (unless specified)
Mease, et al. 2011 Phase II	ABT (3 doses) vs. PCB (N=170)	Day 169	HAQ- DI	ABT 30/10 mg/kg: 1.2 (0.8) ABT 10mg/kg: 1.3 (0.7) ABT 3 mg/kg: 1.1 (0.7) PCB: 1.2 (0.7)	Δ scores not available  % achieved HAQ-DI >0.3 (MCID) ABT 30/10mg/kg = 35% ABT 10mg/kg = 45% ABT 3mg/kg = 36% PCB = 19% (not statistically significant)	NA	Insufficient data for effect size calculation
			SF-36 PCS	NA	<u>Δ score</u> Mean [SE] (calculated SD) ABT 30/10 mg/kg: 7.3 [1.9] (12.31) ABT 10mg/kg: 9.3 [1.9] (12.02) ABT 3 mg/kg: 6.3 [1.8] (11.80) PCB: 0.2 [1.9] (12.17)	NA	SRM at day 169: ABT 30/10 mg/kg: 0.59 ABT 10mg/kg: 0.77 ABT 3 mg/kg: 0.53 PCB: 0.02
Mease, et al. 2017 (ASTRAEA)	ABT vs. PCB (N=424)	ACR20/ Week 24*	HAQ- DI	ABT: 1.3 (0.7) PCB: 1.3 (0.7)	<u>Δ score</u>  Mean [95% CI] (calculated SD) ABT: -0.33 [-0.41, -0.24] (0.48) PCB: -0.20 [-0.29, -0.10] (0.50)  <u>TNFi naïve</u> Mean [95% CI] (calculated SD) ABT: -0.29 [-0.42, -0.16] (0.47) PCB: -0.17 [-0.32, -0.03] (0.44)  <u>TNFi exposed</u> Mean [95% CI] (calculated SD) ABT: -0.35 [-0.47 to -0.22] (0.55) PCB: -0.18 [-0.31 to -0.05] (0.53)	0.097	SRM at Week 24: ABT = -0.69 PCB = -0.40  SRM at Week 24: Subgroup analysis: (TNFi naïve vs. exposed) ABT = -0.62 vs. -0.64 PCB = -0.39 vs. -0.34
			SF-36 PCS	NA	<u>Δ score</u>  Mean [95% CI] (calculated SD)	NS	SRM at Week 24: ABT: 0.72 PCB: 0.53

					ABT: 5.11 [3.86, 6.36] (7.10) PCB: 3.69 [2.30, 5.08] (6.98)		
Mease, et al. 2006	ALC/MTX vs. PCB/MTX (N=185)	ACR20/ Week 12	HAQ- DI	ALC: 1.0 PCB: 1.1	<u>Δ Score</u>  ALC -24.3 PCB: -7.7	NS	% Δ in HAQ-DI at Week 12: ALC/MTX = -24.5% PCB/MTX = -7.7% (NS)
			SF-36 PCS or PF	NA	NA	NA	NA
JAK inhibitors							
Author / year/ study acronyms	Intervention/ comparator (sample size, N)	Primary Outcome/ Time point	PROM	Baseline PROM Mean (SD) (unless specified)	End point PROM/ Δ score Mean (SD) (unless specified)	P value of Δ scores of interventions vs. PCB	Effect sizes at primary endpoint (unless specified)
Mease, et al. 2017 (OPAL Broaden)	TOF (2 doses) vs. PCB vs. ADA (N= 422)	ACR20/ 3 months	HAQ- DI	TOF 10mg: 1.1 (0.6) TOF 5mg: 1.2 (0.6) PCB: 1.1 (0.6) ADA: 1.1 (0.6)	<u>Δ score</u> Mean [SE] (calculated SD) TOF 10mg: -0.40 [0.05] (0.51) TOF 5mg: -0.35 [0.05] (0.51) PCB: -0.18 [0.05] (0.50) ADA: -0.38 [0.05] (0.50)	< 0.001 for TOF 10mg <0.01 for TOF 5mg  NS for ADA	SRM at 3 month: TOF 10mg = -0.78 TOF 5mg = -0.69 PCB = -0.36 ADA = -0.76
			SF-36 PF	NA	<u>Δ score</u>  Mean [SE] (calculated SD) TOF 10mg: 5.2 [0.8] (8.12) TOF 5mg: 5.2 [0.8] (8.08) PCB: 2.1 [0.9] (9.09) ADA: 5.2 [0.9] (9.04)	All NS	SRM at 3 month: TOF 10mg = 0.64 TOF 5mg = 0.64 PCB = 0.23 ADA = 0.58
			SF-36 PCS	TOF 10mg: 36.4 (7.6) TOF 5mg: 35.4 (7.9) PCB: 36.0 (7.4) ADA: 35.9 (8.6)	<u>Δ score</u>  Mean [SE] (calculated SD) TOF 10mg: 5.69 [0.74] (7.55) TOF 5mg: 5.51 [0.73] (7.55) PCB: 2.68 [0.79] (8.10) ADA: 6.23 [0.75] (7.72)	<0.01 for both doses of TOF  <0.001 for ADA	SRM at 3 month: TOF 10mg = 0.75 TOF 5mg = 0.73 PCB = 0.33 ADA = 0.81
Gladman, et al. 2017 (OPAL Beyond)	TOF (2 doses) vs. PCB (N=395)	ACR20, Δ in HAQ- DI/ 3 months	HAQ- DI	TOF 10mg: 1.4 (0.6) TOF 5mg: 1.3 (0.7) PCB: 1.3 (0.8)	<u>Δ score</u>  Mean [SE] (calculated SD) TOF 10mg: -0.35 [0.05] (0.55) TOF 5mg: -0.39 [0.05] (0.56) PCB: -0.14 [0.05] (0.54)	<0.05 for both doses	SRM at 3 month: TOF 10mg = -0.64 TOF 5mg = -0.70 PCB = -0.26

			SF-36 PF	Norm based scores TOF 10mg: 32.1 (9.9) TOF 5mg: 33.5 (10.4) PCB: 34.0 (11.0)	<u>Δ score</u>  Mean [SE] (calculated SD) TOF 10mg: 4.1 [0.7] (7.67) TOF 5mg: 5.0 [0.7] (7.79) PCB: 1.7 [0.7] (7.57)	NS	SRM at 3 month: TOF 10mg = 0.53 TOF 5mg = 0.64 PCB = 0.22
Strand, et al. 2019 (OPAL Beyond)			SF-36 PCS	TOF 10mg: 31.7 (8.9) TOF 5mg: 33.6 (8.5) PCB: 34.7 (9.6)	<u>Δ score</u>  Mean [SE] (calculated SD) TOF 10mg: 5.34 [0.69] (7.93) TOF 5mg: 5.18 [0.68] (7.78) PCB: 1.77 [0.69] (7.90)	<0.001 for both doses	SRM at 3 month: TOF 10mg = 0.67 TOF 5mg = 0.67 PCB = 0.22
Mease, et al. 2018 (EQUATOR) Phase II	FIL 200mg vs. PCB (N=131)	ACR20/ Week 16	HAQ- DI	FIL: 1.43 (0.5) PCB: 1.36 (0.6)	<u>Score at end point</u> FIL: 0.86 (0.6) PCB: 1.09 (0.6)  <u>Δ score</u> FIL: -0.57 (0.5) PCB: -0.28 (0.5)	0.0009	SRM at Week 16: FIL = -1.14 PCB = -0.56  ES <sub>2</sub> at Week 16: FIL = -1.14 PCB = -0.47
			SF-36 PF	NA	NA	0.0009	Insufficient data for effect size calculation
<b>IL6 inhibitors</b>							
<b>Author / year/ (study acronyms)</b>	<b>Intervention/ comparator (sample size, N)</b>	<b>Primary Outcome/ Time point</b>	<b>PROM</b>	<b>Baseline PROM Mean (SD) (unless specified)</b>	<b>End point PROM/ Δ score Mean (SD) (unless specified)</b>	<b>P value of Δ scores of intervention vs. PCB group</b>	<b>Effect sizes at primary end point (unless specified)</b>
Mease, et al. 2016 Phase II	CLZ (3 doses) vs. PCB (N=165)	ACR20/ Week 16	HAQ- DI	CLZ 200mg: 1.4 (0.7) CLZ 100mg: 1.3 (0.7) CLZ 25mg: 1.4 (0.6) PCB: 1.4 (0.6)	<u>Δ score</u>  Mean [95% CI] (calculated SD) CLZ 200mg: -0.26 [-0.42, -0.09] (0.51) CLZ 100mg: -0.40 [-0.56, -0.24] (0.52) CLZ 25mg: -0.44 [-0.61, -0.28] (0.53) PCB: -0.27 [-0.43, -0.11] (0.52)	NS	SRM at Week 16: CLZ 200mg = -0.51 CLZ 100mg = -0.77 CLZ 25mg = -0.83 PCB = -0.52
			SF-36 PCS	NA	<u>Δ score</u>	NS	SRM at Week 16: CLZ 200mg = 0.52



					Mean [95% CI] (calculated SD) CLZ 200mg: 4.1 [1.6, 6.5] (7.81) CLZ 100mg: 4.6 [2.2, 7.0] (7.84) CLZ 25mg: 6.5 [4.0, 8.9] (7.91) PCB: 4.4 [2.0, 6.8] (7.74)		CLZ 100mg = 0.59 CLZ 25mg = 0.82 PCB = 0.57
<b>Others</b>							
<b>Author / year/ (study acronyms)</b>	<b>Intervention/ comparator (sample size, N)</b>	<b>Primary Outcome/ Time point</b>	<b>PROM</b>	<b>Baseline PROM Mean (SD) (unless specified)</b>	<b>End point PROM/ Δ score Mean (SD) (unless specified)</b>	<b>P value of Δ scores of interventions vs. PCB group</b>	<b>Effect sizes at primary end point (unless specified)</b>
Mease, et al. 2018 Phase II	ABT-122 (2 doses) vs. PCB vs. ADA (N=240)	ACR20/ Week 12	HAQ-S	ABT-122 240mg: 1.3 (0.6) ABT-122 120mg: 1.3 (0.6) PCB: 1.2 (0.6) ADA: 1.3 (0.6)	<u>Δ score</u> ABT-122 240mg: -0.56 ABT-122 120mg: -0.55 PCB: -0.28 ADA: -0.58  <u>Improved HAQ-S ≥ 0.5</u> ABT-122 240mg: 54.8% ABT-122 120mg: 43.7% PCB: 25.0% ADA: 41.7%	No statistic test for Δ scores  <u>Statistical test for proportion of HAQ-S ≥ 0.5:</u>  < 0.05 for ABT-122 240mg  NS for ABT- 122 120mg  NS for ADA	ES <sub>1</sub> at Week 12: ABT-122 240mg: -0.93 ABT-122 120mg: -0.92 PCB: -0.47 ADA: -0.97
			SF-36 PCS or PF	NA	NA	NA	NA

ES<sub>1</sub>: Effect size 1 (the mean difference divided by standard deviation of baseline score; ES<sub>2</sub>: Effect size 2 (the mean difference divided by the pooled standard deviation, i.e. Cohen's *d*); SRM: Standardized response mean (mean difference divided by the standard deviation of the differences between baseline and assessment end point); <sup>†</sup> SRM calculated using percentage Δ score and SD of percentage Δ; <sup>§</sup> SRM for improvement, a negative value indicate deterioration; <sup>‡</sup> Effect sizes estimated based on mean and SD of change scores calculated from median and IQR from original publication; \* early escape for patients with inadequate response in the PCB group to active treatment group at Week 16; \*\* option to switch to treatment group at Week 24; Δ: change. Abbreviations: ACR: American College of Rheumatology Response criteria; ABT: abatacept; ADA: adalimumab; ALC: alefacept; BIW: twice a week; BRO: brodalumab; CI: confidence interval; CLZ: clazakizumab; CZP: certolizumab; ETN: etanercept; EULAR: European League Against Rheumatic Diseases; FIL: filgotinib; GOL: golimumab; GUS: guselkumab; HAQ-DI: Health Assessment Questionnaire – Disability index; IFX: infliximab; IXE: Ixekizumab; ivi: intravenous IQR: interquartile range; LS: least squares; MCID: minimally clinically important difference; MTX: methotrexate; NA: not available; NS: not significant; PsARC: Psoriatic arthritis Response Criteria; PASI: psoriasis area and severity index; PCB: placebo; PCS: physical component summary of SF-36; PF: physical functioning domain of SF-36; QW: once a week; Q2W: once every 2-week; SD: standard deviation; SE: standard error; SEC: secukinumab; SF-36:

Medical Outcome Survey Short Form 36 items; SPARCC: *Spondyloarthritis Research Consortium of Canada* enthesitis index; TOF: tofacitinib; TNFi: tumor necrosis factor inhibitors; Ustekinumab: UST; Vs.: versus.

Table A.3. Quality assessment using OMERACT good method checklist for HAQ-DI

1 <sup>st</sup> Author / year/ (study acronyms/ drug)	Appropriate time interval	Expected change in one/ both groups	Hypothesis of change stated a priori	Adequate statistical methods for hypothesis testing	Otherwise good methods	Quality assessment rating
Antoni, 2005 (IMPACT)	Green	Green	Amber	Amber	Green	<b>AMBER (+)</b>
Antoni, 2005 (IMPACT2)	Green	Green	Green	Amber	Green	<b>AMBER (+)</b>
Mease, 2005 (ADEPT)	Green	Green	Green	Green	Green	<b>GREEN (+)</b>
Genovese, 2007 (adalimumab)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Mease, 2000 (etanercept)	Green	Green	Amber	Amber	Green	<b>AMBER (+)</b>
Mease. 2010 (etanercept)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Gniadecki, 2012 (PRESTA)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Mease. 2019 (SEAM-PsA)	Amber	Green	Amber	Amber	Amber	<b>AMBER (+/-)</b>
Kavanaugh, 2009 (GO-REVEAL)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Kavanaugh, 2017 (GO-VIBRANT)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Gladman, 2014 (RAPID-PsA)	Green	Green	Green	Green	Green	<b>GREEN (+)</b>
McInnes, 2014 (secukinumab)	Green	Green	Amber	Amber	Green	<b>AMBER (+)</b>
Mease, 2015 (FUTURE I)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
McInnes, 2015 FUTURE II)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Kavanaugh, 2016 (FUTURE II) -subgroup analysis	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Nash P, 2018 (FUTURE III)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Mease, 2017 (SPIRIT-P1)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Nash, 2017 (SPIRIT-P2)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Mease, 2014 (brodalumab)	Green	Green	Amber	Amber	Green	<b>AMBER (+)</b>
Gottlieb, 2009 (ustekinumab)	Green	Green	Amber	Amber	Green	<b>AMBER (+)</b>
McInnes, 2013 (PSUMMIT I)	Green	Green	Amber	Amber	Green	<b>AMBER (+)</b>
Ritchlin, 2014 (PSUMMIT II)	Green	Green	Amber	Amber	Green	<b>AMBER (+)</b>
Araugo, 2019 (ECLIPSA)	Green	Green	Amber	Amber	Green	<b>AMBER (+)</b>
Deodhar, 2018 (guselkumab)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Mease, 2011 (abatcept)	Green	Green	Amber	Red	Green	<b>RED (+/-)</b>
Mease, 2017 (ASTRAEA)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Mease, 2006 (alefacept)	Green	Green	Amber	Red	Amber	<b>RED (+/-)</b>
Mease, 2017 (OPAL Broaden)	Green	Green	Green	Green	Green	<b>GREEN (+)</b>
Gladman, 2017 (OPAL Beyond)	Green	Green	Green	Green	Green	<b>GREEN (+)</b>
Mease, 2018 (EQUATOR)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Mease, 2016 (clazakizumab)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>

Table A.4. Quality assessment using OMERACT good method checklist for HAQ-S

<b>1<sup>st</sup> Author / year/ (study acronyms/ drug)</b>	<b>Appropriate time interval</b>	<b>Expected change in one/ both groups</b>	<b>Hypothesis of change stated a priori</b>	<b>Adequate statistical methods for hypothesis testing</b>	<b>Otherwise good methods</b>	<b>Quality assessment rating</b>
Mease, 2018 (ABT-122)	Green	Green	Amber	Green	Amber	<b>AMBER (+)</b>

Table A.5. Quality assessment using OMERACT good method checklist for SF-36 PCS

<b>1<sup>st</sup> Author / year/ (study acronyms/ drug)</b>	<b>Appropriate time interval</b>	<b>Expected change in one/ both groups</b>	<b>Hypothesis of change stated a priori</b>	<b>Adequate statistical methods for hypothesis testing</b>	<b>Otherwise good methods</b>	<b>Quality assessment rating</b>
Antoni, 2005 (IMPACT2)	Green	Green	Green	Amber	Green	<b>AMBER (+)</b>
Mease, 2005 (ADEPT)	Green	Green	Green	Green	Green	<b>GREEN (+)</b>
Genovese, 2007 (adalimumab)	Green	Green	Green	Green	Green	<b>GREEN (+)</b>
Mease, 2000 (etanercept)	Green	Green	Amber	Amber	Green	<b>AMBER (+)</b>
Mease. 2010 (etanercept)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Mease. 2019 (SEAM-PsA)	Amber	Green	Amber	Green	Amber	<b>AMBER (+)</b>
Kavanaugh, 2009 (GO-REVEAL)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Kavanaugh, 2017 (GO-VIBRANT)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Gladman, 2014 (RAPID-PsA)	Green	Green	Green	Green	Green	<b>GREEN (+)</b>
McInnes, 2014 (secukinumab)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Mease, 2015 (FUTURE I)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
McInnes, 2015 FUTURE II)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Kavanaugh, 2016 (FUTURE II) -subgroup analysis	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Nash P, 2018 (FUTURE III)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Mease, 2017 (SPIRIT-P1)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Nash, 2017 (SPIRIT-P2)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Mease, 2014 (brodalumab)	Green	Green	Amber	Red	Amber	<b>RED (+)</b>
McInnes, 2013 (PSUMMIT I)	Green	Green	Amber	Amber	Green	<b>AMBER (+)</b>
Ritchlin, 2014 (PSUMMIT II)	Green	Green	Amber	Amber	Green	<b>AMBER (+)</b>
Araugo, 2019 (ECLIPSA)	Green	Green	Amber	Amber	Green	<b>AMBER (+)</b>
Deodhar, 2018 (guselkumab)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>
Mease, 2011 (abatacept)	Green	Green	Amber	Green	Amber	<b>AMBER (+)</b>
Mease, 2017 (ASTRAEA)	Green	Green	Amber	Green	Amber	<b>AMBER (+)</b>
Strand, 2019 (OPAL Broaden)	Green	Green	Green	Green	Green	<b>GREEN (+)</b>
Strand, 2019 (OPAL Beyond)	Green	Green	Green	Green	Green	<b>GREEN (+)</b>
Mease, 2016 (clazakizumab)	Green	Green	Amber	Green	Green	<b>GREEN (+)</b>

Table A.6. Quality assessment using OMERACT good method checklist for SF-36 PF

<b>1<sup>st</sup> Author / year/ (study acronyms/ drug)</b>	Appropriate time interval	Expected change in one/ both groups	Hypothesis of change stated a priori	Adequate statistical methods for hypothesis testing	Otherwise good methods	<b>Quality assessment rating</b>
Kavanaugh, 2006 (IMPACT2)	Green	Amber	Amber	Red	Amber	<b>RED (+)</b>
Mease, 2017 (OPAL Broaden)	Green	Green	Amber	Green	Amber	<b>AMBER (+)</b>
Gladman, 2017 (OPAL Beyond)	Green	Green	Amber	Green	Amber	<b>AMBER (+)</b>
Mease, 2018 (EQUATOR)	Green	Green	Amber	Red	Amber	<b>RED (+)</b>