

Appendix 2

PubMed search strategy

("Severity of Illness Index"[Majr] OR "Disease Progression"[Majr] OR "disease activity"[Title/Abstract] OR "disease severity"[Title/Abstract] OR "remission"[Title/Abstract] OR "flare"[Title/Abstract] OR "Severity of Illness Index "[Title/Abstract] OR "disease progression"[Title/Abstract] OR "disease course"[Title/Abstract])

AND ("Arthritis, Rheumatoid"[Majr] OR rheumatoid arthritis[Title/Abstract])

AND ("questionnaires"[Majr] OR "Patient Participation"[Majr] OR "Self examination"[Majr] OR "Self assessment"[Majr] OR patient index[tiab] OR patient administered[tiab] OR patient derived [tiab] OR patient based [tiab] OR patient assessed[tiab] OR patient assessment[tiab] OR patient assessments[tiab] OR patient reports[tiab] OR patient reported[tiab] OR patient monitor[tiab] OR patient monitoring[tiab] OR patient monitored[tiab] OR patient evaluated[tiab] OR patient evaluation[tiab] OR patient evaluations[tiab] OR patient examination[tiab] OR patient examinations[tiab] OR patient examined[tiab] OR patient rated[tiab] OR patient participation[tiab] OR self administered[tiab] OR self assessment[tiab] OR self assessments[tiab] OR self assessed[tiab] OR self report[tiab] OR self reported[tiab] OR self monitor[tiab] OR self monitoring[tiab] OR self monitored[tiab] OR self evaluate[tiab] OR self evaluation[tiab] OR self evaluations[tiab] OR self evaluated[tiab] OR self examine[tiab] OR self examination[tiab] OR self examinations[tiab] OR self examined[tiab] OR self rate[tiab] OR self rated[tiab]) AND ("instrumentation"[sh] OR "methods"[sh] OR Validation Studies[pt] OR Comparative Study[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR outcome assessment[tiab] OR outcome measure*[tw] OR "observer variation"[MeSH] OR observer variation[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tiab] OR precision[tiab] OR imprecision[tiab] OR "precise values"[tiab] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tiab] OR ((replicab*[tiab] OR repeated[tiab]) AND (measure[tiab] OR measures[tiab] OR findings[tiab] OR result[tiab] OR results[tiab] OR test[tiab] OR tests[tiab])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR factor analysis[tiab] OR factor analyses[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR item discriminant[tiab] OR interscale correlation*[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR ((minimal[tiab] OR minimally[tiab]

OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR meaningful change[tiab] OR “ceiling effect”[tiab] OR “floor effect”[tiab] OR “Item response model”[tiab] OR IRT[tiab] OR Rasch[tiab] OR “Differential item functioning”[tiab] OR DIF[tiab] OR “computer adaptive testing”[tiab] OR “item bank”[tiab] OR “cross-cultural equivalence”[tiab])

AND English[lang]

NOT (Addresses[ptyp] OR Autobiography[ptyp] OR Bibliography[ptyp] OR Biography[ptyp] OR Case Reports[ptyp] OR Clinical Conference[ptyp] OR Congresses[ptyp] OR Consensus Development Conference[ptyp] OR Consensus Development Conference, NIH[ptyp] OR Dictionary[ptyp] OR Directory[ptyp] OR Duplicate Publication[ptyp] OR Editorial[ptyp] OR Electronic Supplementary Materials[ptyp] OR English Abstract[ptyp] OR Festschrift[ptyp] OR Government Publications[ptyp] OR Guideline[ptyp] OR Historical Article[ptyp] OR In Vitro[ptyp] OR Interactive Tutorial[ptyp] OR Interview[ptyp] OR Introductory Journal Article[ptyp] OR Lectures[ptyp] OR Legal Cases[ptyp] OR Legislation[ptyp] OR Letter[ptyp] OR Meta-Analysis[ptyp] OR News[ptyp] OR Newspaper Article[ptyp] OR Overall[ptyp] OR Patient Education Handout[ptyp] OR Periodical Index[ptyp] OR Personal Narratives[ptyp] OR Portraits[ptyp] OR Practice Guideline[ptyp] OR Retracted Publication[ptyp] OR Retraction of Publication[ptyp] OR Review[ptyp] OR Systematic [sb] OR Scientific Integrity Review[ptyp] OR Technical Report[ptyp] OR Twin Study[ptyp] OR Video-Audio Media[ptyp] OR Webcasts[ptyp])

NOT ("infant"[MeSH Terms] OR "child"[MeSH Terms] OR "adolescent"[MeSH Terms])

NOT “animals”[MeSH:noexp]

EMBASE search strategy

1. *Severity of Illness Index/ or *disease severity/ or *disease course/ or *disease activity score/ or disease activity.ti,ab,kw. or disease severity.ti,ab,kw. or remission.ti,ab,kw. or flare.ti,ab,kw. or Severity of Illness Index.ti,ab,kw. or disease progression.ti,ab,kw. or disease course.ti,ab,kw.
2. *rheumatoid arthritis/ or rheumatoid arthritis.ti,ab,kw.
3. *questionnaire/ or *patient participation/ or *patient assessment/ or *self examination/ or *self evaluation/ or *self report/ or *self monitoring/ or *patient monitoring/ or patient index.ti,ab,kw. or questionnaire.kw. or questionnaires.kw. or patient administered.ti,ab,kw. or patient derived.ti,ab,kw. or patient based.ti,ab,kw. or patient assessed.ti,ab,kw. or patient assessment.ti,ab,kw. or patient reports.ti,ab,kw. or patient reported.ti,ab,kw. or patient monitor.ti,ab,kw. or patient monitoring.ti,ab,kw. or patient monitored.ti,ab,kw. or patient evaluated.ti,ab,kw. or patient evaluation.ti,ab,kw. or patient examination.ti,ab,kw. or patient examined.ti,ab,kw. or patient rated.ti,ab,kw. or patient participation.ti,ab,kw. or self administered.ti,ab,kw. or self assessment.ti,ab,kw. or self assessed.ti,ab,kw. or self report.ti,ab,kw. or self reported.ti,ab,kw. or self monitor.ti,ab,kw. or self monitoring.ti,ab,kw. or self monitored.ti,ab,kw. or self evaluate.ti,ab,kw. or self evaluation.ti,ab,kw. or self evaluated.ti,ab,kw. or self examine.ti,ab,kw. or self examination.ti,ab,kw. or self examined.ti,ab,kw. or self rate.ti,ab,kw. or self rated.ti,ab,kw.
4. psychometry/ or outcome assessment/ or outcomes research/ or comparative study/ or intermethod comparison/ or multimethod study/ or feasibility study/ or validation study/ or correlational study/ or correlation analysis/ or replication study/ or validity/ or reliability/ or reproducibility/ or observer variation/ or psychometr*.ti,ab,kw. OR clinimetr*.ti,ab,kw. OR clinometr*.ti,ab,kw. OR outcome assessment.ti,ab,kw. OR outcome measure*.ti,ab,kw. OR observer variation.ti,ab,kw. OR reproducib*.ti,ab,kw. OR reliab*.ti,ab,kw. OR unreliab*.ti,ab,kw. OR valid*.ti,ab,kw. OR coefficient.ti,ab,kw. OR homogeneity.ti,ab,kw. OR homogeneous.ti,ab,kw. OR internal consistency.ti,ab,kw. OR (cronbach*.ti,ab,kw. AND (alpha.ti,ab,kw. OR alphas.ti,ab,kw.)) OR (item.ti,ab,kw. AND (correlation*.ti,ab,kw. OR selection*.ti,ab,kw. OR reduction*.ti,ab,kw.)) OR agreement.ti,ab,kw. OR precision.ti,ab,kw. OR imprecision.ti,ab,kw. OR precise values.ti,ab,kw. OR test retest.ti,ab,kw. OR (test.ti,ab,kw. AND retest.ti,ab,kw.) OR (reliab*.ti,ab,kw. AND (test.ti,ab,kw. OR retest.ti,ab,kw.)) OR stability.ti,ab,kw. OR interrater.ti,ab,kw. OR inter rater.ti,ab,kw. OR intrarater.ti,ab,kw. OR intra rater.ti,ab,kw. OR intertester.ti,ab,kw. OR inter tester.ti,ab,kw. OR intratester.ti,ab,kw. OR intra-tester.ti,ab,kw. OR interobserver.ti,ab,kw. OR inter observer.ti,ab,kw. OR intraobserver.ti,ab,kw. OR intra observer.ti,ab,kw. OR intertechnician.ti,ab,kw. OR inter technician.ti,ab,kw. OR intratechnician.ti,ab,kw. OR intra-technician.ti,ab,kw. OR interexaminer.ti,ab,kw. OR interexaminer.ti,ab,kw. OR intraexaminer.ti,ab,kw. OR intra examiner.ti,ab,kw. OR interassay.ti,ab,kw. OR inter-assay.ti,ab,kw. OR intraassay.ti,ab,kw. OR intra assay.ti,ab,kw. OR interindividual.ti,ab,kw. OR inter individual.ti,ab,kw. OR intraindividual.ti,ab,kw. OR intra individual.ti,ab,kw. OR interparticipant.ti,ab,kw. OR inter participant.ti,ab,kw. OR intraparticipant.ti,ab,kw. OR intra participant.ti,ab,kw. OR kappa.ti,ab,kw. OR kappas.ti,ab,kw. OR repeatab*.ti,ab,kw. OR ((replicab*.ti,ab,kw. OR repeated.ti,ab,kw.) AND (measure.ti,ab,kw. OR measures.ti,ab,kw. OR findings.ti,ab,kw. OR result.ti,ab,kw. OR results.ti,ab,kw. OR test.ti,ab,kw. OR tests.ti,ab,kw.)) OR generaliza*.ti,ab,kw. OR generalisa*.ti,ab,kw. OR concordance.ti,ab,kw. OR (intraclass.ti,ab,kw. AND correlation*.ti,ab,kw.) OR discriminative.ti,ab,kw. OR known group.ti,ab,kw. OR factor analysis.ti,ab,kw. OR factor analyses.ti,ab,kw. OR dimension*.ti,ab,kw. OR subscale*.ti,ab,kw. OR (multitrait.ti,ab,kw. AND scaling.ti,ab,kw. AND (analysis.ti,ab,kw. OR analyses.ti,ab,kw.)) OR item discriminant.ti,ab,kw. OR interscale correlation*.ti,ab,kw. OR

error.ti,ab,kw. OR errors.ti,ab,kw. OR individual variability.ti,ab,kw. OR (variability.ti,ab,kw. AND (analysis.ti,ab,kw. OR values.ti,ab,kw.)) OR (uncertainty.ti,ab,kw. AND (measurement.ti,ab,kw. OR measuring.ti,ab,kw.)) OR standard error of measurement.ti,ab,kw. OR sensitiv*.ti,ab,kw. OR responsive*.ti,ab,kw. OR ((minimal.ti,ab,kw. OR minimally.ti,ab,kw. OR clinical.ti,ab,kw. OR clinically.ti,ab,kw.) AND (important.ti,ab,kw. OR significant.ti,ab,kw. OR detectable.ti,ab,kw.)) AND (change.ti,ab,kw. OR difference.ti,ab,kw.)) OR (small*.ti,ab,kw. AND (real.ti,ab,kw. OR detectable.ti,ab,kw.)) AND (change.ti,ab,kw. OR difference.ti,ab,kw.)) OR meaningful change.ti,ab,kw. OR ceiling effect.ti,ab,kw. OR floor effect.ti,ab,kw. OR Item response model.ti,ab,kw. OR IRT.ti,ab,kw. OR Rasch.ti,ab,kw. OR Differential item functioning.ti,ab,kw. OR DIF.ti,ab,kw. OR computer adaptive testing.ti,ab,kw. OR item bank.ti,ab,kw. OR cross-cultural equivalence.ti,ab,kw.

5. 1 and 2 and 3 and 4

6. limit 5 to (english language)

7. limit 6 to (book or book series or conference abstract or conference paper or conference proceeding or "conference review" or editorial or erratum or letter or note or "review" or short survey or trade journal)

8. limit 6 to (embryo or infant or child or preschool child or school child or adolescent)

9. limit 6 to (concensus development or meta analysis or "systematic review")

10. limit 6 to (phase 1 clinical trial or phase 2 clinical trial or phase 3 clinical trial or phase 4 clinical trial)

11. 7 or 8 or 9 or 10

12. 6 not 11

Appendix 4

Any results that were assessed under the COSMIN structural validity and internal consistency domains were ignored for the following PROMs, (with the reason given, and whether any structural validity or internal consistency analyses were undertaken):

- RADAI requires the pre-calculation of the RADAI tender joint count – Yes;
- PDAS2 requires the pre-calculation of HAQ and the 28-joint swollen joint count – Yes;
- PAS requires the pre-calculation of HAQ – No;
- PAS-II requires the pre-calculation of the 10-item HAQ-II – No;
- RAPID3 requires the pre-calculation of a 10-item modified HAQ – Yes;
- RAPID4 requires the pre-calculation of a 10-item modified HAQ and the RADAI tender joint count – No;
- PRO-CLARA requires the pre-calculation of the 10-item ROAD and the RADAI tender joint count – Yes;
- GAS requires the pre-calculation of the 28-joint tender joint count – No.

Appendix 5

Hypotheses were agreed in consultation between all assessors and with clinical expertise from EC.

The following hypotheses were used:

- **for the COSMIN hypothesis testing for construct validity domain;**
 - *The following are deemed Convergent and thus require correlation ≥ 0.5 ;*
 - Assessor tender joint count (Assessor 28TJ/ Assessor 68TJ);
 - Assessor swollen joint count (Assessor 28SJ/Assessor 80SJ);
 - Assessor global rating (Assessor global/EGA);
 - Patient global rating (Patient global/PGA);
 - Functional ability (HAQ/MDHAQ);
 - Disease activity other than DAS28, CDAI and SDAI (Modified disease activity score/Mallya index);
 - Pain;
 - CRP;
 - ESR;
 - The Physical Component Summary (PCS) of SF-36;
 - The Nottingham Health Profile Physical Abilities subarea (NHP Physical);
 - The Nottingham Health Profile Pain subarea (NHP Pain);
 - *The following are deemed Semi convergent and require correlation ≥ 0.4 ;*
 - Fatigue;
 - Generic quality of life (EQ-5D/EQ-5D VAS);
 - SOFI-Hand;
 - Pulp-to-palm distance ;
 - *The following are deemed Divergent and required correlation ≥ 0.3 and <0.5 ;*
 - The Mental Component Summary (MCS) of SF-36;
 - The Nottingham Health Profile Sleep subarea (NHP Sleep);
 - The Nottingham Health Profile Social Isolation subarea (NHP Social);
 - The Nottingham Health Profile Emotional Reaction subarea (NHP Emotion);
 - Grip strength;
 - Muscle Strength Index;
 - Hemoglobin;
 - *Effect size ≥ 0.6 ;*
 - Effect size to discriminate between flare and non-flare groups;

- Effect size to discriminate between Patient Acceptable Symptom State (PASS) positive and negative groups;
- *AUC ≥ 0.70 and 95% CI Lower Bound > 0.5 ;*
 - Area under the Receiver Operating Characteristic Curve (AU-ROC-C) to discriminate between those active and non-active;
 - AU-ROC-C to discriminate between those in OMERACT minimal disease activity (MDA) and above MDA;
 - AU-ROC-C to discriminate between those in modified ACR MDA and above MDA;
- **for the COSMIN responsiveness domain;**
 - *Correlation ≥ 0.7 (those of criterion validity);*
 - DAS28 (any variant);
 - CDAI;
 - SDAI;
 - *Correlation ≥ 0.5 ;*
 - Assessor swollen joint count (Assessor 28SJ);
 - Assessor swollen joint count (Assessor 28SJ);
 - Assessor global rating (Assessor global);
 - Pain;
 - C-reactive Protein (CRP);
 - The Physical Component Summary (PCS) of SF-36;
 - The Nottingham Health Profile Physical Abilities subarea (NHP Physical);
 - The Nottingham Health Profile Pain subarea (NHP Pain);
 - *Correlation ≥ 0.4 ;*
 - Fatigue;
 - Generic quality of life (EQ-5D);
 - *Effect size ≥ 0.6 ;*
 - Effect size pre-post;
 - Effect size of change score to discriminate between flare and non-flare groups;
 - Effect size of change score to discriminate between placebo and intervention arms;
 - *Effect size ≥ 0.8 ;*
 - Effect size pre-post of responders;

- *Effect size < 0.2;*
 - Effect size pre-post of non-responders;
- *AUC \geq 0.70 and 95% CI Lower Bound > 0.5 (NB. already defined by COSMIN);*
 - AU-ROC-C using change score to discriminate between responders and non-responders;
 - AU-ROC-C using change score to discriminate between those in flare and non-flare;
 - AU-ROC-C using change score to discriminate between those improved and non-improved).

For correlations, the magnitude was relevant for assessment, but the sign (positive or negative) were not considered.