UNIVERSITY OF LEEDS

This is a repository copy of *Patient Reported Outcome Measures for Rheumatoid Arthritis Disease Activity: Rasch measurement theory to identify items and domains*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/225802/

Version: Accepted Version

Article:
Pickles, T., Horton, M. orcid.org/0000-0002-6675-7335, Christensen, K.B. et al. (6 more authors) (2025) Patient Reported Outcome Measures for Rheumatoid Arthritis Disease Activity: Rasch measurement theory to identify items and domains. Rheumatology. keaf189. ISSN 1462-0324 (In Press)

https://doi.org/10.1093/rheumatology/keaf189

This is an author produced version of an article accepted for publication in Rheumatology made available under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Patient Reported Outcome Measures for Rheumatoid Arthritis Disease Activity: Rasch measurement theory to identify items and domains

Tim Pickles[1], Mike Horton[2], Karl Bang Christensen[3], Rhiannon Phillips[4], David Gillespie[1], Neil Mo[5], Janice Davies[6], Susan Campbell[6], Ernest Choy[7]

[1]Centre for Trials Research, Cardiff University, Cardiff, UK;

[2]Psychometric Laboratory for Health Sciences, University of Leeds, Leeds, UK;

[3]Section of Biostatistics, University of Copenhagen, Copenhagen, Denmark;

[4]Cardiff School of Sport and Health Sciences, Cardiff Metropolitan University, Cardiff, UK;

[5]Swansea Bay University Health Board, Port Talbot, UK;

[6]Patient and Public Involvement Stakeholders, UK;

[7]Division of Infection and Immunity, Cardiff University, Cardiff, UK;

Corresponding author details:

Tim Pickles

5th Floor, Neuadd Meirionnydd

Heath Park

Cardiff

CF14 4YS

PicklesTE@cardiff.ac.uk

Abstract

Objectives

Disease Activity (DA) monitoring is a standard of care in Rheumatoid Arthritis (RA). There is demand for achieving this through Patient Reported Outcome Measures (PROMs). The aim of this study was to determine which items could be used to measure the construct of RA DA, by analysing legacy PROMs, using Rasch measurement theory (RMT) analyses.

Methods

Questionnaires including 10 legacy PROMs were sent to people with RA to create original and validation datasets. Items were grouped according to OMERACT domains and analysed using Principal Components Analysis. By domain RMT analyses in original dataset, and domain-level testlets were assessed to determine which measure the construct of RA DA. The result was then replicated in confirmatory factor analyses bifactor models and RMT analyses in the validation dataset. Psychometric properties of legacy PROMs was assessed in the original dataset.

Results

The total sample size was 691 (original: 398, validation: 293). The *Patient Global* domain was split into *General health* and *Disease activity* domains under RMT. *General health* and *Fatigue* domain items measure a separate construct to the construct of RA DA. A set of 12 *Pain*, *Disease activity*, *Tenderness and swelling*, *Physical functioning* and *Stiffness* domain items can be used to measure the construct of RA DA. No legacy PROMs fully fit the Rasch measurement model.

Conclusion

General health and *Disease activity* domain items are not interchangeable. 12 items form an item pool that can be used to measure the construct of RA DA. Legacy PROMs should not be recommended for use.

Keywords (up to 10 – please note that the word count refers to individual words, not phrases)

Rheumatoid Arthritis Disease Activity

Patient Reported Outcome Measures

Measurement properties

Key messages (up to 3, maximum 15 words each)

General health and Disease activity domain items are not inter-changeable

RA DA requires *Tenderness and swelling*, *Pain*, *Disease activity*, *Stiffness* and *Physical functioning* domain items

No legacy PROMs fully fit Rasch measurement model

Introduction

Patient Reported Outcome Measures (PROMs) are critical to research and clinical care, as recognised by the U.S. Food and Drug Administration (FDA), who mandated PROMs to be captured in all randomised controlled trials. Additionally, they have published guidelines on how to develop and validate PROMs. (1, 2) Disease activity (DA) monitoring is a standard of care in Rheumatoid Arthritis (RA), and there is demand for achieving this through PROMs. Although there are many RA DA PROMs, (1) these are currently used as secondary outcomes in clinical trials of rheumatic diseases, but rarely in clinical care. All these PROMs were developed using classical test theory methods and often have various limitations. FDA (2, 3) and COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines (4-6) both recognise item response theory (IRT) and Rasch measurement theory (RMT) as suitable methods to assess the measurement properties of instruments. Validation using these methods requires PROMs to meet stringent measurement criteria, which include unidimensionality, internal consistency, targeting and lack of local dependence and differential item functioning. Thus, IRT and RMT provide a statistical framework where all these measurement criteria can be formulated as testable hypotheses. Specifically, RMT (7-9) allows for these attributes to be formally assessed, as it provides a template to determine PROM score validity.

A systematic review (10) of 10 legacy RA DA PROMS showed that none can be recommended for use according to COSMIN guidelines. (4-6) This justifies the need to collect further data to start the process of determining the domains, and items within those domains, that can be used to measure the construct of RA DA.

The overall aim of this study was to use RMT analyses to determine which items can form an item pool to measure the construct of RA DA to. A secondary aim was to examine the measurement properties of legacy RA DA PROMs and other relevant PROMs.

Methods

This research is reported in line with the Strengthening the reporting of observational studies in epidemiology (STROBE) framework (Supplementary Material 1). (11)

Study design

This was a cross-sectional study that took place in 2020 and 2021. In Cardiff and Vale and Swansea Bay University Health Boards (UHBs), potential participants were identified by NHS staff by searching the electronic health records of the Rheumatology Department for those at least 18 years old with

Rheumatoid Arthritis (RA). In Aneurin Bevan UHB, potential participants were identified by NHS staff as those at least 18 years old with an entry on the British Society for Rheumatology Biologics Registry for Rheumatoid Arthritis (BSRBR-RA) database, Paper questionnaires were sent out as part of study packs to these people living with RA (plwRA) In Cwm Taf Morgannwg UHB, potential participants were identified as those at least 18 years old with RA in clinic by NHS staff and handed the study pack. Inclusion criteria were: at least 18 years old; a diagnosis of RA and signed informed consent. Patients were excluded if they were unable to complete the questionnaire in English. The study was approved by the North West – Preston Research Ethics Committee (20/NW/0039).

Sample Size

To provide item calibrations within ± 0.5 logits within a Rasch measurement theory (RMT) analysis, the advised sample size is 250. (12) Given this, it was decided that a sample size of n ≥ 250 was required, for both an original dataset and a validation dataset.

Questionnaire creation

A questionnaire (see Supplementary Material 2) was created based on the items from 10 legacy Patient Reported Outcome Measures (PROMs) identified and reviewed in a systematic review: (10)

- Rheumatoid Arthritis Disease Activity Index-5 (RADAI5); (13-15)
- Rheumatoid Arthritis Disease Activity Index (RADAI); (16, 17)
- RADAI-SF; (17, 18)
- Patient-based Disease Activity Score 2 (PDAS2); (19, 20)
- Patient Reported Outcome CLinical ARthritis Activity (PRO-CLARA); (21)
- Global Arthritis Score (GAS); (22)
- Patient Activity Score (PAS); (23)
- Patient Activity Score-II (PAS-II); (23)
- Routine Assessment of Patient Index Data 3 (RAPID3); (24)
- Routine Assessment of Patient Index Data 4 (RAPID4). (25)

Also included were the items from two PROMs measuring level of flare:

- Rheumatoid Arthritis Flare Questionnaire (RA-FQ) (26, 27);
- FLARE-RA (which includes FLARE-RA Old, FLARE-RA Arthritis and FLARE-RA General Symptoms). (28-31)

The items of The Rapid Assessment of Disease Activity in Rheumatology (RADAR), (32, 33) PROM-score (34) and the foot-specific RADAI-F5, (35) were included, as were fatigue items included on the

PAS and PAS-II assessments, the Health Assessment Questionnaire (HAQ) (PDAS2, PAS) and the multidimensional Health Assessment Questionnaire (MDHAQ) (used in RAPID3, RAPID4). The HAQ also has an additional pain item. RA-FQ has additional items about having a flare and how long it has been going on.

A draft questionnaire containing these items was discussed with two groups of plwRA: a meeting with J.D. and S.C. and a focus group convened by the National Rheumatoid Arthritis Society (NRAS). From these discussions, items on discomfort when walking, standing and exercising, plus fear of falling when walking were added. These four items used the Copenhagen Hip and Groin Outcome Score (HAGOS) (36) as a template. A focus group attendee also provided a pain scale, which was included. Thus, the total item pool contained 268 items (Supplementary Material 2, which states item codes).

Demographic items relating to current age, age at diagnosis, gender and sex assigned at birth, shielding during the COVID-19 pandemic, whether the participant completed the questionnaire themselves, ethnicity, education level, earlier and accompanying diseases, current or previous disease-modifying antirheumatic drug (DMARD) treatment are also included.

Item grouping

All items in the questionnaire, minus the two homunculi (`G01`, `A02`) and the aids and devices and help from another person items from HAQ (`H10`, `H11`, `H23`, `H24`), were grouped according to Outcome Measures in Rheumatology (OMERACT) domains for RA. (37, 38) 145 items were initially grouped by T.P. (researcher) and then checked by E.C. (Rheumatologist) to ensure correct grouping. Where necessary, additional domains were created (Table 1).

Analyses

Principal component analysis – original dataset only

Principal component analyses (PCA) (39) were undertaken on the 145 items described listed in Table 1. Two PCA were undertaken, one using a polychoric correlation matrix and another using Pearson's correlation coefficients. Within the PCA, the principal-component factor method was used and only factors with a minimum eigenvalue of 1 were retained. Oblique promax rotation was then applied. The purpose was to see if items within the identified domains loaded together onto factors that reflected those domains. If this was the case, the domain, and the items loading to that domain, were carried forward to further RMT analyses.

Rasch measurement theory – original and validation datasets

The Rasch Measurement Model (RMM) is a statistical model (7-9, 40) where the sum score of item responses contains all information about the underlying latent trait, here the construct of RA disease activity (DA), in a statistical concept known as sufficiency. The satisfaction of RMM assumptions therefore provides a prescription for what is necessary for a PROM to deliver fundamental measurement, (41) and therefore RA DA PROMs should be assessed on this basis.

Items were assessed by RMT analyses, which provides results on targeting and item locations, overall and individual item fit to the RMM, internal consistency, local dependency, uni-dimensionality and item threshold ordering. Differential item functioning was investigated by age group (18 to 54, 55 to 74, 75+), age at diagnosis (2 to 36, 37 to 56, 57+), sex (male, female), earlier and accompanying diseases (yes, no), previous DMARD treatment (yes, no), and highest educational qualification (Qualifications below university graduate, University graduate qualification as minimum). Grouping for age group and age at diagnosis were determined by the inter-quartile ranges for these variables.

RMT analyses in the original dataset were undertaken on items grouped by domain, with the purpose to identify potential items within each domain as candidate items for an item pool.

In the validation dataset, RMT analyses were undertaken on the potential items for each domain. Where discrepancies were found, these were reported. If suitable, items within domains were grouped together to form domain-level testlets, which operate as single items that represent a domain. These domain-level testlets were assessed together by RMT analyses to determine whether they could measure the construct of RA DA. If any evidence was found that this was not the case, iterative changes were made to achieve better fit to the RMM.

Structural validity – original and validation datasets

A confirmatory factor analysis (CFA) model is a statistical model used to test whether measures of a construct are consistent with a hypothesised measurement model based on theory and/or previous analytic research. (42, 43) CFA using Mplus (44) was used to calculate a X^2-test, root mean square error of approximation (RMSEA) along with an accompanying 90% confidence interval (CI), comparative fit index (CFI), Tucker-Lewis index (TLI), standardised root mean square residual (SRMR), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

CFA was applied to the validation dataset to examine whether the solution determined by RMT analyses could replicated in CFA using bifactor models. (45)

Legacy PROMs – original dataset only

To assess construct validity, Mann-Whitney U tests (46) were performed to see if there was a difference between those identifying as having a flare and not having a flare, with a Hodges-Lehmann median difference and 95% CI calculated. (47) Spearman's ρ correlation coefficients (48) were calculated between legacy PROM scores, with the hypothesis that all $\rho \geq 0.5$. To assess internal consistency, Cronbach's α (49) values were calculated. In line with COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines, (4-6) internal consistency was indicated by $\alpha > 0.7$. Legacy PROMs in the original dataset were assessed using CFA. In line with COSMIN guidelines, (4-6) structural validity was indicated by RMSEA < 0.06, TLI > 0.95, CFI > 0.95 and SRMR < 0.08. RMT analyses were applied to the legacy PROMs in the original dataset to assess the measurement properties structural validity, internal consistency and measurement invariance.

Results

Descriptives

The total sample size was n = 691, with n = 398 in the original dataset and n = 293 in the validation dataset. Study packs were sent out in batches in September 2020 and June, October and November 2021. The mean current age was 63.8 (SD 12.82), mean age at diagnosis was 46.4 (SD 15.69) and mean disease duration was 17.3 years (SD 13.65). 67.4% (466/691) were female and all were the same as assigned at birth (Table 2). 15.5% (107/691) completed all demographic questions and legacy PROM items of the questionnaire.

Principal component analysis – original dataset

From the results of both principal component analyses, a set of 30 items loaded together with other items in the domains they were grouped in, a priori. These were taken forward for Rasch measurement theory (RMT) analyses. These items were in the *Tenderness and swelling*, *Patient global*, *Pain*, *Fatigue*, *Physical functioning* and *Stiffness* domains (Figure 1).

Rasch measurement theory – original dataset

Tenderness and swelling

The three items (`D02`, `T02`, `Q04`) in the *Tenderness and swelling* domain provided good fit to the Rasch measurement model (RMM) and were retained.

Patient global

Of the 10 *Patient global* domain items, five were general health items and five were disease activity (DA) items. There was evidence of local dependence between general health items and, separately, evidence of local dependence between DA items (Table 3). The residual principal components loadings also showed that all general health items loaded negatively, whilst all DA items loaded positively, on the first component (Table 3). Given this, two new domains were created: *General health* and *Disease activity*.

General health

For the five *General health* domain items, there were four item showing misfit, one item with differential item functioning (DIF) by sex and local dependence between three items. It was decided to retain the other two items alongside one of these locally dependent items and therefore three items (R05, P01, C01) were retained for the *General health* domain.

Disease activity

For the five *Disease activity* domain items, there were two item showing misfit and all items were locally dependent on other items. There was a distinction in local dependence between the three items with a six-month and those with shorter recall periods. These three items were the only items amongst the 30 with a six-month recall so it was decided to retain the other two items (PS1, A01) in the *Disease activity* domain.

Pain

For the eight *Pain* domain items, there were five items showing misfit and only one item was not locally dependent on another item. It was decided to retain three items (one with no local dependence (F01) and two with only minimal evidence of local dependence between them (R04 and P07)) and one of the five locally dependent items. Four items (F01, R04, P07, Q05) were retained in the *Pain* domain that provided the best fit to the RMM.

Fatigue

The four *Fatigue* domain items demonstrated three item showing misfit, and DIF by age group and gender for one item. On retaining the three items without DIF, the analysis showed only a minor issue for item misfit and therefore these three items (F03, PF1, RF1) were retained for the *Fatigue* domain.

Physical functioning

The two *Physical functioning* domain items (F02, F05) provided good fit to the RMM and were retained.

Stiffness

For the three items in the *Stiffness* domain, there was one item showing misfit, all items had disordered thresholds and one item displayed DIF by earlier and accompanying diseases. There were two duration items, one of which had entirely illogical threshold ordering, and one intensity items. Therefore, the single intensity item (F04) was retained in the *Stiffness* domain.

Rasch measurement theory – validation dataset

Discrepancies

There was evidence of DIF by earlier and accompanying diseases for two items in the *General health* domain. For the *Pain* domain, the original item over-discrimination issue remained, and another item also displayed misfit. A pair of items displayed local dependence and unidimensionality could not be evidenced. For the *Fatigue* domain, there was evidence of item misfit and also DIF by highest educational qualification.

There were no discrepancies for the analyses of the *Tenderness and swelling*, *Disease activity* and *Physical functioning* domains, with no analysis for the *Stiffness* domain (only one item retained).

Domain-level testlets

None of the above discrepancies led to any need for changes to be made, therefore seven domain-level testlets representing the *Tenderness and swelling*, *General health*, *Disease activity*, *Pain*, *Fatigue*, *Physical functioning* and *Stiffness* domains were created (using the 18 retained items) and analysed. The *Fatigue* domain-level testlet had an extremely high positive fit residual (indicating under-discrimination) and also displayed extremely large negative residual correlations with the *Tenderness and swelling*, *Disease activity*, *Pain*, *Physical functioning* and *Stiffness* domain-level testlets. This suggested that the *Fatigue* domain-level testlet did not measure the same construct as the other domain testlets (Figure 2; panel (a)) and it was therefore removed.

Analysis of the six remaining domain-level testlets provided a similar picture for the *General health* domain-level testlet: an extremely high positive fit residual (indicating under-discrimination, Figure 2; panel (b)) and also extremely large negative residual correlations with all of the domain testlets. This suggested that the *General health* domain-level testlet did not measure the same construct as the other domain testlets and it was therefore removed.

A final analysis of the five remaining domain-level testlets displayed issues but none that required further change. There was item misfit for the *Disease activity* domain-level testlet with a large negative fit residual (indicating over-discrimination) and a significant F-value. The *Physical functioning* domain-level testlet also had a large positive fit residual (indicating under-discrimination). However, the item characteristic curves did not suggest any issues, so these were determined to be non-problematic. For the *Disease activity* domain-level testlet to exhibit over-discrimination was logical, as this is the same wording as the construct of Rheumatoid Arthritis (RA) DA itself. The *Physical functioning* domain-level testlet is more of a functional status than a symptom status so may under-discriminate in comparison to the other domain-level testlets. There was evidence of local dependence between the *Disease activity* and *Tenderness and swelling* domain-level testlets and the *Physical functioning* and *Stiffness* domain-level testlets. Both of these combinations have conceptual sense in that RA DA inevitably causes tenderness and swelling, and greater levels of stiffness create issues with physical functioning. The *Pain* and *Physical functioning* domain-level testlets displayed DIF by age group, though this DIF was not evident graphically for the *Pain* domain-level testlet. For *Physical functioning* domain-level testlet, it was logical that those 75 and over were at higher levels across the continuum in comparison to the other two age group categories. Also, unidimensionality could not be proven.

The 12 items therefore retained across the *Pain*, *Disease activity*, *Tenderness and swelling*, *Physical functioning* and *Stiffness* domains have their item codes highlighted in green in Supplementary Material 2.

Confirmatory factor analysis – validation dataset

Confirmatory factor analysis (CFA) was used to assess and compare a 1-dimensional bifactor model and a 2-dimensional bifactor model, with a hypothesis that the 2-dimensional bifactor model would produce better summary statistics as it better represented the model created through RMT analyses. This hypothesis was confirmed as all summary values were better for the 2-dimensional bifactor model (Figure 3).

Legacy Patient Reported Outcome Measures **– original dataset**

For all legacy Patient Reported Outcome Measures (PROMs), the median of those having a flare was greater than the median of those not having a flare and, when compared through a Mann-Whitney U test, produce $p < 0.001$. (Supplementary Material 3 Supplementary Table 1). Spearman's ρ correlation coefficients were generally very high ($\rho \geq 0.833$ for RA DA PROMs) (Supplementary Material 3 Supplementary Table 2). Except for the PDAS2 variations, $\alpha \geq 0.802$ across the PROMs

(Supplementary Material 3 Supplementary Table 3). Detail on discretised VAS items is shown in Supplementary Material 3 Supplementary Table 4. The CFA results show that only RADAI5, RADAI-SF and RA-FQ could evidence structural validity (Supplementary Material 3 Supplementary Table 5). RADAI5, RADAI, RADAI-SF, PDAS2, PRO-CLARA, GAS, PAS, PAS-II, RAPID3, RAPID4, PROM-score, RADAI-F5 and FLARE-RA Old did not fit the RMM (Supplementary Material 3 Supplementary Table 6) and all had misfitting items. Local dependence, disordered thresholds and DIF were issues across the majority of legacy PROMs. Unidimensionality could only be evidenced for PROM-score, RADAI-F5, FLARE-RA Arthritis, FLARE-RA General Symptoms and RA-FQ. The Person Separation Index was high for all PROMs suggesting good levels of internal consistency. The measurement properties of the legacy PROMs are summarised in Supplementary Material 3 Supplementary Figure 1.

Discussion

We undertook a cross-sectional study in people living with Rheumatoid Arthritis (RA) (plwRA) to determine which items can form an item pool to measure the construct of RA disease activity (DA) and examine the measurement properties of legacy RA DA Patient Reported Outcome Measures (PROMs) and other relevant PROMs.

In analysing the initial domains under Rasch measurement theory (RMT), *General health* and *Disease activity* were found to be separate domains within the *Patient Global* domain. By analysing domain-level testlets, it was found that 12 items across the *Pain*, *Disease activity*, *Tenderness and swelling*, *Physical functioning* and *Stiffness* domains can be used to form an item pool for a new PROM to measure the construct of RA DA. *Fatigue* and *General health* domain items were shown through RMT analyses to measure a separate construct to the construct of RA DA.

Additionally, whilst all legacy PROMs had good evidence for the internal consistency and hypothesis testing for construct validity, and that many had evidence for the structural validity from confirmatory factor analysis (CFA), no legacy PROMs could fully evidence fit to the Rasch measurement model.

The strength of this study is the novel and detailed strategy for analyses for the construct of RA DA. This was the first use of cross-validation (testing across two datasets) and RMT analyses for such items. This was the first use of CFA to complement RMT analyses, and the first use of bifactor models within CFA to confirm such an item structure. Equally, this was also the first time that RMT analyses were applied to assess the measurement properties of legacy PROMs with adequate sample size to obtain reliable estimates could be obtained through RMT analyses.

Patient and Public Involvement

J.D. and S.C., both plwRA, co-developed the participant information sheets, consent forms and questionnaires. The National Rheumatoid Arthritis Society (NRAS) organised a focus group of 15 plwRA to discuss this research ahead of application.

Limitations

The data collected were from a small, densely populated, area of South Wales, with an assumption that participants were able to understand the English language used in study documents and data collection forms. Collecting data from one geographical area meant that it was not possible to undertake simultaneous external validation with data from another area.

The paper questionnaire was very long at 18 pages: this and other factors contributed to only 15.5% providing a response to all demographic questions and legacy PROM items. These questionnaires were also sent out at varying stages of the lockdowns enforced in Wales as a result of the COVID-19 pandemic. This may have discouraged potential participants from responding to the questionnaire, and possibly in different ways across distinct demographic groups. Further detail is available in Supplementary Material 4.

Future research

The next step is to undertake cognitive interviews with plwRA to assess the content validity measurement property. This will determine whether plwRA believe these items have relevance, comprehensiveness and comprehensibility in order to measure the construct of RA DA. The 12 items have different recall periods, response formats and anchor wordings, so it will be important to explore preferences around these.

If this can be evidenced, then the item pool can be used to develop a computer adaptive test (CAT) or electronic PROM. However, there are only 12 items in the item pool so the CAT will only provide a marginal burden reduction for plwRA, as a minimum of five items must be asked to cover all domains.

Acknowledgements: Kerry Nyland and Liz Griebel from the Research and Development Delivery Team in Cardiff and Vale UHB, Paula Phillips in the rheumatology department at Swansea Bay UHB, Keri Turner at the Research and Development Delivery Team and Ceril Rhys-Dillon and Catrin Margaret Jones in the rheumatology department in Cwm Taf Morgannwg UHB and Anna Roynon the Research and Development Delivery Team in Aneurin Bevan UHB. Marcin Bargiel and Terri Kitson at the Centre for Trials Research for building the study database and for helping with the data entry respectively.

Funding: T.P. was supported by a National Institute of Health Research Doctoral Fellowship, funded by the Welsh Government through Health and Care Research Wales (NIHR-FS-19).

Conflict of interest statement: E.C. has received research grants from Bio-Cancer, Biogen, Pfizer and Sanofi, and honoraria/or served as member of speakers' bureaus from Abbvie, Bio-Cancer, Biocon, Biogen, Eli Lilly, Fresenius Kai, Galapagos, Janssen, Pfizer, Sanofi, UCB and Viatris. EC also receives a stipend as editor in chief of Rheumatology (Oxford). N.M. has received consultancy fees from Novartis and Roche. N.M. has received other financial support from Abbvie, Amgen and UCB.

Data availability statement: Data can be made available on request to the Centre for Trials Research https://www.cardiff.ac.uk/centre-for-trials-research/collaborate-with-us/data-requests

References

1. Hendrikx J, de Jonge MJ, Fransen J, Kievit W, van Riel PL. Systematic review of patient-reported outcome measures (PROMs) for assessing disease activity in rheumatoid arthritis. RMD Open. 2016;2(2):e000202. https://doi.org/10.1136/rmdopen-2015-000202
2. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, the Mayo FDA Patient-Reported Outcomes Consensus Meeting Group. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health. 2007;10 Suppl 2:S94-S105. https://doi.org/10.1111/j.1524-4733.2007.00272.x
3. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes. 2006;4:79. https://doi.org/10.1186/1477-7525-4-79
4. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018;27(5):1171-9. https://doi.org/10.1007/s11136-017-1765-4
5. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147-57. https://doi.org/10.1007/s11136-018-1798-3
6. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159-70. https://doi.org/10.1007/s11136-018-1829-0
7. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danmarks Paedagogiske Institut; 1960.
8. Christensen KB, Kreiner S, Mesbah M. Rasch models in health. London: ISTE & Hoboken, NJ: John Wiley & Sons; 2013.
9. Fischer GH, Molenaar IW. Rasch models: foundations, recent developments, and applications. New York: Springer-Verlag; 1995.
10. Pickles T, Macefield R, Aiyegbusi OL, Beecher C, Horton M, Christensen KB, et al. Patient Reported Outcome Measures for Rheumatoid Arthritis Disease Activity: a systematic review following COSMIN guidelines. RMD Open. 2022;8:e002093. https://doi.org/10.1136/rmdopen-2021-002093
11. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ. 2007;335(7624):806-8. https://doi.org/10.1136/bmj.39335.541782.AD
12. Linacre JM. Sample Size and Item Calibration [or Person Measure] Stability. Rasch Measurement Transactions. 1994;7(4):328. https://www.rasch.org/rmt/rmt74m.htm
13. Leeb BF, Haindl PM, Brezinschek HP, Nothnagl T, Rintelen B. RADAI-5 to monitor rheumatoid arthritis. Clin Exp Rheumatol. 2014;32(5 Suppl 85):S-55-8. https://www.clinexprheumatol.org/abstract.asp?a=8604
14. Leeb BF, Haindl PM, Maktari A, Nothnagl T, Rintelen B. Patient-centered rheumatoid arthritis disease activity assessment by a modified RADAI. J Rheumatol. 2008;35(7):1294-9. https://www.jrheum.org/content/35/7/1294

15. Leeb BF, Sautner J, Mai HT, Haindl PM, Deutsch C, Rintelen B. A comparison of patient questionnaires and composite indexes in routine care of rheumatoid arthritis patients. Joint Bone Spine. 2009;76(6):658-64. https://doi.org/10.1016/j.jbspin.2009.04.001
16. Fransen J, Langenegger T, Michel BA, Stucki G. Feasibility and validity of the RADAI, a self-administered rheumatoid arthritis disease activity index. Rheumatology (Oxford). 2000;39(3):321-7. https://doi.org/10.1093/rheumatology/39.3.321
17. Stucki G, Liang MH, Stucki S, Bruhlmann P, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. Psychometric properties and correlation with parameters of disease activity. Arthritis Rheum. 1995;38(6):795-8.
18. Veehof MM, ten Klooster PM, Taal E, van Riel PL, van de Laar MA. Psychometric properties of the Rheumatoid Arthritis Disease Activity Index (RADAI) in a cohort of consecutive Dutch patients with RA starting anti-tumour necrosis factor treatment. Ann Rheum Dis. 2008;67(6):789-93. https://doi.org/10.1136/ard.2007.081984
19. Choy EH, Khoshaba B, Cooper D, MacGregor A, Scott DL. Development and validation of a patient-based disease activity score in rheumatoid arthritis that can be used in clinical trials and routine practice. Arthritis Rheum. 2008;59(2):192-9. https://doi.org/10.1002/art.23342
20. Leung AM, Farewell D, Lau CS, Choy EH. Defining criteria for rheumatoid arthritis patient-derived disease activity score that correspond to Disease Activity Score 28 and Clinical Disease Activity Index based disease states and response criteria. Rheumatology (Oxford). 2016;55(11):1954-8. https://doi.org/10.1093/rheumatology/kew279
21. Salaffi F, Migliore A, Scarpellini M, Corsaro SM, Lagana B, Mozzani F, et al. Psychometric properties of an index of three patient reported outcome (PRO) measures, termed the CLinical ARthritis Activity (PRO-CLARA) in patients with rheumatoid arthritis. The NEW INDICES study. Clin Exp Rheumatol. 2010;28(2):186-200. https://www.clinexprheumatol.org/abstract.asp?a=543
22. Harrington JT. The uses of disease activity scoring and the physician global assessment of disease activity for managing rheumatoid arthritis in rheumatology practice. J Rheumatol. 2009;36(5):925-9. https://doi.org/10.3899/jrheum.081046
23. Wolfe F, Michaud K, Pincus T. A composite disease activity scale for clinical practice, observational studies, and clinical trials: the patient activity scale (PAS/PAS-II). J Rheumatol. 2005;32(12):2410-5. https://www.jrheum.org/content/32/12/2410
24. Pincus T, Bergman MJ, Yazici Y, Hines P, Raghupathi K, Maclean R. An index of only patient-reported outcome measures, routine assessment of patient index data 3 (RAPID3), in two abatacept clinical trials: similar results to disease activity score (DAS28) and other RAPID indices that include physician-reported measures. Rheumatology (Oxford). 2008;47(3):345-9. https://doi.org/10.1093/rheumatology/kem364
25. Pincus T, Swearingen CJ, Bergman MJ, Colglazier CL, Kaell AT, Kunath AM, et al. RAPID3 (Routine Assessment of Patient Index Data) on an MDHAQ (Multidimensional Health Assessment Questionnaire): agreement with DAS28 (Disease Activity Score) and CDAI (Clinical Disease Activity Index) activity categories, scored in five versus more than ninety seconds. Arthritis Care Res (Hoboken). 2010;62(2):181-9. https://doi.org/10.1002/acr.20066
26. Bartlett SJ, Barbic SP, Bykerk VP, Choy EH, Alten R, Christensen R, et al. Content and Construct Validity, Reliability, and Responsiveness of the Rheumatoid Arthritis Flare

Questionnaire: OMERACT 2016 Workshop Report. J Rheumatol. 2017;44(10):1536-43. https://doi.org/10.3899/jrheum.161145
27. Bykerk VP, Bingham CO, Choy EH, Lin D, Alten R, Christensen R, et al. Identifying flares in rheumatoid arthritis: reliability and construct validation of the OMERACT RA Flare Core Domain Set. RMD Open. 2016;2(1):e000225. https://doi.org/10.1136/rmdopen-2015-000225
28. Berthelot JM, De Bandt M, Morel J, Benatig F, Constantin A, Gaudin P, et al. A tool to identify recent or present rheumatoid arthritis flare from both patient and physician perspectives: the 'FLARE' instrument. Ann Rheum Dis. 2012;71(7):1110-6. https://doi.org/10.1136/ard.2011.150656
29. de Thurah A, Maribo T, Stengaard-Pedersen K. Patient self-assessment of flare in rheumatoid arthritis: criterion and concurrent validity of the Flare instrument. Clin Rheumatol. 2016;35(2):467-71. https://doi.org/10.1007/s10067-014-2849-y
30. Fautrel B, Alten R, Kirkham B, de la Torre I, Durand F, Barry J, et al. Call for action: how to improve use of patient-reported outcomes to guide clinical decision making in rheumatoid arthritis. Rheumatol Int. 2018;38(6):935-47. https://doi.org/10.1007/s00296-018-4005-5
31. Maribo T, de Thurah A, Stengaard-Pedersen K. Patient-self assessment of flare in rheumatoid arthritis: translation and reliability of the Flare instrument. Clin Rheumatol. 2016;35(4):1053-8. https://doi.org/10.1007/s10067-014-2848-z
32. Mason JH, Anderson JJ, Meenan RF, Haralson KM, Lewis-Stevens D, Kaine JL. The rapid assessment of disease activity in rheumatology (radar) questionnaire. Validity and sensitivity to change of a patient self-report measure of joint count and clinical status. Arthritis Rheum. 1992;35(2):156-62. https://doi.org/10.1002/art.1780350206
33. Mason JH, Meenan RF, Anderson JJ. Do self-reported arthritis symptom (RADAR) and health status (AIMS2) data provide duplicative or complementary information? Arthritis Rheum. 1992;5(3):163-72. https://doi.org/10.1002/art.1790050309
34. Hendrikx J, Fransen J, van Riel PL. Monitoring rheumatoid arthritis using an algorithm based on patient-reported outcome measures: a first step towards personalised healthcare. RMD Open. 2015;1(1):e000114. https://doi.org/10.1136/rmdopen-2015-000114
35. Hoque A, Gallagher K, McEntegart A, Porter D, Steultjens M, Woodburn J, et al. Measuring Inflammatory Foot Disease in Rheumatoid Arthritis: Development and Validation of the Rheumatoid Arthritis Foot Disease Activity Index-5. Arthritis Care Res (Hoboken). 2021;73(9):1290-9. https://doi.org/10.1002/acr.24259
36. Thorborg K, Holmich P, Christensen R, Petersen J, Roos EM. The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. Br J Sports Med. 2011;45(6):478-91. https://doi.org/10.1136/bjsm.2010.080937
37. Boers M, Tugwell P, Felson DT, van Riel PL, Kirwan JR, Edmonds JP, et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol Suppl. 1994;41:86-9. https://pubmed.ncbi.nlm.nih.gov/7799394/
38. Kirwan JR, Minnock P, Adebajo A, Bresnihan B, Choy E, de Wit M, et al. Patient perspective: fatigue as a recommended patient centered outcome measure in rheumatoid arthritis. J Rheumatol. 2007;34(5):1174-7. https://www.jrheum.org/content/34/5/1174

39. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 1901;2(11):559–572. https://doi.org/10.1080/14786440109462720
40. Kreiner S. Validity and objectivity: Reflections on the role and nature of Rasch models. Nord Psychol. 2007;59(3):268-98. https://doi.org/10.1027/1901-2276.59.3.268
41. Newby VA, Conner GR, Grant CP, Bunderson CV. The Rasch model and additive conjoint measurement. J Appl Meas. 2009;10(4):348-54. http://jampress.org/abst2009.htm
42. Jöreskog KG. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika. 1969;34(2):183-202. https://doi.org/10.1007/BF02289343
43. Kline RB. Principles and practice of structural equation modeling. 3rd ed. New York, NY: Guilford Press; 2011.
44. Muthen LK, Muthen B. Mplus Version 8 User's Guide. [online]. Muthen & Muthen: Los Angeles, CA; 2017 [Accessed 3 February 2021]. Available from: https://www.statmodel.com/ugexcerpts.shtml
45. Reise SP. The Rediscovery of Bifactor Measurement Models. Multivariate Behav Res. 2012;47(5):667-96. https://doi.org/10.1080/00273171.2012.715555
46. Mann HB, Whitney DR. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. Ann Math Stat. 1947;18(1):50-60. https://doi.org/10.1214/aoms/1177730491
47. Hodges JL, Lehmann EL. Estimates of Location Based on Rank Tests. Ann Math Stat. 1963;34(2):598-611. https://doi.org/10.1214/aoms/1177704172
48. Spearman C. The Proof and Measurement of Association between Two Things. Am J Psychol. 1904;15(1):72-101. https://doi.org/10.2307/1412159
49. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16(3):297–333. https://doi.org/10.1007/BF02310555

(Please ignore the references below – I can't get them to function better in endnote but the ones above are formatted better)

1. Hendrikx J, de Jonge MJ, Fransen J, Kievit W, van Riel PL. Systematic review of patient-reported outcome measures (PROMs) for assessing disease activity in rheumatoid arthritis. RMD Open. 2016;2(2):e000202.
2. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, the Mayo FDA Patient-Reported Outcomes Consensus Meeting Group. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health. 2007;10 Suppl 2:S94-S105.
3. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes. 2006;4:79.
4. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018;27(5):1171-9.
5. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147-57.
6. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159-70.

7. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen ,1960. xiii, 184 p. p.
8. Christensen KB, Kreiner S, Mesbah M. Rasch models in health. London Hoboken, NJ: ISTE John Wiley & Sons; 2013. xvi, 368 p. p.
9. Fischer GH, Molenaar IW. Rasch models : foundations, recent developments, and applications. New York: Springer-Verlag; 1995. viii, 436 p. p.
10. Pickles T, Macefield R, Aiyegbusi OL, Beecher C, Horton M, Christensen KB, et al. Patient Reported Outcome Measures for Rheumatoid Arthritis Disease Activity: a systematic review following COSMIN guidelines. RMD Open. 2022;8(1).
11. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ. 2007;335(7624):806-8.
12. Linacre JM. Sample Size and Item Calibration [or Person Measure] Stability. Rasch Measurement Transactions. 1994;7(4):328.
13. Leeb BF, Haindl PM, Brezinschek HP, Nothnagl T, Rintelen B. RADAI-5 to monitor rheumatoid arthritis. Clin Exp Rheumatol. 2014;32(5 Suppl 85):S-55-8.
14. Leeb BF, Haindl PM, Maktari A, Nothnagl T, Rintelen B. Patient-centered rheumatoid arthritis disease activity assessment by a modified RADAI. J Rheumatol. 2008;35(7):1294-9.
15. Leeb BF, Sautner J, Mai HT, Haindl PM, Deutsch C, Rintelen B. A comparison of patient questionnaires and composite indexes in routine care of rheumatoid arthritis patients. Joint Bone Spine. 2009;76(6):658-64.
16. Fransen J, Langenegger T, Michel BA, Stucki G. Feasibility and validity of the RADAI, a self-administered rheumatoid arthritis disease activity index. Rheumatology (Oxford). 2000;39(3):321-7.
17. Stucki G, Liang MH, Stucki S, Bruhlmann P, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. Psychometric properties and correlation with parameters of disease activity. Arthritis Rheum. 1995;38(6):795-8.
18. Veehof MM, ten Klooster PM, Taal E, van Riel PL, van de Laar MA. Psychometric properties of the Rheumatoid Arthritis Disease Activity Index (RADAI) in a cohort of consecutive Dutch patients with RA starting anti-tumour necrosis factor treatment. Ann Rheum Dis. 2008;67(6):789-93.
19. Choy EH, Khoshaba B, Cooper D, MacGregor A, Scott DL. Development and validation of a patient-based disease activity score in rheumatoid arthritis that can be used in clinical trials and routine practice. Arthritis Rheum. 2008;59(2):192-9.
20. Leung AM, Farewell D, Lau CS, Choy EH. Defining criteria for rheumatoid arthritis patient-derived disease activity score that correspond to Disease Activity Score 28 and Clinical Disease Activity Index based disease states and response criteria. Rheumatology (Oxford). 2016;55(11):1954-8.
21. Salaffi F, Migliore A, Scarpellini M, Corsaro SM, Lagana B, Mozzani F, et al. Psychometric properties of an index of three patient reported outcome (PRO) measures, termed the CLinical ARthritis Activity (PRO-CLARA) in patients with rheumatoid arthritis. The NEW INDICES study. Clin Exp Rheumatol. 2010;28(2):186-200.
22. Harrington JT. The uses of disease activity scoring and the physician global assessment of disease activity for managing rheumatoid arthritis in rheumatology practice. J Rheumatol. 2009;36(5):925-9.
23. Wolfe F, Michaud K, Pincus T. A composite disease activity scale for clinical practice, observational studies, and clinical trials: the patient activity scale (PAS/PAS-II). J Rheumatol. 2005;32(12):2410-5.
24. Pincus T, Bergman MJ, Yazici Y, Hines P, Raghupathi K, Maclean R. An index of only patient-reported outcome measures, routine assessment of patient index data 3 (RAPID3), in

two abatacept clinical trials: similar results to disease activity score (DAS28) and other RAPID indices that include physician-reported measures. Rheumatology (Oxford). 2008;47(3):345-9.
25. Pincus T, Swearingen CJ, Bergman M, Yazici Y. RAPID3 (Routine Assessment of Patient Index Data 3), a rheumatoid arthritis index without formal joint counts for routine care: proposed severity categories compared to disease activity score and clinical disease activity index categories. J Rheumatol. 2008;35(11):2136-47.
26. Bartlett SJ, Barbic SP, Bykerk VP, Choy EH, Alten R, Christensen R, et al. Content and Construct Validity, Reliability, and Responsiveness of the Rheumatoid Arthritis Flare Questionnaire: OMERACT 2016 Workshop Report. J Rheumatol. 2017;44(10):1536-43.
27. Bykerk VP, Bingham CO, Choy EH, Lin D, Alten R, Christensen R, et al. Identifying flares in rheumatoid arthritis: reliability and construct validation of the OMERACT RA Flare Core Domain Set. RMD Open. 2016;2(1):e000225.
28. Berthelot JM, De Bandt M, Morel J, Benatig F, Constantin A, Gaudin P, et al. A tool to identify recent or present rheumatoid arthritis flare from both patient and physician perspectives: the 'FLARE' instrument. Ann Rheum Dis. 2012;71(7):1110-6.
29. de Thurah A, Maribo T, Stengaard-Pedersen K. Patient self-assessment of flare in rheumatoid arthritis: criterion and concurrent validity of the Flare instrument. Clin Rheumatol. 2016;35(2):467-71.
30. Fautrel B, Morel J, Berthelot JM, Constantin A, De Bandt M, Gaudin P, et al. Validation of FLARE-RA, a Self-Administered Tool to Detect Recent or Current Rheumatoid Arthritis Flare. Arthritis Rheumatol. 2017;69(2):309-19.
31. Maribo T, de Thurah A, Stengaard-Pedersen K. Patient-self assessment of flare in rheumatoid arthritis: translation and reliability of the Flare instrument. Clin Rheumatol. 2016;35(4):1053-8.
32. Mason JH, Anderson JJ, Meenan RF, Haralson KM, Lewis-Stevens D, Kaine JL. The rapid assessment of disease activity in rheumatology (radar) questionnaire. Validity and sensitivity to change of a patient self-report measure of joint count and clinical status. Arthritis Rheum. 1992;35(2):156-62.
33. Mason JH, Meenan RF, Anderson JJ. Do self-reported arthritis symptom (RADAR) and health status (AIMS2) data provide duplicative or complementary information? Arthritis Care Res. 1992;5(3):163-72.
34. Hendrikx J, Fransen J, van Riel PL. Monitoring rheumatoid arthritis using an algorithm based on patient-reported outcome measures: a first step towards personalised healthcare. RMD Open. 2015;1(1):e000114.
35. Hoque A, Gallagher K, McEntegart A, Porter D, Steultjens M, Woodburn J, et al. Measuring Inflammatory Foot Disease in Rheumatoid Arthritis: Development and Validation of the Rheumatoid Arthritis Foot Disease Activity Index-5. Arthritis Care Res (Hoboken). 2021;73(9):1290-9.
36. Thorborg K, Holmich P, Christensen R, Petersen J, Roos EM. The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. Br J Sports Med. 2011;45(6):478-91.
37. Boers M, Tugwell P, Felson DT, van Riel PL, Kirwan JR, Edmonds JP, et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol Suppl. 1994;41:86-9.
38. Kirwan JR, Minnock P, Adebajo A, Bresnihan B, Choy E, de Wit M, et al. Patient perspective: fatigue as a recommended patient centered outcome measure in rheumatoid arthritis. J Rheumatol. 2007;34(5):1174-7.
39. Pearson K. LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 1901;2(11):559-72.

40. Kreiner S. Validity and objectivity: Reflections on the role and nature of Rasch models. Nordic Psychology. 2007;59(3):268-98.
41. Newby VA, Conner GR, Grant CP, Bunderson CV. The Rasch model and additive conjoint measurement. Journal of applied measurement. 2009;10(4):348-54.
42. Jöreskog KG. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika. 1969;34(2):183-202.
43. Kline RB. Principles and practice of structural equation modeling, 3rd ed. New York, NY, US: Guilford Press; 2011. xvi, 427-xvi, p.
44. Muthen LK, Muthen B. Mplus Version 8 User's Guide: Muthen & Muthen; 2017.
45. Reise SP. Invited Paper: The Rediscovery of Bifactor Measurement Models. Multivariate Behav Res. 2012;47(5):667-96.
46. Mann HB, Whitney DR. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. The Annals of Mathematical Statistics. 1947;18(1):50-60, 11.
47. Hodges JL, Lehmann EL. Estimates of Location Based on Rank Tests. The Annals of Mathematical Statistics. 1963;34(2):598-611, 14.
48. Spearman C. The Proof and Measurement of Association between Two Things. The American Journal of Psychology. 1904;15(1):72-101.
49. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–333.