Rheumatology, 2024, **63**, 181–189
https://doi.org/10.1093/rheumatology/kead201
Advance access publication 5 May 2023
Original Article

British Society for Rheumatology RHEUMATOLOGY OXFORD

Clinical science

Validation of a patient-reported outcome measure for giant cell arteritis

Mwidimi Ndosi [1,2]*, Celia Almeida[1,2], Jill Dawson [3], Emma Dures [1,2], Rosemary Greenwood [4], Alison Bromhead[1,2], Catherine Guly[5], Steve Stern[1,2], Catherine Hill [6,7], Sarah Mackie [8,9], Joanna C. Robson [1,2,*]

[1]School of Health and Social Wellbeing, University of the West of England, Bristol, UK
[2]Academic Rheumatology Unit, Bristol Royal Infirmary, University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK
[3]Nuffield Department of Population Health (HSRU), University of Oxford, Oxford, UK
[4]NIHR Research Design Service South West, University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK
[5]Bristol Eye Hospital, University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK
[6]Discipline of Medicine, The University of Adelaide, Adelaide, Australia
[7]Rheumatology Unit, Royal Adelaide Hospital, Adelaide, Australia
[8]Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK
[9]Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust, Leeds, UK

*Correspondence to: Mwidimi Ndosi, School of Health and Social Wellbeing, Glenside Campus, Blackberry Hill, Bristol BS16 1DD, UK. E-mail: Mwidimi.Ndosi@uwe.ac.uk; Joanna C. Robson, Academic Rheumatology Unit, Bristol Royal Infirmary, University Hospitals Bristol and Weston NHS Foundation Trust, Level 5, Upper Maudlin Street, Bristol BS2 8HW, UK. E-mail: Jo.Robson@uwe.ac.uk

Abstract

Objectives: GCA is systemic vasculitis manifesting as cranial, ocular or large vessel vasculitis. A prior qualitative study developed 40 candidate items to assess the impact of GCA on health-related quality of life (HRQoL). This study aimed to determine final scale structure and measurement properties of the GCA patient reported outcome (GCA-PRO) measure.

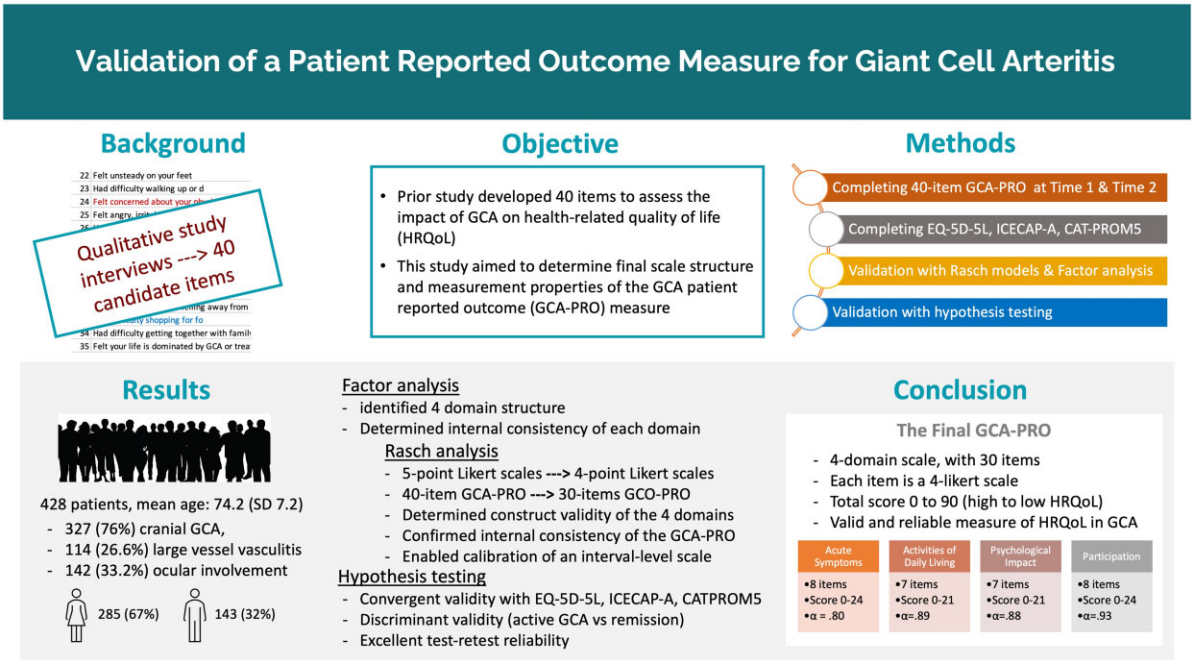
Methods: Cross-sectional study included UK patients with clinician-confirmed GCA. They completed 40 candidate items for the GCA-PRO at times 1 and 2 (3 days apart), EQ-5D-5L, ICECAP-A, CAT-PROM5 and self-report of disease activity. Rasch and exploratory factor analyses informed item reduction and established structural validity, reliability and unidimensionality of the final GCA-PRO. Evidence of validity was also established with hypothesis testing (GCA-PRO *vs* other PRO scores, and between participants with 'active disease' *vs* those 'in remission') and test–retest reliability.

Results: The study population consisted of 428 patients: mean (S.D.) age 74.2 (7.2), 285 (67%) female; 327 (76%) cranial GCA, 114 (26.6%) large vessel vasculitis and 142 (33.2%) ocular involvement. Rasch analysis eliminated 10 candidate GCA items and informed restructuring of response categories into four-point Likert scales. Factor analysis confirmed four domains: acute symptoms (eight items), activities of daily living (seven items), psychological (seven items) and participation (eight items). The overall scale had adequate Rasch model fit ($\chi^2 = 25.219$, degrees of freedom = 24, $P = 0.394$). Convergent validity with EQ5D-5L, ICECAP-A and Cat-PROM5 was confirmed through hypothesis testing. Internal consistency and test–retest reliability were excellent.

Conclusion: The final GCA-PRO is a 30-item, four-domain scale with robust evidence of validity and reliability in measuring HRQoL in people with GCA.

Received: 1 December 2022. **Accepted:** 4 April 2023

Graphical abstract



Keywords: GCA, temporal arteritis, quality of life, patient reported outcome measures, large vessel vasculitis, ocular GCA, Rasch analysis

Rheumatology key messages

- Giant cell arteritis and its treatment can have a negative impact on quality of life.
- A new disease-related patient-reported outcome measure, the GCA-PRO, has been validated.
- The GCA-PRO has been validated for use in clinical trials and clinical practice.

Introduction

GCA is the most common form of systemic vasculitis affecting people over the age of 50 [1]. Granulomatous inflammation of the medium and large extradural arteries causes narrowing or stenosis of the temporal arteries, thoracic aorta and its branches [2, 3]. Patients with cranial GCA present with headache, jaw claudication and scalp tenderness; those with ocular involvement either develop visual changes alongside cranial symptoms or can present with isolated visual symptoms [4]. Large vessel vasculitis is part of the spectrum of GCA, either presenting alongside cranial or visual symptoms or independently with systemic features of weight loss, fevers and raised inflammatory markers [5, 6].

Glucocorticoids (GC) have traditionally been the mainstay of treatment for GCA but can result in a range of adverse effects that can impact on health-related quality of life (HRQoL) [7]. Current recommendations advise use of glucocorticoid-sparing agents such as methotrexate, particularly in relapsing cases or patients with large vessel vasculitis [8, 9].

Patient-reported outcome measures (PROMs) capture the impact of disease on patients' HRQoL in clinical trials and practice [10, 11]. Their use can be key to evaluating the effectiveness of novel treatments in terms of patient benefit. PROMs can be generic, e.g. the Short-Form-36 (SF-36) [12] or Euroqol (EQ-5D-5L) [13], or disease or symptom specific [14, 15]. Using both generic and disease specific PROMS can be useful to ensure that the impact on HRQoL of a particular disease is accurately measured from the patients' perspective [16]. In 2015, the Large Vessel Vasculitis Working Group at the Outcome Measures in Rheumatology (OMERACT) consensus conference reported that a disease-specific PROM for GCA was required [17]. While generic PROMs have the benefit of enabling comparisons across different disease groups, they may be insensitive to disease-related factors. For example the generic SF-36 correlates poorly with ocular involvement in GCA, undermining its sensitivity as an outcome measure if used alone in clinical trials [18].

Patient involvement is key at every stage of clinical research and of critical importance in the development of patient reported outcomes [19–21]. Research has also shown that patients with vasculitis have different perspectives from their clinicians in terms of what is important to their HRQoL [22]. An international steering committee including patient research partners, clinicians and methodologists oversaw the first stage development of a PROM for GCA [23]. In-depth

qualitative interviews with people with GCA were completed in the UK and Australia. Based on this underpinning work, candidate items were developed and revised through cognitive interviewing and piloting, resulting in a 40-item draft questionnaire [23]. The aim of this study was to determine the final scale structure and measurement properties of the GCA patient reported outcome (GCA-PRO) measure.

Methods

Design

A cross-sectional validation study was conducted involving 38 National Health Service (NHS) rheumatology and ophthalmology centres in England and Wales. A steering committee comprising patient research partners, clinicians (rheumatology and ophthalmology), researchers, statisticians and methodologists oversaw the running of the study including review of all patient survey materials.

Patients

Patients were included if they had GCA confirmed by a clinician (rheumatologist or ophthalmologist) and were diagnosed within the previous 3 years or who had a flare within the previous year.

Recruitment strategy

Patients were screened for eligibility as they presented in face-to-face or telephone clinics by research nurses and clinicians at collaborating centres. Collaborators also reviewed their list of GCA registry patients (the UKIVAS registry [24] and the UK GCA Consortium [25]) for eligibility. These patients had previously given consent to being contacted directly about future studies in GCA.

Practical procedures

The study co-ordinator based at the Central Study Office at the Bristol Royal Infirmary sent study packs to collaborating centres and monitored return of completed questionnaires. Participants were able to use an advocate when completing the questionnaire and were asked to record this on the questionnaire.

Each collaborating centre kept a screening log with unique study ID numbers. They informed the study co-ordinator when they sent out the questionnaire packs and the associated study ID numbers. All documentation sent to participants/returned to the Central Study Office, contained only the study ID number with no identifiable patient details. The Central Study Office contacted the collaborating centre when a questionnaire was returned (i.e. the participant had given implied consent). The clinician at the collaborating centre then completed the Clinician Report Form and returned it to the Central Study Office.

Postal survey

The survey comprised two sets of questionnaires, completed twice, 3 days apart. Questionnaire pack A included: (i) the 40-item GCA PROM, comprising the 40 candidate items developed and refined during the qualitative study [23]; (ii) EuroQuol (EQ-5D-5L) [13], a short, generic measure of health status with five different dimensions, which can be used to compare patient states across different diseases; (ii) the ICECAP-A [26], a five-item measure of capability for the general adult (18+) population; (iii) the Cat-PROM5 [27], a five-item PROM capturing participants' quality of eyesight developed in people with cataracts; and (iv) patient self- assessment of disease state (active/remission), flare, treatment and demographics. The three PROMs (EQ-5D-5L [13], Cat-PROM5 [27 and ICECAP-A [26]) were selected for hypothesis testing (described in the analysis section) as they aim to capture relevant aspects of the impact of GCA and its treatment. They were selected on the advice of clinicians in rheumatology (J.C.R., S.M.), ophthalmology (C.G.) and medical statistician (R.G.) and reviewed by patient research partners (A.B. and S.S.).

Questionnaire pack B was sent with pack A but in a separate envelope marked 'IMPORTANT open 3 days after completing the first questionnaire'. Pack B contained the draft 40-item GCA PROM, and a question relating to change in state: 'Overall, how are you NOW (in terms of your GCA and any side effects) compared with three days ago (when you first answered the questionnaire)?' Response options were 'much better', 'slightly better', 'no change', 'slightly worse' or 'much worse'.

Clinician case report form

The clinician case report form contained questions regarding patient age, date of diagnosis, type of GCA, clinical features, diagnostic tests and treatments.

Sample size estimation

The sample size estimation was based on the draft 40-item GCA-PRO version, each with five-point response categories. Assuming retention of all 40 questionnaire items, meaningful results would require 200 completed questionnaires for the exploratory factor analysis (EFA). For a scale with polytomous items (where all items share equivalent rating scale), analysis with Rasch models would require 243 responses to produce statistically stable measures (with the precision of ± half a logit) [28, 29].

Statistical analysis

Data were first analysed descriptively before validation with Rasch models and EFA. A Rasch model provides a formal representation of fundamental measurement, and therefore fit to the model implies construct validity, reliability (internal consistency) and statistical sufficiency of the total score from the scale [30–33]. EFA was used iteratively with Rasch analysis to determine the underlying latent structure in the set of items, thus determining structural validity and internal consistency.

Each item was first tested for fit with the Rasch model, by comparing the difference between observed responses and expected values (null hypothesis: no significant difference between observed and values expected by the model). Fit to the model was supported by non-significant χ^2 probability. Each item was then assessed for 'threshold' ordering—'threshold' being the point between two adjacent categories where either response is equally probable [33]. GCA-PRO items had five response categories, reflecting an ordered continuum from low to high (items 1–13: none = 0, very mild = 1, mild = 2, moderate = 3, severe = 4; items 14–40: never = 0, rarely = 1, sometimes = 2, often = 3, always = 4), higher magnitude corresponding to higher impact. To fit the Rasch model, respondents with high levels of disease impact (low HRQL) would consistently endorse high scores in the continuum. Where

thresholds were disordered (determined graphically), suggesting participants had difficulty in consistently discriminating between response categories) [33], two adjacent categories were collapsed to ensure correct ordering and fit to the Rasch model. Local dependency was assessed in the correlation matrix of the residuals, and locally dependent items (a correlation of ±0.3) [34] were highlighted and discussed for clinical importance and possible discarding (due to redundancy) or combining into a testlet (subscale) [35].

Item reduction decisions were based on clinical importance, lack of fit to the Rasch model and redundancy. Retained items were subjected to EFA with orthogonal (varimax) rotation (null hypothesis: the observed items are not correlated). Factors were extracted if their eigenvalue was >1. The extracted factors (testlets) were then tested for fit with the Rasch model, reliability (internal consistency) and invariance to personal characteristics. Finally, the unidimensionality of the overall scale was tested using the confirmatory principal component analysis and *t*-test procedure proposed by Smith [36], where two sets of items hypothesized to represent low levels and high levels of disease impact are identified (based on correlation between items and the first residual factor), then an independent *t*-test is used to compare the difference in these estimates for each person. Unidimensionality is confirmed if ≤5% of the *t*-tests are significant or if the lower bound of a binomial 95% CI of the observed proportion overlaps 5% [33, 36].

Further evidence of validity using hypothesis testing was determined by (i) comparing the GCA-PRO scores with EQ-5D-5L, ICECAP-A and Cat-PROM5 using univariable Spearman's correlation (R_S)–convergence validity; as these PROMs capture relevant (but not full [ICECAP-A and Cat-PROM5] or specific [EQ-5D-5L]) impact of GCA and its treatment on HRQoL, we hypothesized that they should have moderate correlations with the GCA-PRO; and (ii) comparing the GCA-PRO scores of participants reporting 'active disease' and those 'in remission' using a *t*-test—discriminative (known groups) validity.

Reliability was established by assessing (i) the Person Separation Index (PSI), which estimates the scale's internal consistency, equivalent to Cronbach's α, only using the logit value as opposed to the raw score in the same formulae—a minimum value of 0.7 is acceptable for group use (with scores aggregated) of the questionnaire and 0.85 for individual use [33]; (ii) invariance (differential item functioning—DIF) of the scale, occurring when items are biased against a subgroup of patients based on gender, age, disease subgroups—observed scores should depend only on latent construct being measured and not on group membership [37, 38]; (iii) test–retest reliability between time 1 (questionnaire pack A) and time 2 (questionnaire pack B) completed 3 days later, for patients who reported 'no change' compared with 3 days ago—using intraclass correlation coefficient (ICC) estimates with 95% CI, calculated using absolute-agreement, two-way mixed-effects model [39]; and (iv) calculating the minimum detectable change from the standard error of measurement (S.E.m), obtained from the pooled standard deviation (of the mean, time 1 and time 2) and ICC estimates (of average measures) [40].

A *P*-value of 0.05 was considered significant except where a Bonferroni adjustment was applied to account for multiple testing, i.e. 0.05/number of tests. Analyses were conducted using IBM SPSS Statistics Version 28.0.1.1 (IBM Corp., Armonk, NY, USA) and RUMM2030 software (RUMM Laboratory Pty Ltd, Perth, Australia).

Ethical approval

Ethical approval was given by the South Central—Oxford A Research Ethics Committee (REC reference: 19/SC/0439), Health Research Authority (HRA) and Health and Care Research Wales (HCRW) Approval.

Results

Study sample and characteristics

Postal questionnaires were returned from 428 participants: mean (S.D.) age of 74.2 (7.2), 285 (66.6%) female; type of GCA: 327 (76.4%) cranial GCA, 114 (26.6%) large vessel vasculitis and 142 (33.2%) GCA with visual involvement. Positive diagnostic tests included temporal artery biopsy (167, 39%), temporal artery ultrasound (177, 41.4%) positron emission tomography and computed tomography (PET-CT) (51, 11.9%); 86 (20.1%) had a clinical diagnosis alone. Active disease was reported in 197 (46%), and 108 participants (25%) received second-line immunosuppressants, and 34 (7.9%) anti-IL6 therapy. For full clinical and demographic features see Table 1.

Distribution of item responses

Response rates for all items were very high, ranging 403–422 (Supplementary Fig. S1, available at *Rheumatology* online). Responses were largely distributed across response categories, although eight items had >50% of participants endorsing the lowest (least problems/impact) category (Supplementary Fig. S1, available at *Rheumatology* online). Examination of the person–item threshold distribution showed that all items were well targeted for people with different levels of impact on HRQoL (Supplementary Fig. S2, available at *Rheumatology* online).

Internal validity with Rasch models and factor analysis

Initial analysis of individual items with Rasch, revealed lack of fit in 11/40 items, which affected the overall item–person interaction: χ^2 (degrees of freedom [DF]) = 969.47 (240), $P < 0.001$. For most items (31/40) the five-category structure (none, very mild, mild, moderate, severe) was not working as expected. Amalgamating the first two response categories ('none' and 'very mild') improved the threshold ordering. Supplementary Fig. S3 (available at *Rheumatology* online) shows examples of ordered and disordered thresholds respectively. Ten items were discarded due lack of fit to the model and redundancy (Supplementary Table S1, available at *Rheumatology* online). This improved the overall fit to the model, although significant local dependency suggested multidimensionality in the scale, which was explored in the iterative EFA and Rasch analyses.

Initial EFA had revealed five factors within the scale (acute symptoms, psychological, activities of daily living, sight/stability, and participation; Supplementary Table S2, available at *Rheumatology* online); however, four factors were better supported by Rasch analysis, with four items from sight/stability being redistributed to other domains, guided also, in part, by clinical considerations. It was considered important to have sight/stability-related items in both 'acute symptoms' and

Table 1. Demographic and clinical features of survey participants[a]

Feature	Value
Age, mean (S.D.), years	74.21 (7.2)
≤70	126 (29.4)
>70	302 (70.6)
Sex	
Female	285 (66.5)
Male	135 (31.5)
Type of GCA	
Cranial	327 (76.4)
Ocular	142 (33.2)
Large-vessel vasculitis	114 (26.6)
Flare-ups in the last year ($n = 428$), n (%)	201 (47)
Positive diagnostic test	
Temporal artery biopsy	167 (39)
Temporal artery ultrasound	177 (41.4)
PET-CT	51 (11.9)
MRA	4 (0.9)
CTA	12 (2.8)
Clinical without confirmatory test	86 (20.1)
Duration of disease, median (IQR), years	2 (1–3)
Current glucocorticoid dose, median (IQR), mg	5 (2–10)
Patient assessment of disease activity	
Active disease	197 (51.6)
In remission	185 (48.4)
Steroid sparing treatment	
Currently	108 (25.2)
Previously	124 (29)
Tocilizumab (any other biologics)	
Currently	24 (7.9)
Previously	51 (11.9)
Clinical features of survey participants	
ESR ≥50 mm/h (prior to treatment)	215 (50.2)
CRP ≥10 mg/dl (prior to treatment)	362 (84.6)
New onset localized headache	363 (84.8)
Scalp or temporal artery tenderness	298 (69.6)
Transient visual loss	167 (39)
Optic neuropathy or retinal artery occlusion in one eye	41 (9.6)
Otherwise unexplained mouth or jaw pain upon mastication	237 (55.4)
Polymyalgia rheumatic	152 (35.5)
Certificate of sight impairment?	
Yes registered as severely sight impaired (blind)	9 (2.2)
Yes registered as sight impaired (partially sighted)	7 (1.7)
Educational level	
No formal qualifications	135 (34.6)
One to four GCSEs (or equivalent)	55 (14.1)
Five GCSEs (or equivalent)	47 (12.1)
Apprenticeships	20 (5.1)
Two or more A-levels or equivalent qualifications	34 (8.7)
Bachelors degree or equivalent, higher qualifications	72 (18.5)
Other qualifications including foreign qualifications	27 (6.9)
Employment status	
Employed	27 (6.6)
Self-employed	16 (3.9)
Unemployed	3 (0.7)
Disabled	4 (1)
Retired	359 (83.9)
Carer	3 (0.7)
Ethnicity	
White English/Welsh/Scottish/Northern Irish/British	404 (94.4)
Irish	4 (0.9)
Any other White background	4 (0.9)
Indian	1 (0.2)
Mixed White and Asian	1 (0.2)
Any other Mixed/Multiple ethnic background	1 (0.2)
Arab	1 (0.2)
Any other ethnic group	1 (0.2)
Missing	11 (2.6)

Values are n (%) except where otherwise stated.
[a] The inflammatory markers, clinical features and diagnostic tests were all from time of diagnosis before start of glucocorticoid treatment to describe the clinical presentation of participants, rather than reflecting current disease activity. IQR: interquartile range.

Table 2. Fit statistics of the individual domains

Item	Location	S.E.	Fit residuals	DF	χ^2	*P*-value
Acute	0.270	0.017	2.807	294.96	3.346	0.764
Activities of daily living	−0.047	0.010	−1.969	290.59	9.336	0.156
Psychological	−0.208	0.013	−0.421	297.87	4.271	0.640
Participation	−0.016	0.011	−1.828	282.58	8.267	0.219
Expected values			−2.5 to 2.5			>0.0125[a]

[a] Bonferroni adjusted *P*-value, i.e. 0.05/4 = 0.0125. DF: degrees of freedom.

'impact on ADL' domains. Each factor (or 'domain') resulted in satisfactory fit to the Rasch model (Table 2). The four-domain structure comprised: acute symptoms (eight items), activities of daily living (seven items), psychological (seven items), and participation (eight items). This four-domain structure addressed the local dependency, resulting in overall scale fit to the model: χ^2 (DF) = 37.563 (30), $P = 0.161$. Smith's unidimensionality test revealed the proportion of significant *t*-tests to be 2.9% (95% CI: 0.8%, 5%), supporting the unidimensionality of the overall scale.

Internal consistency

Internal consistency reliability measured by person separation index (PSI) was high from the initial analysis (PSI = 0.949) (Table 3). However, this reliability was superficially inflated due to local dependency of items. Grouping items into respective domains after EFA, addressed the local dependency (and lowered the artificially inflated reliability, from 0.938–0.867). The reliability of the overall scale remained excellent (PSI = 0.867).

The internal consistency values for each domain measured by Cronbach's α (also Cronbach's α-value for each domain if an item is deleted) are presented in Supplementary Table S3, available at *Rheumatology* online. They ranged from 0.802 to 0.927 supporting the internal consistency of each domain.

Further evidence of validity with hypothesis testing

Each domain correlated at least moderately with EQ5D-5L ($R_S = 0.638$–0.786), CAT-PROM5 ($R_S = 0.433$–0.550), and ICACAP-A ($R_S = 0.493$–0.740) scores, supporting evidence of convergent validity of the GCA-PRO with the three measures of HRQoL (Table 4).

All GCA-PRO domain scores differed significantly between patients who self-identified as having 'active disease' *vs* 'in

Table 3. Summary fit statistics for the overall scale

Analysis name	Item mean	S.D.	Person mean	S.D.	χ^2 (DF)	*P*-value[a]	PSI reliability
1. Initial analysis ($n=423$)	−0.019	3.294	−0.081	1.669	969.467 (240)	<0.001	0.949
2. Rescoring items into four categories ($n=428$)	0.129	2.780	−0.107	1.511	623.069 (198)	<0.001	0.938
3. The four-domains (subscales) scale ($n=426$)	0.139	0.891	−0.308	0.934	37.563 (30)	**0.161**	0.867
Expected values for fit to the Rasch model	0	1	0	1		>0.05	>0.7

[a] Non-significant *P*-value suggests adequate fit to (data do not deviate from) the Rasch model. PSI: Person Separation Index.

Table 4. Correlations between GCA-PRO scores with EQ5D-5L and CAT-PROM5

GCA-PRO domain (range of domain scale)	R_S	95% CI	*P*-value
Correlation with EQ5D-5L			
Acute symptoms (0–24)	−0.638	−0.695, −0.574	<0.001
Activities of daily living (0–21)	−0.736	−0.779, −0.686	<0.001
Psychological (0–21)	−0.658	−0.711, −0.597	<0.001
Participation (0–24)	−0.752	−0.793, −0.704	<0.001
Total score (0–90)	−0.786	−0.823, −0.741	<0.001
Correlation with CAT-PROM5			
Acute symptoms (0–24)	0.542	0.464, 0.611	<0.001
Activities of daily living (0–21)	0.541	0.464, 0.610	<0.001
Psychological (0–21)	0.433	0.346, 0.512	<0.001
Participation (0–24)	0.502	0.419, 0.577	<0.001
Total score (0–90)	0.550	0.469, 0.621	<0.001
Correlation with ICECAP-A			
Acute symptoms (0–24)	0.493	0.412, 0.566	<0.001
Activities of daily living (0–21)	0.603	0.535, 0.664	<0.001
Psychological (0–21)	0.604	0.537, 0.664	<0.001
Participation (0–24)	0.740	0.690, 0.784	<0.001
Total score (0–90)	0.713	0.656, 0.762	<0.001

GCA-PRO: GCA patient reported outcome; R_S: Spearman's correlation coefficient.

remission', supporting discriminative (known groups) validity of the GCA-PRO (Table 5).

Test–retest reliability and minimum detectable changes

A total of 413 patients returned the time 2 (retest) GCA-PRO questionnaire. Compared with 3 days ago, 288 (69.7%) reported 'no change' in their condition; 33 (8%) 'much better'; 57 (13.8%) 'slightly better'; 31 (7.5%) 'slightly worse'; and 4 (1%) 'much worse'. All the 95% CI of the ICC estimates of the domain scores at time 1 and time 2 (3 days later), in those whose conditions had not changed, were between 0.932 and 0.967 indicating 'excellent' reliability (Table 6).

The S.E.m for the GCA-PRO domain scores ranged from 0.473 to 0.696, and for the total score was 1.392. The minimum detectable changes (MDC_{90}) for the GCA-PRO domains ranged from 1.601 to 1.940, and for the total score was 3.271.

Calibration of an interval scale

Following fit to the model, the raw scores were mapped against the corresponding logit-based (Rasch-transformed) scores, and were linearly transformed to calibrate an interval scale of the same range to allow transformation of GCA-PRO raw scores to interval scaling when desired [41]. Supplementary Table S4 (available at *Rheumatology* online) presents score transformation tables.

Descriptive statistics of the final scale

The final GCA-PRO score ranges between 0 and 90, with zero representing no impact (good HRQoL) and 90 representing high disease impact (poor HRQoL). The GCA scores suggest that the majority of the participants recorded low disease impact, median (interquartile range) score for the overall scale was 23 (12–38) (Supplementary Table S5, available at *Rheumatology* online). This was consistent with the measures of impact: ICECAP-A Summary 0.321 (0.141–0.461); Cat-PROM5 Total Raw score 4 (2–8); and EQ-5D-5L, where the majority scored level 1 (no problem) across all the five dimensions (Supplementary Table S6, available at *Rheumatology* online).

Discussion

Underpinned by qualitative in-depth interviews and cognitive interviews to develop the 40 GCA-PRO candidate items [23], this study utilized both item response and classical testing theories to reduce items and determine the final scale structure. While the qualitative study ensured that the items were comprehensive and comprehensible (content validity), this validation study has produced the final, 30-item GCA-PRO supported by robust evidence of construct validity (structural validity and validity using hypothesis testing) and reliability (internal consistency, test–retest and measurement error) [20].

This study included patients from 38 rheumatology and ophthalmology centres in England and Wales, with different types of GCA (cranial, ocular and large vessel vasculitis) and different disease activity levels. Analysis with Rasch models showed that the GCA-PRO items were well-targeted for patients with different levels of HRQoL, thus accurately capturing the impact of GCA on patients across different severity levels. Hypothesis testing showed that the tool worked in the same way, for patients with active disease and in remission, and could discriminate between these two groups. These properties suggest that the GCA-PRO has the ability to detect effects of novel treatments on HRQoL in people with GCA.

While generic measures of quality of life are unable to accurately capture specific aspects of GCA impact, convergent validity was observed in GCA-PRO score comparisons with those of general health status (EQ-5D-5L) [13], capability (ICECAP-A) [26] and quality of eyesight (Cat-PROM5) [27]. All were moderately correlated with the GCA-PRO, as expected, as testing true criterion-related validity is not possible in PROMs due to lack of a 'gold standard' (except when a shortened tool is compared with its original long version) [20].

A good response rate across all items suggests that the GCA-PRO is feasible for patients. Validation with Rasch models justified rescoring all items from five-point to four-point response category structure, and reduction of items from 40 to 30. This improved the measurement properties of the GCA-PRO and likely eases completion of the tool.

Table 5. Discriminative (known groups) validity for the four domains of GCA-PRO

Domain (range)	Active disease, mean (S.D.)	Remission, mean (S.D.)	Mean difference	95% CI	*t*-statistic	*P*-value
Acute symptoms (0–24) ($n = 364$)	7.78 (4.441)	4.01 (3.298)	3.765	2.955, 4.576	9.221	<0.001
Activities of daily living (0–21) ($n = 368$)	7.42 (5.204)	3.88 (4.065)	3.545	2.590, 4.499	7.306	<0.001
Psychological (0–21) ($n = 370$)	9.18 (4.373)	6.12 (4.221)	3.509	0.447, 2.179	6.836	<0.001
Participation (0–24) ($n = 353$)	7.97 (6.249)	3.85 (4.809)	4.119	2.956, 5.282	6.964	<0.001
Total score (0–90) ($n = 330$)	31.98 (16.437)	17.68 (14.164)	14.299	10.969, 17.629	8.448	<0.001

GCA-PRO: GCA patient reported outcome.

Table 6. Test–retest reliability and minimum detectable changes

Item	ICC[a]	95%CI	*P*-value	S.E.m	MDC_{68}	MDC_{90}	MDC_{95}
Acute symptoms ($n = 265$)	0.932	0.913, 0.947	<0.001	0.560	1.058	1.741	2.074
Activities of daily living ($n = 264$)	0.967	0.958, 0.974	<0.001	0.473	0.973	1.601	1.907
Psychological ($n = 266$)	0.941	0.923, 0.954	<0.001	0.557	1.055	1.736	2.068
Participation ($n = 254$)	0.949	0.936, 0.960	<0.001	0.696	1.179	1.940	2.312
Total score ($n = 210$)	0.974	0.964, 0.981	<0.001	1.392	1.669	2.745	3.271

[a] ICC estimates based on single-measurement, absolute-agreement, two-way mixed-effects model. MDC: minimum detectable change, calculated as $MDC = \sqrt{}(2 \times S.E.m)$ presented at 68%, 90% and 95% CI levels [40]; ICC: intraclass correlation coefficient; S.E.m: standard error of measurement calculated as S.E.m = Pooled S.D. $\times \sqrt{}(1 - ICC$ of average measures).

This is the first disease-specific PRO for people with GCA that measures the impact of the disease and its treatment, developed using a robust methodology [19, 20, 42], and including patient perspectives at each step. Patients in this validation survey were only included if they had a clinician confirmed diagnosis of GCA in rheumatology and ophthalmology departments; a high percentage had confirmatory tests, with a clinical diagnosis alone in only 20% of patients, reflecting current clinical practice and case mix [43]. Care was taken to include centres across England and Wales with sites in rheumatology and ophthalmology services to capture a range of presentations and patient characteristics (and different subtypes of disease presentation), thus providing a high level of external validity.

Key limitations of this study include, first, that it was not possible to assess responsiveness of the GCA-PRO, which would require a longitudinal study [44]. However, this study established the standard error of measurement and the minimum detectable difference, which are useful to understand the change in scores that represent a real change, useful in estimating study sample sizes. Future longitudinal studies should evaluate the responsiveness of the GCA-PRO [20]. Second, this validation study was based on UK patients, and therefore a cross-cultural validation will be required before the tool can be used for multinational comparisons [20].

The potential uses of the GCA-PRO are twofold. First, it can be used as a communication tool between patients and clinicians to aid remote and in-person consultations and support shared decision making [45, 46]. For this purpose, clinicians can use the GCA-PRO domains or total scale by adding together domain scores to obtain an overall composite score. Second, the GCA-PRO can be used as a validated outcome measure for disease specific HRQoL in research alongside other PROMs. The tool can work at individual and group levels to discriminate between different HRQoL levels. Where a high level of precision is required, such as in clinical trials, the conversion table can be used to produce interval-level measures, allowing parametric analyses, provided sample size and other conditions are sufficient.

In conclusion, this study has validated the new 30-item GCA-PROM as a robust disease-specific measure of HRQoL in patients with GCA. It has excellent construct validity and reliability and can be used with confidence in clinical and research contexts alongside other PROMS for people with different types of GCA.

Supplementary material

Supplementary material is available at *Rheumatology* online.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Contribution statement

J.R. is the Principal Investigator, designed the study, led the grant application, oversaw the project and interpretation of the results, co-drafted and revised the manuscript for intellectual content. MN was the methodologist, co-designed the study, contributed to the grant application, analysed the data and interpretation of the results, co-drafted the manuscript and revised it for intellectual content. J.D. (methodologist) contributed to the study design, interpretation of the results, drafting and revision of the manuscript for intellectual content. C.A. coordinated the study, collected the data, revised the manuscript for intellectual content. R.G. is the senior statistician, contributed to the grant application, interpretation of the results and revised it for intellectual content. E.D. contributed to the grant application and revision of the manuscript for intellectual content. A.B. and S.S. were the Patient Research Partners, contributing to patient perspective in the study design, writing of the patient-facing materials and revision of the manuscript for intellectual content. C.G. contributed to the grant application and revision of the manuscript for intellectual content. C.H. and S.M. contributed to the grant application and revision of the manuscript for

intellectual content. M.N. had access to the data. J.R. and M.N. are responsible for the overall content as guarantors, controlled the decision to publish and accept full responsibility for the finished work and the conduct of the study. All authors gave final approval of the version to be published.

Funding

Funding for this project was from the National Institute for Health and Care Research, Research for Patient Benefit, PB-PG-1217-20017. The views expressed in this manuscript by the co-authors are not necessarily the views of the NIHR.

Disclosure statement: All authors have completed the ICMJE form for Competing Interests Disclosure and report a research grant from National Institute for Health Research for this study. M.N. received research grants from BMS, Vifor Pharma and Sanofi paid to his institution, speaking fees from CCIS—The Conference Company and Eli Lilly, all outside the submitted work. J.R. received research grants from University Hospitals Bristol and Weston NHS Foundation Trust, Vifor Pharma and Sanofi paid to her institution; and speaking fees from Vifor Pharma outside the submitted work. J.R. is on Advisory Body for COMBIVAS, University of Cambridge. C.G. received consulting fees from Eli Lilly outside the submitted work. S.M. carried out consultancy on behalf of her institution for Roche/Chugai, Sanofi, AbbVie, AstraZeneca; was an Investigator on clinical trials for Sanofi, GSK, Sparrow; carried out speaking/lecturing on behalf of her institution for Roche/Chugai, Vifor, Pfizer and Novartis; was chief investigator on STERLING-PMR trial, funded by NIHR; was patron of the charity PMRGCAuk. No personal remuneration was received for any of the above activities. Support from Roche/Chugai to attend EULAR2019 and from Pfizer to attend ACR Convergence 2021. C.A., J.D., E.D., R.G., A.B., C.H. and S.S. report no competing interests.

Acknowledgements

This was a collaborative work and authors wish to thank the following people who recruited patients and completed clinician case report forms: Shabina Sultan, Emma Dooks and Chantel McParland, Airedale NHS Foundation Trust; Professor Rod Hughes and Maggie Walsh, Ashford and St Peter's Hospital NHS Foundation Trust; Dr Victoria Bejarano and Susan Hope, Barnsley Hospital NHS Foundation Trust; Anupama Nandagudi, Angelo Ramos and Moroti Abioye, Basildon and Thurrock University Hospitals NHS Foundation Trust; Dr John Pauling, Royal United Hospitals Bath NHS Foundation Trust; Jonathan Marks, Emma Gunter, Rochelle Hernandez and Alison Pitcher, Royal Bournemouth and Christchurch Hospitals NHS Trust; Ines Marcal and Edel Robbins, University Hospitals Bristol NHS Foundation Trust; Professor David R. W. Jayne and Dr Maria King, Cambridge University Hospitals NHS Foundation Trust; Dr Tim Blake, University Hospitals Coventry and Warwickshire; Dr Mahdi Abusalameh and Jessica Record, Royal Devon and Exeter NHS Foundation Trust; Dr Charles Kwok-Chong Li, Royal Surrey NHS Foundation Trust; Dr Richard Watts and Rebecca Francis, East Suffolk & North Essex NHS Foundation Trust—Ipswich Hospital Site; Dr Damodar Makkuni and Katherine Mackintosh, James Paget University Hospital NHS Foundation Trust; Dr Eoin P. O'Sullivan and Georgina Bird, King's College Hospital NHS Foundation Trust; Prof. Hedley C. A. Emsley, Royal Preston Hospital; Dr Vanessa Quick and Ruth Lovelock, Bedfordshire Hospitals NHS Foundation Trust; Dr Arabella Waller and Ruby Einosas, Maidstone and Tunbridge Wells NHS Trust; Dr Ajit Menon, Sebastian Meighan-Davies and Kathlyn Prado, Midlands Partnership NHS Foundation Trust; Dr Syed M. H. Bilgrami, Dr Marwan A. S. Bukhari, Lynda Fothergill and Jackie Toomey, University Hospitals of Morecambe Bay NHS Foundation Trust; Dr Sarah Emerson and Wendy Wilmot, North Bristol NHS Trust (Southmead Hospital); Dr Alaa Hassan and Katherine Davidson, North Cumbria Integrated Care NHS Foundation Trust (Cumberland Infirmary); Dr Chetan Mukhtyar, Georgina Ducker and James Kennedy, Norfolk and Norwich University Hospital; Dr Frances Rees, Paige Draper, Alan Thomas and Marie-Josèphe Pradère, Nottingham University Hospitals NHS Trust; Dr Sharma Poonam and Stephanie Diaz, North West Anglia NHS Foundation Trust; Dr Sanjeet Kamath and Melissa Shaw, Salford Royal NHS Foundation Trust; Prof. Bhaskar Dasgupta, Prisca Gondo and Bridgett Masunda, Southend University Hospital; Dr Julie Dawson, Denise Graham and Nicola Hornby, St Helens & Knowsley NHS Teaching Hospital NHS Trust; Dr Kanchan Manchegowda and Louise Fairlie, Sunderland Royal Hospital; Dr Elizabeth Price and Suzannah Pegler, The Great Western Hospitals NHS Trust; Dr Kirsten Mackay, Tracey Camden-Woodley, Joan Redome and Lorraine Thornton, Torbay Hospital; David Hutchinson, Daniel Mynors-Wallis, Zoe Berry and Fiona Hammonds, Royal Cornwall Hospitals NHS Trust; Dr Anna Ciechomska, Karen Black and Suzanne Clements, University Hospital Wishaw (Lanarkshire); Dr James Bateman, Lawrence Phiri, Georgina Falagan-Garman and Leon Mcdonald, Institute of Clinical Sciences, University of Birmingham Royal Wolverhampton NHS Trust; Dr Shweta Bhagat, Lily John and Helen Cockerill, West Suffolk NHS Foundation Trust; and Dr Sally Knights, Ben Mulhearn and Alison Lewis, Yeovil District Hospital NHS Foundation Trust.

The authors also wish to thank the Treatment According to Response in Giant cEll arTeritis (TARGET) UK research network for supporting this research and the patient charity PMRGCAuk for opportunity to present and receive feedback on the underpinning qualitative stages of the GCA PRO.

Patient and public involvement: Patients were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

References

1. Petri H, Nevitt A, Sarsour K *et al.* Incidence of giant cell arteritis and characteristics of patients: data-driven analysis of comorbidities. Arthritis Care Res (Hoboken) 2015;67:390–5.
2. Maleszewski JJ, Younge BR, Fritzlen JT *et al.* Clinical and pathological evolution of giant cell arteritis: a prospective study of follow-up temporal artery biopsies in 40 treated patients. Mod Pathol 2017;30:788–96.
3. Jennette JC, Falk RJ, Bacon PA *et al.* 2012 revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides. Arthritis Rheum 2013;65:1–11.
4. Vodopivec I, Rizzo JF 3rd. Ophthalmic manifestations of giant cell arteritis. Rheumatology (Oxford) 2018;57:ii63–72.

5. Stone JH, Klearman M, Collinson N. Trial of tocilizumab in giant-cell arteritis. N Engl J Med 2017;377:1494–5.
6. Ponte C, Grayson PC, Robson JC *et al.*; DCVAS Study Group. 2022 American College of Rheumatology/EULAR classification criteria for giant cell arteritis. Ann Rheum Dis 2022;81:1647–53.
7. Cheah JTL, Robson JC, Black RJ *et al.* The patient's perspective of the adverse effects of glucocorticoid use: a systematic review of quantitative and qualitative studies. From an OMERACT working group. Semin Arthritis Rheum 2020;50:996–1005.
8. Koster MJ, Yeruva K, Crowson CS *et al.* Efficacy of methotrexate in real-world management of giant cell arteritis: a case-control study. J Rheumatol 2019;46:501–8.
9. Maz M, Chung SA, Abril A *et al.* 2021 American college of rheumatology/vasculitis foundation guideline for the management of giant cell arteritis and takayasu arteritis. Arthritis Rheumatol 2021; 73:1349–65.
10. Dawson J, Doll H, Fitzpatrick R *et al.* The routine use of patient reported outcome measures in healthcare settings. BMJ 2010;340: c186.
11. Fitzpatrick R, Davey C, Buxton MJ *et al.* Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2:1–74.
12. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.
13. Feng YS, Kohlmann T, Janssen MF *et al.* Psychometric properties of the EQ-5D-5L: a systematic review of the literature. Qual Life Res 2021;30:647–73.
14. Juniper EF, Guyatt GH, Epstein RS *et al.* Evaluation of impairment of health related quality of life in asthma: development of a questionnaire for use in clinical trials. Thorax 1992;47:76–83.
15. Yi H, Shin K, Shin C. Development of the sleep quality scale. J Sleep Res 2006;15:309–16.
16. Kirwan JR, Hewlett SE, Heiberg T *et al.* Incorporating the patient perspective into outcome assessment in rheumatoid arthritis–progress at OMERACT 7. J Rheumatol 2005;32:2250–6.
17. Aydin SZ, Direskeneli H, Sreih A *et al.* Update on outcome measure development for large vessel vasculitis: report from OMERACT 12. J Rheumatol 2015;42:2465–9.
18. Kupersmith MJ, Speira R, Langer R *et al.* Visual function and quality of life among patients with giant cell (temporal) arteritis. J Neuroophthalmol 2001;21:266–73.
19. 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.
20. Mokkink LB, Terwee CB, Knol DL *et al.* The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010;10:22.
21. Sacristán JA, Aguarón A, Avendaño-Solá C *et al.* Patient involvement in clinical research: why, when, and how. Patient Prefer Adherence 2016;10:631–40.
22. Herlyn K, Hellmich B, Seo P *et al.* Patient-reported outcome assessment in vasculitis may provide important data and a unique perspective. Arthritis Care Res (Hoboken) 2010;62:1639–45.
23. Robson JC, Almeida C, Dawson J *et al.* Patient perceptions of health-related quality of life in giant cell arteritis: international development of a disease-specific patient-reported outcome measure. Rheumatology (Oxford) 2021;60:4671–80.
24. UKIVAS. UKIVAS Registry. Oxford: University of Oxford, 2023. https://ukivas.ndorms.ox.ac.uk/ (8 March 2023, date last accessed).
25. Leeds Institute for Data Analytics. UK GCA Consortium. Leeds: University of Leeds; 2023. https://lida.leeds.ac.uk/target/research-projects/gcatregistry/uk-gca-consortium/ (8 March 2023, date last accessed).
26. Al-Janabi H, Flynn TN, Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. Qual Life Res 2012;21:167–76.
27. Sparrow JM, Grzeda MT, Frost NA *et al.* Cat-PROM5: a brief psychometrically robust self-report questionnaire instrument for cataract surgery. Eye (Lond) 2018;32:796–805.
28. Azizan NH, Mahmud Z, Rambli A. Rasch rating scale item estimates using maximum likelihood approach: effects of sample size on the accuracy and bias of the estimates. Int J Adv Sci Technol 2020;29:2526–31.
29. Linacre JM. Sample size and item calibration stability. Rasch Measur Trans 1994;7:328–31.
30. Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. London: Lawrence Arlbaum Associates, 2001.
31. Rosenbaum PR. Criterion-related construct-validity. Psychometrika 1989;54:625–33.
32. Anderen E. Sufficient statistics and latent trait models. Psychometrika 1977;42:69–81.
33. Tennant A, Conaghan P. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper. Arthritis Rheum 2007;57:1358–62.
34. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 2007;46:1–18.
35. Guemin L, Robert LB, David AF. Incorporating the testlet concept in test score analyses. Educ Measur Issues Pract 2000;19:9–15.
36. Smith E. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Measur 2002;3:205–31.
37. Smith RM, Suh KK. Rasch fit statistics as a test of the invariance of item parameter estimates. J Appl Meas 2003;4:153–63.
38. Brodersen J, Meads D, Kreiner S *et al.* Methodological aspects of differential item functioning in the Rasch model. J Med Econ 2007; 10:309–24.
39. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–63.
40. Polit DF. Getting serious about test-retest reliability: a critique of retest research and some recommendations. Qual Life Res 2014; 23:1713–20.
41. Wright BD, Linacre JM. Observations are always ordinal – measurements, however, must be interval. Arch Phys Med Rehabil 1989;70:857–60.
42. Mokkink LB, Terwee CB, Patrick DL *et al.* The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19:539–49.
43. Mahr A, Belhassen M, Paccalin M *et al.* Characteristics and management of giant cell arteritis in France: a study based on national health insurance claims data. Rheumatology (Oxford) 2020;59: 120–8.
44. Copay AG, Subach BR, Glassman SD *et al.* Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 2007;7:541–6.
45. Greenhalgh J, Gooding K, Gibbons E *et al.* How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. J Patient Rep Outcomes 2018;2:42.
46. Field J, Holmes MM, Newell D. PROMs data: can it be used to make decisions for individual patients? A narrative review. Patient Relat Outcome Meas 2019;10:233–41.