VALIDATION OF THE ANCA-ASSOCIATED VASCULITIS PATIENT-REPORTED OUTCOMES (AAV-PRO) QUESTIONNAIRE

[1]Joanna Robson*. PhD MRCP. Consultant Senior Lecturer in Rheumatology, Faculty of Health and Applied Sciences, University of the West of England, Bristol, & Hon Senior Lecturer, School of Clinical Sciences, University of Bristol, & Hon Consultant in Rheumatology, University Hospitals Bristol NHS Trust. Bristol Royal Infirmary, Upper Maudlin Street, Bristol, BS2 8HW. Jo.Robson@uwe.ac.uk. Corresponding author.

[2]Jill Dawson DPhil*, Associate Professor, Nuffield Department of Population Health (HSRU), University of Oxford, Old Road Campus, Oxford, UK. jill.dawson@dph.ox.ac.uk

Helen Doll DPhil, Senior Principal, ICON Clinical Outcomes Assessments, Abingdon, UK. Helen.Doll@iconplc.com

Peter F. Cronholm, MD MSCE, Department of Family Medicine and Community Health, University of Pennsylvania, Philadelphia, PA, USA. peter.cronholm@uphs.upenn.edu

Nataliya Milman, MD, Department of Rheumatology, University of Ottawa, Ottawa, Canada. nmilman@toh.on.ca

Katherine Kellom, PolicyLab, Children's Hospital of Philadelphia, Philadelphia, US. kellomk@email.chop.edu

Susan Ashdown, RN, Oxfordshire, UK. sueashdown63@hotmail.com

Ebony Easley, Department of Family Medicine and Community Health, Mixed Methods Research Laboratory, University of Pennsylvania, Philadelphia, PA, USA. easley.ebony@gmail.com

Don Gebhart, MD, Columbus, Ohio, USA. Dgeb18@gmail.com

Georgia Lanier, Boston, MA, USA. Georgia.lanier@gmail.com

John Mills, Derbyshire, Vasculitis UK, UK. jandsmills@btinternet.com

Jacqueline Peck RN, Oxfordshire, UK. jackiepeck@btinternet.com

Raashid A. Luqmani, DM FRCP, FRCP(E), Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK. raashid.luqmani@ndorms.ox.ac.uk

Judy A. Shea, PhD, School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. sheaja@mail.med.upenn.edu

Gunnar Tomasson, MD, Department of Public Health Sciences, University of Iceland, Reykjavik, Iceland. Gunnar.tomasson@gmail.com

Peter A. Merkel, MD MPH, Division of Rheumatology, Department of Medicine and Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA, USA. pmerkel@upenn.edu

[1 and 2] These two authors contributed equally.

ABSTRACT

Objectives: To finalize and validate a disease-specific patient-reported outcome (PRO) measure: the ANCA-associated vasculitis patient-reported outcome (AAV-PRO) questionnaire. Using a 35-item candidate questionnaire developed following 50 qualitative interviews in the UK, USA and Canada, a longitudinal survey was conducted to determine the final scale structure and validate the AAV-PRO.

Methods: Participants were recruited via Vasculitis UK and the Vasculitis Patient-Powered Research Network. The 35-item candidate questionnaire was completed at baseline and three-months; UK participants completed the EuroQol-5D-5L (EQ-5D-5L), whilst US participants completed a test-retest exercise, three-to-five days after baseline. Scale structure was defined using exploratory factor analysis (EFA) and Rasch analysis. Convergent and known groups validity, test-retest reliability, and longitudinal construct validity were assessed.

Results: There were 626 participants with AAV**;** >25% reporting “active disease”. EFA and Rasch analysis supported a 29-item profile measure comprising 6 domains: “Organ-Specific Symptoms”, “Systemic Symptoms”, “Treatment Side Effects”, “Social and Emotional Impact”, “Concerns about the Future”, and “Physical Function”. Mean domain scores were higher for participants with “active disease” versus “remission” ($p<0.001$). Construct validity was demonstrated by correlations between domain scores and the EQ-5D-5L (range $r=-0.55$ to 0.78), all $p<0.0001$. In participants reporting “no change” (n=97) during the test re-test, Intraclass Correlation Coefficient values were high (range 0.89-0.96) for each domain.

Conclusions: The AAV-PRO, a new disease-specific PRO measure for AAV, has good face and construct validity, is reliable, feasible, and discriminates among disease states.

Key words

Granulomatosis with polyangiitis

Systemic vasculitis

Patient perspective

Outcomes research

Corticosteroids

Glucocorticoids

Granulomatosis with polyangiitis (GPA, Wegener's), eosinophilic granulomatosis with polyangiitis (EGPA, Churg-Strauss), and microscopic polyangiitis (MPA) are life- and organ-threatening disorders affecting the lungs, kidneys, ear, nose, throat, nerves, skin, and quality of life of affected patients and are collectively known as the ANCA-associated vasculitides (AAV) [1 2]. Despite improvement in mortality and morbidity with newer treatment regimens, the risk of relapse in AAV is 35% over five years [3]. Many patients experience persistent disease activity, long-term exposure to toxic therapies [4], and the psychosocial impact of a serious illness [5].

Health related quality of life (HRQoL) is impaired in AAV [6-8]. A quarter of patients experience depression and more than 40% anxiety [6]. Work disability is high with a quarter unemployed due to AAV [9], and 50% reported their careers had been hindered [10]. Fatigue and pain are important symptoms [6 11]. The opinions of patients and clinicians on the relative importance of outcomes often differ [12 13].

The Outcome Measures in Rheumatology (OMERACT) core set of outcome measurements for use in clinical trials in AAV included the generic Short-Form 36 (SF-36) patient reported outcome (PRO) measuring HRQoL [14-16]. Generic PROs can lack specificity [17] and the OMERACT Vasculitis Working Group identified the need for an AAV- specific PRO to fully capture the patient's perspective [18]. An international steering committee comprising patient partners, methodologists, statisticians and clinicians from the UK, USA, and Canada have been developing a new disease-specific PRO, in line with guidance from the United States Food & Drug Administration (FDA) [19]. The project received critical scrutiny and feedback at three successive Vasculitis Workshop sessions at OMERACT conferences [20 21].

A three-stage approach has been followed. Stage 1: Qualitative analysis, item production and testing in the UK, USA and Canada resulting in a 35-iten candidate AAV-PRO questionnaire (completed) [22]. Stage 2: Large-scale parallel survey of people with AAV in the UK and USA, to investigate the underlying scale structure of the AAV-PRO. Stage 3: Assessment and validation of the AAV-PRO's measurement properties, including construct validity, reliability, discriminatory ability and ability to detect change. Stages 2 and 3 are reported here.

METHODS AND MATERIALS

An international steering committee, including four patient partners, had oversight of the patient survey materials, working with the patient groups Vasculitis UK and the Vasculitis Patient-Powered Research Network (the VPPRN).

Ethical approval was given by the Medical Sciences IDREC, University of Oxford, Oxford, Ref: MS-IDREC-C1-2015-087 for the UK and US survey. In the US, approval was given by the Institutional Review Boards at the University of Pennsylvania and the University of South Florida, Ref: Pro00018514. Patients were recruited between June and October 2015.

***Inclusion / exclusion criteria*:**

Participants were required to have AAV, English speaking, aged ≥18 years and to fulfill the following:

1) Affirm that they had ANCA-associated vasculitis (AAV); and

2) Received either a positive test result for ANCA OR diagnostic biopsy OR angiogram; and

3) Currently or previously taken glucocorticoids or another immunosuppressant/s.

Participants with AAV were sent a pack by post (Vasculitis UK) or email (VPPRN in the US) containing a covering letter, information sheet, and forms for demographic (date of birth, location, sex, race, highest educational level, employment status) and disease-related data (type of AAV, date of diagnosis, positive ANCA test, current disease state, immunosuppressant medications), and the 35-item candidate AAV-PRO questionnaire. The first 12 items addressed symptom severity; the remaining 23 items addressing the impact of AAV, or its treatment, on HRQoL. Each item has five

ordinal integer response options (three formats applying to different items: symptom severity, level of difficulty, frequency of experiencing a problem), scored 0 to 4; higher scores denoting greater severity or impact. UK participants were also sent an EQ-5D-5L questionnaire [23] at baseline. This five-item generic measure assesses mobility, self-care, usual activities, pain/discomfort and anxiety/depression on a five-point scale.EQ-5D-5L index values were calculated using the cross-walk method [24].

Three-to-five days after they provided baseline responses the US participants were sent a repeat 35-item AAV-PRO candidate questionnaire (test-retest), disease state questions and a transition question concerning change in disease state since baseline questionnaire completion: 'Overall, how are you NOW (in terms of your vasculitis and any treatment side effects), compared with 5 days ago (when you first answered the questionnaire).

After three months, all UK and US participants were sent the same 35-item candidate AAV-PRO questionnaire and the transition item used for test retest, but with comparison made with '3 months ago'.

Sample size

Sample size for health status questionnaire development requires at least three respondents per questionnaire item tested [25]. The aim was to recruit at least 500 patients (250 from each country).

Statistical methods

Data were analysed using SPSS release 20 (PASW Statistics 20© 2015 SPSS Inc. SPSS (Hong Kong) Ltd). To minimize type-I error, the significance level for all analyses was set at two-sided $p<0.01$.

Criteria for questionnaire item reduction

1) Missing responses >3% [26]; 2) Distribution of responses exhibiting ceiling or floor effects (≥50% responses to an item taking either of the 2 most extreme response categories); 3) High inter-item correlation (≥0.80) or Cronbach's alpha (≥ 0.93) suggesting redundancy; 4) Items poorly correlated with their overall domain/scale score (i.e. item-to-total correlations <0.3); 5) Cross-loading during factor analysis, 6) Particularly poor fit to the model (Item Trait Interaction $p<0.01$) on fitting to a Rasch unidimensional model to any identified domains.

Scale structure and dimensionality

Conceptual Framework

The AAV-PRO Conceptual Framework indicated that the PRO was likely to be multidimensional, i.e., containing items addressing symptom severity, and differing aspects of HRQoL (physical, psychological, social and global impact on health).

Factor structure

The formal process of item reduction and determination of scale structure was guided by Exploratory Factor Analysis [27] (EFA), Rasch analysis [28] (RUMM2010 software; RUMM Laboratory Pty Ltd; Western Australia 6023) and from insights from the Conceptual Framework [26]. EFA was conducted using FACTOR [27], based on a polychoric correlation matrix, using Principal Axis Factoring extraction, with oblique

rotation method. Items correlating with a factor of >0.4 were considered to significantly load and the item was assigned to that factor [29].

Individual item functioning

The polytomous Rasch model (for items with >2 responses) is equivalent to a test of the theoretical construct validity and adequacy of a scale [30 31], assessing the unidimensionality of items in a scale [28 32 33].

Scale/domain properties

Internal consistency

Cronbach's alpha coefficients were calculated to assess the internal consistency of questionnaire domains. An alpha ≥0.70 is recommended to claim internal consistency [34 35], alpha > 0.90 may suggest redundancies, requiring item reduction [31], with 0.80 to 0.90 considered optimal [36].

Convergent validity.

It was hypothesised that a large Pearson's correlation (r≥0.5) would be obtained between the AAV-PRO domains and generic EQ-5D-5L index scores (UK baseline sample only). It was anticipated that negative correlations would be seen, as the two measures are scored in opposite directions.

Test retest reliability

Intraclass Correlation Coefficients (ICC) were used to compare baseline AAV-PRO domain scores, with scores obtained three-to-five days later (US sample only) in those individuals whose condition had remained stable. ICC values >0.60 are recommended [37].

Meaningful change

The Standard Error of the Mean (SEM) was the error estimated for a single use of the questionnaire and is directly related to the reliability of the scale. The Minimal Detectable Change (MDC) was defined as the smallest amount of change between two time points that indicated a real change in the patient's health status [38]. The MDC_{90} was set to indicate that 90% of stable patients demonstrated random variation of less than this magnitude when assessed on multiple occasions [39-41].

Known groups validity [26].

It was hypothesised that AAV-PRO domain scores would differ significantly between patients self-identifying at baseline as having 'Active disease' versus patients 'In remission'.

Longitudinal construct validity: responsiveness

Responsiveness was assessed where respondents provided relevant outcomes data at baseline and 3-months. Change scores were calculated as the baseline score minus the 3-month follow-up score for each AAV-PRO domain. Effect sizes (ES) were calculated as the difference between the sample's mean baseline score and mean 3-month follow-up score, divided by the standard deviation (SD) of baseline score. ES calculates the magnitude of change measured by an instrument in a standardised way allowing comparison between instruments [42]. Change scores and ES were compared with responses on a 3-month transition item regarding change in patients' condition.

RESULTS

Study sample and characteristics

The baseline survey response rate was 74% (n=662/900). Of the 662 respondents, 626 were eligible for inclusion (95%). Demographic and clinical characteristics of participants shown in (Table 1 and online supplementary Table S1)). The mean age was 60.4 years (SD 13.2) and participants were predominantly female (397, 64%). The sample represented the UK (348/626) and the US (278/626) with 45% and 46% of the sample respectively; UK respondents were older (mean 63 vs 57 years, $p<0.001$), and more likely to be retired (59% versus 32%, $p<0.001$).

Item response distribution - candidate AAV-PRO items

Candidate questionnaire items and baseline distribution of their responses are shown in Figure 1. Item response rates were high overall (maximum 1.6% missing data), supporting the feasibility of the questionnaire. One exception concerned 'difficulties with sexual activity or desire' (6.2% missing; 11.8% missing in age-group >65). Responses were generally evenly spread across responses, although >50% of respondents endorsed an extreme ('No difficulty') response on two items ('Using hands for small careful movements' and 'Washing/drying/ dressing unaided').

Final dimensionality and scale structure of the AAV-PRO

The final AAV-PRO including 29 individual questionnaire items is shown in Figure 2 Details of the Rasch and EFA analyses are given in online supplementary Figures S1-3 and Table S2. The AAV-PRO is a profile measure containing 6 different domains: 'Organ Symptoms Severity (OSS)', 'Systemic Symptoms Severity (SSS)', 'Treatment Side-Effects (TSE)', 'Social and Emotional Impact (SEI)', 'Concerns About the Future (CAF)' and 'Physical Function (PF)'. The identified domains each fit the Rasch

unidimensional model (Item Trait Interaction p>0.01) and had good internal consistency (Cronbach's alphas 0.77- 0.96) (Figure 2). Patient partners on the steering committee reviewed the items within each domain and developed the domain titles used above.

Six questionnaire items were identified for rejection based on failure to fit within the Rasch model for a particular domain, plus insights from the Conceptual Framework, EFA and clinical input : "nerve pain or numbness" reflected damage and felt not suitable for a PRO as would not capture change; "sexual activity …" obtained poor response rate; "worried about income" was considered too contextual with responses influenced by differing healthcare; "using hands for small tasks" response distribution indicating a ceiling effect (skewed towards 'no difficulty'); "… social life is limited" had strong overlap in responses with other better fitting items indicating redundancy; and "… activities essential to your day", "walking around shops" and "walking up-stairs" were all highly correlated indicating redundancy (exact meaning of "essential" flagged as problematic by patient partners).

Scoring of the 29-item AAV-PRO profile measure

Scores for each domain are calculated as the sum of each individual item score, (online supplementary Figure S4). Examples of items with response categories are shown in supplementary Figure S5.

Measurement properties

Convergent validity

Correlations (Pearson) between baseline AAV-PRO domains and EQ-5D-5L index-scores (UK sample only) were all large (≥0.50): OSS r= -0.55, SSS r= -0.67, TSE r= -0.65, SEI r= -0.73, CAF r= -0.68, and PF r= -0.78 (all p<0.001).

Test-retest reliability

All Intraclass Correlation Coefficient (ICC) values between domain scores at baseline and three-to-five days later (US sample) were very good: OSS ICC = 0.89 (95%CI I 0.84 to 0.93); SSS ICC= 0.91 (95%CI 0.86 to 0.94); TSS= 0.95 (95%CI 0.93 to 0.97); SEI= 0.96 (95%CI 0.94 to 0.97); CAF= 0.95 (0.92 to 0.97); PF= 0.96 (0.94 to 0.97). (Table 2)

Meaningful change

The Standard Error of the Mean (SEM), Minimal Detectable Change (MDC_{90}) estimate were calculated based on the ICC and the Standard Deviation of the baseline score (Table 2).

Known groups validity

AAV-PRO domain scores all differed significantly (p<0.001) between patients self-identifying as having 'Active disease' versus 'In remission' (see Table 3) as was also the case for the EQ-5D-5L.

Longitudinal construct validity

Mean change scores and ES for the AAV-PRO domains were mapped to level of response to the 3-month transition item (Table 4). Results showed that respondents

reporting ‘no change’ in their condition exhibited appropriate ES, close to zero, while positive ES range 0.21 to 0.28 were associated with the response ‘Much better’ for all domains. The response ‘slightly better’ had ES lying between zero and the value associated with ‘Much better’ responses. In general, responses indicating a worsening health state were associated with negative ES of a magnitude that mirrored results associated with positive responses/improvement. The exception here was the *organ symptom severity* domain with scores lacking a significant linear trend across the transition item responses (all other domains’ linear trend $p \leq 0.003$).

Comparison between the AAV-PRO domain scores and demographic and clinical features

There were no differences in mean scores between each of the three AAV (GPA, MPA, and EGPA) ($p<0.01$) and no correlation between length of time from diagnosis and any of the AAV-PRO scales ($p<0.01$). There were differences between i) USA and UK respondents, with UK scores higher (i.e. worse) ($p \leq 0.001$) on all scales, ii) Male and female mean scores, with women scoring higher on all scales ($P<0.01$), and iii) Younger and older respondents with higher mean scores on the Social and Emotional Impact Subscale in those in the ≤ 65 age-group compared with older participants ($p<0.01$) (see online supplements S3-7).

DISCUSSION

The AAV-PRO is a new 29-item, disease-specific PRO measure for use in ANCA-associated vasculitis. It has good face, content, and construct validity, is reliable, feasible, and discriminates among disease states. Patients have played a key role within every stage of development [22]. This manuscript describes the underlying structure of the final AAV-PRO and its validation in terms of reliability, feasibility, discrimination and construct validity.

The final 29-item questionnaire comprises six subscales/domains: “Organ-Specific Symptoms”, “Systemic Symptoms”, “Treatment Side Effects”, “Social and Emotional Impact”, “Concerns about the Future”, and “Physical Function”. The identified domains offer a comprehensive profile of the impact of AAV on patients’ everyday life and were felt by the patient partners to represent “what AAV was to them”. Each domain is unidimensional and has good measurement properties including good internal consistency (Cronbach’s alphas range 0.77 to 0.92) and test-retest reliability (Intraclass Correlation Coefficients (ICCs) range 0.89 to 0.96); plus evidence supporting concurrent validity, with moderate to high correlations (range r-0.55 to -0.78, all $p<.0001$) with EQ-5D-5L index scores, as hypothesised. All AAV-PRO domain scores distinguished between patients who self-reported having active disease versus disease in remission ($p<0.0001$), providing support for known groups validity. Length of time from diagnosis alone was not correlated with worse scores, indicating that disease activity, rather than duration of disease, is a key correlate to AAV-PRO scores. There were also no differences in mean scores between the different subtypes of AAV.

Characteristics of the UK and USA survey populations differed slightly, participants in the USA were on average younger, with shorter duration of disease, and higher educational level. This may reflect the different methods of data collection and may account for the differences seen in subscale scores between countries. Age, educational level, and socioeconomic status are associated with computer usage [43 44].

Women scored higher (i.e. worse) on all six subscales of the AAV-PRO. Health-related quality of life is reduced in females in other conditions [47], and trends towards higher scores for women have been reported in AAV [15]. Younger people (<65) scored higher on the Social and Emotional Impact subscale of the AAV-PRO, and lower on mental health, a trend also seen in other chronic diseases in this age group [47].

The design of the survey was to identify the scale structure and measurement properties of the AAV-PRO. As predicted, participants were generally stable regarding self-reported disease activity, with around 70% describing themselves as “in remission”. Follow up was 3-months. This somewhat limited the assessment of responsiveness and minimally important change, which are usually assessed over a longer time-period in participants expected to change in clinical state, e.g., within the context of a clinical trial [19]. The study produced evidence of longitudinal construct validity. Among participants who reported “no change” effect sizes were appropriately close to zero, and the few participants who reported their condition as “much better” demonstrated a small amount of change in AAV-PRO scores (ES range 0.21 to 0.28). Distribution-based estimates of minimal change (SEM and MDC_{90}), which relate to the reliability (ICC) of each scale were all appropriate and will be useful for calculating sample sizes in future studies [41].Future studies will provide more robust estimates of

minimal important differences (MID), further longitudinal construct validity [48] and determine whether summary component scores can be derived.

Validated PROs are an important way of accurately measuring the impact and value of new drug treatments on HRQoL by measuring outcomes of importance to patients themselves [17]. PROs can be part of evidence submitted for new drug approvals and can also provide valuable information to clinicians and policymakers asked with making decisions about the use of new treatments[49]. The involvement of patients with AAV-PRO at every stage of development should ensure its face validity and relevance. In addition, it has also been shown that disease-specific instruments may be more responsive to change than generic instruments, which is a crucial characteristic for detecting treatment effect within randomized controlled trials[50]. The AAV-PRO is, therefore, presented as complementary to the SF-36 or EQ5D, which allow comparison with other conditions and population controls, but are not specific to AAV.

The AAV-PRO, a new disease-specific PRO measure for ANCA-associated vasculitis, has good face and construct validity, is reliable, feasible, and discriminates among disease states. The AAV-PRO is ready for inclusion within clinical trials and research studies as part of its ongoing validation and exploration of its measurement properties within different populations. The AAV-PRO provides the means to ensure patients' perspectives on their disease are represented in the study of AAV.

Acknowledgements

We would like to thank all of the patients who contributed their views and valuable time to assist us with this study. We also thank Vasculitis UK, our collaborators on the paper-based UK survey, for their assistance with identifying patients, and for organising and funding postage and packaging. We also thank the Vasculitis Patient-Powered Research Network for its collaboration on the US-based online survey, including identifying patients, building the online version of the questionnaire, and compiling a report on patient responses. Authors 1 and 2 contributed equally to this manuscript.

Grants and other financial support:

Sponsored by the University of Oxford and the Vasculitis Clinical Research Consortium (VCRC). Funding for the development of the PRO was received from the Medical Research Fund, Oxford, the Oxfordshire Health Services Research Committee Ref. 1098, and a Patient-Centered Outcomes Research Institute Pilot Project Grant. The VCRC has received support from the US National Institute of Arthritis and Musculoskeletal and Skin Diseases (U54 AR057319 and U01 AR51874), the National Center for Research Resources (U54 RR019497); and the Office of Rare Diseases Research. and the National Center for Advancing Translational Science. The VCRC is part of the Rare Diseases Clinical Research Network (RDCRN). Dr. Robson and Professor Luqmani were supported in part by the National Institute for Health Research Musculoskeletal Biomedical Research Unit, Oxford, UK. Dr Robson was supported by a National Institute for Health Research (NIHR) clinical lectureship. Dr. Milman was supported by a UCB/Canadian Rheumatology Association/Arthritis Society postgraduate rheumatology fellowship award and a research fellowship from

the Department of Medicine at the Ottawa Hospital. Oxford University Innovation provided funding of translatability assessment.

Conflicts of interest

Nil declared.

Table 1. Demographic and clinical characteristics of survey participants.

Demographic characteristics		UK N=348 (%)	USA N=278 (%)	All N=626 (%)	X^2	p=
Sex (n=623)	Male	135 (38.9)	90 (32.7)	225 (36.7)	*2.54*	0.11
	Female	212 (61.1)	185 (67.3)	397 (63.8)		
Age group (years) (n=608)	≤45	25 (7.3)	51 (19.1)	76 (12.5)		
	>45 ≤60	95 (27.9)	90 (33.7)	185 (30.4)	40.64	0.00
	>60 ≤75	166 (48.7)	116 (43.4)	282 (46.4)		
	>75	55 (16.1)	10 (3.7)	65 (10.7)		
Ethnicity (n=624)	Asian	5 (1.4)	7 (2.5)	12 (1.9)		
	Black or African/American	1 (0.3)	2 (0.7)	3 (0.5)	*3.77*	0.71
	Black African or Caribbean British	1 (0.3)	0 (0)	1 (0.2)		
	White	333 (95.7)	259 (93.8)	592 (94.9)		
	American Indian or Alaska Native	0 (0)	1 (0.4)	1 (0.2)		
	Multiple	3 (0.9)	3 (1.1)	6 (1.0)		
	Other	5 (1.4)	4 (1.4)	9 (1.4)		
Qualifications (highest) (n=623)	Degree	157 (45.4)	204 (73.6)	361 (57.9)		
	Vocational/employment related	71 (20.5)	26 (9.4)	97 (15.6)	*59.46*	0.00
	School/high school qualifications	90 (26.0)	46 (16.6)	136 (21.8)		
	None	28 (8.1)	1 (0.4)	29 (4.7)		
Employment status (n=623)	Disabled	50 (14.5)	48 (17.3)	98 (15.7)		
	Employed with income	78 (22.6)	112 (40.3)	190 (30.5)		
	Retired	204 (59.1)	88 (31.7)	292 (46.9)	*56.68*	0.00
	Employed without income	3 (0.9)	3 (1.1)	6 (1.0)		
	Homemaker/carer	4 (1.2)	9 (3.2)	13 (2.1)		
	Unemployed	6 (1.7)	7 (2.5)	13 (2.1)		
	Other (e.g. student, employed & student)	0 (0.0)	11 (4.0)	11 (1.8)		
Type of AAV	EGPA	47 (13.5)	48 (17.3)	95 (15.2)		
	GPA	251 (72.1)	184 (66.2)	435 (69.5)	*20.37*	0.00
	MPA	28 (8.0)	43 (15.5)	71 (11.3)		
	Unspecified AAV	22 (6.3	3 (1.1)	25 (4.0)		
Positive ANCA test	Yes	270 (78.3)	222 (79.9)	492 (79.9)		
	No	15 (4.3)	31 (11.2)	46 (7.4)	*17.66*	0.00
	Don't know	60 (17.4)	25 (9.0)	85 (13.6)		
Current disease status	Active disease	100 (29.8)	75 (27.0)	175 (28.5)	*0.58*	0.45
	Remission	236 (70.2)	203 (73.0)	439 (71.5)		
Flare within the last 2 years	Yes	135 (40.2)	129 (46.4)	264 (43.0)		
	No	157 (46.7)	112 (40.3)	269 (43.8)	*5.09*	0.17
	Don't know	32 (9.5)	21 (7.6)	53 (8.6)		
	Never had a flare	12 (3.6)	16 (5.8)	28 (4.6)		
Organs affected by AAV	Lungs	215 (61.8)	205 (73.7)	420 (67.1)	*10.01*	0.00
	ENT	249 (71.6)	215 (77.3)	464 (74.1)	*2.70*	0.10
	Eyes	135 (38.8)	124 (44.6)	259 (41.4)	*2.15*	0.14
	Kidneys	185 (53.2)	153 (55.0)	338 (54.6)	*0.22*	0.64
	Nerves	139 (39.9)	91 (32.7)	230 (36.7)	*3.46*	0.06
	Skin	128 (36.8)	123 (44.2)	251 (40.1)	*3.58*	0.06
	Joints	192 (55.2)	151 (53.6)	341 (54.5)	*0.16*	0.70
Time from diagnosis (yrs)	Mean (SD)	10.6 (7.5)	7.6 (7.4)	9.3 (7.5)	*t=4.89*	0.00
	[range]	[0.2-38.8]	[0.1-44.5]	[0.1 – 44.5]		

Figure 1. Survey responses at baseline of 35 candidate questionnaire items (N=626). A. Symptom severity; B. Difficulties with everyday life and C. Social and Emotional Impact. (n=individual response rate for each candidate item).

Figure 2. The AAV-PRO. A profile measure containing 6 different domains which all individually fit the Rasch model and have good internal consistency. A. Domains of the AAV-PRO, B. Distribution of the 29 items of the AAV-PRO across the 6 domains.

Table 2. Test-retest reliability and estimates of meaningful change. Intraclass correlation coefficients (ICCs), standard error of the measure (SEM) and minimal detectable change (MDC_{90}) for the 6 AAV-PRO scales/domains.

AAV-PRO Scale (number of items)	**ICC**[a]	**95%CI**	**Baseline mean raw score (SD)**	**SEM**[b] Raw score	**SEM**[b] 0-100 scale	**MDC_{90}**[c] Raw score	**MDC_{90}**[c] 0-100 scale
Organ Specific Symptoms (5)	0.89	0.84 to 0.93	6.91 (4.70)	1.56	7.80	3.64	18.20
Systemic Symptoms (4)	0.91	0.87 to 0.94	7.21 (4.40)	1.32	8.25	3.08	22.75
Treatment side effects (5)	0.95	0.93 to 0.97	7.17 (4.43)	0.99	4.95	2.31	11.55
Social and Emotional Impact (6)	0.96	0.94 to 0.97	9.88 (6.24)	1.25	5.21	2.91	12.13
Concerns about the Future (5)	0.95	0.92 to 0.97	8.83 (5.35)	1.20	6.00	2.79	13.95
Physical Function (4)	0.96	0.94 to 0.97	5.22 (4.17)	0.83	5.19	1.94	12.13

[a] ICC based on USA sample only, while SEM uses baseline scores from all USA/UK respondents combined.

[b] SEM = SD x $\sqrt{1}$ – ICC. Computed using raw scores with scale converted to 0-100 metric at the end.

Example - for Organ Symptom Severity scale: SEM = 4.70 x $\sqrt{1}$-0.89 = 1.56 then Convert to 0-100 scale: 1.56 x 100/20 = 7.80

[C] MDC_{90} = 1.65 x ($\sqrt{2}$) X SEM. Computed using raw scores with scale converted to 0-100 metric at the end.

Example - for Systemic Symptom Severity scale: MDC_{90} = 1.65 x 1.41 x 1.32 = 3.08 then Convert to 0-100 scale: 3.64 x 100/16 = 22.75

Table 3. Known groups validity. Comparison (using t-tests) of baseline AAV-PRO domain scores according to patient-reported current disease state "active" versus "in remission".

	Current disease state	N	Mean	Std. Deviation	t	p
ORGAN SPECIFIC SYMPTOMS	Active	167	47.28	22.55	8.898	<0.0001
	Remission	425	29.35	21.86		
SYSTEMIC SYMPTOMS	Active	168	60.75	25.37	9.525	<0.0001
	Remission	426	38.53	25.70		
TREATMENT SIDE-EFFECTS	Active	171	48.54	22.12	9.565	<0.0001
	Remission	422	30.59	20.09		
SOCIAL & EMOTIONAL IMPACT	Active	172	53.54	24.17	8.079	<0.0001
	Remission	430	35.65	24.69		
CONCERNS ABOUT THE FUTURE	Active	170	56.76	24.39	7.999	<0.0001
	Remission	431	38.50	25.52		
PHYSICAL FUNCTION	Active	172	44.08	25.76	7.370	<0.0001
	Remission	432	27.56	24.49		

Table 4. Longitudinal construct validity. Mean changes (SD), (using 0-100 metric), and effect sizes for the AAV-PRO domains in relation to patients' responses to a transition item on 3-month follow-up survey.

Transition item: *"How are you NOW (in terms of your vasculitis and any treatment side-effects) compared with 3 months ago (when you first answered the questionnaire)?"* ***Responses:***	**AAV-PRO DOMAINS**					
	ORGAN SYMPTOM SEVERITY Mean (SD) [ES]	**SYSTEMIC SYMPTOM SEVERITY** Mean (SD) [ES]	**TREATMENT SIDE-EFFECTS** Mean (SD) [ES]	**SOCIAL & EMOTIONAL IMPACT** Mean (SD) [ES]	**CONCERNS ABOUT THE FUTURE** Mean (SD) [ES]	**PHYSICAL FUNCTION** Mean (SD) [ES]
Much better	n=38 3.68 (13.79) [0.22]	n=40 6.09 (13.68) [0.21]	n=39 4.62 (12.16) [0.28]	n=40 5.31 (12.27) [0.21]	n=40 6.13 (14.57) [0.23]	n=41 5.79 (15.08) [0.24]
Slightly better	n=69 0.72 (14.86) [0.01]	n=68 2.76 (11.85) [0.07]	N=69 2.03 (12.44) [0.10]	n=72 2.37 (13.36) [0.09]	n=71 3.87 (16.52) [0.17]	N=73 5.05 (10.66) [0.19]
No change/worse	n=186 0.54 (15.09) [0.01]	n=186 0.44 (15.37) [0.00]	n=185 1.24 (13.12) [0.09]	n=190 1.58 (11.18) [0.06]	n=187 1.66 (13.62) [0.05]	n=193 0.32 (9.43) [0.01]
Slightly worse	n=64[a] -1.64 (16.67) [-0.07]	n=63 -5.46 (17.84) [-0.19]	n=62 -0.81 (15.92) [-0.01]	n=65[a] 0.26 (16.02) [0.04]	n=63 -0.87 (17.36) [-0.02]	n=65 -3.08 (18.52) [-0.17]
Much worse	n=14 1.43 (11.17) [0.06]	n=15 -6.25 (18.15) [-0.23]	n=14 -10.71 (16.04) [-0.45]	n=15[a] -7.78 (15.42) [-0.26]	n=15 -13.67 (16.85) [-0.43]	n=15 -9.58 (27.84) [-0.31]
P-value for linear trend	0.185	<0.001	0.001	0.003	<0.001	<0.001
Total	n=374 0.64 (15.06) [0.02]	n=377 0.22 (15.43) [-0.01]	n=373 0.95 (13.71) [0.07]	n=387 1.53 (12.90) [0.06]	n=381 1.44 (15.41) [0.05]	n=392 0.91 (13.69) [0.03]

[a]The 'n' in each cell varies slightly across the rows reflecting an occasional missing patient's response to an item within the health status scale (hence the particular overall scale would not have been computed for that individual).

Online supplementary Figures and Tables

Table S1. Geographical representation of eligible UK and USA (N= 607)

UK	N (%)	USA	N (%)		N (%)		N (%)		N (%)
East Anglia	40 (6.6)	Alabama	1 (0.2)	Indiana	2 (0.3)	New Jersey	11 (1.8)	Texas	8 (1.3)
East Midlands	37 (6.1)	Alaska	2 (0.3)	Iowa	4 (0.7)	New Mexico	1 (0.2)	Virginia	3 (0.5)
North	13 (2.1)	Arizona	2 (0.3)	Kansas	1 (0.2)	New York	4 (0.7)	Washington	6 (1.0)
North east	20 (3.3)	Arkansas	1 (0.2)	Kentucky	4 (0.7)	North Carolina	8 (1.3)	Wisconsin	3 (0.5)
North west	50 (8.2)	California	20 (3.3)	Maryland	3 (0.5)	North Dakota	3 (0.5)	Wyoming	1 (0.2)
London & South east	74 (12.2)	Colorado	8 (1.3)	Massachusetts	7 (1.2)	Ohio	9 (1.5)	District of Columbia	3 (0.5)
South west	43 (7.1)	Connecticut	4 (0.7)	Michigan	6 (1.0)	Oklahoma	1 (0.2)	USA Other	84 (13.8)
West midlands	31 (5.1)	Florida	6 (1.0)	Minnesota	2 (0.3)	Oregon	5 (0.8)		
Northern Ireland	2 (0.3)	Georgia	2 (0.3)	Missouri	3 (0.5)	Pennsylvania	11 (1.8)		
Scotland	24 (4.0)	Idaho	1 (0.2)	Montana	2 (0.3)	South Carolina	5 (0.8)		
Wales	15 (2.5)	Illinois	9 (1.5)	Nevada	4 (0.7)	Tennessee	2 (0.3)		

Supplementary Table S2. Rasch Analysis

	Un-recoded				Recoded (0=0, 1=1, 2=1, 3=2, 4=3)			
Items	**Item-trait interaction**	**N misfit**	**N FitResid**	**N reversed**	**Item-trait interaction**	**N misfit**	**N FitResid**	**N reversed**
		(p<0.01)	**>|2.5|**	**thresholds**		**(p<0.01)**	**>|2.5|**	**thresholds**
All 35 items	1309.7, df=315, p<0.000001	24 (69%)	24 (69%)	20 (57%)	1065.5, df=315, p<0.000001	17 (49%)	23 (66%)	3 (9%)
12 Symptom severity (SS) items	201.8, df=108, p<0.000001	4 (30%)	4 (30%)	11 (92%)	188.1, df=108, p=0.000003	4 (30%)	4 (30%)	0
23 Health-related Quality of Life (HRQo) items	771.0, df=207, p<0.000001	11 (48%)	14 (61%)	6 (26%)	614.11, df=207, p<0.000001	9 (39%)	12 (52%)	-13%
Organ symptoms scale:	52.97, df=45, p=0.194	0	0	5 (100%)	40.7, df=40, p=0.438	0	0	0
5 items: nose/sinuses, ears, chest, mouth/throat, eyes								
7 Systemic symptoms (nb these items loaded >.5 in EFA)	140.2, df=63, p<0.000001	3 (43%)	2 (29%)	6 (86%)	117.7, df=63, p=0.00004	1 (14%)	2 (29%)	0
Systemic symptoms scale:	38.5, df=36, p=0.356	0	0	1	36.53, df=36, p=0.444	0	0	0
4 Items: Fever, fatigue/tired, muscle, joint								
Treatment side-effects scale:	67.90, df=45, p=0.015	1 (20%)	0	4 (80%)	54.29, df=40, p=0.065	1 (20%)	1 (20%)	0
5 items: Skin, indigestion, sleep, weight , appearance								
8 Difficulty items	402.4, df=72, p<0.000001	8 (100%)	8 (100%)	1 (13%)	299.3, df=72, p<0.000001	5 (63%)	7 (88%)	2 (25%)
Physical Function scale:	58.29, df=36, p=0.011	1 (25%)	1 (25%)	0	107.4, df=32, p<0.000001	3 (75%)	1 (25%)	0
4 items: Walking, stairs, physical activities, washing/dressing								
15 Social Emotional items	386.1, df=135, p<0.000001	7 (47%)	10 (67%)	2 (13%)	293.7, df=135, p<0.000001	3 (20%)	9 (60%)	0
9 Social items	195.8, df=81, p<0.000001	4 (44%)	4 (44%)	2 (22%)	142.4, df=81, p=0.00003	1 (11%)	5 (56%)	0
Concerns about the future scale:	58.07, df=45, p=0.091	0	1 (20%)	1 (20%)	91.57, df=45, p=0.000005	1 (20%)	1 (20%)	0
5 items: Future, dependency, long-term plans, travel, treatment								
6 Emotional	194.6, df=54, p<0.000001	2 (33%)	2 (33%)	2 (22%)	179.6, df=54, p<0.000001	3 (50%)	3 (50%)	0
Social and emotional impact scale:								
6 items: Anxious/stressed, felt down, concentrating, let people	77.38, df=54, p=0.020	0	1 (17%)	0	56.19, df=54, p=0.393	0	0	0
focused on coping, felt upset as unable to work etc								
Further analyses:								
8 Difficulty	402.43, df=72, p<0.000001	8 (100%)	8 (100%)	1 (13%)	339.3, df=72, p<0.000001	8 (100%)	8 (100%)	0
16 Social Emotional + Sleep	477.46, df=144, p<0.000001	8 (50%)	10 (63%)	4 (25%)	412.72, df=144, p<0.000001	7 (44%)	9 (56%)	0
7 Difficulty (minus sleep)	383.74, df=63, p<0.000001	7 (100%)	6 (86%)	1 (14%)	292.10, df=63, p<0.000001	5 (71%)	5 (71%)	1 (14%)
6 Difficulty (minus sleep and sex life)	238.16, df=54, p<0.000001	3 (50%)	3 (50%)	0	247.53, df=54, p<0.000001	3 (50%)	3 (50%)	0

Table S2. Rasch Report. The 12 SS items did not fit the Rasch model (item trait interaction (ITI) 201.8 df=108, p<0.000001), with 30% of items misfitting; however, 5 of the 12 SS items (AAVNOSE1, AAVEARS1, AAVCHEST1, AAVMOUTH1, AAVEYES1) did fit (ITI 40.7, df=40, p=0.438; best following minor recode of one response category). These items made sense clinically as an 'Organ symptoms' scale and were internally consistent (α 0.77). Seven remaining SS items did not fit the Rasch model (item trait interaction (ITI) 140.2, df=63, p<0.000001), with 3 items (AAVNERV1, AAVINDIG1, AAVSKIN1) misfitting. Removal of these left 4 items (AAVHOT1, AAVFATIG1AAVMUSC1, AAVJOINTS1), which together fitted the Rasch model (ITI 36.53, df=36, p=0.444; best following minor recode of one item response). These 4 items made sense clinically as a 'Systemic symptoms severity' scale and had good internal consistency (α 0.84). The 'Nerve pain/numbness' item was rejected, as it was reasoned that this symptom related to damage that would improve over years. Two remaining SS items are addressed below. All 23 HRQoL items (initially retaining AAVsex1) did not fit the Rasch model (item trait interaction (ITI) 1309.7, df=315, p<0.000001), with 11 (48%) of items misfitting. Eight of the HRQoL items concerned with (physical) functional difficulties (AAVSTAIRS1, AAVPHYS1, AAVHANDS1, AAVSEX1, AAVSLEEP1, AAVSHOPS1, AAVWASH1, AAVESSACT1) also did not fit the Rasch model (ITI 402.4, df=72, p<0.000001). Continued misfitting of 'sexual activity/desire' item (within subsets of other items), despite recoding, further supported its rejection (also given unacceptably high level of missing responses). Four items concerned

with Physical Function (AAVSHOPS1, AAVSTAIRS1, AAVPHYS1, AAVWASH1) fitted the Rasch model (item trait interaction (ITI) 58.29, df=36, p=0.011), made sense conceptually, and were adopted as a 'Physical function' scale. This had excellent internal consistency (α 0.89). Fifteen items concerned with Social/Emotional concerns did not fit the Rasch model (ITI 386.1, df=135, p<0.000001) with 7 (47%) items misfitting. (9 of these – particularly 'social' items - also did not fit the Rasch model) However, 5 items (AAVPLANS1, AAVFUTURE1, AAVRXEFX1, AAVDEPEND1, AAVTRAV1) did fit the Rasch model (ITI 58.07, df=45, p=0.091) and made sense conceptually as a 'Concerns about the future' scale. This also had excellent internal consistency (α 0.89). A further 6 of the social/emotional items (AAVANX1, AAVDEPR1, AAVCONC1, AAVLETDN1, AAVCOPING1, AAVFRUST1) also fitted the Rasch model (ITI 56.19, df=54, p=0.393, particularly following minor recode of one item response), corresponded to a 'Social and emotional impact scale' (α 0.93). Prior qualitative work noted that the SS items 'Indigestion/heartburn/ nausea' and 'Skin problems' had frequently been reported in the context of treatment side-effects, especially in association with steroids. This had also been the case for 3 HRQoL items: 'concern about weight' 'embarrassment about appearance/ symptoms' and 'getting enough good sleep'. Together, these 5 items fitted the Rasch model (ITI 54.29, df=40, p=0.065 following minor recode of one item response) and corresponded to a 'Treatment side-effects' scale that was internally consistent (α 0.73).

Table S3. Comparison of baseline AAV-PRO scale scores among different AAV types' (EGPA, GPA, MPA) baseline AAV-PRO scale scores. Abbreviations: EGPA, Eosinophilic granulomatosis; GPA, Granulomatosis with polyangiitis; MPA: Microscopic polyangiitis. *Analysis of variance (ANOVA) used to analyze the differences among group means. Higher scores on the AAV-PRO subscales indicate worse Health Related Quality of Life.

Type of		Organ Specific	Systemic	Treatment	Social and	Concern about	Physical

vasculitis		Symptoms	Symptoms	side-effects	Emotional Impact	the Future	Function
		(0-100 scale)	(0-100 scale)	(0-100 scale)	(0-100 scale)	(0-100 scale)	(0-100 scale)
EGPA	Mean	34.6	43.6	37.2	40.3	43.3	30.8
	N	94	94	94	93	92	93
	SD	21.9	27.3	23.3	26.0	27.4	24.5
	Median	30	37.5	35	37.5	40	25
GPA	Mean	35.8	43.9	34.8	40.2	43.2	32.4
	N	416	417	420	426	426	429
	SD	23.9	27.3	21.9	26.1	26.7	26.3
	Median	35	37.5	35	37.5	40	25
MPA	Mean	26.6	49.6	37.4	45.4	46.7	34.4
	N	70	68	66	69	70	68
	SD	22.4	28.8	22.4	25.5	26.1	27.7
	Median	25	53.1	37.5	50	50	31.3
Total	Mean	34.5	44.5	35.5	40.8	43.7	32.4
	N	580	579	580	588	588	590
	SD	23.6	27.5	22.2	26.0	26.7	26.2
	Median	30	43.8	35	41.7	40	25
Differences between group means*							
ANOVA F value		4.65	1.34	0.71	1.21	0.51	0.37
Sig. (p-value)		0.01	0.26	0.49	0.3	0.6	0.69

Table S4. Comparison of baseline AAV-PRO subscales between male and female participants. *T-test for Equality of Means performed to determine if two population means are equal. Higher scores on the AAV-PRO subscales indicate worse Health Related Quality of Life.

							95% Confidence Interval			
	Sex	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**	**Mean Difference**	**Lower**	**Upper**	**t-value**	**Sig.***
Organ Specific Symptoms	male	216	30.3	23.1	1.57	-6.52	-10.4	-2.62	-3.29	0.001
(0-100 scale)	female	384	36.8	23.5	1.2					
Systemic Symptoms	male	214	40	27.3	1.86	-7.91	-12.5	-3.34	-3.40	0.001
(0-100 scale)	female	385	47.9	27.3	1.38					
Treatment side-effects	male	215	29.1	19.8	1.35	-10.4	-13.9	-6.94	-5.88	<0.001
(0-100 scale)	female	386	39.6	22.5	1.14					
Social and Emotional Impact	male	220	36.5	25.4	1.71	-7.41	-11.7	-3.14	-3.40	0.001
(0-100 scale)	female	389	43.9	26	1.32					
Concern about the Future	male	223	39.5	26.1	1.75	-7.41	-11.8	-3.02	-3.32	0.001
(0-100 scale)	female	386	46.9	26.8	1.36					
Physical Function	male	220	29	25.8	1.74	-5.82	-10.1	-1.53	-2.66	0.008
(0-100 scale)	female	391	34.8	26.1	1.32					

Table S5. Comparison of baseline AAV-PRO scale scores by age-groups (≤65 versus >65 years). *T-test for Equality of Means performed to determine if two population means are equal. Higher scores on the AAV-PRO subscales indicate worse Health Related Quality of Life.

	Age group	N	Mean	Std. Deviation	Std. Error Mean	Mean Difference	Lower	Upper	t-value	Sig.*
Organ Specific Symptoms	≤65	347	36.3	23.6	1.3	3.96	0.11	7.8	2.02	0.04
(0-100 scale)	>65	245	32.3	23.2	1.48					
Systemic Symptoms	≤65	346	45.3	28.3	1.52	0.48	-4.04	4.99	0.21	0.83
(0-100 scale)	>65	245	44.8	26.4	1.69					
Treatment side-effects	≤65	346	37.5	23.4	1.26	4.26	0.75	7.78	2.39	0.018
(0-100 scale)	>65	247	33.2	20	1.27					
Social and Emotional Impact	≤65	345	43.6	27.1	1.46	6.16	2.04	10.3	2.94	0.003
(0-100 scale)	>65	256	37.5	24	1.5					
Concern about the Future	≤65	345	45.5	27.2	1.47	3.57	-0.76	7.9	1.62	0.102
(0-100 scale)	>65	256	42	26.1	1.63					
Physical Function	≤65	349	30.8	25.9	1.39	-4.51	-8.72	-0.3	-2.10	0.036
(0-100 scale)	>65	255	35.3	26.2	1.64					

Table S6. Comparison of baseline AAV-PRO scale scores for UK versus USA participants. *T-test for Equality of Means performed to determine if two population means are equal. Higher scores on the AAV-PRO subscales indicate worse Health Related Quality of Life.

								95% Confidence Interval			
		N	Mean	Std. Deviation	Std. Error Mean	Mean Difference	Std. Error Difference	Lower	Upper	t-value	Sig.*
Organ Specific Symptoms	UK	334	38.2	24	1.32	7.98	1.88	4.29	11.7	4.24	<0.001
(0-100 scale)	USA	270	30.1	22.1	1.35						
Systemic Symptoms	UK	334	50.7	28	1.53	12.6	2.17	8.3	16.8	5.78	<0.001
(0-100 scale)	USA	269	38.1	25.3	1.54						
Treatment side-effects	UK	337	38.4	22.4	1.22	5.8	1.8	2.27	9.34	3.23	0.001
(0-100 scale)	USA	268	32.6	21.4	1.31						
Social and Emotional Impact	UK	347	44.2	26.5	1.42	6.98	2.1	2.84	11.1	3.32	0.001
(0-100 scale)	USA	266	37.2	24.9	1.53						
Concern about the Future	UK	345	47.2	27.5	1.5	7.05	2.14	2.85	11.2	3.30	0.001
(0-100 scale)	USA	268	40.2	25.2	1.54						
Physical Function	UK	343	37.6	27.4	1.48	11.2	2.03	7.23	15.2	5.53	<0.001
(0-100 scale)	USA	272	26.4	22.8	1.38						

Table T7. Relationship (Pearson correlation) between length of time since diagnosis and AAV-PRO subscale scores. * No significant correlation was seen.

		Organ Specific Symptoms	Systemic Symptoms	Treatment side-effects	Social and Emotional Impact	Concern about the Future	Physical Function
		(0-100 scale)	(0-100 scale)	(0-100 scale)	(0-100 scale)	(0-100 scale)	(0-100 scale)
Time (years) since diagnosis	Correlation*	0.077	0.015	0.018	-0.057	-0.055	0.016
	Sig. (2-tailed)	0.062	0.724	0.665	0.167	0.176	0.7
	N	586	585	586	595	595	597

REFERENCES

1. Jennette JC, Falk RJ, Bacon PA, et al. 2012 revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides. *Arthritis and rheumatism* 2013;65(1):1-11. doi: 10.1002/art.37715
2. Flossmann O, Berden A, de Groot K, et al. Long-term patient survival in ANCA-associated vasculitis. *Annals of the rheumatic diseases* 2011;70(3):488-94. doi: 10.1136/ard.2010.137778
3. Rhee RL, Hogan SL, Poulton CJ, et al. Trends in Long-Term Outcomes Among Patients With Antineutrophil Cytoplasmic Antibody-Associated Vasculitis With Renal Disease. *Arthritis Rheumatol* 2016;68(7):1711-20. doi: 10.1002/art.39614
4. Robson J, Doll H, Suppiah R, et al. Glucocorticoid treatment and damage in the anti-neutrophil cytoplasm antibody-associated vasculitides: long-term data from the European Vasculitis Study Group trials. *Rheumatology* 2015;54(3):471-81. doi: 10.1093/rheumatology/keu366
5. Mooney J, Poland F, Spalding N, et al. 'In one ear and out the other - it's a lot to take in': a qualitative study exploring the informational needs of patients with ANCA-associated vasculitis. *Musculoskeletal Care* 2013;11(1):51-9. doi: 10.1002/msc.1030
6. Koutantji M, Harrold E, Lane SE, et al. Investigation of quality of life, mood, pain, disability, and disease status in primary systemic vasculitis. *Arthritis and rheumatism* 2003;49(6):826-37. doi: 10.1002/art.11471
7. Basu N, McClean A, Harper L, et al. The characterisation and determinants of quality of life in ANCA associated vasculitis. *Annals of the rheumatic diseases* 2014;73(1):207-11. doi: 10.1136/annrheumdis-2012-202750
8. Fardet L, Flahault A, Kettaneh A, et al. Corticosteroid-induced clinical adverse events: frequency, risk factors and patient's opinion. *Br J Dermatol* 2007;157(1):142-8. doi: 10.1111/j.1365-2133.2007.07950.x
9. Basu N, McClean A, Harper L, et al. Markers for work disability in anti-neutrophil cytoplasmic antibody-associated vasculitis. *Rheumatology* 2014;53(5):953-6. doi: 10.1093/rheumatology/ket483
10. Benarous L, Terrier B, Laborde-Casterot H, et al. Employment, work disability and quality of life in patients with ANCA-associated vasculitides. The EXPOVAS study. *Clin Exp Rheumatol* 2016
11. Basu N, McClean A, Harper L, et al. Explaining fatigue in ANCA-associated vasculitis. *Rheumatology* 2013;52(9):1680-5. doi: 10.1093/rheumatology/ket191
12. Herlyn K, Hellmich B, Seo P, et al. Patient-reported outcome assessment in vasculitis may provide important data and a unique perspective. *Arthritis care & research* 2010;62(11):1639-45. doi: 10.1002/acr.20276
13. Seo P, Jayne D, Luqmani R, et al. Assessment of damage in vasculitis: expert ratings of damage. *Rheumatology* 2009;48(7):823-7. doi: 10.1093/rheumatology/kep103
14. Merkel PA, Aydin SZ, Boers M, et al. The OMERACT core set of outcome measures for use in clinical trials of ANCA-associated vasculitis. *The Journal of rheumatology* 2011;38(7):1480-6. doi: 10.3899/jrheum.110276
15. Walsh M, Mukhtyar C, Mahr A, et al. Health related quality of life in patients with newly diagnosed anti-neutrophil cytoplasm antibody associated vasculitis. *Arthritis care & research* 2011 doi: 10.1002/acr.20471 [published Online First: 2011/04/01]
16. Tomasson G, Boers M, Walsh M, et al. Assessment of health-related quality of life as an outcome measure in granulomatosis with polyangiitis (Wegener's). *Arthritis care & research* 2012;64(2):273-9. doi: 10.1002/acr.20649
17. Fitzpatrick R, Davey C, Buxton MJ, et al. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998;2(14):i-iv, 1-74.
18. Robson JC, Milman N, Tomasson G, et al. Exploration, Development, and Validation of Patient-reported Outcomes in Antineutrophil Cytoplasmic Antibody-associated Vasculitis Using the

OMERACT Process. *The Journal of rheumatology* 2015;42(11):2204-9. doi: 10.3899/jrheum.141143
19. Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value in Health* 2007;10:S125-S37. doi: 10.1111/j.1524-4733.2007.00275.x
20. Merkel PA, Aydin SZ, Boers M, et al. Current status of outcome measure development in vasculitis. *The Journal of rheumatology* 2014;41(3):593-8. doi: 10.3899/jrheum.131248
21. Robson JC, Milman N, Tomasson G, et al. Exploration, Development, and Validation of Patient-reported Outcomes in Antineutrophil Cytoplasmic Antibody-associated Vasculitis Using the OMERACT Process. *J Rheumatol* 2015 doi: 10.3899/jrheum.141143
22. Robson J, Dawson J, Cronholm PF, et al. Health related quality of life in ANCA associated vasculitis and item generation for a disease specific patient reported outcome measure. *Patient Related Outcome Measures* 2017;In Press
23. EuroQol G. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16(3):199-208.
24. van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 2012;15(5):708-15. doi: 10.1016/j.jval.2012.02.008
25. Barrett P, Kline P. The Observation to Variable Ratio in Factor Analysis. *Personality Study and Group Behaviour* 1981;1(1):23-33.
26. Fayers P, Machin D. Quality of Life- Assessment, Analysis and Interpretation. Chichester: John Wiley & Sons Ltd. 2000.
27. Lorenzo-Seva U, Ferrando PJ. FACTOR: a computer program to fit the exploratory factor analysis model. *Behav Res Methods* 2006;38(1):88-91.
28. Andrich D. Rasch Models for Measurement. Newbury Park, CA: Sage 1988.
29. Kaiser HF. The application of electronic computers to factor analysis *Educ psychol measurement* 1960;20:141-51.
30. Avlund K, Kreiner S, Schultz-Larsen K. Construct validation and the Rasch model: functional ability of healthy elderly people. *Scand J Soc Med* 1993;21(4):233-46.
31. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess* 2003;80(1):99-103. doi: 10.1207/S15327752JPA8001_18
32. R.K H, Jones RW. Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement Issues and Practice* 1993;12:38-47.
33. Prieto L, Alonso J, Lamarca R. Classical Test Theory versus Rasch analysis for quality of life questionnaire reduction. *Health Qual Life Outcomes* 2003;1:27. doi: 10.1186/1477-7525-1-27
34. Nunally JC, Bernstein IH. Psychometric theory. New York: Columbus OH. 1994.
35. Kline P. A Handbook of Psychological Testing. London: Routledge 1993.
36. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993;118(8):622-9.
37. Andrews F, ., Withey S. Social Indicators of Well-Being: American's Perceptions of Life Quality. New York: Plenum 1976.
38. Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther* 1996;76(10):1109-23.
39. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10(4):407-15.
40. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999;52(9):861-73.

41. de Vet HC, Terwee CB, Ostelo RW, et al. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes* 2006;4:54. doi: 10.1186/1477-7525-4-54
42. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;27(3 Suppl):S178-89.
43. Nulty DD. The adequacy of response rates to online and paper surveys: what can be done? *Assessment & Evaluation in Higher Education* 2008;33(3):301-14.
44. File T, Ryan C. Computer and Internet Use in the United States: 2013. American community survey reports. In: Bureau USC, ed. www.census.go/comtent/dam/Census/library/publications/2014/acs/acs-28.pdf (last accessed 29/09/2016, 2014.
45. Muehlhausen W, Doll H, Quadri N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. *Health Qual Life Outcomes* 2015;13:167. doi: 10.1186/s12955-015-0362-x
46. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health* 2008;11(2):322-33. doi: 10.1111/j.1524-4733.2007.00231.x
47. Matcham F, Scott IC, Rayner L, et al. The impact of rheumatoid arthritis on quality-of-life assessed using the SF-36: a systematic review and meta-analysis. *Semin Arthritis Rheum* 2014;44(2):123-30. doi: 10.1016/j.semarthrit.2014.05.001 [published Online First: 2014/06/30]
48. Revicki DA, Cella D, Hays RD, et al. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes* 2006;4:70. doi: 10.1186/1477-7525-4-70
49. Willke RJ. Measuring the value of treatment to patients: patient-reported outcomes in drug development. *Am Health Drug Benefits* 2008;1(1):34-40.
50. Wiebe S, Guyatt G, Weaver B, et al. Comparative responsiveness of generic and specific quality-of-life instruments. *J Clin Epidemiol* 2003;56(1):52-60. [published Online First: 2003/02/19]