Validation of the German version of the ANCA-associated vasculitis patient-reported outcome questionnaire

A. Niesmann[1,2], M. Jacobs[1,2], T. Filla[1,2], C. Düsing[1,2], J. Mucke[1,2,3], J.G. Richter[1,2], N. Ruffer[4], M.C. van Saan[1,2], O. Sander[1,2], A. Stütz[1,2], M. Schneider[1,2], M. Schmitz[5], G. Chehab[1,2,6], A. Kernder[1,2,3]

[1]Department of Rheumatology, University Hospital Düsseldorf, Medical Faculty of Heinrich Heine University Düsseldorf; [2]Hiller Research Center, University Hospital Düsseldorf, Medical Faculty of Heinrich Heine University Düsseldorf; [3]Ruhr University Bochum, Rheumazentrum Ruhrgebiet, Herne; [4]III Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg; [5]Städtisches Klinikum Solingen GmbH, Solingen; [6]Department of Rheumatology and Clinical Immunology, KEM Klinken Essen Mitte, Essen, Germany.

Abstract

Objective

This study aimed to validate the German version of the ANCA-associated vasculitis patient-reported outcome (AAV-PRO) measure by evaluating its psychometric properties.

Methods

Our study was a prospective cohort study of German speaking AAV patients, recruited in the rheumatology outpatient clinics at the University Hospital Düsseldorf. The questionnaire includes 29 items across six domains, each rated from 0 to 4 points. Participants completed the AAV-PRO (German) at three time points. They also completed the EuroQol-5D-5L (EQ-5D-5L) and self-reported their vasculitis status. The Birmingham Vasculitis Activity Score version 3 (BVASv3) and Vasculitis Damage Index (VDI) were completed by the treating physicians.

Results

70 AAV patients (48.5% female) participated. Overall, fatigue was the most prevalent symptom. Female AAV patients reported higher symptom burden across all domains and had a higher total score ($Z=-2.55$, $p=0.011$). The AAV-PRO (German) demonstrated good internal consistency in the domains "social and emotional impact", "concerns about the future" and "physical function" (Cronbach's alpha≥0.83). The domain "systemic symptoms severity" demonstrates acceptable to good consistency. The other domains demonstrated insufficient consistency. All domains showed high test-retest reliability (Pearson r≥0.8). BVASv3 and VDI showed limited correlations with the AAV-PRO (German) domains. Construct validity was supported by moderate to strong correlations with the EQ-5D-5L in five domains. Patients with self-reported active disease had statistically significantly higher mean scores than those with self-reported inactive disease in three domains.

Conclusion

Our study supports the use of the AAV-PRO (German). However, two domains need to be interpreted carefully.

Key words

AAV-PRO, ANCA, vasculitis, patient-reported outcome, PRO-questionnaire, validation

Amelie Niesmann
Miriam Jacobs
Tim Filla, PhD, MSc
Christina Düsing, MD
Johanna Mucke, MD
Jutta G. Richter, MD
Nikolas Ruffer, MD
Marie Celine van Saan, MD
Oliver Sander, MD
Ayla Stütz, MD
Matthias Schneider, MD
Michael Schmitz, MD
Gamal Chehab, MD
Anna Kernder, MD

Please address correspondence to:
Anna Kernder
Rheumazentrum Ruhrgebiet,
Ruhr University Bochum
Claudiusstraße 45,
44649 Herne, Germany.
E-mail: anna.kernder@elisabethgruppe.de

Received on August 29, 2025; accepted in revised form on December 23, 2025.

Competing interests: none declared.

Introduction

Anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis (AAV) comprises three small-vessel vasculitis: granulomatosis with polyangiitis (GPA), eosinophilic granulomatosis with polyangiitis (EGPA) and microscopic polyangiitis (MPA) (1). These systemic diseases can affect multiple organ systems (1).

Improved treatments have transformed AAV from acute, severe illness into a chronic disease. Despite better outcomes, patients often face long-term effects from both the disease and its therapy, which can cause serious side effects or even damage (2). This burden does not necessarily have to correspond to the medically objectifiable findings on disease activity or disease damage (3). To better understand these effects, patient-reported outcome (PRO) questionnaires are used. They capture patients' own experiences and provide valuable insights into how AAV impacts their lives. In response to the OMERACT Vasculitis Working Group's identified need for a disease-specific PRO, an international steering committee developed the AAV-PRO, a questionnaire designed to assess symptom burden in patients with AAV (4). It includes 29 items, grouped into six domains (Fig. 1, 2).

Although the AAV-PRO (German) was professionally translated by the OMERACT Vasculitis Working Group, it has not yet been formally validated. Currently, the English (4) and Italian (5) versions have completed formal validation.

Methods

The objective of this study was to validate the AAV-PRO (German) as a reliable instrument. Psychometric properties were assessed, alongside disease activity and quality of life, including correlations with established outcome measures and clinical instruments.

Study design

Our study was a prospective observational cohort.

Inclusion criteria and data collection

To be eligible, participants had to be diagnosed and classified as AAV, be German speakers and be at least 18 years old. The diagnosis of AAV was based on the expert opinion of the treating rheumatologist. The subtype was classified according to the 2012 Chapel Hill Consensus Conference nomenclature (1) and the respective classification criteria (6-8).

Participants were recruited from the rheumatology outpatient clinic at University Hospital Düsseldorf. Recruiting patients with this rare disease at a single centre limited the cohort size to 70 participants. All participants were informed orally from a team member in a 5-minute conversation and via patient information leaflet in detail about the objectives, procedures, and purpose of the study prior to giving their written informed consent to participate. At their first clinical visit (t1), the participants completed the 29-item AAV-PRO (German) candidate questionnaire in the waiting area or treatment room, either prior to or following their consultation with the physician. Three to five days later (t2), the participants completed the provided questionnaire again at home and returned it by mail to assess test-retest reliability. After three to six months (t3), they completed the questionnaire one more time during their next outpatient visit in the clinic. Each item of the AAV-PRO (German) can be answered using one of five options, scored from 0 to 4 points. Domain scores were calculated by summing the corresponding items for each domain. The Birmingham Vasculitis Activity Score version 3 (BVASv3) (9) and the Vasculitis Damage Index (VDI) (10) were applied to measure disease activity and disease-related organ damage. Subsequently, correlations with the AAV-PRO (German) were calculated. At both baseline and follow-up (t1 and t3), participants also filled out the EuroQol-5D-5L (EQ-5D-5L) questionnaire. Correlations with the EQ-5D-5L were examined to verify the instrument's convergent validity. The EQ-5D-5L is a standardised measure of health-related quality of life encompassing five dimensions (11). EQ-5D-5L index values were derived using the crosswalk method.

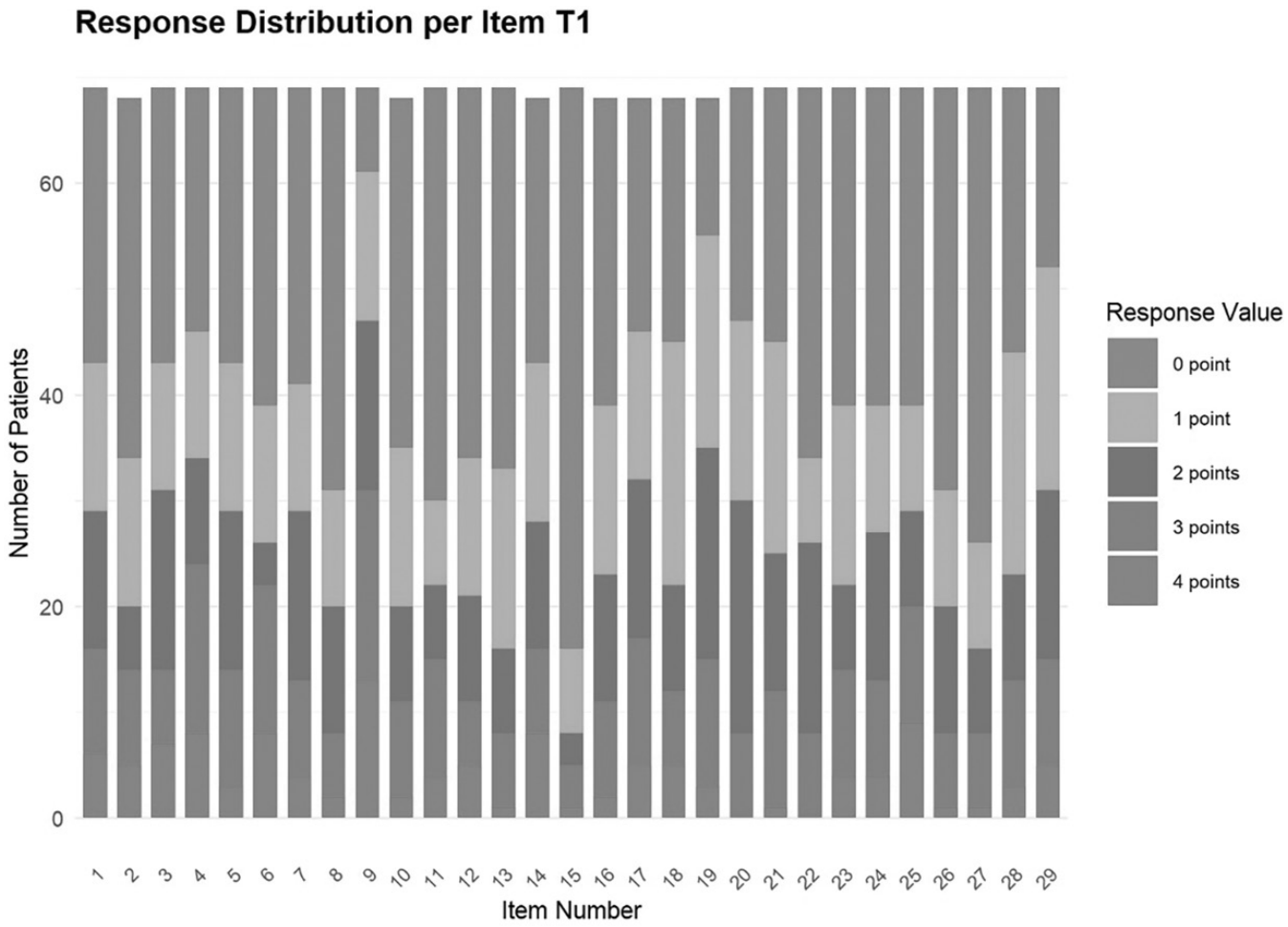

Fig. 1. Response distribution for the 29 AAV-PRO (German) items at baseline.
The y-axis shows the distribution of ratings among participants. The x-axis shows the questionnaire item number.
T1: first clinical visit.

Statistical analysis

Descriptive statistics included frequency analyses for categorical variables, as well as mean, standard deviation and median for continuous variables. Internal consistency was assessed using Cronbach's alpha, while test-retest reliability was evaluated by calculating the Pearson correlation coefficient with 95% confidence intervals (CI) between scores at t1 and t2. The Wilcoxon rank-sum test was performed to examine potential differences between two independent groups with respect to an ordinal or non-normally distributed metric variable. The Kruskal-Wallis test was used to assess differences in questionnaire scores among more than two independent groups. If the test indicated statistically significant differences, pairwise *post-hoc* comparisons were performed using Dunn's test to identify which specific groups differed.

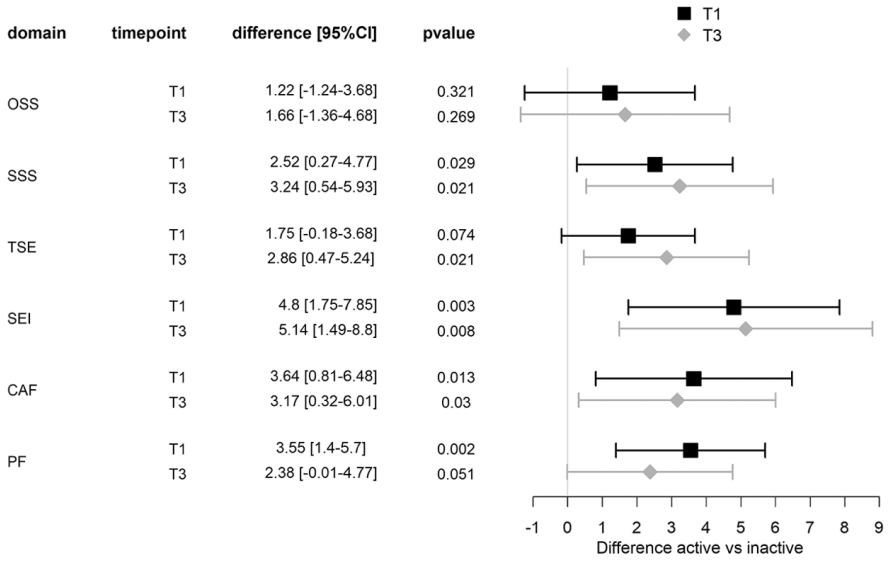

Fig. 2. Comparison of mean domain-specific scores by self-reported disease status (active *vs*. inactive). The y-axis shows the AAV-PRO (German) domain. The x-axis shows the mean difference of the domains score between active and inactive.
OSS: organ-specific symptoms; SSS: systemic symptoms severity; TSE: treatment side effects; SEI: social and emotional impact; CAF: concerns about the future; PF: physical function; CI: confidence interval; T1: first clinical visit; T3: 3-6 months after T1.

Table I. The demographic and clinical characteristics of survey participants.

Demographic characteristics	t1	t3
Participants, n (%)	70 (100)	61 (100)
Female, n (%)	34 (48.5)	31 (51)
Male, n (%)	36 (51.5)	30 (49)
Age, years mean (SD)	60.1 (13.7)	60.4 (14.1)
Age groups, n (%)		
≤ 65 years	45 (64.3)	38 (62.3)
> 65 years	25 (35.7)	23 (37.7)
Type of AAV, n (%)		
GPA	45 (64.3)	40 (65.6)
EGPA	16 (22.9)	14 (23)
MPA	9 (12.9)	7 (11.5)
Vasculitis-activity (patient perspective), n (%)		
Active	21 (30)	16 (26.2)
Inactive	44 (62.9)	37 (60.7)
Missing information	5 (7.1)	8 (13.1)
Clinical instruments		
BVASv3, mean (SD)	1.68 (2.43)	2.13 (3.21)
BVASv3 ≥ 1, n (%)	16 (22.9)	9 (14.8)
BVASv3 = 0, n (%)	22 (31.4)	15 (24.6)
BVASv3 missing information, n (%)	32 (45.7)	37 (60.7)
VDI, mean (SD)	2.45 (2.03)	3.36 (2.94)
VDI missing information, n (%)	39 (55.8)	50 (82)
Education, n (%)		
Low education	17 (24.3)	15 (24.6)
Medium education	22 (31.4)	20 (32.8)
High education	12 (17.1)	10 (16.4)
Missing information	19 (27.1)	16 (26.2)
Ethnicity, n (%)		
Europe	64 (91.4)	56 (91.8)
Asian	3 (4.3)	2 (3.3)
Africa	2 (2.6)	2 (3.3)
missing information	1 (1.4)	1 (1.6)
Therapy, n (%)		
Glucocorticoid therapy	24 (34.9)	25 (41)
Immunosuppressive therapy	56 (80)	36 (59)
Rituximab	20 (28.6)	14 (23)
Methotrexate	11 (15.8)	6 (9.8)
Mycophenolate mofetil	4 (5.7)	2 (3.3)
Azathioprine	11 (15.7)	9 (14.8)
Others	10 (14.3)	5 (8.2)

Correlations between continuous variables and main scores were assessed using the Spearman rank correlation coefficient or the Pearson correlation coefficient with 95% CI. All statistical analyses were conducted using R Studio (v. 2024.12.0+467).

Compliance with ethical standards

The data collection was approved by the Ethics Committee of the Medical Faculty of the University of Düsseldorf (Study number 2019-602).

Results

Study population at baseline

A total of 70 participants were included in the study. The mean age was 60.1 years (SD 13.7), gender distribution (biological sex) was balanced (female n=34, 48.5%; male n=36, 51.5%). GPA (n=45, 64.3%) was the most prevalent subtype followed by EGPA (n=16, 22.9%) and MPA (n=9, 12.9%). The mean BVASv3 at baseline was 1.7 (SD 2.4). Of the 70 respondents, 21 (30%) self-reported an active disease, whereas 44 (62.9%) reported an inactive disease state (missing information n=5, 7.1%). The majority of the patients (n=56, 80%) received immunosuppressive therapy at t1 (Table I).

Measurement properties of the AAV-PRO (German) questionnaire

The AAV-PRO (German) questionnaire was self-completed by 70 participants at baseline. The response rate for the questionnaires at t2 was 65 of 70 (92.9%). At t3 61 (87.1%) patients were recorded again. A total of 196 AAV-PRO (German) questionnaires were analysed.

Survey responses at baseline

Of the symptoms surveyed, fatigue (item 9) was most prevalent as 61 patients (88.4%) indicated the presence. Fatigue was classified as moderate or severe by 31 patients (severe: n=13, 18.8%). Other clinically relevant symptoms included upper airway symptoms (item 4, 5). Furthermore, 43 (62.3%) participants reported ocular symptoms (item 3) and respiratory symptoms (item 1). More than half of the participants (n=39, 56.5%) reported arthralgia (item 6).

In relation to physical functioning, 28 (41.2%) patients reported that participating in sports or physical activity (item 14) was at least moderately difficult (n=12, 17.6%), extremely difficult (n=8, 11.8%) or impossible (n=8, 11.8%). Item 15, "washing or drying yourself or getting dressed" was described by most patients as unproblematic (n=53, 76.8%) (Supplementary Table S2).

The AAV-PRO (German) responses revealed that patients expressed concerns about the future, general life issues and long-term treatment. At baseline, 35 (51.5%) patients indicated being sometimes, often, or always worried about the future (item 19). Using the same response scale, 31 (44.9%) patients expressed concern about long-term treatment (item 29) and 30 (43.5%) patients reported general stress (item 20) (Fig. 1).

Comparison between the AAV-PRO (German) domain scores and demographic and clinical characteristics

The mean scores for female patients (biological sex) were higher than those for male patients across all domains at t1. However, a statistically significant difference was observed only in the CAF domain (t1: Z=-2.33; p=0.02), as determined by the Wilcoxon rank-sum test. In addition to the domain-specific comparisons, the Wilcoxon rank-sum test also revealed a statistically significant difference in the total score (t1: Z=-2.55; p=0.011).

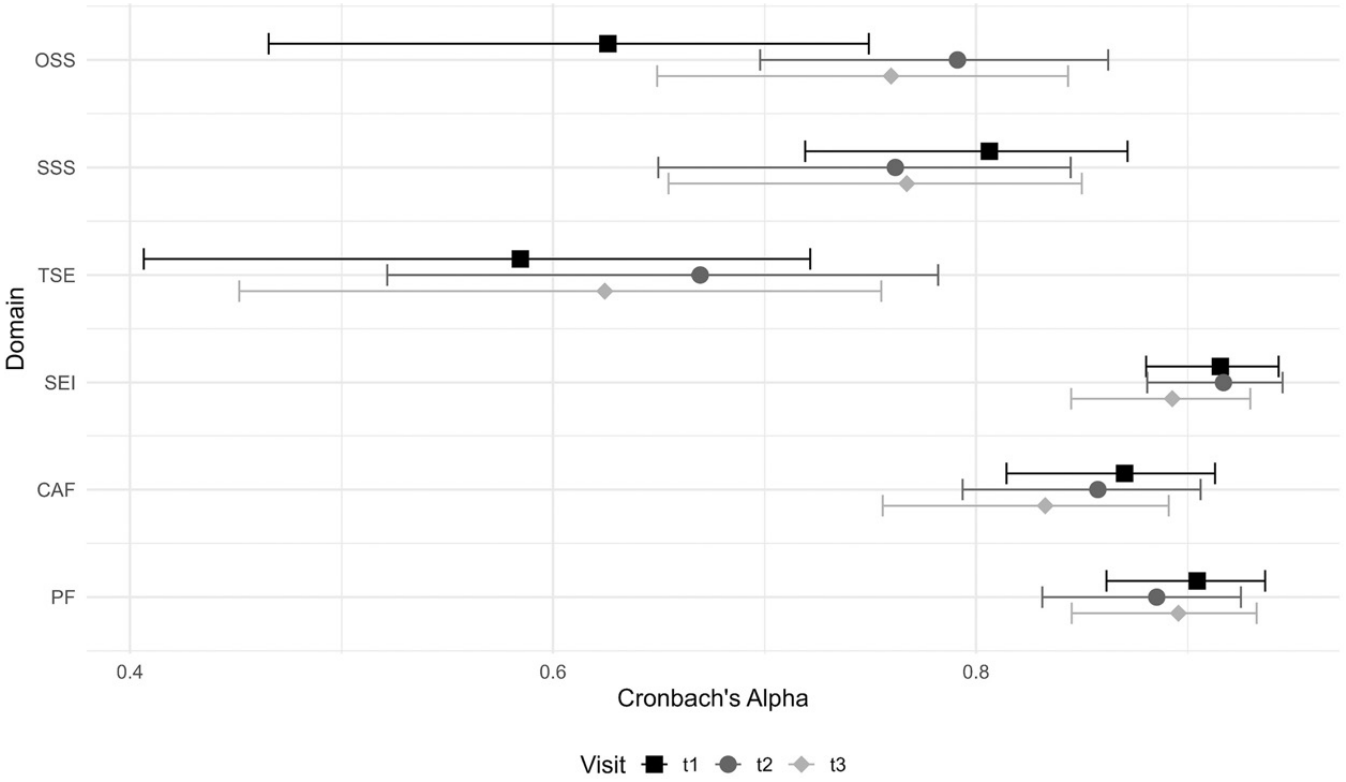

Fig. 3. Confidence interval plot of Cronbach's alpha by time point.
OSS: organ-specific symptoms; SSS: systemic symptoms severity; TSE: treatment side effects; SEI: social and emotional impact; CAF: concerns about the future; PF: physical function; t1: first clinical visit; t2: 3-5 days after t1; t3: 3-6 months after t1.

No relevant correlation (Spearman) was found between age as a continuous variable and AAV-PRO (German) domain scores.

The Kruskal-Wallis test showed no statistically significant differences between the three diagnostic groups (GPA/EGPA/MPA) in any of the six examined domains. In all cases, the p-values were above the significance level of 0.05 suggesting that the distributions of responses did not differ significantly between groups (OSS: χ^2=1.33, p=0.52; SSS: χ^2=0.91, p=0.64; TSE: χ^2=0.38, p=0.83; SEI: χ^2=1.53, p=0.47; CAF: χ^2=4.58, p=0.10; PF: χ^2=1.73, p=0.42).

AAV-PRO (German) domain scores were compared between active (n=21, 30%) and inactive (n=44, 62.9%) patients based on self-reported disease status at t1. On average higher scores were observed across all six AAV-PRO (German) domains among patients who self-reported an active vasculitis compared to those who reported an inactive disease (Fig. 3). A Kruskal-Wallis test revealed statistically significant differences between the three groups (active/inactive/missing information) in the domains SEI (χ^2=9.38, p<0.01), CAF (χ^2=6.52, p=0.04) and PF (χ^2=14.67, p<0.01). *Post-hoc* analyses showed that the group with self-reported active disease had statistically significantly higher mean scores than the group with inactive disease (SEI: mean difference 4.8 [95% CI 1.75–7.85]; CAF: mean difference 3.64 [95% CI 0.81–6.48]; PF: mean difference 3.55 [95% CI 1.4–5.7]) (Fig. 2).

Internal consistency

To assess internal consistency, Cronbach's alpha was calculated for each domain at three time points. The domain TSE showed the lowest consistency, with a Cronbach's alpha of 0.58 [95% CI 0.41–0.72] at baseline. OSS had a Cronbach's alpha of 0.63 [95% CI 0.47–0.75] at t1. These two domains showed variations in Cronbach's alpha across the three time points. In particular, the domain OSS exhibited lower internal consistency at t1 compared to t2 (0.79 [95% CI 0.70–0.86]) and t3 (0.76 [95% CI 0.65–0.84]). Similarly, in domain TSE, internal consistency was lower at t1 than at the other assessment points. For further analysis, factor analysis showed that in OSS and TSE four of five items exhibited low loadings on their respective constructs (Suppl. Table S1). The domain SSS achieved a Cronbach's alpha of 0.81 [95% CI 0.72–0.87] at t1, with comparable values observed at t2 (Cronbach's alpha 0.76 [95% CI 0.65–0.84]) and t3 (Cronbach's alpha 0.77 [95% CI 0.65–0.85]). The remaining three domains, SEI, CAF, and PF, demonstrated high reliability, with Cronbach's alpha values ranging from 0.83 to 0.92 across all time points (Fig. 3).

Test-retest reliability

To calculate the test-retest reliability, the sum scores of the domains at t1 were correlated with those at t2. All six domains showed high reproducibility, with Pearson correlation coefficients of 0.8 or higher. The domain SEI showed the highest correlation with a Pearson correlation coefficient of 0.9 [95% CI 0.84–0.94; p=0.001] (Table II).

Table II. Pearson's correlation coefficient of the AAV-PRO (German) domains of Visit 1 and Visit 2.

Domain of AAV-PRO	Pearson's correlation coefficient	95% CI
Organ-specific symptoms (OSS)	0.80	[0.69, 0.88]
Systemic symptoms severity (SSS)	0.86	[0.77, 0.91]
Treatment side effects (TSE)	0.84	[0.74, 0.90]
Social and emotional impact (SEI)	0.90	[0.84, 0.94]
Concerns about the future (CAF)	0.86	[0.77, 0.91]
Physical function (PF)	0.86	[0.78, 0.91]

AAV: ANCA-associated vasculitis; PRO: patient-reported outcome; CI: confidence interval.

Convergent validity: correlation between the AAV-PRO (German) domain scores and current clinical instruments

The analysis of convergent validity revealed predominantly weak to moderate correlations (Pearson) between disease activity assessed by the treating physicians using the BVASv3 and the AAV-PRO (German) domains at t1. Two domains showed statistically significant moderate correlations: OSS (r=0.50 [95% CI 0.21–0.70; $p<0.01$]) and SSS (r=0.33 [95% CI 0.02–0.60; $p=0.04$]). The remaining domains showed weaker, non-significant correlations: TSE (r=0.2 [95% CI -0.13–0.49; $p=0.24$]); SEI (r=0.28 [95% CI -0.05–0.55; $p=0.1$]); CAF (r=0.06 [95% CI -0.27–0.37; $p=0.74$]) and PF (r=0.3 [95% CI -0.02–0.57; $p=0.07$]).

At t3 a similar pattern was observed. While most correlations remained weak and non-significant, two domains again showed statistically significant moderate associations: SSS (r=0.45 [95% CI 0.05-0.72; $p=0.03$]) and SEI (r=0.43 [95% CI 0.03–0.71; $p=0.04$]). Correlations for the other domains were non-significant: OSS (r=0.23 [95% CI -0.2–0.58; $p=0.28$]); TSE (r=0.37 [95% CI -0.04–0.68; $p=0.07$]); CAF (r=0.25 [95% CI -0.18–0.60; $p=0.25$]) and PF (r=0.34 [95% CI -0.08–0.66; $p=0.11$]). Comparable results were observed in the correlations between the total score and the BVASv3 at t1 (r=0.34 [95% CI 0.02-0.59; $p=0.04$]) and t3 (r=0.38 [95% CI -0.03–0.68; $p=0.07$]).

However, when comparing physician-assessed disease activity by BVASv3 (0 = remission, ≥1 active disease), AAV-PRO (German) scores were higher in 5 of 6 AAV-PRO (German) domains among patients with active disease, except for the CAF domain. A statistically significant difference was only found in the OSS domain (mean difference 3.73 [95% CI 0.67–6.8; $p=0.02$]).

In line with these findings, correlations between the VDI and the AAV-PRO (German) domains at t1 varied in strength. The OSS and SSS domains demonstrated statistically significant moderate correlations (OSS: r=0.42 [95% CI 0.08–0.67; $p=0.02$]; SSS: r=0.47 [95% CI 0.13–0.70; $p=0.01$]). The remaining four domains showed non-significant correlations ranging from very weak to moderate (TSE: r=0.18 [95% CI -0.19–0.50; $p=0.33$]; SEI: r=0.21 [95% CI -0.16–0.52; $p=0.27$]; CAF: r=0.36 [95% CI 0.01–0.63; $p=0.05$]; PF: r=0.25 [95% CI -0.11–0.56; $p=0.17$]), indicating generally limited associations with the VDI. The total score of the AAV-PRO (German) showed a weak-to-moderate but statistically significant positive correlation with the VDI (t1: r=0.38 [95% CI 0.03-0.65; $p=0.04$].

Convergent validity: correlation between the AAV-PRO (German) domain scores and the EQ-5D-5L

In addition, correlations between AAV-

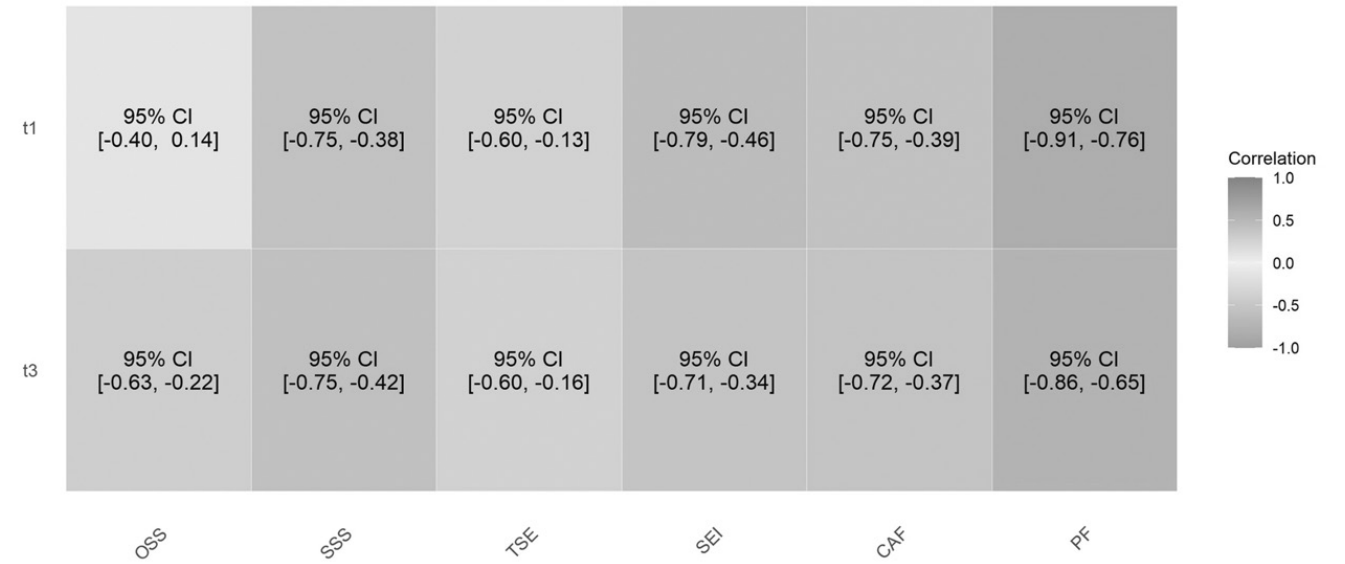

Fig. 4. Heatmap of Spearman Correlations (with 95 % CI) between AAV-PRO (German) domains and the EQ-5D-5L index value (t1 and t3).
The y-axis shows the time point. The x-axis shows the AAV-PRO (German) domain. The colour indicates the strength of the correlation.
OSS: organ-specific symptoms; SSS: systemic symptoms severity; TSE: treatment side effects; SEI: social and emotional impact; CAF: concerns about the future; PF: physical function; CI: confidence interval; t1: first clinical visit; t3: 3-6 months after t1.

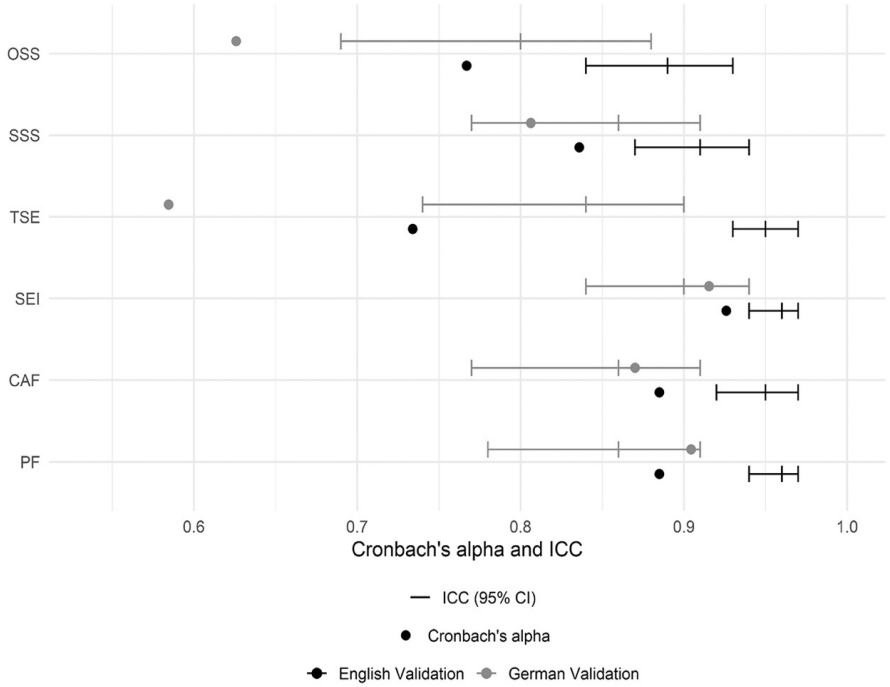

Fig. 5. Comparison of Cronbach's alpha and ICC in the English and German AAV-PRO Validation Studies. The y-axis shows the AAV-PRO domains. The x-axis shows Cronbach's Alpha and ICC with 95%CI at baseline.
AAV: ANCA-associated vasculitis; PRO: patient-reported outcome; ICC: intraclass correlation coefficient; CI: confidence interval; OSS: organ-specific symptoms; SSS: systemic symptoms severity; TSE: treatment side effects; SEI: social and emotional impact; CAF: concerns about the future; PF: physical function.
(The figure presents results obtained from the following source: Robson *et al*.; Validation of the ANCA-associated vasculitis patient-reported outcomes (AAV-PRO) questionnaire. *Ann Rheum Dis* 2018; 77(8): 1157-64).

PRO (German) domain scores and the EQ-5D-5L were analysed to test convergent validity. Correlations (Spearman) between baseline AAV-PRO (German) domains and EQ-5D-5L index-score showed a weak correlation in domain OSS [Spearman's Rho 95% CI -0.4–0.14] while domain TSE [95% CI -0.6–(-0.13)] demonstrated a moderate correlation. Domains SSS [95% CI -0.75-(-0.38)], SEI [95% CI -0.79–(-0.46)] and CAF [95% CI -0.75–(-0.39)] each exhibited a strong correlation, and domain PF [95% CI -0.91-(-0.76)] showed a very strong correlation. Negative correlations were expected, as the two measures are scored in opposite directions (Fig. 4).

Discussion

The AAV-PRO (German) is a recently adapted, disease-specific PRO measure intended for use in AAV research in German language.

Fatigue (item 9) was rated as the most severe symptom, highlighting its relevant impact on the disease. Our results support the findings of Basu *et al*. who identified fatigue as a primary complaint and a key contributor to reduced quality of life (12). In line with this, Thiele *et al*. emphasise the importance of fatigue in rheumatic diseases and recommend its regular assessment (13). It should be noted that the ceiling effect in item 9 (4 points: n=13, 18.8%) limits the instrument's ability to detect worsening, thereby reducing its sensitivity. Similarly, the floor effect in item 15 (0 point: n=53, 76.8%) constrains the detection of further improvements and may mask meaningful positive changes. Across all six domains, female participants (biological sex) reported higher scores, with a statistically significant difference observed in the CAF domain. In the English (4) and Italian (5) validation studies, women consistently reported higher scores than men, with statistically significant differences observed across all domains. These findings underscore the importance of considering sex differences in AAV research and patient management (14). Addressing psychological aspects, such as CAF, could support a sex-sensitive treatment approach. Incorporating psychological support may help patients cope more effectively with these challenges.

A very weak association was found between age and the AAV-PRO (German) domain scores. This finding is consistent with the Italian validation study (5). In contrast, the English validation study showed a statistically significantly higher mean score in the SEI domain among younger patients (4).

There were no differences observed in the mean domain scores between the AAV subtypes, just like in the other validations (4, 5).

The mean domain scores across all domains were higher among patients reporting themselves as active disease compared to those reporting themselves as inactive disease. This finding supports the AAV-PRO's (German) ability to reflect perceived disease activity. Additionally, the SEI, CAF and PF domains showed statistically significant differences between patients reporting active and inactive vasculitis, supporting group validity in these domains. Robson *et al*. (4) found statistically significant differences for all six domains.

Three domains (SEI, CAF, PF) exhibited good internal consistency with a Cronbach's alpha greater than 0.82 at all time points, indicating a reliable measurement of these dimensions. SSS showed a Cronbach's alpha ranging from 0.76 to 0.81, indicating acceptable to good internal consistency and suggesting a satisfactory level of item coherence. Notably, the internal consistency within the other two domains varied across the three time points, indicating potential fluctuations in the reliability of responses over time. The internal consistency of the TSE domain at t1 (Cronbach's alpha 0.58) indicated low reliability and limited coherence among the items. For OSS, Cronbach's alpha was 0.63 at t1, reflecting questionable internal consistency. In both domains, the internal consistency at time points t2 and t3 showed slight improvement,

yet remained within the questionable range for TSE and the acceptable range for OSS. The low factor loadings observed in OSS and TSE suggest that the items may not adequately represent their intended constructs, which could account for the low internal consistency and could be addressed in future revisions (Suppl. Table S1).

Interestingly, the TSE and OSS domains exhibited higher internal consistency in the validation studies conducted in other countries (4, 5). One possible explanation relates to cultural differences in perceiving and reporting physical symptoms and side effects. In some cultures, side effects may be perceived as more severe or reported more readily, while in others they may be downplayed or stigmatised (15). Moreover, linguistic and conceptual challenges in translation may have influenced results, as some items or symptom terms may differ in connotation (16). Future qualitative studies could provide insights and support refinement to improve cross-cultural comparability.

All AAV-PRO domains had good test-retest reliability, with a Pearson correlation coefficient higher than 0.80. This indicates that the instrument yields stable and consistent results, reflecting temporal stability and measurement precision.

In the English AAV patient cohort, each domain showed good measurement properties with Cronbach's alpha between 0.77 and 0.92 and test-retest reliability with an ICC from 0.89 to 0.96 (Fig. 5) (4). The Italian AAV-PRO demonstrated similar internal consistency (Cronbach's alpha 0.81–0.93) and test-retest reliability (ICC 0.94–0.95), but was evaluated using three broader domains instead of the original six (5).

Both BVASv3 and VDI showed weak to moderate correlations with the AAV-PRO (German) domains and total score, suggesting limited and inconsistent associations that may reflect a divergence between physician-reported measures of disease activity and damage and patient-reported outcomes (17). It is important to note that the AAV-PRO (German) does not explicitly assess vasculitis activity in any of its items, nor does it capture disease-related damage. Consistent with our findings, the validation of the Italian AAV-PRO also revealed limited correlations with BVASv3 or VDI (5). Nevertheless, physician-assessed active disease (BVASv3≥1) was reflected in higher domain scores compared to remission, with the difference reaching statistical significance only in the OSS domain (mean difference 3.73; [95% CI 0.67–6.8; p=0.02]). Discrepancies between patient and physician perceptions of disease activity have been documented in other rheumatic diseases. In rheumatoid arthritis, patients tend to evaluate disease activity based on subjective symptoms such as pain, fatigue and functional limitations, whereas physicians primarily rely on objective indicators, including swollen joint counts and inflammatory markers (18). Comparable observations have been made in systemic lupus erythematosus, where patients' assessments incorporate subjective experiences, while physicians focus mainly on objective measures of disease activity (19). The absence of strong correlations between the AAV-PRO (German) and clinician-derived measures underscores the potential of the AAV-PRO (German) as a complementary tool, offering greater emphasis on patient-reported aspects of the disease (20-22). In this context, the integration of PROs represents an important step towards personalised medicine, with the overarching goal of further optimising long-term disease control and minimising adverse effects (23).

Five AAV-PRO (German) domains showed moderate to very strong correlations with the EQ-5D-5L. In contrast, statistical analysis revealed a weak correlation between the OSS domain and the EQ-5D-5L. A possible explanation is that this domain captures organ-specific manifestations not addressed by the generic EQ-5D-5L. As expected, correlations were observed between conceptually related domains, these findings support the construct validity of the AAV-PRO (German). In the English validation study, all correlations between baseline AAV-PRO domains and EQ-5D-5L index scores were large (r≥0.50) (4). In line with our findings, the OSS domain showed the weakest correlation with the EQ-5D-5L index value.

This study has several limitations. Due to the limited sample size, CI were consistently reported to provide a transparent representation of the associated uncertainty.

The restriction to a single-centre setting may limit the generalisability of the results, therefore, the study population was analysed with reference to the Joint Vasculitis Registry in German-speaking countries (GeVas) (24). Gender distribution and median age were comparable and GPA was the most prevalent subtype in both datasets (64.3 % *vs*. 61 %) (24). The relative frequencies of EGPA and MPA differed: in the registry, MPA was second most common (26 %) and EGPA least common (14 %), whereas in the cohort EGPA was second (22.9 %) and MPA least frequent (12.9 %) (24). Overall, these findings indicate that the cohort can only be partially generalised to other German-speaking populations.

Moreover, the possibility of selection bias due to the high participation rate cannot be excluded. Additionally, some missing responses and the fact that not all participants completed the AAV-PRO (German) three times introduce certain biases into the analysis, although the overall rate of missing data was low. In addition, leaving questions unanswered has been reported as an important matter of choice (25, 26). The high proportion of missing BVAS and VDI values reflects practical and organisational constraints in the outpatient setting. Importantly, the missing data occurred completely at random, and our findings remain consistent with other validation studies (5, 21), supporting their strength. As the questionnaires were not consistently completed before or after the physician consultation, a systematic bias in patient responses cannot be ruled out. New medical findings during the consultation may have influenced patients' subjective perception of their disease, potentially affecting their responses and the comparability of results (27). For future studies, a standardised protocol should be established to minimise the risk of such bias.

Conclusion

Our study supports the use of the AAV-PRO (German), showing results comparable to the English and Italian validations and indicating that it can be used by AAV patients to self-assess their quality of life and disease burden. However, low internal consistency in domains TSE and OSS requires careful interpretation. Unlike physician-reported indices, the AAV-PRO (German) reflects the patient's subjective experience of disease activity and thus complements traditional clinical assessments, supporting its use in research and routine patient management.

Acknowledgements

We thank the AGJR for their support of this research project.

References

1. JENNETTE JC, FALK RJ, BACON PA *et al*.: 2012 revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides. *Arthritis Rheum* 2013; 65(1): 1-11. https://doi.org/10.1002/art.37715
2. QUARTUCCIO L, TREPPO E, VALENT F, DE VITA S: Healthcare and economic burden of ANCA-associated vasculitis in Italy: an integrated analysis from clinical and administrative databases. *Intern Emerg Med* 2021; 16(3): 581-89. https://doi.org/10.1007/s11739-020-02431-y
3. HANLY JG: Disease activity, cumulative damage and quality of life in systematic lupus erythematosus: results of a cross-sectional study. *Lupus* 1997; 6(3): 243-47. https://doi.org/10.1177/096120339700600305
4. ROBSON JC, DAWSON J, DOLL H *et al*.: Validation of the ANCA-associated vasculitis patient-reported outcomes (AAV-PRO) questionnaire. *Ann Rheum Dis* 2018; 77(8): 1157-64. https://doi.org/10.1136/annrheumdis-2017-212713
5. TREPPO E, ISOLA M, DE MARTINO M *et al*.: Validation of the Italian version of the ANCA-associated vasculitis patient-reported outcome (AAV-PRO) questionnaire. *Rheumatol Adv Pract* 2024; 8(1): rkae001. https://doi.org/10.1093/rap/rkae001
6. SUPPIAH R, ROBSON JC, GRAYSON PC *et al*.: 2022 American College of Rheumatology/European Alliance of Associations for Rheumatology classification criteria for microscopic polyangiitis. *Ann Rheum Dis* 2022; 81(3): 321-26. https://doi.org/10.1136/annrheumdis-2021-221796
7. ROBSON JC, GRAYSON PC, PONTE C *et al*.: 2022 American College of Rheumatology/European Alliance of Associations for Rheumatology classification criteria for granulomatosis with polyangiitis. *Ann Rheum Dis* 2022; 81(3): 315-20. https://doi.org/10.1136/annrheumdis-2021-221795
8. GRAYSON PC, PONTE C, SUPPIAH R *et al*.: 2022 American College of Rheumatology/European Alliance of Associations for Rheumatology classification criteria for eosinophilic granulomatosis with polyangiitis. *Ann Rheum Dis* 2022; 81(3): 309-14. https://doi.org/10.1136/annrheumdis-2021-221794
9. MUKHTYAR C, LEE R, BROWN D *et al*.: Modification and validation of the Birmingham Vasculitis Activity Score (version 3). *Ann Rheum Dis* 2009; 68(12): 1827-32. https://doi.org/10.1136/ard.2008.101279
10. EXLEY AR, BACON PA, LUQMANI RA *et al*.: Development and initial validation of the Vasculitis Damage Index for the standardized clinical assessment of damage in the systemic vasculitides. *Arthritis Rheum* 1997; 40(2): 371-80. https://doi.org/10.1002/art.1780400222
11. HERDMAN M, GUDEX C, LLOYD A *et al*.: Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011; 20(10): 1727-36. https://doi.org/10.1007/s11136-011-9903-x
12. BASU N, JONES GT, FLUCK N *et al*.: Fatigue: a principal contributor to impaired quality of life in ANCA-associated vasculitis. *Rheumatology (*Oxford) 2010; 49(7): 1383-90. https://doi.org/10.1093/rheumatology/keq098
13. THIELE K, ALBRECHT K, CALLHOFF J, KARBERG K, KÖHLER B, RICHTER J: ABS0752 Fatigue across 12 inflammatory rheumatic diseases: long-term patient documentation from the National database of the German Collaborative Arthritis Centres from 2007 to 2023. *Ann Rheum Dis* 2025; 84: 1526-27. https://doi.org/10.1016/j.ard.2025.06.889
14. ZHU Q, LI F, XIE X *et al*.: Relationship between gender and 1-year mortality in ANCA-associated vasculitis patients: a single-center retrospective analysis and meta-analysis. *Front Med* (Lausanne) 2022; 9: 945011. https://doi.org/10.3389/fmed.2022.945011
15. HERDMAN M, FOX-RUSHBY J, BADIA X: 'Equivalence' and the translation and adaptation of health-related quality of life questionnaires. *Qual Life Res* 1997; 6(3): 237-47. https://doi.org/10.1023/a:1026410721664
16. BEHR D: What to consider and look out for in questionnaire translation. (1st Ed.) Mannheim, GESIS-Leibnitz-Institute for the Social Sciences (GESIS-Survey Guidelines), 2023. https://doi.org/10.15465/gesis-sg_en_043
17. QUARTUCCIO L, TREPPO E, URSO L *et al*.: Unmet needs in ANCA-associated vasculitis: Physicians' and patients' perspectives. *Front Immunol* 2023; 14: 1112899. https://doi.org/10.3389/fimmu.2023.1112899
18. DESTHIEUX C, HERMET A, GRANGER B, FAUTREL B, GOSSEC L: Patient-physician discordance in global assessment in rheumatoid arthritis: a systematic literature review with meta-analysis. *Arthritis Care Res* (Hoboken) 2016; 68(12): 1767-73. https://doi.org/10.1002/acr.22902
19. YEN JC, ABRAHAMOWICZ M, DOBKIN PL, CLARKE AE, BATTISTA RN, FORTIN PR: Determinants of discordance between patients and physicians in their assessment of lupus disease activity. *J Rheumatol* 2003; 30(9): 1967-76.
20. BERTI A, BOLETO G, MERKEL PA *et al*.: Psychometric properties of outcome measurement instruments for ANCA-associated vasculitis: a systematic literature review. *Rheumatology* (Oxford) 2022; 61(12): 4603-18. https://doi.org/10.1093/rheumatology/keac175
21. MAUNZ A, JACOBY J, HENES J, ROBSON JC, HELLMICH B, LOFFLER C: Association of the AAV-PRO questionnaire with established outcome measures in AAV. *Rheumatology* (Oxford) 2024; 63(1): 174-80. https://doi.org/10.1093/rheumatology/kead199
22. TREPPO E, MONTI S, DELVINO P *et al*.: Systemic vasculitis: one year in review 2024. *Clin Exp Rheumatol* 2024; 42(4): 771-81. https://doi.org/10.55563/clinexprheumatol/gkve60
23. DELVINO P, BALDINI C, BONACINI M *et al*.: Systemic vasculitis: one year in review 2025. *Clin Exp Rheumatol* 2025; 43(4): 553-62. https://doi.org/10.55563/clinexprheumatol/oyqz1p
24. ARNOLD S, WALLMEIER P, TAIS A *et al*.: The Joint Vasculitis Registry in German-speaking countries (GeVas): subgroup analysis of 266 AAV patients. *Clin Exp Rheumatol* 2024; 42(4): 852-58. https://doi.org/10.55563/clinexprheumatol/suxkyq
25. BISCHOFF-FERRARI HA, VONDECHEND M, BELLAMY N, THEILER R: Validation and patient acceptance of a computer touch screen version of the WOMAC 3.1 osteoarthritis index. *Ann Rheum Dis* 2005; 64(1): 80-84. https://doi.org/10.1136/ard.2003.019307
26. RICHTER JG, WEISS A, BUNGARTZ C *et al*.: Mobile Responsive App - a useful additional tool for data collection in the German Pregnancy Register Rhekiss? *Front Med* (Lausanne) 2021; 8: 773836. https://doi.org/10.3389/fmed.2021.773836
27. PEETERS MCM, ZWINKELS H, KOEKKOEK JAF, VOS MJ, DIRVEN L, TAPHOORN MJB: The impact of the timing of health-related quality of life assessments on the actual results in glioma patients: a randomized prospective study. *Cancers* (Basel) 2020; 12(8). https://doi.org/10.3390/cancers12082172