






Measurement Properties of Outcome Instruments for Large-Vessel Vasculitis: A Systematic Literature Review

Gonçalo Boleto¹ , Alvise Berti² , Peter A. Merkel³, Sibel Zehra Aydin⁴ , Haner Direskeneli⁵, Christian DeJaco⁶, Leslie C. Hassett⁷ , Loreto Carmona⁸ , and Sofia Ramiro⁹,
for the OMERACT Vasculitis Working Group

ABSTRACT. **Objective.** To systematically review the measurement properties of outcome instruments used in large-vessel vasculitis (LVV).

Methods. MEDLINE, Embase, Cochrane, and Scopus databases were searched for studies published from inception to July 14, 2020, that addressed measurement properties of instruments used in giant cell arteritis (GCA) and Takayasu arteritis (TA). The measurement properties of the instruments identified were collected following the Outcome Measures in Rheumatology (OMERACT) and Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) frameworks. Instruments were grouped according to the following domains measured: disease activity/damage, organ function, and health-related quality of life (HRQOL)/health status.

Results. From 3534 articles identified, 13 met the predefined criteria. These studies addressed 12 instruments: 4 specific to TA, 2 designed for all types of systemic vasculitis, and 6 non-disease-specific instruments. No instruments specific to GCA were identified. Regarding TA, the Indian Takayasu Clinical Activity Score (ITAS) showed very good consistency, adequate reliability, but doubtful validity for disease activity. The Disease Extent Index-Takayasu (DEI-Tak) showed adequate construct validity but doubtful discriminating validity for disease activity/damage. Instruments, including the Vasculitis Damage Index and the Birmingham Vasculitis Activity Score, were poorly assessed for disease activity/damage. In total, 6 non-vasculitis-specific patient-reported outcome (PRO) instruments showed inadequate validity in GCA/TA.

Conclusion. The measurement properties of 12 outcome instruments for LVV covering the OMERACT domains of disease activity/damage, organ function, and HRQOL were assessed. The ITAS and the DEI-Tak were the instruments with the most adequate measurement properties for disease activity/damage in TA. Disease activity/damage instruments specific to GCA, as well as validated PROs for both GCA and TA, are lacking.

Key Indexing Terms: giant cell arteritis, large-vessel vasculitis, measurement properties, OMERACT, outcome measures, Takayasu arteritis

¹G. Boleto, MD, Université de Paris, Hôpital Cochin, Department of Rheumatology, Paris, France; ²A. Berti, Santa Chiara Regional Hospital and Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Department of Rheumatology, Trento, Italy; ³P.A. Merkel, MD, PhD, Division of Rheumatology, and Department of Medicine, Division of Epidemiology, Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, Pennsylvania, USA; ⁴S.Z. Aydin, MD, PhD, University of Ottawa Faculty of Medicine, Department of Rheumatology, Ottawa, Canada; ⁵H. Direskeneli, MD, PhD, Marmara University, Internal Medicine, Division of Rheumatology, Istanbul, Turkey; ⁶C. DeJaco, MD, PhD, Medical University of Graz, Department of Rheumatology, Graz, Austria; ⁷L.C. Hassett, Mayo Clinic Libraries, Mayo Clinic, Rochester, Minnesota, USA; ⁸L. Carmona, Instituto de Salud Musculoesquelética (InMusc), Department of Rheumatology, Madrid, Spain; ⁹S. Ramiro, Leiden University Medical Center, Department of Rheumatology, Leiden, and Zuyderland Medical Center, Rheumatology, Heerlen, the Netherlands.

The authors declare no conflicts of interest relevant to this article.

Address correspondence to Dr. G. Boleto, AP-HP, Department of Rheumatology, Paris Descartes University Cochin Hospital, 27 Rue du Faubourg Saint-Jacques, Paris 75014, France. Email: goncalo.boleto@aphp.fr

Accepted for publication October 12, 2022.

Giant cell arteritis (GCA) and Takayasu arteritis (TA) are 2 major subtypes of chronic, progressive large-vessel vasculitis (LVV) of unknown etiology. In LVV, disease flares are common and disease burden is high.¹⁻⁵ Standardized measurement of LVV outcomes is of the utmost importance for understanding the course of the disease and to measure efficacy of treatment in clinical trials. Not all outcome measurement instruments currently used in LVV research are fully validated, and the study of their measurement performance, as well as the development of new instruments as needed, should be a priority.⁶ In response to this need, in 2016, the Outcome Measures in Rheumatology (OMERACT) Vasculitis Working Group, through an international Delphi exercise, developed a preliminary core set of domains for LVV (ie, GCA and TA).⁷ The group highlighted the importance of having a common set of domains and outcome measurement instruments for GCA and TA, supplemented with disease-specific elements. A draft core set of domains for LVV was proposed, which included pain, fatigue, mortality, organ involvement, arterial function, and biomarkers. Future steps

have been identified, including the formulation of core contextual factors and the formulation of core adverse events. The ultimate goal of the OMERACT Vasculitis Working Group LVV Task Force is to develop an OMERACT-endorsed core set of outcome measures for LVV for use in clinical trials. Over the past several years, more data about the use of outcome measure instruments in LVV have accumulated, and a heterogeneous set of instruments has been used in trials of LVV (ie, instruments assessing different domains and with different performances). A first definition of disease activity in TA was proposed by Kerr et al⁸ based on the presence of constitutional symptoms, new bruits, acute-phase reactants, and angiographic features. More recently, the European Alliance of Associations for Rheumatology (EULAR) proposed new consensus definitions for disease activity in LVV. These included the presence of typical signs or symptoms of active LVV, activity on imaging or biopsy, ischemic complications attributed to LVV, and elevated inflammatory markers.⁴ On the other hand, there is no consensus to define disease damage in LVV. Indeed, according to the report from OMERACT 2018, the disease states in LVV are not well-defined, and the complexity of the disease makes it difficult to differentiate “activity” from “damage.”⁹ Nevertheless, it is accepted that damage consists mainly of the presence of irreversible lesions—stenotic or aneurismatic—that have occurred since the onset of the disease.¹⁰

OMERACT uses a staged process to establish core sets by first establishing the key domains of illness and then identifying validated instruments to assess the domains. Systematic reviews of clinical trials help catalog the outcome measures used and the domains targeted. Groups then seek agreement on the core domains, review the measurement properties of instruments measuring each domain, and hold a final vote on the core set of domains and matched instruments.¹¹ This process aligns with the principles of the Core Outcome Measures in Effectiveness Trials (COMET) initiative¹²; in addition, the process uses the OMERACT filter to critically appraise the instruments identified¹³ and uses a reduced version of the COSMIN (Consensus-Based Standards for the Selection of Health Measurement Instruments) checklist.¹⁴

The EULAR Outcome Measures Library (OML) is an international collaborative initiative that acts as an open access repository of outcome measures in rheumatology.¹⁵ One approach to populate the EULAR-OML is to conduct systematic reviews of existing instruments for any given disease or domain and to appraise their measurement properties. The OML uses the COSMIN checklist to appraise the instruments.

Based on the interest of the vasculitis community to better understand measurement properties of outcome instruments used to measure core domains of vasculitis, this systematic review was designed—in collaboration with the OMERACT Vasculitis Working Group and the EULAR-OML—to evaluate the measurement properties of all available outcome instruments used in LVV.

METHODS

A protocol was registered in the International Prospective Register of

Systematic Reviews (PROSPERO) prior to the initiation of this systematic review (PROSPERO No. CRD42020181949). This review is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.¹⁶

Search strategy, eligibility criteria, and selection process. The research clinical question was formulated according to the Population, Instrument of interest, Measurement properties (PIM) framework of OMERACT.¹⁷ The population included patients with GCA or TA without any age restrictions; the instruments of interest were those measuring disease progression, disease exacerbation, and disease severity indices; those measuring treatment outcomes, physician global assessment, and patient global assessment; and any instrument measuring any of the following domains: disease activity, disease damage, organ function, and health-related quality of life (HRQOL)/health status. The measurement properties of interest were validity, interobserver reliability, intraobserver reliability, sensitivity to change, and feasibility. Studies where miscellaneous types of vasculitis were analyzed and did not provide separate information on GCA or TA were excluded. A comprehensive systematic literature search was undertaken based on the PIM framework from inception of each of the following databases to July 14, 2020: Ovid MEDLINE, Ovid Embase, Ovid Cochrane Central Register of Controlled Trials, Ovid Cochrane Database of Systematic Reviews, and Scopus.

With the supervision of an expert librarian (LCH) and with input from the study's principal investigators, the search strategies for the different databases were generated (see Supplementary Tables S1 and S2 for full details of the search strategy, available from the authors upon request).

Titles and abstracts of the retrieved studies were screened by 2 independent reviewers (GB and AB). The full-text articles were retrieved where abstracts were felt to be relevant. Any duplicate articles were excluded. Reference lists of relevant articles were screened to ensure that no relevant publications were overlooked.

Data extraction. Two reviewers (GB and AB) collected the data independently in predesigned and tested extraction forms. Once the data had been collected, data extraction sheets were compared; these were checked for discrepancies with the original article, if needed, and verified with 2 reviewers experienced in outcome measures (SR and LC) when in doubt.

The data were collected within 2 categories: (1) studies, including design and sample description, validation-related objectives, and risk of bias, and (2) instruments, where information was compiled from several studies. Data elements collected included the following:

1. Study design and population: type (ie, list options), country, diseases (ie, GCA or TA), sample size, ages, and sex distribution.
2. Instrument: type (ie, questionnaires, index, or scales). Biomarkers and imaging instruments were excluded.
3. Practical applications: method of administration; score interpretation; cutpoints; smallest detectable change, if described; completion time by assessor; strengths; and limitations.
4. Instrument measurement properties: validity, reliability, responsiveness, and feasibility (ie, OMERACT Filter of Truth, Feasibility, and Discrimination).¹³ Qualitative evaluations of each measurement property were based on the method by Streiner and Kottner.¹⁸

After reviewing all studies and the properties per instrument, we had meetings to decide, by consensus, the final rating of each property based on the studies and their results. The decisions were systematically based on the results of the studies with lower risk of bias.

We considered Pearson and Spearman correlation values < 0.5 indicative of inadequate validity, values between 0.5 and 0.75 indicative of doubtful validity, values between 0.75 and 0.9 indicative of adequate validity, and values > 0.90 indicative of very good validity. Regarding reliability, we considered intraclass correlation coefficient (ICC) and Cronbach α < 0.5 indicative of inadequate reliability, values between 0.5 and 0.75 indicative of doubtful reliability, values between 0.75 and 0.9 indicative of adequate reliability, and values > 0.90 indicative of very good reliability.¹⁹

Risk of bias (quality) assessment. Risk of bias (ie, quality) for each study was assessed according to the COSMIN checklist.²⁰ The checklist contains 10 boxes. These boxes contain standards for the included measurement properties: patient-reported outcome (PRO) measure development, content validity, structural validity, internal consistency, cross-cultural validity, reliability, measurement error, criterion validity, hypotheses testing, and responsiveness. The studies were evaluated by rating each property, when present, with the following levels of quality: “not available,” “inadequate,” “doubtful,” “adequate,” and “very good.” Risk of bias (ie, quality) for each study was assessed separately and independently by 2 reviewers (GB and AB). When the same instrument was evaluated in different studies, an average rating was given for each measurement property after taking the quality of each study into account. After reviewing all studies and properties per instrument, we had meetings to decide, by consensus, the final rating of each property based on the studies and their results. The decisions were systematically based on the results of the studies with lower risk of bias.

RESULTS

In total, 3534 references were identified in the initial search strategy; the full text of 129 of them were reviewed, of which 13 were included^{21–33} (Supplementary Figure S1 and Table S3, available from the authors upon request). Of these, one was a development study,²¹ 11 were validation studies,^{22–33} and 1 was both a development and validation study.²² All studies included only adults, except for 1 study³² that assessed patients with TA who had a median age of 12 years.

The characteristics of the included studies and the risk of bias (ie, study quality) of each measurement property are shown in Table 1. Baseline characteristics of the populations of the 13 studies included are shown in Supplementary Table S4 (available from the authors upon request). These studies provided information on 12 instruments:

- Four instruments were specific to TA: the Indian Takayasu Clinical Activity Score 2010 (ITAS2010),²² the Indian Takayasu Clinical Activity Score A (ITAS.A),²² the Disease Extent Index-Takayasu (DEI-Tak),²⁴ and the Takayasu Arteritis Damage Score (TADS).^{31,32}
 - Two studies were designed to study all forms of systemic vasculitis: the Vasculitis Damage Index (VDI)²¹ and the Birmingham Vasculitis Activity Score (BVAS).²⁶
 - Six studies were general, non-disease-specific instruments: the Brief Illness Perception Questionnaire (BIPQ),³² the Activities of Daily Vision Scale (ADVS),²⁷ the 36-item Short Form Health Survey (SF-36),^{29,30} the Vision Core Measurement 1 (VCM1),²⁸ the patient global assessment (PtGA) of disease activity,³⁰ and the Multidimensional Fatigue Inventory (MFI).³⁰
- No instruments specific to GCA were identified. A detailed description of each included instrument is shown in Supplementary Table S5 (available from the authors upon request).

Regarding the domains of LVV assessed, 4 instruments evaluated disease activity exclusively (ie, ITAS2010, ITAS.A, BVAS, and PtGA), 2 instruments evaluated disease damage exclusively (ie, VDI and TADS), 1 instrument evaluated both disease activity and disease damage (ie, DEI-Tak), 3 instruments evaluated HRQOL (ie, BIPQ, SF-36, and MFI), and 2 instruments evaluated visual function (ie, ADVS and VCM1).

Measurement properties of the instruments. There was a high degree of heterogeneity in the measurement properties assessed

for each instrument. Validity was the property most frequently assessed among the instruments (11/12, 92%), followed by responsiveness (4/12, 33%). In terms of study quality (ie, risk of bias), among the measurement properties assessed, 34% (11/32) of the instruments had very good quality, 38% (12/32) had adequate quality, 3% (1/32) had doubtful quality, and 25% (8/32) had inadequate quality (Table 1).

Table 2 gives a summary of the adequacy of measurement properties of the instruments identified, and Table 3 gives a detailed overview of the results retrieved.

Disease activity. Regarding TA, the ITAS2010 showed very good internal consistency ($\alpha = 0.97$), doubtful intraobserver reliability (ICC = 0.60), and adequate interobserver reliability (ICC = 0.92). This score showed doubtful construct and discriminating validity, with moderate correlations with the PtGA ($r = 0.73$) and the BVAS ($r = 0.75$) and poor correlation with the National Institutes of Health (NIH) score ($\kappa = 0.35$). The ITAS.A showed doubtful intraobserver reliability (ICC = 0.59) and very good interobserver reliability (ICC = 0.92). The ITAS.A showed very good correlation with the ITAS2010 ($r = 0.98$) and poor correlations with the PtGA ($\kappa = 0.29$) and the NIH score ($\kappa = 0.35$). Responsiveness was, however, poor for both instruments. Validity was poorly tested (ie, few measurement property studies) for the PtGA.

Concerning instruments designed for use in all types of vasculitis, the BVAS showed doubtful construct validity in the assessment of disease activity in GCA (physician global assessment, $r = 0.50$). One PRO instrument—the PtGA—showed doubtful construct validity, with moderate correlations with erythrocyte sedimentation rate and C-reactive protein ($\rho = 0.71$) and poor correlation with the positron emission tomography vascular activity score (PETVAS; $\rho = -0.21$ to -0.32) in GCA/TA.

Among all instruments assessed, the ITAS2010 and ITAS.A performed best, but with a limited assessment of validity and with demonstrated poor responsiveness. However, it is important to realize that no assessment of responsiveness of instruments measuring further disease activity was found in the literature.

Disease damage. In TA, the DEI-Tak showed adequate construct validity ($\kappa = 0.85$ with NIH score). The TADS showed inadequate discriminative validity with poor correlation with disease duration ($r = 0.19$) and poor correlation with cumulative corticosteroid dose ($r = 0.19$).

Concerning instruments designed for use in all types of vasculitis, the VDI showed inadequate discriminating validity (correlation with disease duration, $r = 0.25$; cumulative glucocorticoid dose, $r = 0.19$). Neither reliability nor responsiveness was assessed.

Overall, the DEI-Tak was the instrument with the best measurement properties for disease damage.

HRQOL/health status. In total, 3 nonspecific instruments assessing HRQOL/health status were analyzed. The SF-36 physical component score (PCS) and the SF-36 mental component score (MCS) were poorly correlated with the VDI ($r = -0.34$ and $r = -0.23$, respectively). However, higher VDI values were

Table 1. Characteristics of the 13 studies included and quality of the measurement property assessment (risk of bias).

Study No., Author, Year	Country	Disease	Instrument	Type of Study	Objective	Patients, n ^a	Age, yrs ^a	Women, %	Measurement Property Assessed	Study Quality
1. Exley, 1997 ²¹	UK	GCA	VDI	Longitudinal	Development	7	–	–	Discriminating validity	Doubtful
2. Kermani, 2016 ²⁶	USA	GCA	BVAS	Longitudinal	Validation	136	72 (9)	74	Construct validity	Adequate
3. Kupersmith, 2001 ²⁷	USA	GCA	ADVS	Longitudinal	Validation	20	–	–	Construct validity	Inadequate
	USA	GCA	ADVS	Longitudinal	Validation	20	–	–	Discriminating validity	Inadequate
	USA	GCA	SF-36	Longitudinal	Validation	20	–	–	Discriminating validity	Inadequate
4. Ní Mhéalóid, 2021 ²⁸	Ireland	GCA	VCM1	Longitudinal	Validation	116	70 (9)	64	Construct validity	Very good
	Ireland	GCA	VCM1	Longitudinal	Validation	116	70 (9)	64	Discriminating validity	Very good
	Ireland	GCA	VCM1	Longitudinal	Validation	116	70 (9)	64	Responsiveness	Adequate
	Ireland	GCA	SF-36	Longitudinal	Validation	116	70 (9)	64	Construct validity	Adequate
	Ireland	GCA	SF-36	Longitudinal	Validation	116	70 (9)	64	Discriminating validity	Adequate
5. Misra, 2013 ²²	India	TA	ITAS2010	Longitudinal	Development, validation	177	–	–	Reliability	Very good
	India	TA	ITAS2010/ITAS.A	Longitudinal	Development, validation	177	–	–	Construct validity	Very good
	India	TA	ITAS2010/ITAS.A	Longitudinal	Development, validation	177	–	–	Responsiveness	Very good
6. Alibaz-Oner, 2015 ²³	Turkey	TA	ITAS2010/ITAS.A	Longitudinal	Validation	144	41 (12)	85	Construct validity	Very good
	Turkey	TA	ITAS2010/ITAS.A	Longitudinal	Validation	144	41 (12)	85	Discriminating validity	Very good
	Turkey	TA	ITAS2010/ITAS.A	Longitudinal	Validation	144	41 (12)	85	Responsiveness	Adequate
7. Aydin, 2010 ²⁴	Turkey	TA	DEI.Tak	Cross-sectional	Validation	145	38 (13-73)	79	Construct validity	Very good
	Turkey	TA	DEI.Tak	Cross-sectional	Validation	145	38 (13-73)	79	Discriminating validity	Adequate
8. Fritsch, 2019 ²⁵	Brazil	TA	ITAS2010/ITAS.A	Cross-sectional	Development, validation	42	54 (13)	93	Reliability	Very good
	Brazil	TA	ITAS2010/ITAS.A	Cross-sectional	Development, validation	42	54 (13)	93	Construct validity	Very good
	Brazil	TA	ITAS2010/ITAS.A	Cross-sectional	Development, validation	42	54 (13)	93	Discriminating validity	Very good
9. Omma, 2017 ²⁹	Turkey	TA	VDI	Cross-sectional	Validation	114	35 (13)	89	Discriminating validity	Adequate
10. Kaymaz-Tahra, 2020 ³¹	Turkey	TA	VDI	Longitudinal, validation	Validation	165	33 (12)	89	Discriminating validity	Adequate
	Turkey	TA	TADS	Longitudinal, validation	Validation	165	33 (12)	89	Discriminating validity	Adequate
11. Goel, 2014 ³²	India	TA	ITAS2010	Longitudinal	Validation	40	13 (1-16)	65	Construct validity	Inadequate
	India	TA	DEI.Tak	Longitudinal	Validation	40	13 (1-16)	65	Construct validity	Inadequate
	India	TA	TADS	Longitudinal	Validation	40	13 (1-16)	65	Construct validity	Inadequate
12. Rimland, 2020 ³⁰	USA	GCA, TA	PtGA	Longitudinal	Validation	112	55 (34-71)	80	Construct validity	Adequate
	USA	GCA, TA	SF-36	Longitudinal	Validation	112	55 (34-71)	80	Construct validity	Adequate
	USA	GCA, TA	MFI	Longitudinal	Validation	112	55 (34-71)	80	Responsiveness	Adequate
13. Schwartz, 2020 ³³	USA	GCA	BIPQ	Longitudinal	Validation	47	68 (11)	75	Discriminating validity	Inadequate
		TA	BIPQ	Longitudinal	Validation	47	34 (14)	85	Discriminating validity	Inadequate
	USA	GCA	BIPQ	Longitudinal	Validation	47	68 (11)	75	Responsiveness	Inadequate
		TA	BIPQ	Longitudinal	Validation	47	34 (14)	85	Responsiveness	Inadequate ^a

^aAge is reported as mean (SD) for studies 2, 4, 6, 8, 9, 10, and 13; age is reported as median (IQR) for studies 7, 11, and 12. ADVS: Activities of Daily Vision Scale; BIPQ: Brief Illness Perception Questionnaire; BVAS: Birmingham Vasculitis Activity Score; DEI.Tak: Disease Extent Index-Takayasu; GCA: giant cell arteritis; ITAS2010: Indian Takayasu Clinical Activity Score 2010; ITAS.A: Indian Takayasu Clinical Activity Score A; MFI: Multidimensional Fatigue Inventory; PtGA: patient global assessment; SF-36: 36-item Short Form Health Survey; TA: Takayasu arteritis; TADS: Takayasu Arteritis Damage Score; VCM1: Vision Core Measurement 1; VDI: Vasculitis Damage Index.

detected in patients with PCS values of less than 50. Both the PCS and MCS were not significantly correlated to the PETVAS ($\rho = -0.05$ and $\rho = -0.12$, respectively). The BIPQ was significantly correlated to the MFI, the PtGA, the SF-36 PCS, and the SF-36 MCS ($\rho = 0.50-0.70$, $P < 0.001$), but it did not correlate with the physician global assessment ($\rho = 0.13$, $P = 0.13$). The

MFI was significantly negatively correlated with the PETVAS ($\rho = -0.23$).

Organ function. Visual function was exclusively assessed in GCA through the ADVS and the VMC1. The former showed poor correlation with cumulative steroid dose ($r = 0.45$). The latter showed poor correlations with all of the SF-36 subscales, except

Table 2. Summary of the adequacy of evidence for measurement properties of the LVV instruments identified.

Domain	Instrument	Studies, n	Reliability			Validity				Feasibility	Responsiveness
			Internal Consistency	Intraobserver	Interobserver	Face	Content	Construct	Discriminating		
Disease activity	ITAS2010	3	Very good	Doubtful	Adequate	–	–	Doubtful	Doubtful	–	Inadequate
	ITAS.A	2	–	Doubtful	Very good	–	–	Doubtful	Doubtful	–	Inadequate
	BVAS	1	–	–	–	–	–	Doubtful	–	–	–
	PtGA	1	–	–	–	–	–	Doubtful	–	–	–
Disease activity and damage	DEI-Tak	1	–	–	–	–	–	Adequate	Doubtful	–	–
Disease damage	VDI	3	–	–	–	–	–	–	Very good	–	–
	TADS	2	–	–	–	–	–	–	Inadequate	–	–
HRQOL/health status	SF-36	2	–	–	–	–	–	Inadequate	Inadequate	–	–
	BIPQ	1	–	–	–	–	–	Doubtful	–	–	–
	MFI	1	–	–	–	–	–	–	–	–	Inadequate
	ADVS	1	–	–	–	–	–	–	Inadequate	–	–
Organ function	VCM1	1	–	–	–	–	–	Inadequate	Inadequate	–	–

ADVS: Activities of Daily Vision Scale; BIPQ: Brief Illness Perception Questionnaire; BVAS: Birmingham Vasculitis Activity Score; DEI-Tak: Disease Extent Index-Takayasu; HRQOL: health-related quality of life; ITAS2010: Indian Takayasu Clinical Activity Score 2010; ITAS.A: Indian Takayasu Clinical Activity Score A; LVV: large-vessel vasculitis; MFI: Multidimensional Fatigue Inventory; PtGA: patient global assessment; SF-36: 36-item Short Form Health Survey; TADS: Takayasu Arteritis Damage Score; VCM1: Vision Core Measurement 1; VDI: Vasculitis Damage Index.

bodily pain ($r = -0.22$ to -0.40), and inadequate discriminating validity, with similar median VCM1 scores between GCA (4.0, IQR 1–14.5) and non-GCA groups (2.0, IQR 0.25–8.5).

DISCUSSION

To our knowledge, this is the first systematic review summarizing the measurement properties of instruments developed or validated for LVV. In this study, the measurement properties of 12 outcome measurement instruments for GCA and TA covering the domains of disease activity, damage, organ function, and HRQOL/health status were assessed. The domains identified in this systematic review and endorsed by OMERACT as the core set of outcomes for randomized controlled trials of antineutrophil cytoplasmic antibody-associated vasculitis (AAV)³⁴ have been suggested as potential domains for future clinical investigation in LVV.^{7,35} Our study identified the ITAS2010, ITAS.A, and DEI-Tak as the instruments with the most adequate measurement properties for disease activity and/or damage in TA and, therefore, could be recommended for research and/or clinical practice.

In TA, despite the identification of specific outcome measurement instruments for disease activity (ie, ITAS2010 and ITAS.A) and disease damage (ie, DEI-Tak) with adequate measurement properties, a combination of clinical symptoms, acute-phase reactants, imaging and glucocorticoid-sparing effects, as well as other composite scores (eg, NIH score) were still largely used in recent clinical trials.^{36–39} However, since some of the identified tools (ie, ITAS2010, ITAS.A, and DEI-Tak) have been developed and validated by the same group, one should be cautious when interpreting the results.

Contrary to TA, no instruments measuring disease activity or damage specific for GCA were identified. The OMERACT 2016 workshop cited the NIH score, the BVAS, the DEI-Tak, and the ITAS2010 as the main outcome instruments used in

clinical research for TA and highlighted that similar disease-specific tools do not exist for GCA.⁷ The BVAS or a combination of clinical symptoms, acute-phase reactants, glucocorticoid dose, and duration and imaging have been used in previous GCA clinical trials.^{40–43} This systematic review identified 1 study that evaluated the performance of the BVAS in the assessment of GCA disease activity.²⁶ This study revealed a substantial limited utility for the use of the BVAS in GCA, with a considerable number of patients (11%) with active disease having a BVAS of 0. Moreover, ischemic symptoms secondary to vasculitis are not included in the BVAS. This contrasts with the good performance of the BVAS in AAV reported in the literature, which might be related to the fact that this tool was better designed and validated in small-vessel vasculitis than in LVV.⁴⁴

Regarding PROs reflecting disease activity, only the widely used generic PtGA was identified in this review. This instrument was used in both GCA and TA trials; however, measurement properties were insufficiently assessed for this instrument. Although patients' evaluation of disease activity/damage is usually a difficult goal to reach, developing a composite measure combining the perspectives of patients and physicians, as has been done for other systemic inflammatory diseases,⁴⁵ could improve the evaluation of LVV.

There has been growing interest in the importance of integrating patient perspectives regarding the effects of their disease, and this has been proposed by OMERACT as a mandatory area to be assessed in LVV clinical trials. The measurement properties of instruments measuring HRQOL were assessed using generic instruments not specifically validated for LVV. We acknowledge that HRQOL is not a domain that is part of the LVV preliminary/draft core domains, but as we were aiming at being inclusive in this systematic literature review, we collected all data on all existing instruments and matched them to the OMERACT LVV draft core domains, where possible. The instruments

Table 3. Summary of the results on the psychometric properties of the 12 LVV instruments identified.

Instrument	Items, n	Range of Scores	Reliability		Responsiveness	Construct validity	Discriminating Validity	Overall Rating for a Good Result
			Internal Consistency, α (95% CI)	Intraobserver, ICC (95% CI)	Interobserver, ICC (95% CI)			
ITAS2010	44	0-51	0.97 (0.95-0.99)	0.60 (0.29-0.79)	0.92 (0.83-0.96)	<ul style="list-style-type: none"> PGA: $r = 0.73$ BVAS: $r = 0.75$ NIH score: 82.8%, $\kappa = 0.35$ 	Significantly higher in clinically active disease assessed by PGA (mean 1.27, SD 1.26; $P < 0.001$)	Sufficient (+)
ITAS-A	47	0-54	-	0.59 (0.29-0.79)	0.92 (0.83-0.96)	<ul style="list-style-type: none"> ITAS2010: $r = 0.98$ ITAS-CRP: $r = 0.97$ PGA: 67%, $\kappa = 0.29$ NIH score: 86.3%, $\kappa = 0.35$ 	ITAS-ESR and ITAS-CRP scores were both significantly higher with active disease ($P = 0.001$)	Sufficient (+)
BVAS	66	0-66	-	-	-	PGA past 28 days: <ul style="list-style-type: none"> BVAS1: $r = 0.50$ BVAS2: $r = -0.20$ PGA today: <ul style="list-style-type: none"> BVAS1: $r = 0.01$ BVAS2: $r = -0.08$ ESR: <ul style="list-style-type: none"> BVAS1: $r = 0.02$ BVAS2: $r = -0.12$ CRP: <ul style="list-style-type: none"> BVAS1: $r = 0.02$ BVAS2: $r = -0.01$ Disease activity past 28 days: <ul style="list-style-type: none"> BVAS1: $r = 0.46$ BVAS2: $r = -0.18$ Disease activity today: <ul style="list-style-type: none"> BVAS1: $r = 0.05$ BVAS1: $r = -0.04$ 	-	Insufficient (-)
PtGA	1	0-10	-	-	-	PtGA, MFI, BIPQ, Neg SF-36 PCS, and Neg SF-36 MCS: <ul style="list-style-type: none"> $\rho = 0.35$-0.60 ESR and CRP: $\rho = 0.71$ <ul style="list-style-type: none"> PtGA standardized β estimate: 0.59 PETVAS: $\rho = -0.21$ to -0.32 	-	Insufficient (-)
DEL-Tak	59	-	-	-	-	ESR: $r = 0.17$ <ul style="list-style-type: none"> NIH score: 94%, $\kappa = 0.85$ 	Higher DEL-Tak in active disease (mean 1.2, SD 1.4) vs inactive disease (mean 0.2, SD 0.6) as per PGA ($P < 0.003$)	Sufficient (+)

Table 3. Continued.

Instrument	Items, n	Range of Scores	Reliability		Responsiveness	Construct validity	Discriminating Validity	Overall Rating for a Good Result
			Internal Consistency, α (95% CI)	Intraobserver, ICC (95% CI)				
TADS	42	-	-	-	-	-	Differences between groups: <ul style="list-style-type: none">With and without relapsing disease: $P = 0.89$With and without peripheral involvement (pulse loss and bruit) at baseline (96% vs 55.6%; $P = 0.01$) Correlation between TADS and the following: <ul style="list-style-type: none">Disease duration: $r = 0.19$Cumulative CS dose: $r = 0.19$	Insufficient (-)
VDI	63	0-63	-	-	-	-	Correlation between VDI and the following: <ul style="list-style-type: none">Disease duration: $r = 0.25$Cumulative CS dose: $r = 0.19$ Other measures: <ul style="list-style-type: none">Cumulative dose of GC in active disease (median 11 g, IQR 0-45) vs inactive disease (median 5.7 g, IQR 0-43); $P = 0.006$Disease resistance: median 1.362 (IQR 0.708-2.017)	Insufficient (-)
ADVS	25	-	-	-	-	<ul style="list-style-type: none">Worse eye acuity: $r = 0.53$Worse eye central impairment: $r = -0.54$% Binocular impairment: $r = -0.59$Better eye acuity: $r = 0.62$% Binocular impairment at 1 year: $r = -0.52$Cumulative steroid dose: $r = 0.45$	-	Insufficient (-)
SF-36	36	0-100	-	-	-	PCS: <ul style="list-style-type: none">Age: $r = -0.20$ (95% CI -0.35 to -0.05)VDI: -1.60 (95% CI -2.45 to -0.76) MCS: <ul style="list-style-type: none">VDI: -0.92 (95% CI -1.72 to -0.12)	Correlations with the following: <ul style="list-style-type: none">VDI: $r = -0.23$ (MCS) and $r = -0.34$ (PCS)PGA: $\rho = 0.18$ (PCS) and $\rho = 0.09$ (MCS)CRP: $\rho = 0.07$ (PCS) and $\rho = 0.09$ (MCS)PETVAS: $\rho = -0.05$ (PCS) and $\rho = -0.12$ (MCS) Higher VDI scores in patients with PCS < 50 (median 5, IQR 1-12) vs ≥ 50 (median 2, IQR 1-6)	Insufficient (-)

Table 3. Continued.

Instrument	Items, n	Range of Scores	Reliability		Responsiveness	Construct validity	Discriminating Validity	Overall Rating for a Good Result
			Internal Consistency, α (95% CI)	Intraobserver, ICC (95% CI)				
VCM1	10	0.0-5.0	-	-	-	<ul style="list-style-type: none"> Worse eye VA: $r = 0.42$ SF-36 subscales, except bodily pain: $r = -0.22$ to -0.40 PCS: $r = -0.69$ MCS: $r = -0.92$ 	VCM1 score ($P = 0.431$): GCA: median 4.0 (IQR 1-14.5) non-GCA: median 2.0 (IQR 0.25-8.5)	Insufficient (-)
MFI	-	-	-	-	-	<ul style="list-style-type: none"> Correlation with CRP: $\rho = 0.07$ Correlation with PETVAS: $\rho = -0.23$ 		Insufficient (-)

ADVS: Activities of Daily Vision Scale; BIPQ: Brief Illness Perception Questionnaire; BVAS: Birmingham Vasculitis Activity Score; CRP: C-reactive protein; DEI-Tak: Disease Extent Index-Takayasu; ESR: erythrocyte sedimentation rate; GC: ICC: intraclass correlation coefficient; ITAS: Indian Takayasu Clinical Activity Score; ITAS2010: Indian Takayasu Clinical Activity Score 2010; ITAS-A: Indian Takayasu Clinical Activity Score A; LVV: large-vessel vasculitis; MCS: mental component score; MFI: Multidimensional Fatigue Inventory; Neg: negative; NIH: National Institutes of Health; PCS: physical component score; PETVAS: positron emission tomography vascular activity score; PGA: physician global assessment; PtGA: patient global assessment; SF-36: 36-item Short Form Health Survey; TADS: Takayasu Arteritis Damage Score; VA: visual assessment; VCM1: Vision Core Measurement 1; VDI: Vasculitis Damage Index.

measuring HRQOL are, therefore, an example of instruments that do not cover any of the OMERACT domains, so HRQOL is a core theme of those instruments but not an OMERACT domain. Indeed, HRQOL has not been investigated extensively in patients with LVV, and whether generic PROs are sensitive to change in GCA and TA has not been demonstrated. Although the SF-36 does not strictly measure HRQOL, but is rather an indicator of overall health status, and since we believe that HRQOL is intimately related to health status, we decided to include this instrument in our analysis. Indeed, the SF-36 is a widely used generic PRO that covers 8 different domains, including physical and social functioning and mental health,⁴⁶ and has been widely used in trials of rheumatic and musculoskeletal diseases^{47,48} and AAV.³⁴ However, this review demonstrated that there is currently inadequate evidence to suggest using the SF-36 as a generic instrument in LVV clinical trials. Further, the ADVS and the VCM1 have been used for subjectively evaluating visual function in GCA but with inadequate construct and discriminating validity according to our analysis.

Surprisingly, the number of studies identified as assessing measurement properties of instruments was limited. This might be a consequence of the selection criteria chosen, which excluded some studies where miscellaneous types of vasculitis were analyzed and did not provide separate information on GCA or TA. This review opted for the most conservative and cleanest approach to specifically collect data on the measurement properties of GCA and TA measurement instruments. Another limitation of this systematic review is the heterogeneity of the included studies, with no hierarchy or a settled format of properties required to properly validate a specific instrument, as well as the heterogeneity in which similar properties were assessed, limiting a direct comparability of the performance of different instruments. When assessing construct validity, as opposed to recommendations, most studies did not specifically report, or did not formulate, a priori hypotheses. One should also emphasize that most of the included studies reported on 1 or 2 measurement properties only, making it difficult to have an overall assessment of the instrument. It is important to notice that reliability or validity are not fixed properties of a scale, and they depend on the testing situation. Indeed, these properties are limited to the results obtained with an evaluation instrument and not to the instrument itself.¹⁸

In addition, it is challenging to directly compare the measurement properties of different instruments, since the comparators are not always the same. Therefore, we collected and reported all the data in the papers in a systematic fashion to provide an overview of what has been published, avoiding direct or head-to-head comparisons.

Finally, even though OMERACT and COSMIN frameworks might function as the backbone of these assessments, the heterogeneity, together with the scarcity of the studies, led to limited available evidence on the measurement properties of the instruments analyzed.

In conclusion, this systematic review demonstrated that specific tools for the assessment of outcome domains in LVV are lacking, particularly for GCA. GCA and TA are both very rare

conditions; distinguishing and separately measuring damage vs activity, which are often related, is not straightforward. With recent advances in imaging, incorporating composite scores might help to better assess these conditions. Moreover, most of the instruments included in the analysis were only partially validated. Our systematic review also highlighted the need for specific PROs to evaluate both disease activity/damage and HRQOL for both GCA and TA.

REFERENCES

- Hunder GG, Bloch DA, Michel BA, et al. The American College of Rheumatology 1990 criteria for the classification of giant cell arteritis. *Arthritis Rheum* 1990;33:1122-8.
- Gulati A, Bagga A. Large vessel vasculitis. *Pediatr Nephrol* 2010;25:1037-48.
- Nesher G. The diagnosis and classification of giant cell arteritis. *J Autoimmun* 2014;48-49:73-5.
- Hellmich B, Agueda A, Monti S, et al. 2018 Update of the EULAR recommendations for the management of large vessel vasculitis. *Ann Rheum Dis* 2020;79:19-30.
- Quinn KA, Gribbons KB, Carette S, et al. Patterns of clinical presentation in Takayasu's arteritis. *Semin Arthritis Rheum* 2020;50:576-81.
- Aydin SZ, Direskeneli H, Merkel PA, International Delphi on Disease Activity Assessment in Large-vessel Vasculitis. Assessment of disease activity in large-vessel vasculitis: results of an international Delphi exercise. *J Rheumatol* 2017;44:1928-32.
- Sreih AG, Alibaz-Oner F, Kermani TA, et al. Development of a core set of outcome measures for large-vessel vasculitis: report from OMERACT 2016. *J Rheumatol* 2017;44:1933-7.
- Kerr GS, Hallahan CW, Giordano J, et al. Takayasu arteritis. *Ann Intern Med* 1994;120:919-29.
- Aydin SZ, Robson JC, Sreih AG, et al. Update on outcome measure development in large-vessel vasculitis: report from OMERACT 2018. *J Rheumatol* 2019;46:1198-201.
- Cimmino MA. SP0151 Clinical diagnosis of large vessel vasculitis [abstract]. *Ann Rheum Dis* 2016;75: 37.
- Aydin SZ, Direskeneli H, Sreih A, et al. Update on outcome measure development for large vessel vasculitis: report from OMERACT 12. *J Rheumatol* 2015;42:2465-9.
- Prinsen CAC, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline. *Trials* 2016;17:449.
- Beaton DE, Maxwell LJ, Shea BJ, et al. Instrument selection using the OMERACT Filter 2.1: the OMERACT methodology. *J Rheumatol* 2019;46:1028-35.
- Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539-49.
- Castrejon I, Carmona L, Agrinier N, et al. The EULAR Outcome Measures Library: development and an example from a systematic review for systemic lupus erythematosus instruments. *Clin Exp Rheumatol* 2015;33:910-6.
- Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* 2021;372:n160.
- Boers M, Beaton DE, Shea BJ, et al. OMERACT Filter 2.1: elaboration of the conceptual framework for outcome measurement in health intervention studies. *J Rheumatol* 2019;46:1021-7.
- Streiner DL, Kottner J. Recommendations for reporting the results of studies of instrument and scale development and testing. *J Adv Nurs* 2014;70:1970-9.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155-63.
- Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1171-9.
- 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;40:371-80.
- Misra R, Danda D, Rajappa SM, et al. Development and initial validation of the Indian Takayasu Clinical Activity Score (ITAS2010). *Rheumatology* 2013;52:1795-801.
- Alibaz-Oner F, Aydin SZ, Akar S, et al. Assessment of patients with Takayasu arteritis in routine practice with Indian Takayasu Clinical Activity Score. *J Rheumatol* 2015;42:1443-7.
- Aydin SZ, Yilmaz N, Akar S, et al. Assessment of disease activity and progression in Takayasu's arteritis with Disease Extent Index-Takayasu. *Rheumatology* 2010;49:1889-93.
- Fritsch S, Copes RM, Savioli B, et al. Translation and validation of the Indian Takayasu Clinical Activity Score (ITAS2010) for the Brazilian Portuguese language. *Adv Rheumatol* 2019;59:43.
- Kermani TA, Cuthbertson D, Carette S, et al. The Birmingham Vasculitis Activity Score as a measure of disease activity in patients with giant cell arteritis. *J Rheumatol* 2016;43:1078-84.
- Kupersmith MJ, Speira R, Langer R, et al. Visual function and quality of life among patients with giant cell (temporal) arteritis. *J Neuroophthalmol* 2001;21:266-73.
- Ní Mhéalóid Á, Conway R, O'Neill L, Clyne B, Molloy E, Murphy CC. Vision-related and health-related quality of life in patients with giant cell arteritis. *Eur J Ophthalmol* 2021;31:727-33.
- Omma A, Erer B, Karadag O, et al. Remarkable damage along with poor quality of life in Takayasu arteritis: cross-sectional results of a long-term followed-up multicentre cohort. *Clin Exp Rheumatol* 2017;35 Suppl 103:77-82.
- Rimland CA, Quinn KA, Rosenblum JS, et al. Outcome measures in large vessel vasculitis: relationship between patient-, physician-, imaging-, and laboratory-based assessments. *Arthritis Care Res* 2020;72:1296-304.
- Kaymaz-Tahra S, Alibaz-Oner F, Direskeneli H. Assessment of damage in Takayasu's arteritis. *Semin Arthritis Rheum* 2020; 50:586-91.
- Goel R, Kumar TS, Danda D, et al. Childhood-onset Takayasu arteritis -- experience from a tertiary care center in South India. *J Rheumatol* 2014;41:1183-9.
- Schwartz MN, Rimland CA, Quinn KA, et al. Utility of the Brief Illness Perception Questionnaire to monitor patient beliefs in systemic vasculitis. *J Rheumatol* 2020;47:1785-92.
- Merkel PA, Aydin SZ, Boers M, et al. The OMERACT core set of outcome measures for use in clinical trials of ANCA-associated vasculitis. *J Rheumatol* 2011;38:1480-6.
- Direskeneli H, Aydin SZ, Kermani TA, et al. Development of outcome measures for large-vessel vasculitis for use in clinical trials: opportunities, challenges, and research agenda. *J Rheumatol* 2011;38:1471-9.
- Mertz P, Kleinmann J-F, Lambert M, et al. Infliximab is an effective glucocorticoid-sparing treatment for Takayasu arteritis: results of a multicenter open-label prospective study. *Autoimmun Rev* 2020;19:102634.
- Nakaoka Y, Isobe M, Tanaka Y, et al. Long-term efficacy and safety of tocilizumab in refractory Takayasu arteritis: final results of the randomized controlled phase 3 TAKT study. *Rheumatology* 2020;59:2427-34.
- Régner P, Le Joncour A, Maciejewski-Duval A, et al. Targeting JAK/STAT pathway in Takayasu's arteritis. *Ann Rheum Dis* 2020;79:951-9.

39. Mekinian A, Saadoun D, Vicaud E, et al. Tocilizumab in treatment-naïve patients with Takayasu arteritis: TOCITAKA French prospective multicenter open-labeled trial. *Arthritis Res Ther* 2020;22:218.
40. De Silva M, Hazleman BL. Azathioprine in giant cell arteritis/polymyalgia rheumatica: a double-blind study. *Ann Rheum Dis* 1986;45:136-8.
41. Seror R, Baron G, Hachulla E, et al. Adalimumab for steroid sparing in patients with giant-cell arteritis: results of a multicentre randomised controlled trial. *Ann Rheum Dis* 2014;73:2074-81.
42. Villiger PM, Adler S, Kuchen S, et al. Tocilizumab for induction and maintenance of remission in giant cell arteritis: a phase 2, randomised, double-blind, placebo-controlled trial. *Lancet* 2016;387:1921-7.
43. Stone JH, Tuckwell K, Dimonaco S, et al. Trial of tocilizumab in giant-cell arteritis. *N Engl J Med* 2017;377:317-28.
44. Merkel PA, Aydin SZ, Boers M, et al. Current status of outcome measure development in vasculitis. *J Rheumatol* 2014;41:593-8.
45. Landewé RBM, van der Heijde D. Use of multidimensional composite scores in rheumatology: parsimony versus subtlety. *Ann Rheum Dis* 2021;80:280-5.
46. Ware Jr JE, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
47. Russak SM, Croft Jr JD, Furst DE, et al. The use of rheumatoid arthritis health-related quality of life patient questionnaires in clinical practice: lessons learned. *Arthritis Rheum* 2003;49:574-84.
48. Salaffi F, Carotti M, Gasparini S, Intorcchia M, Grassi W. The health-related quality of life in rheumatoid arthritis, ankylosing spondylitis, and psoriatic arthritis: a comparison with a selected sample of healthy people. *Health Qual Life Outcomes* 2009;7:25.