External validation of the giant cell arteritis probability score in the Netherlands

L.M. Neuman[1], M. van Nieuwland[1], M. Vermeer[2], D. Boumans[1], E.M. Colin[1], C. Alves[1]

[1]Ziekenhuisgroep Twente, Hospital Group Twente, Department of Rheumatology, Almelo/Hengelo; [2]Ziekenhuisgroep Twente, Hospital Group Twente, ZGT Academy, Almelo, The Netherlands.

Abstract

Objective

To prevent complications of giant cell arteritis (GCA), early and accurate diagnosis is essential. Recently, Laskou et al. (2019) developed the giant cell arteritis probability score (GCAPS) which allows physicians to assess the likelihood of GCA at an early stage. The aim of this study was to validate the GCAPS in a Dutch hospital.

Methods

A retrospective cohort of patients with suspected GCA between January 1st, 2017 and October 1st, 2019 was used. As the variable extra-cranial artery abnormality was not measured, a modified GCAPS was used (m-GCAPS). Clinical diagnosis of the rheumatologist after six months follow-up was used as reference. The m-GCAPS was assessed for discrimination and calibration. We applied risk stratification according to Sebastian et al. (2020) (low, intermediate and high-risk groups based on the median and 75th percentile).

Results

Our study included 209 suspected GCA patients. 135 patients had complete records. Of these patients, 40 had GCA. The m-GCAPS had an area under the curve of 0.83, a sensitivity of 80.0% and specificity of 75.8% at the optimal cut-off value >10.5. The Hosmer-Lemeshow test was non-significant. Using risk stratification, GCA prevalence was 12.5% in the low (score<10), 23.3% in the intermediate (10-14) and 78.6% in the high-risk group (>14).

Conclusion

The m-GCAPS showed good discrimination and calibration in a Dutch retrospective cohort and can aid early recognition of GCA. Stratification into low, intermediate and high-risk is promising, but might need optimisation.

Key words

giant cell arteritis, temporal arteritis, validation studies, early diagnosis, diagnostic test.

Lize M. Neuman, MD
Marieke van Nieuwland, MSc
Marloes Vermeer, PhD
Dennis Boumans, MD
Edgar M. Colin, MD, PhD
Celina Alves, MD, PhD

Please address correspondence to:
Celina Alves,
Department of Rheumatology,
Hospital Group Twente,
Zilvermeeuw 1, 7609 PP Almelo,
The Netherlands.
E-mail: c.alves@zgt.nl

Received on April 2, 2021; accepted in revised form on July 9, 2021.

Competing interests: none declared.

Introduction

Giant cell arteritis (GCA) is a vasculitis of the medium and large sized arteries and may result in severe complications, such as blindness and stroke (1). Rapid initiation of glucocorticoid (GC) treatment reduces the risk of these complications. However, many patients (86%) suffer from GC induced side effects such as diabetes mellitus, increased infection risk and osteoporosis (2). Early and accurate diagnosis is necessary to prevent GCA-related complications in GCA patients and avoid GC related side effects in those without GCA.

To date, early and accurate diagnosis of GCA is difficult due to its low incidence, generic symptoms, non-distinctive blood tests and suboptimal sensitivity (39%) of the (traditionally) most commonly used diagnostic test: temporal artery biopsy (TAB) (1, 3). A viable alternative to biopsy is colour Doppler ultrasound (CDUS) which is inexpensive, real-time and non-invasive (3). Current European League Against Rheumatism recommendations identify CDUS as first-choice diagnostic test to either confirm or exclude GCA in high and low pre-test probability cases respectively (4).

Recently Laskou *et al*. (2019) developed a tool that may aid in quantification of this pre-test probability: the giant cell arteritis probability score (GCAPS) (5). In Southend University Hospital (United Kingdom), the GCAPS had good diagnostic value and was promising in the stratification of suspected GCA patients into low, intermediate and high-risk groups (5, 6). However, external validation of the GCAPS is needed before application in clinical practice is possible (7).

The aim of the present study was to validate the GCAPS in a retrospective cohort of patients with suspected GCA in a Dutch general hospital.

Methods

Cohort

This retrospective cohort study was conducted at the Rheumatology Department of Hospital Group Twente (ZGT), the Netherlands. GCA suspected patients between January 1st, 2017 and October 1st, 2019 were included. All patients were referred by a general practitioner or medical specialist.

GCAPS

The GCAPS was generated on the basis of expert opinion, supported by relevant literature (5). It includes 17 variables divided over 5 domains: patients' demographic characteristics, symptoms, signs, laboratory markers and (possible) alternative diagnosis at presentation.

Data collection

The variables of the GCAPS were collected from the patients' medical records at presentation. One variable, extra-cranial artery abnormality, was not noted in any of the records. Therefore, a modified version of the GCAPS was used, *i.e*. m-GCAPS (GCAPS without extra-cranial artery abnormality).

Reference diagnosis

The reference diagnosis was the rheumatologists' clinical diagnosis six months after initial assessment, similar to the studies of Laskou *et al*. and Luqmani *et al*. (3, 5).

Statistical analysis

The baseline characteristics are described as mean with standard deviation (SD) or as number with corresponding percentages. Differences in the variables of the m-GCAPS between patients with and without GCA were tested using Chi-square tests (or Fisher's exact tests when appropriate).

The validity of the m-GCAPS was assessed using the area under the curve (AUC) of the receiver operating characteristic curve (ROC) for discrimination and the Hosmer-Lemeshow test for calibration. Sensitivity and specificity were calculated for each sum score of the m-GCAPS. To minimalise overtreatment and undertreatment, the optimal cut-off value was considered as the score at which the product of sensitivity and specificity was optimal. Risk stratification of Sebastian *et al*. (2020) was applied, based on the median and 75th percentile scores (*i.e*. low risk <9, intermediate risk 9-12, high risk >12 group) (6). To compare the accuracy of risk stratification, the median

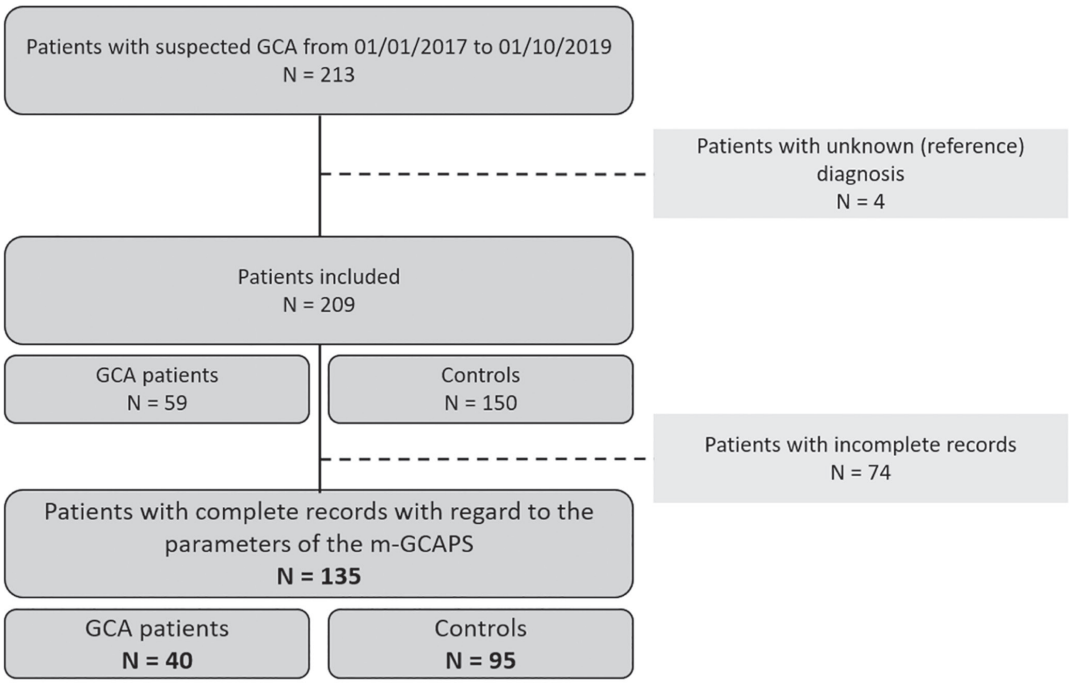

Fig. 1. Flowchart inclusion and exclusion of patients.

and 75th percentile of the m-GCAPS (in our data) were calculated.

Complete case analysis was made. Multiple imputation was applied by Fully Conditional Specification (iterative Markov Chain Monte Carlo method) with ten imputations and results were pooled using Rubin's Rules (8).

All statistical analyses were carried out in SPSS (Inc., Chicago VS), version 24. A $p<0.05$ was considered statistically significant.

Ethical approval

The study is limited to retrospective use of data previously collected during normal clinical care with no patient identifier recorded. Therefore, this study did not require Research Ethics Committee review or formal patient consent. The study complies with the Declaration of Helsinki.

Results

Baseline characteristics

In total 213 suspected GCA patients were eligible for inclusion (Fig. 1). Four patients were excluded because follow-up data was missing. The mean (SD) age in the study population (n=209) was 70.0 (12.0) years and 116 (55.5%) patients were female.

After six months of follow-up, 59 (26.2%) patients had a clinical GCA diagnosis. Fifty-five of these 59 patients had a positive CDUS, TAB or 18F-FDG positron emission tomography scan. The remaining four patients were diagnosed based on high clinical suspicion, positive response to steroids and lack of other diagnosis.

The baseline values on the variables of the m-GCAPS in patients with and without GCA are summarised in Table I. In total 40 patients with and 95 patients without the clinical diagnosis of GCA had complete data with regard to the m-GCAPS variables.

Patients without GCA were diagnosed with one or more of the following: infection (n=16), head or neck pathology (n=34), malignancy (n=4), polymy-

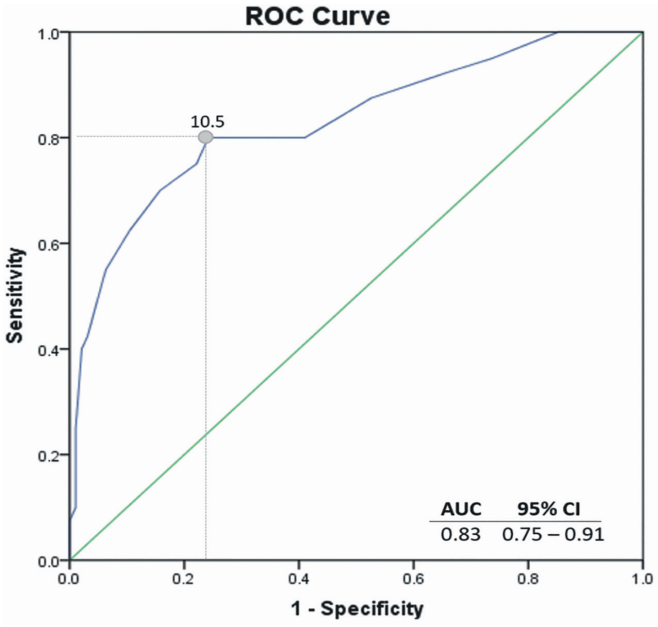

Fig. 2. Receiver operating characteristic-curve (ROC) of the m-GCAPS with the area under the curve (AUC) and the proposed cut-off value of 10.5.
ROC: receiver operating characteristic; AUC: area under the curve; CI: confidence interval.

Table I. Characteristics of patients with and without GCA on the variables of the m-GCAPS.

	Patients with GCA[a] n = 59	Patients without GCA[a] n = 150	*p*-value
Demographics			
Female; n (%)	40 (67.8)	76 (50.7)	0.03
Age (years); n (%)			0.08
≤49	1 (1.7)	6 (4.0)	
50 – 60	6 (10.2)	36 (24.0)	
61 – 65	6 (10.2)	10 (6.7)	
≥66	46 (78.0)	98 (65.3)	
Onset			
Time since onset of symptoms[b] (weeks); n (%)			0.20
<6	35 (61.4)	83 (56.5)	
6 – 11	10 (17.5)	15 (10.2)	
12 – 23	7 (12.3)	22 (15.0)	
≥24	5 (8.8)	27 (18.4)	
data missing	*2*	*3*	
Symptoms			
Cranial symptoms[c]; n (%)			0.91
present	46 (82.1)	114 (81.4)	
absent	10 (17.9)	26 (18.6)	
data missing	*3*	*10*	
Polymyalgia rheumatica[d]; n (%)			0.75
present	21 (42.9)	49 (40.2)	
absent	28 (57.1)	73 (59.8)	
data missing	*10*	*28*	
Constitutional symptoms[e]; n (%)			0.02
none	35 (59.3)	110 (73.3)	
one	14 (23.7)	32 (21.3)	
more than one	10 (16.9)	8 (5.3)	
Ischaemic symptoms[f]; n (%)			0.005
present	29 (50.9)	43 (29.9)	
absent	28 (49.1)	101 (70.1)	
data missing	*2*	*6*	
Signs			
Visual impairment (caused by AION, CRAO or RAPD or diplopia); n (%)	14 (23.7)	21 (14.0)	0.10
Temporal artery abnormality[g]; n (%) – *data missing*			
tenderness	17 (31.5) – *5*	25 (19.7) – *23*	0.02
thickening	25 (54.3) – *13*	6 (5.8) – *46*	<0.001
pulse loss	5 (10.6) – *12*	3 (2.5) – *29*	0.02
thickening and tenderness	13 (24.1) – *5*	2 (1.5) – *18*	<0.001
pulse loss, tenderness and thickening	3 (5.5) – *4*	0 (0.0) – *12*	0.02
Cranial nerve palsy; n (%)	0 (0.0)	2 (1.3)	1.0
Laboratory			
CRP; n (%)			<0.001
0 – 5	3 (5.2)	59 (41.0)	
>5 – 10	2 (3.4)	11 (7.6)	
>10 – 25	5 (8.6)	21 (14.6)	
>25	48 (82.8)	53 (36.8)	
data missing	*1*	*6*	
Alternative diagnosis at presentation			
Presence alternative diagnosis[h]; n (%)	20 (33.9)	114 (76.0)	<0.001

Percentages are calculated on the basis of known values only.
[a]Clinical diagnosis of the rheumatologist after six months follow-up; [b]Time between onset of symptoms and first visit to the rheumatology department; [c]Headache or scalp pain; [d]Pain and stiffness in shoulder- or hipgirdle; [e]Fever (temperature > 38.5°C), night sweats or weight loss; [f]Jaw claudication or unilateral diplopia, blurring of vision or sight loss permanent or transient; [g]At least one side; [h]Determined by the rheumatologist (after ultrasound).
AION: anterior ischaemic optic neuropathy; CRAO: central retinal arterial occlusion; CRP: C-reactive protein; GCA: giant cell arteritis; RAPD: relative afferent pupil defect.

algia rheumatica (n=31) or other illnesses (n=25). In 55 patients, the exact cause of their symptoms was unknown (in our department).

Validity

The m-GCAPS had an AUC of 0.83 (95% CI: 0.75 – 0.91) in our retrospective cohort (Fig. 2). Since a modified GCAPS was used, the optimal cut-off value was re-assessed. This resulted in a cut-off value of 10.5 with a sensitivity of 80.0% and specificity of 75.8% (Table II). The m-GCAPS revealed good fit between observed and predicted values as the Hosmer-Lemeshow test was not statistically significant (p=0.41). Multiple imputation resulted in a pooled AUC of 0.84 (95% CI: 0.77-0.90).

Risk stratification

Using stratification scores proposed by Sebastian *et al*., GCA prevalence was 10.0% in the low (score<9), 16.7% in the intermediate (9-12) and 65.1% in the high (>12) risk group (6). By applying the median and 75th percentile of our data in risk stratification, the GCA prevalence was 12.5% in the low (score<10), 23.3% in the intermediate (10-14) and 78.6% in the high (>14) risk group (Table III).

Discussion

To our knowledge, this is the first study that has externally validated the GCAPS. Despite not being a data-driven model, the (m-)GCAPS is currently the best validated tool in assessing the risk of GCA at an early stage. The (m-)GCAPS showed an AUC of 0.94 in the internal validation of Laskou *et al*. and an AUC of 0.83 in our external validation (5). El-Dairi *et al*., Ing et al. and Weis *et al*. have also developed and internally validated prediction models for GCA (9-11). Although these models have reasonably high AUC's (ranging from 0.80 to 0.81), they were not externally validated and not evaluated for calibration.

There are two main advantages of the (m-)GCAPS in comparison other models. Firstly, the (m-)GCAPS independently predicts the risk of GCA prior to TAB. The American College of Rheu-

Table II. Different m-GCAPS cut-off values with corresponding sensitivity, specificity, positive and negative predictive value.

m-GCAPS		GCA+ (n)	GCA- (n)	Sens. (%)	Spec. (%)	PPV (%)	NPV (%)
5.5	<5.5	0	14	100	14.7	33.1	100
	>5.5	40	81				
6.5	<6.5	2	25	95	26.3	35.2	92.6
	>6.5	38	70				
7.5	<7.5	3	32	92.5	33.7	37	91.4
	>7.5	37	63				
8.5	<8.5	5	45	87.5	47.4	41.2	90
	>8.5	35	50				
9.5	<9.5	8	56	80	58.9	45.1	87.5
	>9.5	32	39				
10.5	<10.5	8	72	80	75.8	58.2	90
	>10.5	32	23				
11.5	<11.5	10	74	75	77.9	58.8	88.1
	>11.5	30	21				
12.5	<12.5	12	80	70	84.2	65.1	87
	>12.5	28	15				
13.5	<13.5	15	85	62.5	89.5	71.4	85
	>13.5	25	10				
14.5	<14.5	18	89	55	93.7	78.6	83.2
	>14.5	22	6				
15.5	<15.5	23	92	42.5	96.8	85	80
	>15.5	17	3				
16.5	<16.5	24	93	40	97.9	88.9	79.5
	>16.5	16	2				
17.5	<17.5	30	94	25	98.9	90.0	75.8
	>17.5	10	1				
18.5	<18.5	36	9	10	98.9	80	72.3
	>18.5	4	1				
19.5	<19.5	37	95	7.5	100	100	72
	>19.5	3	0				
20.5	<20.5	39	95	2.5	100	100	70.9
	>20.5	1	0				

Sens.: sensitivity; spec.: specificity; PPV: positive predictive value; NPV: negative predictive value.

matology criteria and other prediction models require TAB results (which can take up to two weeks) to estimate the likelihood of GCA (12, 13). Secondly, the (m-)GCAPS uses a combination of demographic, clinical and laboratory data to predict the risk of GCA. Some models only used nonspecific laboratory results such as C-reactive protein, erythrocyte sedimentation rate or number of thrombocytes without combining these with other clinical data (11, 14).

A strength of our study is that we simultaneously evaluated discrimination and calibration.

Furthermore, we reduced the misclassification rate by choosing the rheumatologists' clinical diagnosis after six months as our reference. As there is no gold standard for GCA, taking a follow-up period of six months ensures that the manifestation of alternative pathology is taken into account.

Our study also has limitations. Although the retrospective nature of our study prevents bias from prior knowledge of the model, it comes with missing data. Nevertheless, multiple imputation showed a similar AUC, suggesting that the missing data had no or minimal effect on our outcomes. Given the relatively low prevalence of GCA a multi-centred prospective study is recommended to verify our study results. Another limitation of our study is that we could not validate the GCAPS completely since one variable was not noted in the clinical records. However, the variable extra-cranial abnormality is rarely crucial in the diagnostic process since research has shown that only a small proportion of GCA patients has extra-cranial vascular bruits or abnormal pulses (15).

The GCAPS was developed by Laskou *et al*. with the intention to stratify patients into those with high and low probability of GCA (5). As our study and Sebastian *et al*. have shown, risk stratification by (m-)GCAPS looks promising (6). However, scores should be handled with caution and might need optimisation based on the population.

In conclusion, the (m-)GCAPS showed good discrimination and calibration in a Dutch retrospective cohort and can aid in quantification of pre-test GCA probability in patients with suspected GCA. In this way, complications such as blindness and stroke in GCA patients may be prevented and unfortunate side effects of GC treatment in patients without GCA can be avoided.

Table III. Risk stratifications for the diagnosis of GCA proposed by Sebastian *et al*. (2020) (based on median and 75th percentile cut-off values of Sebastian *et al*. and Neuman *et al*.).

	Low-risk		Intermediate-risk		High-risk	
	GCA, n (%)	Controls, n (%)	GCA, n (%)	Controls, n (%)	GCA, n (%)	Controls, n (%)
Sebastian *et al*. *	5 (12.5%)	45 (47.4%)	7 (17.5%)	35 (36.8%)	28 (70.0%)	15 (15.8%)
Neuman *et al*. **	8 (20.0%)	56 (58.9%)	10 (25.0%)	33 (34.7%)	22 (55.0%)	6 (6.3%)

*Low-risk <9, intermediate-risk 9-12, high-risk >12; **Low-risk <10, intermediate-risk 10-14, high-risk >14.

References

1. BARRACLOUGH K, MALLEN CD, HELLIWELL T, HIDER SL, DASGUPTA B: Diagnosis and management of giant cell arteritis. *Br J Gen Pract* 2012; 62: 329-30.
2. PROVEN A, GABRIEL SE, ORCES C, O'FALLON WM, HUNDER GG: Glucocorticoid therapy in giant cell arteritis: duration and adverse outcomes. *Arthritis Rheum* 2003; 49: 703-8.
3. LUQMANI R, LEE E, SINGH S *et al*.: The Role of Ultrasound Compared to Biopsy of Temporal Arteries in the Diagnosis and Treatment of Giant Cell Arteritis (TABUL): a diagnostic accuracy and cost-effectiveness study. *Health Technol Assess* 2016; 20: 1-238.
4. DEJACO C, RAMIRO S, DUFTNER C *et al*.: EULAR recommendations for the use of imaging in large vessel vasculitis in clinical practice. *Ann Rheum Dis* 2018; 77: 636-43.
5. LASKOU F, COATH F, MACKIE SL, BANERJEE S, AUNG T, DASGUPTA B: A probability score to aid the diagnosis of suspected giant cell arteritis. *Clin Exp Rheumatol* 2019; 37 (Suppl. 117): S104-8.
6. SEBASTIAN A, TOMELLERI A, KAYANI A *et al*.: Probability-based algorithm using ultrasound and additional tests for suspected GCA in a fast-track clinic. *RMD Open* 2020; 6: 1297.
7. FELICETTI M, TREPPO E, POSARELLI C *et al*.: One year in review 2020: vasculitis. *Clin Exp Rheumatol* 2020; 38 (Suppl. 124): S3-14.
8. RUBIN DB: Multiple imputation for non-response in surveys. New York, John Wiley & Sons, 1987.
9. EL-DAIRI MA, CHANG L, PROIA AD, CUMMINGS TJ, STINNETT SS, BHATTI MT: Diagnostic Algorithm for Patients With Suspected Giant Cell Arteritis. *J Neuroophthalmol* 2015; 35: 246-53.
10. ING EB, LAHAIE LUNA G, TOREN A *et al*.: Multivariable prediction model for suspected giant cell arteritis: development and validation. *Clin Ophthalmol* 2017; 11: 2031-42.
11. WEIS E, TOREN A, JORDAN D, PATEL V, GILBERG S: Development of a predictive model for t emporal artery biopsies. *Can J Ophthalmol* 2017; 52: 599-605.
12. 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.
13. GONZALEZ-LOPEZ JJ, GONZALEZ-MORALEJ J, BURDASPAL-MORATILLA A, REBOLLEDA G, NUNEZ-GOMEZ-ALVAREZ MT, MUNOZ-NEGRETE FJ: Factors associated to temporal artery biopsy result in suspects of giant cell arteritis: a retrospective, multicenter, case-control study. *Acta Ophthalmol* 2013; 91: 763-8.
14. KERMANI TA, SCHMIDT J, CROWSON CS *et al*.: Utility of erythrocyte sedimentation rate and C-reactive protein for the diagnosis of giant cell arteritis. *Semin Arthritis Rheum* 2012; 41: 866-71.
15. MURATORE F, KERMANI TA, CROWSON CS *et al*.: Large-vessel giant cell arteritis: a cohort study. *Rheumatology* (Oxford) 2015; 54: 463-70.