RMD Open
Rheumatic & Musculoskeletal Diseases

ORIGINAL RESEARCH

OMERACT GCA phantom project: validation of a 3D-printed ultrasound training phantom for diagnosis of giant cell arteritis

Tobias Schremmer ,[1] Christian Dejaco ,[2,3] Florian Recker,[4] Markus Aschwanden,[5] Dennis Boumans,[6] George A Bruyn,[7] Stavros Chrysidis ,[8] Thomas Daikeler,[9] Berit Nielsen,[10,11] Eugenio De Miguel ,[12] Andreas P Diamantopoulos,[13] Uffe Moller Dohn,[14] Christina Duftner ,[15] Giuseppe Germano,[16] Cees Haagsma,[6] Wolfgang Hartung,[17] Alojzija Hocevar ,[18] Annamaria Iagnocco ,[19] Aaron Juche,[20] Pantelis Karakostas,[21] Rositsa Karalilova ,[22] Kresten Krarup Keller,[10,23] Borsha Sarker,[24] Kate Smith,[24] Verena Tischler,[25] Pierluigi Macchioni,[26] Marcin Milchert ,[27] Sara Monti ,[28] Chetan Mukhtyar ,[29] Esperanza Naredo ,[30] Wolfgang A Schmidt ,[31] Luca Seitz,[32] Lene Terslev ,[33] Richard J Wakefield,[34,35] Valentin Sebastian Schäfer [36]

To cite: Schremmer T, Dejaco C, Recker F, *et al*. OMERACT GCA phantom project: validation of a 3D-printed ultrasound training phantom for diagnosis of giant cell arteritis. *RMD Open* 2025;**11**:e005316. doi:10.1136/rmdopen-2024-005316

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/rmdopen-2024-005316).

TS and CD are joint first authors.

ACR Annual Scientific Meeting 2024, 2023.

Received 1 December 2024
Accepted 27 January 2025

For numbered affiliations see end of article.

Correspondence to
Dr Valentin Sebastian Schäfer; valentin.s.schaefer@hotmail.de

ABSTRACT

Objective Ultrasonography is crucial for diagnosing giant cell arteritis (GCA); however, training opportunities are rare. This study tested the reliability of ultrasonography findings and measurements of the intima-media thickness (IMT) among ultrasonography experts by using phantoms of the axillary (AA) and temporal arteries (TA) created with high-resolution 3D printing.

Methods Twenty-eight participants from 12 European countries received eight sets of phantoms of the AA and the superficial TA (including common, frontal and parietal branches), which were examined in a blinded fashion according to a predefined protocol and evaluated based on Outcome Measures in Rheumatology (OMERACT) GCA ultrasound definitions. Due to difficulties with the delineation of the intima-media complex, the parietal branch of the phantoms was modified, and a second round was conducted. The IMT was measured, and phantoms were classified as normal or vasculitic.

Results In both rounds, the phantoms were correctly classified as normal/abnormal in >81% of cases yielding a Fleiss' kappa of 0.80 (95% CI 0.78 to 0.81) in round 1 and 0.74 (95% CI 0.72 to 0.75) in round 2. IMT measurements revealed an intraclass correlation coefficient (ICC 1.1) of 0.98 (95% CI 0.98 to 0.99) in both rounds. Intrarater reliability was good with a median Cohens Kappa of 0.83 and median ICC of 0.78.

Conclusion The study demonstrated high reliability among ultrasound experts in applying the OMERACT ultrasound definitions for GCA and in measuring the IMT using a 3D-printed phantom of the AA and TA. This phantom could assist clinicians in training to assess the large arteries of patients with suspected or established GCA.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ 3D printing is used in medical education but lacks high-fidelity training models for ultrasonography in diagnosing giant cell arteritis (GCA).
⇒ Ultrasonography is vital for GCA diagnosis, yet there is a need for realistic phantoms to standardise training.

WHAT THIS STUDY ADDS

⇒ This study introduces the first high-resolution 3D-printed ultrasonography phantoms for GCA, tested by international ultrasonography experts.
⇒ The phantoms accurately mimic normal and pathological GCA conditions, showing good to almost perfect reliability in expert assessments.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ These phantoms can help with ultrasonography training of temporal and axillary arteries for diagnostic and monitoring purposes.
⇒ They offer a consistent training tool, reducing variability in ultrasonography assessments.
⇒ The study supports the development of 3D-printed phantoms for other rheumatic diseases, improving ultrasonography education and practice.

INTRODUCTION

Giant cell arteritis (GCA) is the most common primary systemic vasculitis in adults over the age of 50. It is characterised by transmural inflammation of large and medium-sized

arteries, most commonly affecting the superficial temporal artery (TA). Extracranial involvement, such as the axillary artery (AA), is observed in about half of the patients.[1] Symptoms vary depending on the vascular territory, ranging from headaches and jaw claudication to non-specific signs like fever, fatigue and weight loss. The most severe complication is permanent visual loss caused by anterior ischaemic optic neuropathy. This underscores the critical need for early diagnosis and timely treatment initiation.[2–4]

Ultrasonography of the TA and AAs has become crucial in diagnosis of GCA, endorsed by the European Alliance of Associations for Rheumatology recommendation for imaging in GCA.[5] The technique not only aids in early detection but also serves as a monitoring tool. However, the reliability of ultrasonography heavily depends on operator expertise, making proper training indispensable.[6 7] Despite this, opportunities for hands-on training remain scarce. The urgency of treatment for patients with active GCA limits their availability for educational purposes, and existing courses, such as the European Large Vessel Vasculitis Imaging Course and Southend GCA Sonography Workshops, cannot meet the growing demand.

The lack of accessible and consistent training models is a significant barrier to improving diagnostic reliability. The large-vessel vasculitis subgroup of the Outcome Measures in Rheumatology (OMERACT) ultrasonography working group defined the ultrasonography appearance of elementary lesions in GCA.[8] These include the halo sign (homogeneous, concentric, hypoechoic thickening of the intima-media complex (IMC) and compression sign (loss of IMC compressibility).[9] Cut-off values for the intima-media thickness (IMT of the axillary and TA aid diagnosis, though adjustments for age and gender seem to be necessary.[10–15]

Follow-up studies show that the IMC of the AA often remains thickened under treatment, whereas the TA typically resolves, adding complexity to the training process.[6 16 17]

To address these challenges, the development of reliable training models is paramount. Currently, no commercially available models exist for GCA, leaving a critical gap in education and standardisation.[18]

Our group previously developed and established a viable 3D-printed phantom for both normal and pathological TA and AA.[19] The new phantoms investigated in this study differ significantly from earlier models in several key aspects. Externally, they are identical to one another, ensuring consistency in appearance. Internally, enhancements have been made to better replicate the halo sign—a critical diagnostic feature—through improved echogenic properties. Additionally, the phantoms now encompass a broader range of pathological variations, including the introduction of chronic pathological models, further enhancing their diagnostic relevance.

This phantom leverages 3D printing technology, which creates physical objects layer by layer from digital models, offering precision and adaptability for medical applications. Specifically, the stereolithography (SLA) technique was used to construct vessel-mimicking phantoms since its higher resolution in comparison to other techniques.[20 21]

In the present study, the phantom was refined and tested within the Large Vessel Vasculitis Subgroup of the OMERACT Ultrasound Working Group, with the aim of testing the reliability and feasibility of these ultrasonography phantoms.

METHODS

In the following paragraphs, the production of the phantoms will be briefly described.

Design of the phantoms

We designed different phantoms of the TA and the AA using the computer-aided design software Fusion360 (Autodesk, San Rafael, California). As a template, we used 60 ultrasound images (12 for normal AA, 12 for chronic AA, 12 for acute AA, 12 for normal TA, 12 for pathological TA) of GCA, the published IMT cut-off values and the OMERACT definitions.[10] The TA phantoms depict the common superficial TA and its frontal and parietal branches. The AA phantoms portray the middle and distal segment of the AA beyond the pectoralis minor muscle with the posterior circumflex artery in between.[22] We used PreForm (Formlabs, Somerville, Massachusetts) for the conversion into a stereolithography file (.stl) and for the preparation of the prints.

3D printing

We used Formlabs Form3+ and Form 3BL 3D printers (Formlabs, Somerville, Massachusetts), based on low-force stereolithography, offering a resolution of up to 25 µm per layer. High Temp resin from Formlabs was used. Each model was washed in isopropanol and then UV cured. The phantoms were embedded in a specific ballistic gelatine, which mimics human muscle tissue's ultrasound propagation velocity.[23]

Study protocol

Two rounds were carried out. The first round featured AA and TA phantoms, the second round featured only TA phantoms. Before each round, online meetings were held with instructions and examples. The online meeting provided comprehensive instructions to ensure consistency among participants. The study design was discussed in detail, with visual examples illustrating the imaging process and a step-by-step demonstration of the examination protocol. The questionnaire was reviewed thoroughly, and participants had the opportunity to ask questions to clarify any uncertainties. Additionally, a training video was produced to show the participants how to conduct the examination. The material was available via cloud.

Phantom sets of the AA contained three arteries: one with acute changes, one with chronic changes and one

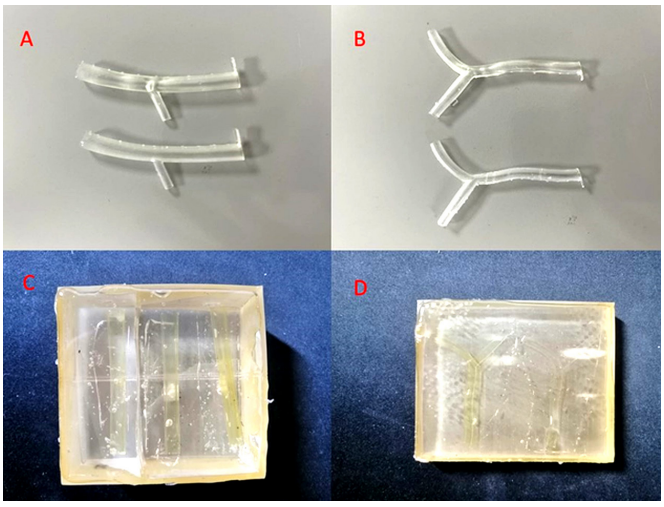

Figure 1 Phantoms. Phantoms of the axillary (left) and temporal artery (right)-embedded and without gelatine (A: axillary phantoms; B: temporal phantoms; C: axillary phantoms embedded as a set; D: temporal phantoms embedded as a set).

normal, while sets of the TA contained two arteries: one with acute changes and one normal (see figure 1). In total, we produced eight distinct versions of each, resulting in eight unique sets, corresponding to eight patients with real GCA.

The phantoms were sent by mail to selected members of the OMERACT ultrasound subgroup of LVV in 12 European countries (Germany, The Netherlands, United Kingdom, Norway, Denmark, Poland, Bulgaria, Slovenia, Austria, Italy, Spain and Switzerland). The participants were selected based on their expertise, interest and their residing country. Participants were required to have at least 1 year of experience and to have performed ultrasound examinations on more than 50 patients with suspected GCA. Each participant received eight sets each of AA and TA phantoms, which they examined schematically and in a blinded manner. Pathological changes were confined to the vessel wall and could not be detected from the outside (ie, without ultrasonography) of the phantom. The ultrasonography examination included longitudinal and transversal scans of every branch. Various ultrasound machines were used for these examinations (see in online supplemental file). Doppler was not used because the models do not simulate blood flow.

The following settings were recommended for the examination: B-mode frequency 18 MHz, image depth 1.5 cm and one focus point about 0.5 cm below the surface. IMT measurements were performed at the thickest point in millimetres in longitudinal sections on the vessel wall distal to the probe. The phantoms were then individually classified as normal (AA and TA), acute pathology (AA and TA), chronic pathology (only AA) or none of the above (AA and TA). Classification should be conducted applying the OMERACT ultrasound definitions for GCA.[8 24]

Because of the common feedback from experts that the parietal branch of several phantoms could not be delineated clearly because of reverberation artefacts, the parietal branch of the phantoms was revised for the second round. The examination protocol remained the same.

We used SurveyMonkey (Momentive.ai, San Mateo, California) as an electronic case report form (eCRF) (see online supplemental 1). Examiners were asked to copy their locally documented data into that platform.

The eCRF was constructed in the following way: (1) question about experience with GCA ultrasonography, (2) evaluation of whether a case corresponds to acute like lesions, chronic change or normal (yes/no) and the confidence with this decision, (3) IMT measurement, (4) questions about the overall quality of the phantoms. Questions referring to the overall quality of the phantoms were about whether the phantoms fulfil the OMERACT definition of acute pathology, chronic pathology or the definitions of a normal vessel, whether the cut-off values were fulfilled and whether the phantoms correlate to clinical practice. Furthermore, participants were able to report issues encountered during the examinations using an open-text field.

Microscopy examination

Simultaneously, four phantoms were randomly selected for examination by the Institute of Clinical Pathology at the University of Bonn. This examination aimed to compare the microscopically detected IMT with the ultrasound-measured IMT. For this examination, phantoms were stripped from the gelatine and were embedded in paraffin, as done previously in TA biopsies, and subsequently sectioned using a microtome. The resulting slices were microscopically analysed to determine the thickness of the phantom wall.

Statistical analysis

All statistical analyses were performed using the SPSS software (V.28.0.1.1; IBM, Armonk, New York). Means and SD were calculated for IMT measurements as well as medians and ranges for confidence of participants in the classification using the Likert scales. Normality of the IMT measurements was tested using the Shapiro-Wilk test. Inter-rater reliability of the classification (acute, chronic, normal for AA or acute, chronic for TA) was calculated using Fleiss-Kappa. Fleiss-Kappa was interpreted according to Landis and Koch.[25]Inter-rater reliability of the IMT measurements was calculated using intraclass coefficient (ICC) (1.1) and interpreted according to Rosner.[26]

Intrarater reliability for the temporal phantoms was calculated using Cohens-Kappa. It was interpreted according to Landis and Koch.[25] For the IMT measurements, the ICC (2.1) was calculated and interpreted according to Rosner.[26] Since only the parietal and frontal branches of the phantoms were modified, they were not included in the calculations, and only the common branch was included.

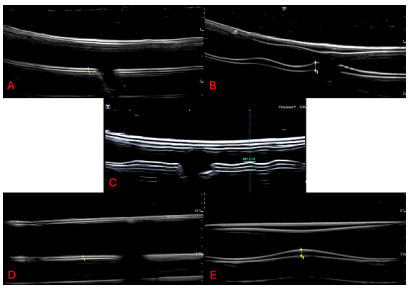

Figure 2 Ultrasound image of the phantoms. Ultrasonography pictures of the phantoms; A=normal axillary artery; B=axillary artery with acute GCA pathology; C=axillary artery with chronic GCA pathology; D=normal temporal artery; E=temporal artery with acute GCA pathology. GCA, giant cell arteritis.

Furthermore, cross tables for the classifications of the phantoms were produced and a true-positive and false-negative rate for classification were calculated.

RESULTS

When comparing the phantoms with real GCA cases, the morphology resembles a halo sign; however, the anechoic surroundings and lack of flow are notable differences (see figure 2).

Participants

Twenty-eight members of the OMERACT LVV subgroup participated in the project (see online supplemental 2). Two participants were unable to examine one set of AAs due to material damage. A distal branch had broken off and the whole set broke from the gelatin structure during handling or transportation. One participant examined only the TA, as the package had been forwarded to the next participant before the full examination could be completed. The other participants completed all study tasks. Each participant has been practising ultrasound on patients with GCA for >1 year, with >70% having experience of >8 years. Only two participants had less than 3 years of experience. Sixty per cent of the participants are conducting >100 scans for suspected vasculitis per year, 92.5% performed at least 25 examinations. All participants except one took part in both rounds. In the second round, an additional participant (CH) joined the study because one participant retired.

Round 1

A total of 416 measurements and ratings for the TA and 558 for the AA were analysed. The data did not appear to deviate significantly from a normal distribution (p>0,05). The overall inter-rater reliability of the IMT measurements of the TA and AA was excellent with an ICC 1.1 of 0.98 (95% interval 0.977 < ICC <0.990). Subanalyses were done on all branches individually (table 1). In each branch, excellent agreement among all the participants was demonstrated with an ICC in the range from 0.96 to 0.99. The mean IMTs of almost each phantom corresponded to the published cut-off values.[10 24 27] The parietal branch of the TA was an exception, as these phantoms did not conform to the cut-off values (see table 1) . The normal AA phantoms were correctly classified in 87.4% of cases, the acute pathology AA phantoms in 97.9% and the chronic AA pathology phantoms in 84.7%. For the TA, normal phantoms were correctly identified in 94% of cases and the acute pathology TA phantoms in 93.4%. Most of the misidentified branches were frontal and parietal branches of the TA and AA phantoms with chronic pathology. Nevertheless, with a Fleiss-Kappa of 0.80 (95% interval 0.780 < κ <0.813), the overall inter-rater reliability of the TA and AA phantoms classification suggested a substantial agreement. Each subanalysis for the branches separately indicated a substantial agreement between the participants (see table 1).

Table 1 Results of the IMT measurements and inter-rater reliability

Vessel	Measurements	Mean (SD) normal (in mm)	Mean (SD) acute (in mm)	Mean (SD) chronic (in mm)	ICC 1.1 (95%CI)	Fleiss Kappa (95% CI)
Round 1						
Common superficial temporal arteries	416	0.35 (0.10)	0.72 (0.14)	–	0.99 (0.98 to 0.99)	0.70 (0.68 to 0.73)
Frontal branches	416	0.32 (0.09)	0.63 (0.11)	–	0.99 (0.98 to 0.99)	0.82 (0.79 to 0.85)
Parietal branches	416	0.35 (0.10)	0.66 (0.08)	–	0.98 (0.97 to 0.99)	0.85 (0.82 to 0.87)
Axillary arteries	558	0.66 (0.14)	1.26 (0.19)	1.19 (0.27)	0.96 (0.93 to 0.98)	0.75 (0.73 to 0.77)
Round 2						
Common superficial temporal arteries	416	0.39 (0.13)	0.73 (0.16)	–	0.98 (0.96 to 0.99)	0.7 (0.67 to 0.72)
Frontal branches	416	0.34 (0.10)	0.62 (0.14)	–	0.98 (0.97 to 0.99)	0.70 (0.65 to 0.70)
Parietal branches	416	0.29 (0.09)	0.63 (0.13)	–	0.99 (0.98 to 0.99)	0.72 (0.69 to 0.74)

ICC, intra-class correlation coefficient; IMT, intima-media thickness.

Table 2 True positive and negative rate

Vessel	Number of segments measured	True negative rate Normal	True positive rate Acute	True positive rate Chronic
Round 1				
Common superficial temporal arteries	416	90.8%	97.4%	–
Frontal branches	416	94.6%	83.3%	–
Parietal branches	416	96.6%	99.5%	–
Axillary arteries	558	87.4%	97.9%	84.7%
Round 2				
Common superficial temporal arteries	416	81.9%	98.1%	
Frontal branches	416	83.1%	89.1%	
Parietal branches	416	89.3%	95.1%	

The true positive and negative rates for the classification of each branch were>80% (see table 2).

Round 2

416 measurements and ratings for the TA were analysed in this round. There seemed to be no significant deviation from a normal distribution in the subgroups (p>0,05).

In the second round, only the TA phantoms were examined. In accordance with the first round, with a Fleiss-Kappa of 0.74 (95% interval 0.72 < κ <0.75) the overall inter-rater reliability of the classification suggested a substantial agreement. Each of the subanalyses indicated substantial agreement between the examiners for the different branches (see table 1). It must also be noted that the qualitative agreement between the experts was lower in the second round compared with the first round. Additionally, like in the first round, qualitative agreement was lower compared with the ICC.

A closer look at the data showed that there was a better identification rate of the phantoms with acute pathology and higher agreement between the participants. The normal phantoms were correctly classified in 84.6%, and the acute pathology phantoms in 94.1%. In contrast to the first round, the normal phantoms were identified less frequently in a correct way. The true positive and negative rates for each classification were >80% (see table 2). It was noticeable that, compared with the first round, more investigators chose the third option 'none of the above'. In the second round, all the branches corresponded to their respective published cut-off values.[10] The inter-rater reliability for the IMT measurement was excellent. For all the measurements of the IMT, the ICC 1.1 was 0.98 (95% interval 0.977 < ICC <0.990). Taking a closer look at the branches separately, the participants showed excellent agreement for all of them (see table 1). Detailed results for the IMT measurements are summarised in table 1.

Subjective statements

Participants generally agreed that the phantoms for acute GCA and normal arteries met the OMERACT ultrasonography definitions. However, they neither agreed nor disagreed regarding the chronic phantoms' adherence to these definitions and cut-off values. Additionally, the sonographic image of the phantoms did not seem to correspond well to clinical practice of the experts (see online supplemental 3). The comments remained the same in the second round.

Several challenges were encountered during the study that impacted the evaluation process. Differentiating chronic cases of GCA remained difficult, some phantoms lacked extensive thickening in the TA and occasionally deviated from real-life pathological appearances. Artefacts, including gas bubbles and issues with embedding, sometimes disrupted imaging quality. Low IMT values in the AA and challenges in obtaining reliable transverse measurements added further complexity.

Comparison between measured and designed IMT

When comparing the means of the IMT measurements to the designed IMT, there were noticeable variations (see figure 3). The IMT measurements of the normal phantoms of both TA and AA were larger than the design, while those of the acute phantoms of the TA and AA were smaller. The chronic phantoms of the AA were measured much smaller than designed. On average, there was 0.25 mm (SD 0.1 mm) difference for the AA phantoms and 0.13 mm (SD 0.15 mm) difference for the TA phantoms. During the second round, this relation seemed to have a smaller impact on the results. Here, the mean difference between the design and the measurements of the TA phantoms was only 0.08 mm (SD 0.1 mm).

Intrarater reliability

On examining intra-rater reliability between rounds 1 and 2, we found that most investigators demonstrated good agreement, with Cohen's Kappa ranging from 0.65 to 1 with a median of 0.83. Additionally, the measurements showed generally good to excellent reliability, with an ICC ranging from 0.59 to 0.93 with a median of 0.78 (see table 3).

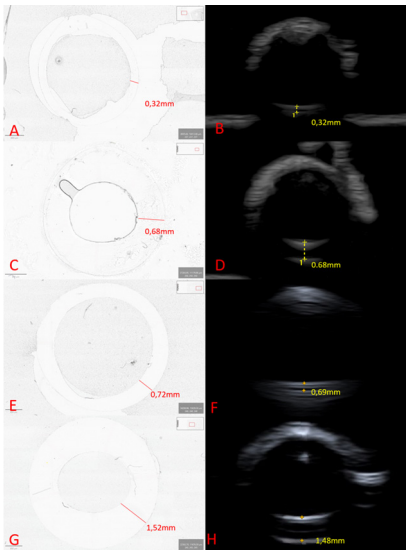

Figure 3 Difference between mean IMT measurement and designed IMT of the phantoms. The figure illustrates the difference between the mean measured IMT of each phantom and the designed IMT. The data are color-coded: purple represents round 1 of the temporal artery, blue represents axillary artery and green represents round 2 of the temporal artery. The x-axis shows the difference in millimetres, while the y-axis lists the names of the phantoms (eg; 2.2.F is the second phantom in the second set, frontal branch). F: frontal branch, P: parietal branch, C: common temporal artery. IMT, intima-media thickness.

Microscopic examination

The TA GCA phantom representing acute pathology exhibited a wall thickness of 0.683 mm, whereas the normal phantom had a wall thickness of 0.323 mm. Compared with the mean measured IMT of the respective phantom, these wall thicknesses fall within 1 SD. However, the designed wall thickness lies outside this range (see table 4 and figure 4). Comparing the measured IMT, the design and the wall thickness of the axillary phantoms, we observed, that the measurements are within 1 SD.

DISCUSSION

This study represents the first effort to evaluate the use of 3D-printed high-resolution ultrasonography phantoms for diagnosing GCA among an international cohort of ultrasonography experts. We developed phantoms replicating GCA pathologies as defined by OMERACT. Quantitative and qualitative reliability of the phantoms was demonstrated to be good to almost perfect when assessed by experts, even though the models were designed to mimic clear pathological and normal conditions. This underscores their validity, as shown by consistent ultrasound and microscopic measurements.

To our understanding, this is the first working ultrasonography phantom for educational purposes in the field of rheumatology. This technology proved to mimic different clinical situations and can even help with the teaching of complicated procedures.[28] Most of the time, 3D-printed models are simply used for surgical planning or for anatomy, although this technology, as previously demonstrated, can also be used for imaging techniques such as ultrasonography as well as for training in invasive methods.[28–30]

When comparing the wall thickness, measured by microscopy, to the average measured values of the IMT by ultrasonography, it can be inferred that the wall of the phantom simulates the IMT. Although this is a departure from clinical reality, it allows for effective representation of pathological conditions. While this diverges slightly from clinical reality, it allows for a realistic representation of pathological conditions. It is important to consider that paraffin embedding and subsequent tissue processing may have influenced IMT measurements. However, the choice of resin with high structural integrity was intended to minimise both thermal and mechanical impacts during the embedding and processing stages.

Participants largely agreed that the acute and normal phantoms met OMERACT criteria, but chronic phantoms were less satisfactory due to reverberation artefacts

Table 3 Intrarater reliability

Rater	Cohens kappa (SE)	Intraclass correlation coefficient (2.1) (SD)
1	0.83 (0.08)	0.73 (0.52–0.86)
2	0.79 (0.10)	0.71 (0.50–0.85)
3	0.83 (0.08)	0.90 (0.80–0.95)
4	1.00 (0.00)	0.93 (0.85–0.96)
5	0.83 (0.08)	0.59 (0.32–0.73)
6	0.81 (0.09)	0.73 (0.52–0.86)
7	0.76 (0.11)	0.62 (0.45–0.81)
8	0.92 (0.04)	0.88 (0.77–0.92)
9	0.65 (0.18)	0.68 (0.48–0.82)
10	1.00 (0.00)	0.90 (0.79–0.95)
11	0.83 (0.08)	0.88 (0.77–0.91)
12	1.00 (0.00)	0.93 (0.86–0.96)
13	0.69 (0.13)	0.60 (0.35–0.75)
14	0.79 (0.10)	0.65 (0.42–0.79)
15	0.87 (0.07)	0.75 (0.51–0.82)
16	0.89 (0.06)	0.90 (0.79–0.95)
17	0.67 (0.14)	0.72 (0.50–0.86)
19	1.00 (0.00)	0.64 (0.42–0.79)
20	0.95 (0.02)	0.81 (0.75–0.88)
22	0.92 (0.04)	0.85 (0.77–0.92)
23	0.79 (0.10)	0.77 (0.57–0.88)
24	0.89 (0.06)	0.93 (0.79–0.95)
25	0.65 (0.18)	0.66 (0.42–0.79)
26	0.70 (0.13)	0.87 (0.77–0.92)
27	0.76 (0.11)	0.83 (0.60–0.92)
28	0.69 (0.13)	0.73 (0.52–0.86)

Table 4 Mean measured IMT, designed IMT and measured IMT by microscopy examination

Vessel	Mean measured IMT (SD) in mm	Designed IMT in mm	Measured IMT by microscopy in mm
Normal axillary	0.69 (0.22)	0.7	0.72
Acute axillary	1.36 (0.28)	1.5	1.52
Normal common temporal	0.32 (0,08)	0.30	0.32
Acute common temporal	0.69 (0,16)	0.70	0.68

Comparison between the mean measured IMT, the designed IMT and the by microscopy examination measured IMT of different phantoms.
IMT, intima-media thickness.

and indistinct edges. These artefacts, along with issues like air bubbles and closely spaced vessel walls, led to some diagnostic challenges and misclassifications. One possible reason for this phenomenon could be material issues. To mimic the several visible lines of the chronic pathology of GCA in ultrasonography, we needed to put more walls inside the vessels. Because of the little space between the different walls, this artefact occurred here. This was also the problem with the other phantoms as the vessel wall was only 2 mm or less.

Additionally, comparing the designed IMT with the measured IMT revealed notable deviations. This discrepancy may stem from measurement point issues or errors, though the second round showed significant improvement in these effects. Another possible explanation could be the 'regression to the mean' effect.[31]

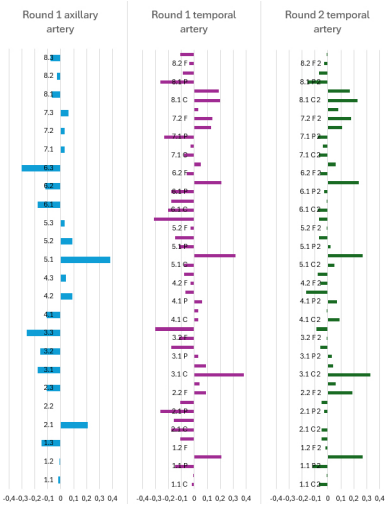

Figure 4 Comparison between the microscopy and ultrasound imaging of the phantoms. Comparison between microscopy and ultrasound imaging of the phantoms with the measurements of the intima-media thickness (yellow) and the measurement of the vessel wall (red) (A: normal common branch temporal artery phantom in microscopy; B: normal common branch temporal artery in ultrasonography; C: acute pathology common branch phantom of the temporal artery in microscopy; D: acute pathology common branch phantom of the temporal artery in ultrasonography; E: normal axillary artery phantom in microscopy; F: normal axillary artery phantom in ultrasonography; G: acute pathology phantom of the axillary artery in microscopy; H: acute pathology phantom of the axillary artery in ultrasonography).

The current phantom represents a significant advancement over the previously published model, addressing key limitations and introducing several improvements.[19] The inclusion of a broader range of pathological variations, such as different IMT values and chronic pathological features, has expanded its utility for comprehensive diagnostic training. The production process has been simplified and standardised, allowing for more consistent results and easier scalability. The results of this study underline the former findings and strengthen them by validation on an international level.

One notable limitation of this study is the lack of intrarater reliability data for the AA. The individual models used in the study differed, and within a single round of assessments, no repeated measurements by the same rater were conducted to evaluate intrarater reliability. This gap in the study design arose partly due to cost considerations for packaging and shipping multiple phantoms as well as the satisfactory results achieved in the initial round. However, this limitation highlights an area of improvement in future studies to better assess the consistency.

Several other limitations of the current phantoms should be noted. The limited elasticity of current materials hinders the replication of features such as the compression sign. Additionally, the phantoms are primarily designed for assessing the halo sign but lack integration with colour Doppler technology and flow simulation, which are critical for dynamic training. The use of transparent materials also limits their realism, as real tissues are non-transparent. These limitations, combined with the selection of clear-cut patient image templates, reduce the ability to simulate borderline or ambiguous cases often encountered in clinical practice.

Scalability and accessibility also remain crucial considerations. While the current cost of approximately €5 per model is low, achieving widespread adoption—particularly in resource-limited settings—requires further reductions in production costs.

Additionally, variations in machine settings could introduce subtle differences that were not assessed in this study. Additionally, the level of expertise among participants was not explicitly correlated with the reliability

results, leaving room to investigate whether diagnostic accuracy and measurement consistency vary with operator experience or equipment-specific factors.

Ongoing efforts aim to address these challenges. Strategies include using flexible materials to replicate the compression sign. A more flexible material would not only enable to simulate the compression sign but also used to simulate other vascular territories like the subclavian arteries, which are more difficult to assess on real patients and also need to be examined in cases of extracranial GCA.[6] Furthermore, incorporating flow-simulating systems to mimic vascular blood flow and adopting materials that mimic the optical properties of non-transparent tissues would enhance the phantom realism. One way to mimic this would be the usage of organic materials like psyllium husk or flour. This method enhances the surrounding environment's realism customises echogenicity and reduces reverberation artefacts. Exploring higher print volumes, alternative materials, and more efficient production methods could address problems with cost reduction.

Broadening the range of patient templates to include diverse disease presentations is also a priority. These advancements will make the phantoms more realistic and clinically relevant while retaining their low production cost and accessibility.

Future perspectives for this project are promising. A key priority is testing these phantoms with less experienced ultrasonographers to assess their utility in broader training scenarios. This would help determine their effectiveness in building foundational skills and improving diagnostic consistency across varying levels of expertise. Such advancements would provide trainees with a more comprehensive experience and facilitate practice in dynamic conditions closer to real-life scenarios.

The long-term educational impact of this work is significant. Reduction of diagnostic variability of GCA ultrasonography could be achieved with the possibility of standardised training conditions. They enable trainees to practice safely on consistent, customisable models. Furthermore, the phantoms' versatility supports decentralised and online learning platforms, making high-quality training accessible globally.

Various approaches for training courses have been implemented in the past, demonstrating essential learning success multiple times.[32–34] In this context, our phantoms should be considered as a complement to these courses. The phantoms help in preparation of these courses or be used within the courses to minimise the number of required subjects or to teach fundamental principles. However, it must be stated that the phantoms cannot completely replace patient-based teaching.

The findings from this study could be used to start a prospective controlled study to evaluate the usage in training.

With further validation and integration into structured ultrasonography programmes, this technology could improve the way clinicians are trained, ultimately enhancing patient care and diagnostic outcomes in GCA.

Author affiliations
[1]Rheumatology, Universitätsklinikum Bonn, Bonn, Germany
[2]Rheumatology, Medical University Graz, Graz, Austria
[3]Rheumatology, Hospital of Bruneck, Bruneck, Italy
[4]Department for Gynecology and Obstetrics, University Hospital Bonn, Bonn, Germany
[5]Angiology, University Hospital Basel, Basel, Switzerland
[6]Department of Rheumatology and Clinical Immunology, Ziekenhuisgroep Twente (Hospital Group Twente), Almelo, The Netherlands
[7]Department of Rheumatology, Tergooi Medical Centre, Hilversum, The Netherlands
[8]Rheumatology, Sydvestjysk Sygehus Esbjerg, Esbjerg, Denmark
[9]Rheumatology, University Hospital Basel, Basel, Switzerland
[10]Department of Rheumatology, Aarhus University Hospital, Aarhus, Denmark
[11]Regional Hospital Horsens, Horsens, Denmark
[12]Hospital Universitario La Paz, Madrid, Spain
[13]Sorlandet sykehus HF, Kristiansand, Norway
[14]Department of Clinical Medicine, University of Copenhagen, Kobenhavn, Denmark
[15]Department of Internal Medicine, Clinical Division of Internal Medicine II, Medical University Innsbruck, Innsbruck, Austria
[16]Department of Specialistic Medicine, Azienda USL, Reggio Emilia, Italy
[17]Asklepios Medical Center, Bad Abbach, Germany
[18]Department of Rheumatology, University Medical Centre Ljubljana, Division of Internal Medicine, Ljubljana, Slovenia
[19]Università degli Studi di Torino, Torino, Italy
[20]Immanuel Hospital Berlin-Wannsee Branch, Berlin, Germany
[21]Oncology, Hematology, Rheumatology and Clinical Immunology, University Hospital of Bonn, Bonn, Germany
[22]Clinic of Rheumatology, Medical University Plovdiv, Plovdiv, Bulgaria
[23]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[24]NIHR Leeds Biomedical Research Centre, Leeds, UK
[25]Department of Clinical Pathology, University Hospital Bonn, Bonn, Germany
[26]Azienda USL IRCCS Reggio Emilia, Reggio Emilia, Italy
[27]Department of Rheumatology, Internal Medicine, Geriatrics and Clinical Immunology, Pomeranian Medical University in Szczecin, Szczecin, Poland
[28]Rheumatology, University of Pavia; early Arthritis Clinic, IRCCS Policlinico S. Matteo Foundation, Pavia, Italy
[29]Vasculitis Service, Rheumatology Department, Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, UK
[30]Rheumatology, University Hospital Fundación Jiménez Díaz, Madrid, Spain
[31]Immanuel Krankenhaus Berlin, Berlin, Germany
[32]Department of Rheumatology and Immunology, University of Bern, Bern, Switzerland
[33]Rigshospitalet, Center for Rheumatology and Spine Diseases, Glostrup, Denmark
[34]Leeds Biomedical Research Centre, Leeds, UK
[35]University of Leeds, Leeds Institute of Rheumatic and Musculoskeletal Medicine, Leeds, UK
[36]Department of Rheumatology, University Hospital Bonn, Bonn, Germany

X Marcin Milchert @MarcinMilchert and Chetan Mukhtyar @cmukhtyar

Contributors TS: conception, design, execution, designing the phantoms, overseeing the 3D printing process, conducting the ultrasonography assessments, compiling and analysing the data, initial draft of the manuscript and editing. CDe: statistical analysis, overall conceptual framework, interpretation of data, revising of the manuscript. VT: microscopic examination, revising of the manuscript. VSS: senior oversight, study design, execution, designing the phantoms, conducting the ultrasonography assessment, ultrasonography examination, statistical analysis, initial draft and editing and revising. This author is the guarantor. The other authors: key participants in the study, ultrasonography examination, revising of the manuscript. All authors reviewed and approved the final manuscript.

Funding Funding was received by Hexal Pharma.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer-reviewed.

Data availability statement Data are available upon reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information. The demographic data of the participants can be found in the appendix. The analyses of the IMT measurements are in the manuscript.

ORCID iDs
Tobias Schremmer http://orcid.org/0009-0009-8642-9070
Christian Dejaco http://orcid.org/0000-0002-0173-0668
Stavros Chrysidis http://orcid.org/0000-0001-8583-6517
Eugenio De Miguel http://orcid.org/0000-0001-5146-1964
Christina Duftner http://orcid.org/0000-0003-3137-8834
Alojzija Hocevar http://orcid.org/0000-0002-7361-6549
Annamaria Iagnocco http://orcid.org/0000-0001-5592-724X
Rositsa Karalilova http://orcid.org/0000-0001-8030-0529
Marcin Milchert http://orcid.org/0000-0002-0943-8768
Sara Monti http://orcid.org/0000-0002-1800-6772
Chetan Mukhtyar http://orcid.org/0000-0002-9771-6667
Esperanza Naredo http://orcid.org/0000-0003-0017-0096
Wolfgang A Schmidt http://orcid.org/0000-0001-7831-8738
Lene Terslev http://orcid.org/0000-0001-8193-9471
Valentin Sebastian Schäfer http://orcid.org/0000-0002-6591-5936

REFERENCES

1 Buttgereit F, Dejaco C, Matteson EL, *et al*. Polymyalgia Rheumatica and Giant Cell Arteritis: A Systematic Review. *JAMA* 2016;315:2442–58.
2 Peral-Cagigal B, Pérez-Villar Á, Redondo-González L-M, *et al*. Temporal headache and jaw claudication may be the key for the diagnosis of giant cell arteritis. *Med Oral Patol Oral Cir Bucal* 2018;23:e290–4.
3 Dasgupta B. Concise guidance: diagnosis and management of giant cell arteritis. *Clin Med (Northfield)* 2010;10:381–6.
4 Bardi M, Diamantopoulos AP. EULAR recommendations for the use of imaging in large vessel vasculitis in clinical practice summary. *Radiol Med* 2019;124:965–72.
5 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.
6 Dejaco C, Ramiro S, Bond M, *et al*. EULAR recommendations for the use of imaging in large vessel vasculitis in clinical practice: 2023 update. *Ann Rheum Dis* 2024;83:741–51.
7 Nielsen BD, Therkildsen P, Keller KK, *et al*. Ultrasonography in the assessment of disease activity in cranial and large-vessel giant cell arteritis: a prospective follow-up study. *Rheumatology (Sunnyvale)* 2023;62:3084–94.
8 Chrysidis S, Duftner C, Dejaco C, *et al*. Definitions and reliability assessment of elementary ultrasound lesions in giant cell arteritis: a study from the OMERACT Large Vessel Vasculitis Ultrasound Working Group. *RMD Open* 2018;4:e000598.
9 Schmidt WA. Ultrasound in the diagnosis and management of giant cell arteritis. *Rheumatology (Sunnyvale)* 2018;57:ii22–31.
10 Schäfer VS, Juche A, Ramiro S, *et al*. Ultrasound cut-off values for intima-media thickness of temporal, facial and axillary arteries in giant cell arteritis. *Rheumatology (Oxford)* 2017;56:1479–83.
11 Sebastian A, van der Geest KSM, Tomelleri A, *et al*. Development of a diagnostic prediction model for giant cell arteritis by sequential application of Southend Giant Cell Arteritis Probability Score and ultrasonography: a prospective multicentre study. *Lancet Rheumatol* 2024;6:e291–9.
12 Milchert M, Fliciński J, Brzosko M. Intima-media thickness cut-off values depicting “halo sign” and potential confounder analysis for the best diagnosis of large vessel giant cell arteritis by ultrasonography. *Front Med* 2022;9:1055524.
13 Ješe R, Rotar Ž, Tomšič M, *et al*. The cut-off values for the intima–media complex thickness assessed by colour Doppler sonography in seven cranial and aortic arch arteries. *Rheumatology (Sunnyvale)* 2021;60:1346–52.
14 López-Gloria K, Castrejón I, Nieto-González JC, *et al*. Ultrasound intima media thickness cut-off values for cranial and extracranial arteries in patients with suspected giant cell arteritis. *Front Med* 2022;9:981804.
15 Seitz P, Lötscher F, Bucher S, *et al*. Ultrasound intima-media thickness cut-off values for the diagnosis of giant cell arteritis using a dual clinical and MRI reference standard and cardiovascular risk stratification. *Front Med* 2024;11:1389655.
16 Aschwanden M, Schegk E, Imfeld S, *et al*. Vessel wall plasticity in large vessel giant cell arteritis: an ultrasound follow-up study. *Rheumatology (Oxford)* 2019;58:792–7.
17 Dejaco C, Ponte C, Monti S, *et al*. The provisional OMERACT ultrasonography score for giant cell arteritis. *Ann Rheum Dis* 2023;82:556–64.
18 Schäfer VS, Chrysidis S, Dejaco C, *et al*. Assessing Vasculitis in Giant Cell Arteritis by Ultrasound: Results of OMERACT Patient-based Reliability Exercises. *J Rheumatol* 2018;45:1289–95.
19 Recker F, Jin L, Veith P, *et al*. Development and Proof of Concept of a Low-Cost Ultrasound Training Model for Diagnosis of Giant Cell Arteritis Using 3D Printing. *Diagnostics (Basel)* 2021;11:1106.
20 AbouHashem Y, Dayal M, Savanah S, *et al*. The application of 3D printing in anatomy education. *Med Educ Online* 2015;20:29847.
21 Ho D, Squelch A, Sun Z. Modelling of aortic aneurysm and aortic dissection through 3D printing. *J Med Radiat Sci* 2017;64:10–7.
22 Schmidt WA, Schäfer VS. Diagnosing vasculitis with ultrasound: findings and pitfalls. *Ther Adv Musculoskelet Dis* 2024;16.
23 Phillips H, Franklin C, Brearley J, *et al*. Natural ballistic gelatine ultrasound phantoms are suitable to be used for student education and can be produced cheaply and effectively. *Vet Radiol Ultrasound* 2023;64:733–9.
24 Schäfer VS, Chrysidis S, Schmidt WA, *et al*. OMERACT definition and reliability assessment of chronic ultrasound lesions of the axillary artery in giant cell arteritis. *Semin Arthritis Rheum* 2021;51:951–6.
25 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
26 Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* 1979;86:420–8.
27 Czihal M, Schröttle A, Baustel K, *et al*. B-mode sonography wall thickness assessment of the temporal and axillary arteries for the diagnosis of giant cell arteritis: a cohort study. *Clin Exp Rheumatol* 2017;35 Suppl 103:128–33.
28 Recker F, Schremmer T, Berg C, *et al*. Advancement of 3D printing technology for the development of a training model in US-guided vesicoamniotic shunting for early LUTO therapy. *Acta Obstet Gynecol Scand* 2024;103:1550–7.
29 Ridgers A, Li J, Coles-Black J, *et al*. Teaching Radial Endobronchial Ultrasound with a Three-Dimensional-printed Radial Ultrasound Model. *ATS Sch* 2021;2:606–19.
30 Squelch A. 3D printing and medical imaging. *J Med Radiat Sci* 2018;65:171–2.
31 Bland JM, Altman DG. Regression towards the mean. *BMJ* 1994;308:1499.
32 Ducker G, Mukhtyar CB. Specialist nurse training in ultrasonography for the diagnosis of giant cell arteritis. *Rheumatology (Sunnyvale)* 2023;62:e142–3.
33 De Miguel E, Castillo C, Rodríguez A, *et al*. Learning and reliability of colour Doppler ultrasound in giant cell arteritis. *Clin Exp Rheumatol* 2009;27:S53–8.
34 Chrysidis S, Terslev L, Christensen R, *et al*. Vascular ultrasound for the diagnosis of giant cell arteritis: a reliability and agreement study based on a standardised training programme. *RMD Open* 2020;6:e001337.