ORIGINAL RESEARCH

Measuring treatment outcomes and change in disease activity in giant cell arteritis: a systematic literature review informing the development of the EULAR-ACR response criteria on behalf of the EULAR-ACR response criteria in giant cell arteritis task force

Catalina Sanchez-Alvarez 1, Milena Bond 2, Medha Soowamber 3, Dario Camellino 4, Melanie Anderson 5, Carol A Langford 6, Christian Dejaco 7, Zahi Touma 8, Sofia Ramiro 9,10

ABSTRACT

Objectives To identify criteria and descriptors used to measure response to treatment and change in disease activity in giant cell arteritis (GCA).

Methods A systematic literature review (SLR) to retrieve randomised controlled trials (RCTs) and longitudinal observational studies (LOS). Criteria and descriptors of active disease, remission, response, improvement, worsening and relapse were extracted. RCTs, LOS with >20 subjects, and qualitative research studies were included.

Results 10,593 studies were retrieved, of which 116 were included (11 RCTs, 104 LOS, 1 qualitative study). No unified definition of response to therapy was found. Most RCTs used composite endpoints to assess treatment outcomes. Active disease was described in all RCTs and 19% of LOS; and was largely defined by a combination of clinical and laboratory components. Remission was reported in 73% of RCTs and 42% of LOS; it was predominantly defined as the combination of clinical and laboratory components. Remission was reported in 73% of RCTs and 42% of LOS; it was predominantly defined as the combination of clinical and laboratory components. One LOS reported response with a definition resembling the definition of remission from other studies. Improvement was rarely used as an endpoint and it was mostly a surrogate of remission. No study specifically defined worsening. Relapse was reported in all RCTs and 86% of LOS. It was predominantly defined as the combination of clinical, laboratory and treatment components.

Conclusions The results of this SLR demonstrate that definitions of response used in clinical studies of GCA are scant and heterogeneous. RCTs and LOS mainly used remission and relapse as treatment outcomes. The descriptors identified will inform the development of new response criteria for GCA similar to what has been done for other rheumatic and musculoskeletal diseases.

INTRODUCTION

Giant cell arteritis (GCA) is the most common form of large vessel vasculitis.1 For several decades, treatment of GCA was primarily based on glucocorticoids (GC), with the main goal of achieving remission. Although remission remains a key goal, there is an increasing
interest in minimising GC to reduce treatment-related adverse events. However, the minimisation of GC should be balanced against the risk of relapses and the subsequent risk of potential accrual damage. While in the past, inflammatory markers were helpful parameters in determining disease activity and change thereof, the added value of erythrocyte sedimentation rate (ESR) and C reactive protein (CRP) to define response to therapy may be limited when using drugs directly inhibiting the interleukin-6 (IL-6) pathway. In addition, imaging has been applied to detect disease activity and damage, hence, this technique might be considered to assess treatment response, particularly when laboratory markers of inflammation are not reliable.

There is a need to standardise the definition of response to therapy in GCA, which would allow the detection of treatment effects in a more uniform manner, enabling better comparisons of study results across trials. Under the auspices of European Alliance of Associations for Rheumatology (EULAR) and American College of Rheumatology (ACR), an international task force was established with the goal of addressing this unmet need in GCA. This project aiming to define EULAR-ACR GCA response criteria will involve a multistep approach including (1) a systematic literature review (SLR) to retrieve criteria and descriptors used in previous studies on GCA to define a response to therapy, (2) a Delphi exercise to complete the descriptors selection, (3) a multicriteria decision analysis methods with 1000Minds exercise to reduce and weight the descriptors and (4) validation of the newly developed criteria in an observational study and, subsequently, a randomised controlled trial (RCT).

We here present an SLR with the objective to identify criteria and descriptors used to measure response to treatment and change in disease activity in GCA in RCTs and Longitudinal Observational Studies (LOS).

**METHODS**

This SLR was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidance. The scientific committee of the project (CAL, CD, ZT, SR, CS-A, MB and MS) formulated the key questions which were eventually transformed according to the PICO framework (Patients, Intervention, Comparator or Control, Outcome). These PICO questions were used as the basis of the SLR (online supplemental table S1). The population included patients with GCA, any intervention or comparator was considered and the outcomes addressed were: (1) disease activity state, namely active disease and remission; (2) change of disease activity state, namely response, improvement, worsening and relapse. Criteria or descriptors reflecting disease activity state and change thereof were collected together with their definitions. The individual descriptors were grouped into categories, for example, clinical components or laboratory components. When multiple articles on the same study population were identified, the study population was only counted once, however, all the reported outcomes of each manuscript were reviewed and all data were collected.

The definition of active disease was collected from all the included studies; in those RCTs where a precise definition of active disease was not provided, inclusion criteria were used as a surrogate for active disease. Any specific definitions of remission, for instance, sustained remission or any other ‘remission’ outcome, were also recorded. The definitions of the outcomes could also include imaging. A separate analysis of the studies Search strategies were developed by an experienced health science librarian (MA) who conducted the literature search using Ovid Medline, Embase and Cochrane Central Register of Controlled Trials from inception to 9 June 2022, with terms suggested by the team (CAL, CD, ZT, SR, CS-A, MB and MS) (online supplemental table S2). No language limits were applied to the searches. Draft searches were performed in Medline and tested against a sample of 16 citations provided by the steering committee. The final search captured all 16 citations confirming the validity of the search strategy. In Embase, conference materials were excluded except for those from ACR and EULAR abstracts (2019, 2020, 2021; relevant abstracts presented longer than 3 years earlier were expected to have been published during that interval and would therefore be included through that search).

All citations were uploaded to Covidence software. Duplicates were removed. Two independent reviewers (CS-A and DC) performed the title and abstract screening to assess eligibility according to the inclusion criteria. COVID-19 led to a change of duties of one of the investigators (DC), so two new investigators were included (MB and MS). The full-text review was subsequently performed by three reviewers (CS-A, MB and MS). Discordant cases were discussed among reviewers until a final consensus was achieved; if no agreement was accomplished, the two methodologists (SR and ZT) were consulted. The reliability of article selection and data extraction of the three reviewers (CS-A, MB and MS) was tested in a sample of 20% of the included articles. Given an agreement of >95% between the reviewers measured by comparing extracted data, the remaining articles were divided among the three reviewers and full-text review and data extraction were conducted independently.

Data extraction (CS-A, MB and MS) from selected studies was performed using a predefined data extraction form which was tested on a couple of studies. Items of interest included the population, demographics, intervention and individual components used to define: (1) disease activity state—active disease and remission; (2) change of disease activity state—response, improvement, worsening and relapse. Criteria or descriptors reflecting disease activity state and change thereof were collected together with their definitions. The individual descriptors were grouped into categories, for example, clinical components or laboratory components. When multiple articles on the same study population were identified, the study population was only counted once, however, all the reported outcomes of each manuscript were reviewed and all data were collected.

The definition of active disease was collected from all the included studies; in those RCTs where a precise definition of active disease was not provided, inclusion criteria were used as a surrogate for active disease. Any specific definitions of remission, for instance, sustained remission or any other ‘remission’ outcome, were also recorded. The definitions of the outcomes could also include imaging. A separate analysis of the studies
primarily focusing on imaging techniques was conducted to describe the imaging related outcome measurement instruments.

As we collected descriptors of the outcome measures used in studies, the risk of bias of the individual studies did not impact the conduct and conclusions of this SLR. Consequently, the risk of bias of the individual studies was not assessed. Outcomes and descriptors were analysed and are presented per category according to the study design (RCT, LOS and qualitative studies separately).

## RESULTS

The study selection process yielded 10,593 references. After removal of duplications, 7,457 remained for the title and abstract screening, and 488 were eventually selected for full-article review. One hundred and sixteen studies with individual study populations were included in our SLR: 11 RCTs, 104 LOS, 1 qualitative study (online supplemental figure S1).

Out of the 11 RCTs, 11/11 (100%) reported a definition of active disease; 10, 19–22, 25 and in 5 articles (45%), inclusion criteria were used as a surrogate of active disease. 16–18, 20–21

Eight out of 11 studies (73%) defined remission, 10, 19–22, 25 1/11 (9%) improvement 16 and 11/11 (100%) relapse (figures 1 and 2). Among observational studies, 20/104 (19%) provided definitions of active disease, 26–44 44/104 (42%) of remission, 15, 16, 18, 20, 21, 23–26, 28, 29, 31, 32, 34–57 4/104 (4%) of improvement, 34, 49, 58–59 and 90/104 (86%) of relapse (figures 1 and 3).

No study specifically defined ‘worsening’.

### Active disease in RCTs

Active disease was predominantly defined by a combination of clinical and laboratory components, both present in 10/11 (91%) of the studies (figure 1). The clinical component was described broadly in 5 of the 10 studies as ‘presence of signs and symptoms of GCA or polymyalgia rheumatica (PMR)’ whereas the other five RCTs specified the presence of cranial, PMR, ischaemic or constitutional symptoms as outlined in figure 2. The laboratory component was always associated with a clinical component in order to be considered a marker of disease activity. For the laboratory component, 1/10 articles applied a broad definition ‘abnormal inflammatory markers’ whereas the other nine RCTs used either the ESR alone (5/9, all published before 2017) or the ESR or CRP (4/9). Thresholds for abnormal ESR decreased from >40–50 before 2017 to >30 in more recent trials. Imaging was used as an optional component in 45% of the definitions of active disease: 4/5 (80%) used a general definition including new, recurrent or worsening imaging abnormalities compatible with vasculitis. Only one study relied on a specific definition for imaging (as an optional component) that included the presence of new vascular stenosis or aneurysm, in a new vascular territory seen by MR angiography (MRA), CT angiography (CTA) or conventional angiography. 23

### Active disease in LOS

In LOS, active disease was mostly defined as a combination of clinical and laboratory components (8/20, 40%) (figure 1, online supplemental table S3). However, 4/20 (20%) studies, 38, 40, 42–45 all published after 2020, used a combination of clinical, laboratory and imaging parameters.

The clinical component of active disease mainly applied a broad definition of symptoms (eg, clinical symptoms directly attributed to ongoing vasculitis), with only 4/17 studies providing specific symptoms (figure 3). Unlike RCTs, the definition of laboratory activity was non-specific (eg, elevated acute phase reactants, raised ESR), with only one (12%) study reporting cut-offs for ESR.
Active disease n=11 (100%)

- Clinical component n=10 (91%)
  - Specific criteria n=5 (50%), including cranial (headache, scalp tenderness, temporal artery abnormalities), PMR, ischemic (ischemic vision loss, stroke, abnormal pulses, claudication), constitutional symptoms (musculoskeletal pain, fever, fatigue), claudication.
  - General definition n=5 (50%), Presence of clinical signs and symptoms of GCA or PMR
- Laboratory component n=10 (91%)
  - Specific laboratory parameters n=9 (90%), ESR (≥ 30-50 mm/hr, CRP ≥ 10 mg/dL)
  - General definition n=1 (10%), Abnormalities in laboratory parameters.
- Imaging component n=5 (45%)
  - General definition n=4 (80%). New, recurrent, worsening imaging abnormalities or stenosis compatible with vasculitis.
  - Specific definition n=1 (20%). New vascular stenosis or aneurysm in a new vascular territory seen by MRA, CTA or conventional angiography.

Remission n=8 (73%)

- Clinical component n=8 (100%)
  - General definition n=8 (100%). Resolution/absence of clinical signs and symptoms attributable to GCA, absence of flare, absence of disease activity.
  - Laboratory component n=7 (88%)
  - General definition= Normalization of inflammatory markers n=7 (88%). ESR (<20-40 mm/hr), CRP (<10-15 mg/dL).
  - Imaging component n=1 (13%)
  - Definition included absence of disease activity (Defined as pre-established clinical and imaging criteria), no description of the specific characteristics of imaging findings for remission encountered.

Improvement n=1 (9%)

- Clinical component n=1 (100%).
- General definition= Absence of symptoms of GCA.
- Laboratory component n=1 (100%)
- General definition= Normalization of inflammatory markers.

Relapse n=11 (100%)

- Clinical component n=11 (100%)
  - General definition n=7 (64%). Recurrence of clinical symptoms attributable to GCA
  - Recurrence of specific symptoms of GCA n=3 (27%). Symptoms included cranial symptoms, ischemic symptoms (ischemic vision loss, TIA, stroke), PMR symptoms, constitutional symptoms.
  - Presence of disease activity n=1 (9%). Disease activity included the presence of cranial symptoms, ischemic symptoms (vision loss, TIA, Stroke), PMR and constitutional symptoms.
- Laboratory component n=9 (82%)
  - Elevation of inflammatory markers. ESR (>30-40 mm/hr), CRP (>10-15 mg/dL).
- Imaging component n=5 (45%)
  - New, recurrent, or worsening angiographic abnormalities compatible with vasculitis of the aorta or its primary branches n=2 (40%).
  - New or worsening imaging abnormalities suggestive of active vasculitis n=1 (20%).
  - MRA abnormalities n=1 (20%).
  - Presence of active disease n=1 (20%). Imaging features of active disease were the development of new vascular stenosis or aneurysm in new vascular territories as seen by magnetic resonance, computed tomography, or conventional dye arteriography
- Treatment component n=6 (55%)
  - Symptom resolution/favourable response with increase of steroids. Symptoms attributable to GCA necessitating reinstitution of steroids.

Figure 2 Components assessed in the definitions of active disease, remission, improvement and relapse across the included RCTs. CRP, C reactive protein; CTA, CT angiography; ESR, erythrocyte sedimentation rate; GCA, giant cell arteritis; MRA, MR angiography; PET, positron emission tomography; PMR, polymyalgia rheumatica; RCTs, randomised controlled trials; TIA, transient ischemic attack.

(≥30 mm/hour), and CRP (≥1 mg/dL). As in the RCTs, the imaging component was mostly applied in recent studies (published after 2020), that included new, recurrent or worsening angiographic abnormalities (CTA or MRA or positron emission tomography (PET)/CT).

Remission in RCTs

Overall, remission was defined by a combination of clinical and laboratory parameters (7/8, 88%). The clinical component was described as the absence or resolution of clinical signs and symptoms attributable to GCA, with two studies using ‘absence of flare’ to define remission.10 25 and one study using the ‘absence of disease activity’ without further specification.23 The laboratory component always required clinical remission and was never used as an isolated category. Laboratory remission was the normalisation of inflammatory markers. Cut-off
values for CRP ranged from 10 to 15 mg/L and for ESR from 20 to 40 mm/hour. The thresholds of ESR and CRP were lower in newer compared with older studies. The imaging component was only specified in the Langford 2017 study, which described remission as the absence of disease activity and included specific imaging features as reported above in the active disease section (figure 2 and figure 4).

Of the eight RCTs describing remission, three additionally specified sustained remission, and one complete remission. Sustained remission was considered as the maintenance of remission over time (with time spans ranging from 12 to 52 weeks) in combination with the adherence to a prespecified treatment protocol. Complete remission was the maintenance of clinical remission for 12 weeks after discontinuation of GC therapy.

Remission in LOS

Thirty-seven out of 44 (84%) studies reported a general definition of remission, and 5/44 (11%) of both (figure 3 and online supplemental figure S4). As in RCTs, sustained remission was mainly characterised by the maintenance of remission over time (with different time frames stipulated, ranging between 12 weeks and more than 2 years). Intriguingly, only one study specifically defined complete response, even though it resembled definitions of remission used in other studies (normalisation of both clinical, serological and imaging parameters after 12 months of therapy).

As in RCTs, remission was mostly defined by a combination of clinical (39/44, 89%) and laboratory parameters (26/44, 59%). While the clinical component (defined as either absence or resolution of clinical signs and symptoms attributable to GCA) was sometimes used as single category (8/39, 21%), the laboratory component (mostly defined as the normalisation of both ESR and CRP in 17/26, 65%) was never used alone to define remission.
A treatment component was described in 15/44 studies (34%), with discontinuation of treatment being required to define remission in 7/15 (47%) LOS. In the remaining 5/15 studies, a specific GC target dose had to be achieved to qualify for remission. Assessment by the investigator was sometimes used as the only parameter to define remission. In three out of these 6 (50%), it was objectively defined as a Physician Global Assessment of 0 (scale: 0–10). Unlike RCTs, some observational studies defined the imaging component of remission as either stability/no progression of imaging abnormalities (5/6, 83%) or normalisation (1/6, 17%) of the lesions, irrespective of the imaging technique used (US, fluorodeoxyglucose FDG-PET, CT, MRI and conventional angiography). As in RCTs, the maintenance of remission (time component, 13/44, 29%) was mostly applied to define sustained remission (11/13, 85%), with time spans ranging from 1 week to 1 year (online supplemental table S4).

### Improvement in RCTs

One RCT specifically defined improvement (figure 2), which was similar to the definitions of remission in other trials. It included a clinical component stipulating the absence of symptoms of GCA and a laboratory component described as the normalisation of inflammatory values. This study did not separately define remission.

### Improvement in LOS

Four LOS described improvement (figure 3). As in RCTs, definitions of improvement were very vague and difficult to differentiate from remission. The clinical (4/4, 100%) and laboratory (2/4, 50%) components were the only two categories used.

### Relapse in RCTs

A definition of relapse was encountered in all 11 RCTs. It was generally defined as a combination of clinical and laboratory parameters, mostly in combination with a treatment component. The term relapse was used in 8/11 (73%) RCTs, and flare in 3/11 (27%). These terminologies were used interchangeably across different studies.

The clinical component was defined as the recurrence of clinical symptoms attributable to GCA in all trials.
RCT studies. Villiger 2016\textsuperscript{22} classified relapse as major ‘if cranial symptoms were present’ and minor ‘in all other situations’. The laboratory component of relapse was encountered in 82% of studies and most often defined as an elevation of inflammatory markers. 4/9 (44%) considered elevation of ESR alone, 4/9 (44%) elevation of ESR and/or CRP, and 1/9 (11%) of CRP alone. Cut-offs for ESR ranged between 30 and 40 mm/hour and for CRP between 10 and 15 mg/L. The imaging component in relapse (5/11, 45%) was primarily defined by the presence of new, recurrent or worsening angiographic abnormalities compatible with vasculitis.\textsuperscript{23} Langford 2017 described relapse as the presence of disease activity including imaging features as described above in the active disease section.\textsuperscript{23} Treatment component (6/11, 55%) was referred to as symptoms or signs that necessitated reinstitution or an increase in GC (2/6, 33%) or the resolution or favourable improvement of these symptoms by increasing GC in 3/6 (50%) of the studies. In Stone 2017, the definition of a disease flare required an increment of the GC dose.\textsuperscript{10} The dose of the GC increase was not specified in any of the RCTs. Two studies\textsuperscript{18,23} specified that a relapse had to be confirmed by two independent physicians (figure 2 and figure 5).

**Relapse in LOS**

Relapse was mainly defined by a combination of clinical and laboratory parameters, often in conjunction with a

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Clinical</th>
<th>Laboratory</th>
<th>Imaging</th>
<th>Treatment increase/restart GC</th>
<th>Assessed by Investigator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jover 2001\textsuperscript{16}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Symptoms improve with GC</td>
</tr>
<tr>
<td>Spiera 2001\textsuperscript{17}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Symptoms improve with GC</td>
</tr>
<tr>
<td>Hoffman 2002\textsuperscript{18}</td>
<td></td>
<td>ESR ≥40/h</td>
<td>Angiographic abnormalities compatible with vasculitis</td>
<td></td>
<td>Features judged by 2 evaluating physicians and confirmed by the monitoring advisor committee.</td>
</tr>
<tr>
<td>Mazumzadeh 2006\textsuperscript{19}</td>
<td>ESR or CRP above normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoffman 2007\textsuperscript{20}</td>
<td>ESR≥40/h</td>
<td></td>
<td>New, recurrent, worsening angiographic abnormalities</td>
<td>Increase in dose of GC</td>
<td></td>
</tr>
<tr>
<td>Seror 2014\textsuperscript{21}</td>
<td>CRP &gt;15 mg/L</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Villiger 2016\textsuperscript{22}</td>
<td>ESR≥40 and CRP≥10 mg/L</td>
<td>MRA abnormalities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reichenbach 2018\textsuperscript{138}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Langford 2017\textsuperscript{23}</td>
<td>MSK symptoms or fatigue in combination with ESR &gt;40 mm/h</td>
<td>New vascular stenosis/aneurysm in new vascular territories seen by MRA, CTA, or arteriography</td>
<td>Symptoms/signs attributed to GCA that necessitate reinstitution or increase in GC</td>
<td>Determination of relapse assessed by both site investigator and study principal investigator and reaffirmed by study team at the end of trial.</td>
<td></td>
</tr>
<tr>
<td>Stone 2017\textsuperscript{10}</td>
<td>ESR≥30</td>
<td></td>
<td></td>
<td>Definition included the necessity for an increase in the GC. \textsuperscript{*}Except stone 2021: Part 2 extension GIACTA</td>
<td></td>
</tr>
<tr>
<td>Schmidt 2020\textsuperscript{24}</td>
<td>ESR≥30 or CRP ≥1 mg/dL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cid 2022\textsuperscript{25}</td>
<td>ESR≥30 or CRP ≥1 mg/dL</td>
<td>Imaging suggestive of disease activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>11/11 (100%)</td>
<td>9/11 (82%)</td>
<td>5/11 (45%)</td>
<td>6/11 (55%)</td>
<td>2/11 (18%)</td>
</tr>
</tbody>
</table>

\textbf{Figure 5} Components used in the definition of relapse in RCTs, reported separately for each study. CRR, C reactive protein; CTA, CT angiography; ESR, erythrocyte sedimentation rate; GC, glucocorticoids; GCA, giant cell arteritis; GIACTA, the Giant-Cell Arteritis Actemra (GiACTA) trial; MRA, MR angiography; RCTs, randomised controlled trials.
treatment change (38/90, 38%) (figure 3). With respect to terminology, 84 of the 90 (93%) studies used the word relapse, 11/90 (12%) flare, 8/90 (9%) recurrence and 1 reactivation. Similar to RCTs, these terminologies were applied synonymously across studies.

The clinical component of relapse was mostly described in broad terms, referring to the reappearance of disease-related signs and symptoms of GCA; specific clinical manifestations were reported only in four studies. The laboratory component of relapse frequently used general descriptors such as ‘elevation of inflammatory markers’ (35/69, 51%). Others were more specific including an increase in either ESR or CRP in 14/69 (20%) of studies or an isolated increase in CRP (6/69, 9%). Cut-offs were heterogeneous when compared with RCTs, but overall, ESR>40 mm/hour and CRP>5 mg/L were mostly used. Similar to RCTs, the treatment component always referred to either an increase or restart of GC, with only 6/58 (10%) studies establishing a specific cut-off dose (increase >10 mg of prednisone equivalent). The imaging component of relapse was overall under-represented with broad terminologies applied such as new or worsening lesions, irrespective of the imaging technique (CTA, MRA, US and FDG-PET).

**Qualitative studies**

There was only one qualitative study included (figure 1). The objective of this study was to explore the impact of GCA and its treatment on health-related quality of life and patient-specific reported outcome measures. The definition of remission (stable disease) was described by patients as having a resolution of GCA symptoms, stable blood monitoring and successful GC reduction. The definition of relapse included solely a clinical component, defined as an increase in disease activity.

**Imaging studies**

In total, 16 (14%) studies focused primarily on imaging. Most of them reported an imaging-based definition of active disease, while imaging-based definitions of remission or relapse were rarely considered. One study included an imaging-based definition of improvement.

PET: All of the PET studies used scoring systems to assess disease activity which included the PET Vascular Activity Score, the Total Vascular Score (TVS) or the target-to-background ratio (TBR). Active and inactive PET results were defined according to a visual inspection in nine studies by comparing the fluoro-deoxyglucose (FDG) uptake of the vasculature with the liver. Remission according to PET was described only in one study defining it as a TVS of 0 and a TBR of <1.34 at the thoracic aorta.

MRA: Disease activity was evaluated based on an adopted cerebral vasculitis grading scale in the study of Reichenbach 2018 (this was same study population as in Villiger 2016). In another study, disease activity was based on the general interpretation of vascular changes by MRI.

US: A scoring system was used in two studies. Bergner 2021 evaluated the intimal wall thickness and uptake of US contrast agent into the vessel wall. Das 2022 used a halo scoring system (0–3) to define improvement. Czihal 2015 did not use a grading scale and considered active vasculitis as the presence of a circumferential, hypoechochogenic wall thickening.

CTA: Hommada 2017 evaluated both PET and CTA. For CTA, active aortitis was defined as circumferential aortic wall thickness of ≥3 mm without calcification.

**DISCUSSION**

In GCA, several different definitions were used to assess the efficacy of treatments in RCTs and LOS. Only a single study provided a specific definition of response, while the majority of studies relied on remission and relapse as primary or secondary outcomes. The criteria and descriptors identified in this SLR will inform the next phases of this project to develop new EULAR-ACR response criteria, and clearly highlight the need to homogenise the primary endpoint of future clinical trials.

Remission was mainly defined as a combination of the resolution of signs and symptoms of GCA as well as the normalisation of inflammatory markers. Interestingly, some studies made a distinction between remission, sustained remission, and complete remission which reflected the same concept but with the main difference being the inclusion of different time and treatment components; this demonstrates again the lack of standardisation of definitions. The only definition of response in one LOS resembled the notion of remission used in other studies; this isolated definition does not directly establish response criteria. Relapse was mainly defined by recurrence of signs and symptoms of GCA, the elevation of inflammatory markers, and the reinstitution or increment of GC. The imaging component was considered infrequently in LOS but used in 45% of RCTs as an optional component to assess active disease and relapse. RCTs and LOS mainly used remission and relapse as study outcomes, and infrequently defined incomplete change in disease status, such as partial response in one study.

The clinical component was the most common element used across studies when defining the different disease states in GCA. This highlights that it is fundamental to accomplish clinical improvement when determining the response to therapy in patients with GCA. The definitions of clinical remission and relapse were typically general, leaving it up to the interpretation of the investigator to determine whether the symptoms were associated with GCA. For example, relapse was broadly defined as reappearance of GCA-related clinical manifestations, and remission was defined as absence of symptoms attributable to GCA.

The laboratory component was also commonly included in the definition of study outcomes in GCA. Inflammatory markers have long been fundamental to the assessment of disease activity in GCA; however,
these parameters are not specific and can be elevated by other causes (eg, infection). Other limitations include the occurrence of relapses despite normal acute phase reactants, and the development of medications targeting directly or indirectly the IL-6 pathway, rendering these markers of inflammation unreliable for the assessment of disease activity.6,140

An imaging component defining or contributing toward the assessment of treatment efficacy was used less frequently. One of the biggest limitations encountered was in determining an imaging definition of remission, which was not clearly outlined neither in RCTs nor in LOS. Imaging techniques have improved our understanding of GCA allowing us to determine the extent of disease; however, imaging has mainly been used to detect the appearance of new lesions rather than to evaluate the improvement of existing findings. The main limitations of imaging lie in the lack of standardisation on how to interpret findings across studies, in the question of whether active lesions always reflect ongoing inflammation (also despite clinical remission), and to what extent damage or remodelling contribute to a positive imaging result.

One interesting observation is the low frequency of patient-reported outcomes as study endpoints in GCA trials and LOS. In other diseases, such as Rheumatoid Arthritis or Spondyloarthritis, Patient Reported Outcomes (PROs) are not only included in the composite scores used as primary outcomes, but dedicated questionnaires have been developed to assess quality of life and disease impact experienced by patients.141 In GCA, the development of PROs is still ongoing and in GCA trials, a few generic PROs have been incorporated as secondary endpoints only.

One of the limitations of this SLR is that different individuals were involved in the different phases of this SLR; this is certainly not the standard approach and was a consequence of changed duties and personal developments during the COVID-19 crisis. This was mitigated by the fact that every step was conducted under the supervision of the methodologists who guaranteed the homogeneous conduct of all procedures and the consistency of one fellow throughout all phases of the project. Therefore, we do not think that this has impacted the results of our study. Another limitation is the absence of data on measurement properties of instruments in GCA as well as the absence of direct comparisons of the performance of different criteria in terms of measuring a response. We can, therefore, provide the EULAR-ACR task force only with a list of the individual descriptors as a basis for the further phases of the project, while at least some evidence on the psychometric properties of the descriptors would have been desirable. However, the literature on measurement properties of (validated) instruments in GCA is limited, so even less data is expected if we are considering individual criteria/descriptors and not only validated instruments.142 In the next phases of the project, this limitation will be taken into account and the task force will seek the best performing response criteria, that will then need to be validated and from which, measurement properties will be assessed.

In conclusion, the criteria and descriptors identified in this SLR provide insights into the current understanding of treatment outcomes and change in disease activity in RCTs and LOS of GCA and will be incorporated in future phases of the development of EULAR-ACR response criteria for GCA. These include a Delphi exercise to complete the descriptors identified by the SLR and, a multicriteria decision analysis methods with 1000Minds exercise to refine and weight the descriptors. Finally, a prospective study will be designed to validate the new response criteria.

Author affiliations
1Division of Rheumatology & Clinical Immunology, Department of Internal Medicine, University of Florida, Gainesville, Florida, USA
2Department of Rheumatology, Hospital of Bruneck, (ASAA-SABES), Teaching hospital of the Paracelsus University, Bruneck, Italy
3Department of Rheumatology, Mount Sinai Hospital, Toronto, Ontario, Canada
4Department of Rheumatology, Local Health Trust, Genoa, Italy
5Department of Library and Information Services, University Health Network, Toronto, Ontario, Canada
6Department of Rheumatic and Immunologic Diseases, Cleveland Clinic, Cleveland, Ohio, USA
7Rheumatology, Medical University Graz, Graz, Austria
8Department of Medicine, Division of Rheumatology, University of Toronto, Toronto, Ontario, Canada
9Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands
10Department of Rheumatology, Zuyderland Medical Center, Heerlen, The Netherlands
11Department of Rheumatology, Zuyderland Medical Center, Heerlen, The Netherlands

Twitter Catalina Sanchez-Alvarez @catysanchez08, Zahi Touma @ZahiTouma and Sofia Ramiro @sofiaramiro82

Acknowledgements The authors thank Andy Abril, Sibel Aydin, Frank Buttgeirer, Maria Cid, Bhaskar Dasgupta, Peter Grayson, William Lichliter, Rashid Luqmani, Bernhard Hellmich, Tanaz Kermani, Nader Khalidi, Sarah Mackie, Alfred Mahar, Eric Matteson, Mehrdad Maz, Peter Merkel, Paul Monach, Lorna Neill, Cristina Ponte, Carlo Salvarani, Wolfgang Schmitt, Peter Villiger, Ken Warrington and Madeline Whitlock for their valuable discussion about the challenges related to the development of response criteria in giant cell arteritis.


Contributors CS-A, MB and MS wrote the first version of the manuscript. All authors reviewed the manuscript, made extensive comments and appropriate changes to it and approved the final version of the manuscript. SR acted as the guarantor.

Funding This work was supported by a grant from the ACR and EULAR (grant number QoC 004).

Competing interests Dario Camellino, MD, PhD, MHA: consulting/speaker bureau: AbbVie, AbbVie. Boehringer Ingelheim, Galapagos, GSK, Novartis, Carol A. Langford, MD, MHS: Grant support from Bristol-Myers Squibb, GlaxoSmithKline, AstraZeneca and National Institutes of Health and served as a non-paid consultant to Bristol-Myers Squibb, AbbVie and AstraZeneca. Christian Dejaco, MD, PhD, MBI: Grant support by AbbVie and consulting/speaker’s fees from Abbvie, Eli Lilly, Janssen, Galapagos, Novartis, Pfizer, Sparrow, Roche and Sanofi. Zahi Touma, MD, PhD: Consulting and speaker fee for GSK, AstraZeneca, Merck, AbbVie, Eli Lilly and UCBI. Sofia Ramiro, MD PhD: Research grants from AbbVie, Galapagos, MSD, Novartis, Pfizer and UCBI and consultancy fees from AbbVie, Eli Lilly, MSD, Novartis, Pfizer, Sanofi and USB.

Patient consent for publication Not applicable.
REFERENCES


