HEALTH TECHNOLOGY ASSESSMENT

VOLUME 20 ISSUE 90 NOVEMBER 2016
ISSN 1366-5278

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

Raashid Luqmani, Ellen Lee, Surjeet Singh, Mike Gillett, Wolfgang A Schmidt, Mike Bradburn, Bhaskar Dasgupta, Andreas P Diamantopoulos, Wulf Forrester-Barker, William Hamilton, Shauna Masters, Brendan McDonald, Eugene McNally, Colin Pease, Jennifer Piper, John Salmon, Allan Wailoo, Konrad Wolfe and Andrew Hutchings

DOI 10.3310/hta20900

National Institute for Health Research

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

Raashid Luqmani,[1]* Ellen Lee,[2] Surjeet Singh,[1] Mike Gillett,[2] Wolfgang A Schmidt,[3] Mike Bradburn,[2] Bhaskar Dasgupta,[4] Andreas P Diamantopoulos,[5] Wulf Forrester-Barker,[1] William Hamilton,[6] Shauna Masters,[1] Brendan McDonald,[7] Eugene McNally,[1] Colin Pease,[8] Jennifer Piper,[1] John Salmon,[9] Allan Wailoo,[2] Konrad Wolfe[10] and Andrew Hutchings[11]

[1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[2]School of Health and Related Research, University of Sheffield, Sheffield, UK
[3]Immanuel Krankenhaus Berlin, Medical Centre for Rheumatology Berlin-Buch, Berlin, Germany
[4]Department of Rheumatology, Southend University Hospital NHS Foundation Trust, Southend, UK
[5]Department of Rheumatology, Hospital of Southern Norway, Kristiansand, Norway
[6]Primary Care Diagnostics, University of Exeter Medical School, Exeter, UK
[7]Department of Neuropathology and Ocular Pathology, Oxford University Hospitals NHS Trust, Oxford, UK
[8]Department of Rheumatology, Leeds Teaching Hospitals NHS Trust, Leeds, UK
[9]Oxford Eye Hospital, Oxford University Hospitals NHS Trust, Oxford, UK
[10]Department of Pathology, Southend University Hospital NHS Foundation Trust, Southend, UK
[11]Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

*Corresponding author

Declared competing interests of authors: Raashid Luqmani received honoraria from GlaxoSmithKline (GSK), Nordic and Chemocentryx for training in the use of the Birmingham Vasculitis Activity Score and Vasculitis Damage Index, and personal fees from Roche outside the submitted work. Raashid Luqmani received grants from Fundação para a Ciência e Tecnologia (Portugal), Canadian Institute of Health Research, Arthritis Research UK, Patient Centered Outcomes Research Institute, Oxford University Hospitals NHS Trust Innovation Challenge Competition and Vasculitis UK. Raashid Luqmani has patents pending for a mechanical arm to automate acquisition of ultrasound images and analysis for reviewing ultrasound images. Bhaskar Dasgupta received personal fees from GSK, Servier, Roche, Merck, and Mundipharma and grants from Napp outside the submitted work. Andrew Hutchings was funded by a Medical Research Council special training fellowship in health services research during the development of the study. Jennifer Piper has a patent pending for an ultrasound arm.

Published November 2016
DOI: 10.3310/hta20900

This report should be referenced as follows:

Luqmani R, Lee E, Singh S, Gillett M, Schmidt WA, Bradburn M, *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**(90).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents®*/ Clinical Medicine.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.058

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 08/64/01. The contractual start date was in January 2010. The draft report began editorial review in June 2015 and was accepted for publication in July 2016. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Abstract

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

Raashid Luqmani,[1]* Ellen Lee,[2] Surjeet Singh,[1] Mike Gillett,[2] Wolfgang A Schmidt,[3] Mike Bradburn,[2] Bhaskar Dasgupta,[4] Andreas P Diamantopoulos,[5] Wulf Forrester-Barker,[1] William Hamilton,[6] Shauna Masters,[1] Brendan McDonald,[7] Eugene McNally,[1] Colin Pease,[8] Jennifer Piper,[1] John Salmon,[9] Allan Wailoo,[2] Konrad Wolfe[10] and Andrew Hutchings[11]

[1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[2]School of Health and Related Research, University of Sheffield, Sheffield, UK
[3]Immanuel Krankenhaus Berlin, Medical Centre for Rheumatology Berlin-Buch, Berlin, Germany
[4]Department of Rheumatology, Southend University Hospital NHS Foundation Trust, Southend, UK
[5]Department of Rheumatology, Hospital of Southern Norway, Kristiansand, Norway
[6]Primary Care Diagnostics, University of Exeter Medical School, Exeter, UK
[7]Department of Neuropathology and Ocular Pathology, Oxford University Hospitals NHS Trust, Oxford, UK
[8]Department of Rheumatology, Leeds Teaching Hospitals NHS Trust, Leeds, UK
[9]Oxford Eye Hospital, Oxford University Hospitals NHS Trust, Oxford, UK
[10]Department of Pathology, Southend University Hospital NHS Foundation Trust, Southend, UK
[11]Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

*Corresponding author raashid.luqmani@ndorms.ox.ac.uk

Background: Giant cell arteritis (GCA) is a relatively common form of primary systemic vasculitis, which, if left untreated, can lead to permanent sight loss. We compared ultrasound as an alternative diagnostic test with temporal artery biopsy, which may be negative in 9–61% of true cases.

Objective: To compare the clinical effectiveness and cost-effectiveness of ultrasound with biopsy in diagnosing patients with suspected GCA.

Design: Prospective multicentre cohort study.

Setting: Secondary care.

Participants: A total of 381 patients referred with newly suspected GCA.

Main outcome measures: Sensitivity, specificity and cost-effectiveness of ultrasound compared with biopsy or ultrasound combined with biopsy for diagnosing GCA and interobserver reliability in interpreting scan or biopsy findings.

Results: We developed and implemented an ultrasound training programme for diagnosing suspected GCA. We recruited 430 patients with suspected GCA. We analysed 381 patients who underwent both ultrasound and biopsy within 10 days of starting treatment for suspected GCA and who attended a follow-up assessment (median age 71.1 years; 72% female). The sensitivity of biopsy was 39% [95% confidence interval (CI) 33% to 46%], which was significantly lower than previously reported and inferior to ultrasound (54%, 95% CI 48% to 60%); the specificity of biopsy (100%, 95% CI 97% to 100%) was superior to ultrasound (81%, 95% CI 73% to 88%). If we scanned all suspected patients and performed biopsies only on negative cases, sensitivity increased to 65% and specificity was maintained at 81%, reducing the need for biopsies by 43%. Strategies combining clinical judgement (clinician's assessment at 2 weeks) with the tests showed sensitivity and specificity of 91% and 81%, respectively, for biopsy and 93% and 77%, respectively, for ultrasound; cost-effectiveness (incremental net monetary benefit) was £485 per patient in favour of ultrasound with both cost savings and a small health gain. Inter-rater analysis revealed moderate agreement among sonographers (intraclass correlation coefficient 0.61, 95% CI 0.48 to 0.75), similar to pathologists (0.62, 95% CI 0.49 to 0.76).

Limitations: There is no independent gold standard diagnosis for GCA. The reference diagnosis used to determine accuracy was based on classification criteria for GCA that include clinical features at presentation and biopsy results.

Conclusion: We have demonstrated the feasibility of providing training in ultrasound for the diagnosis of GCA. Our results indicate better sensitivity but poorer specificity of ultrasound compared with biopsy and suggest some scope for reducing the role of biopsy. The moderate interobserver agreement for both ultrasound and biopsy indicates scope for improving assessment and reporting of test results and challenges the assumption that a positive biopsy always represents GCA.

Future work: Further research should address the issue of an independent reference diagnosis, standards for interpreting and reporting test results and the evaluation of ultrasound training, and should also explore the acceptability of these new diagnostic strategies in GCA.

Funding: The National Institute for Health Research Health Technology Assessment programme.

Contents

List of tables

List of figures

Glossary

Adventitia The outer layer of medium-sized and large arteries.

Amaurosis fugax A transient loss of vision, typically caused by a small embolic occlusion to the arterial supply to the retina or other parts of the visual pathway.

Arteriosclerosis Chronic changes in the arterial wall with thickening, fatty change and calcification typically associated with longstanding hypertension or cigarette smoking.

Axillary arteries Large arteries that are branches of the subclavian artery or innominate artery; they provide arterial supply to the arms and are detectable in the axillae (armpits).

Calcification The presence of deposits of calcium, typically detected in the larger arteries of patients with atherosclerosis or diabetes mellitus. They can be found in temporal arteries. On ultrasound they reflect sound, giving a bright image that is very different from a halo.

Claudication (of the jaw or tongue) Pain in the tongue or the masseter muscles of the jaw which is induced by exercise and is a result of a reduced blood supply. The pain should resolve with rest and is similar to angina in its mechanism.

Fragmentation The break up and duplication of the internal elastic lamina of the temporal artery as a result of giant cell arteritis or ageing. The histological appearance is best seen by staining the elastic, which is a major component of the internal elastic lamina.

Giant cell A large multinucleate cell found in sites of chronic inflammation. The presence of giant cells indicates that granulomatous inflammation is present but is not specific. The same cells are found in giant cell arteritis, other forms of vasculitis, tuberculosis and other chronic infections. If they are found in a biopsy from a patient with suspected giant cell arteritis, however, pathologists are very likely to diagnose giant cell arteritis.

Giant cell arteritis (also known as temporal arteritis) A disease that is characterised by inflammation of large and medium-sized blood vessels. An alternative name for this condition is 'temporal arteritis', as the blood vessels in the temple area of the head (sides of the forehead) are commonly affected. The giant cells referred to are specific collections of immune system cells seen in the areas of inflammation if a biopsy is performed.

Glucocorticoids Potent immunosuppressive corticosteroid therapy, which is used to treat many forms of inflammation. They are currently the main treatment for giant cell arteritis.

Halo An ultrasound finding of a dark shadow adjacent to a blood vessel, which may represent inflammation in the vessel. It is the strongest single indicator of the presence of vessel wall inflammation seen in patients with large vessel vasculitis such as giant cell arteritis or Takayasu's arteritis.

Immunosuppressive agents Drugs that suppress the immune system, typically for treatment of patients with inflammatory conditions such as giant cell arteritis. They include glucocorticoid, methotrexate, azathioprine, cyclophosphamide, ciclosporin and leflunomide.

Internal elastic lamina The histological structural layer in medium-sized and large arteries that separates the innermost layer (intima) from the middle layer (media).

Intima The innermost layer of medium-sized and large arteries.

Intimal hyperplasia Increased numbers of cells (usually inflammatory) present in the intima. The intima is often swollen (increased in thickness) because of accompanying oedema in this layer, as seen histologically. This is a typical feature in patients with active giant cell arteritis but may also be seen in patients with resolving disease, as well as in otherwise healthy older adults.

Occlusion (biopsy) Complete blockage of blood flow through a vessel, usually because of significant intima hyperplasia and or thrombus. This is a typical histological finding of vessel wall inflammation.

Occlusion (ultrasound) A lack of colour Doppler flow through an artery, which is attributed to occlusion.

Reduplication The increased number of apparent layers of internal elastic lamina seen on histology in temporal arteries. The finding is typical in giant cell arteritis, but can also occur in otherwise healthy elderly people.

Stenosis Narrowed sections of an artery as demonstrated on ultrasound. It is characterised by visible narrowing but also by an accelerated rate of colour Doppler flow through the area of stenosis. It is found in patients with giant cell arteritis but is also seen in other conditions such as arteriosclerosis.

Systemic vasculitis A group of diverse and unusual conditions characterised by inflammation of the vessel wall leading to organ or tissue infarction, affecting multiple organs or occurring throughout the vasculature.

Temporal arteries Branches of the external carotid arteries, which supply the scalp with blood. Branches of this artery also supply the retina, which means that narrowing of this artery can lead to critical ischaemia of the central part of the retina, which may result in permanent visual loss.

Vasculitis Inflammation of blood vessels leading to organ or tissue damage as a result of ischaemia.

List of abbreviations

ACR	American College of Rheumatology
AE	adverse event
ANCA	antineutrophil cytoplasm antibody
BSR	British Society for Rheumatology
BVAS	Birmingham Vasculitis Activity Score
CI	confidence interval
CRF	case record form
CRP	C-reactive protein
CT	computerised tomography
DCVAS	Diagnostic and Classification Criteria for Vasculitis Study
EGPA	eosinophilic granulomatosis with polyangiitis
EQ-5D	EuroQol-5 Dimensions
ESR	erythrocyte sedimentation rate
GCA	giant cell arteritis
GP	general practitioner
GPA	granulomatosis with polyangiitis
ICC	intraclass correlation coefficient
ICER	incremental cost-effectiveness ratio
IL	interleukin
IQR	interquartile range
MRI	magnetic resonance imaging
NICE	National Institute for Health and Care Excellence
NMB	net monetary benefit
PMR	polymyalgia rheumatica
QALY	quality-adjusted life-year
R&D	research and development
SD	standard deviation
TAB	temporal artery biopsy
TABUL	The Role of Ultrasound Compared to Biopsy of Temporal Arteries in the Diagnosis and Treatment of Giant Cell Arteritis
TTO	time trade-off
VDI	Vasculitis Damage Index

Plain English summary

Giant cell arteritis (GCA) is a disease causing blood vessel inflammation which, if left untreated, can cause permanent blindness. Patients with suspected GCA usually have a minor surgical procedure that involves taking a biopsy from one of the arteries on the side of the head. A positive biopsy confirms the diagnosis, but many patients with negative biopsies are eventually diagnosed with GCA. We compared the accuracy and cost of an alternative test for GCA, namely an ultrasound scan of arteries, with taking a biopsy. We scanned and biopsied 381 patients with suspected GCA and followed them for up to 6 months to see who actually had GCA; 257 (67%) patients were eventually diagnosed with GCA. Ultrasound was better than biopsy at identifying patients who did have GCA: it identified 54% of these patients compared with 39% identified from biopsy. Biopsy performed better than ultrasound in the patients who did not have GCA: none of these patients had a positive biopsy, whereas 19% had a positive scan.

We also looked at different testing strategies combined with a doctor's assessment of the patient. A strategy that involves scanning all patients with suspected GCA identified 93% of those patients with GCA. This strategy was also cheaper (by £485 per patient) than the current practice of relying on a doctor's assessment and biopsy alone.

Scientific summary

Background

Giant cell arteritis (GCA) is a relatively common form of blood vessel inflammation, which usually affects people over the age of 50 years. GCA typically causes headaches and systemic upset, but can be associated with sudden and irreversible sight loss. For this reason, if a general practitioner (GP) sees a patient with suspected disease, the patient will usually be commenced on high doses of glucocorticoids, often before the diagnosis has been confirmed by further testing. Therefore, it is important to make the diagnosis correctly in order to decide on the need to continue high-dose glucocorticoid therapy to improve the condition and reduce the risk of visual loss. However, it is also important to avoid treating those without the condition, because there is a very high incidence of side effects associated with long-term glucocorticoid therapy. Temporal artery biopsy is the current gold standard test for establishing the diagnosis, with a high specificity but low sensitivity. It can be misleading in a significant number of cases. Up to 44% of patients with clinical features of GCA have a negative biopsy. There are many reasons for this, including the adequacy of the specimen obtained, the duration of glucocorticoid treatment prior to biopsy and the presence of skip lesions (intermittent, dispersed areas of abnormality in the artery that might be missed because not all areas of the artery will be sectioned for examination). Ultrasound and other imaging techniques are emerging as alternative tests to biopsy but have not been taken up widely. Ultrasound imaging can be used to assess both temporal arteries as well as both axillary arteries, which has been shown to increase the diagnostic yield.

Objectives

We aimed to test the clinical effectiveness and cost-effectiveness of ultrasound as an alternative to biopsy in the diagnosis of patients with a new presentation of possible GCA. The primary objectives of the study were (1) to evaluate the diagnostic performance (sensitivity and specificity) of ultrasound as an alternative to biopsy for diagnosing GCA in patients who are referred with suspected GCA and in whom a biopsy was going to be carried out; and (2) to perform a cost-effectiveness analysis to compare different potential investigation strategies for diagnosing GCA, incorporating either or both ultrasound and biopsy.

The secondary objectives in the study were to evaluate:

- sequential diagnostic performance and cost-effectiveness of biopsy following ultrasound in patients who have a negative ultrasound, compared with either ultrasound or biopsy alone
- the clinical effectiveness and cost-effectiveness of providing ultrasound results or biopsy results alone on treatment decisions proposed by participating clinicians
- the diagnostic performance of ultrasound in specific subgroups such as individuals at high or low risk of disease and the level of variation in ultrasound appearance in terms of halo size and degree of stenosis.

Methods

We conducted a multicentre, prospective study of new cases of suspected GCA. In order to ensure standardisation of the new technique of ultrasound scanning of temporal arteries, we needed to develop a training programme for sonographers to standardise the performance and interpretation of ultrasound assessment of temporal and axillary arteries. The training consisted of a Microsoft PowerPoint® version 97–2003 (Microsoft Corporation, Redmond, WA, USA) presentation, an online assessment in which sonographers were required to correctly identify video images of scans as showing or not showing features

consistent with GCA (75% pass mark required), the provision of video and still images from 10 control individuals and one patient with active GCA and evidence of ultrasound abnormalities consistent with the diagnosis. All images were reviewed by an expert panel. If scan techniques were suspected to be inadequate, the sonographers underwent retraining. We compared the standard of care in the investigation of GCA [clinical evaluation, measurement of acute phase response and temporal artery biopsy (TAB) from the most affected artery] with ultrasound. All patients underwent both tests in sequence (ultrasound first, followed by biopsy) within 7 days of commencing high-dose steroids for the suspected diagnosis of new-onset GCA. We did not provide any training in TAB or in the interpretation of the results by pathologists, because these are established techniques in NHS care. All patients received normal care as decided by their clinician. The clinician reviewed the patient at baseline, 2 weeks later (after the scan and biopsy had been performed) and 6 months later. We used the clinician diagnosis made at 2 weeks as the primary outcome measure for the study. We also established a reference standard diagnosis based on the clinician's submitted diagnosis, any revisions by 6 months, or any revisions by the expert panel who reviewed all case records. The scan result was kept blinded from the clinician treating the patient to avoid the results influencing the decision on diagnosis. The clinician was supplied with all other results including the biopsy because this would reflect standard of care. After the clinician submitted their diagnosis at 2 weeks, they were provided with the scan results on request if they were considering withdrawal of therapy if they concluded that the patient did not have GCA. We excluded patients with a previous diagnosis of GCA or a previous TAB; we also excluded patients who had been treated with high-dose steroids for any other reason for more than 7 days prior to the ultrasound and biopsy. We created 30 case vignettes from the data obtained from patients in the study who either did or did not have a reference standard diagnosis of GCA and in whom the biopsy or scan could have been either positive or negative. We undertook two interobserver rater exercises to assess variation in diagnosis based on interpretation of ultrasound and biopsy images.

Results

We recruited 35 centres for the study and provided ultrasound training for 49 sonographers. Only 26 sonographers from 22 sites completed the training. Two of these sites did not take any further part in the study, leaving 24 sonographers at 20 recruiting sites. Seven sonographers passed all three components of the training at first attempt; 13 required further attempts to pass; and four were exempted from some parts of the training because they had already demonstrated expertise.

We enrolled 430 patients with suspected GCA into the study; 44 withdrew or did not have their scan and biopsy within 10 days of starting glucocorticoid therapy and five withdrew before completing a follow-up assessment. The remaining 381 patients were included in the primary analysis. The median age [interquartile range (IQR)] was 71 years (64–78 years); 72% were female. The median time between first symptom onset and baseline was 31 days (IQR 10–93 days, $n = 377$); the median time between symptom onset and starting steroids was 33 days (IQR 13–99 days, $n = 379$). The reference diagnosis was based on the 2-week and 6-month clinical diagnosis, as well as the opinion of an expert review panel, which assessed all of the patient data apart from the ultrasound results. In total, 257 out of 381 patients were considered to have GCA on the basis of the reference standard diagnosis. Twenty-one patients had a change of diagnosis from the original clinician's final available diagnosis following an expert review: in eight patients, the diagnosis was changed from GCA to not GCA; in a further 13 patients, the diagnosis was switched from not GCA to GCA. We compared the relative performance of ultrasound and biopsy in 257 patients with a reference standard diagnosis of GCA and 124 patients diagnosed as not having GCA. The incidence of polymyalgia rheumatica was 28 cases at baseline, and ischaemic optic neuropathy was reported in 9.7% of cases with GCA at baseline (compared with 4.8% in the non-GCA group), 6.2% of cases at 2 weeks (compared with 0.8%) and 4% of cases at 6 months (compared with 2.8%). Baseline comorbidity included hypertension in 52.5% of patients, diabetes mellitus in 14.2% of patients, ischaemic heart disease in 7.3% of patients, heart failure in 5% of patients, malignancy in 2.4% of patients and fracture in 0.3% of patients. At 6 months after diagnosis, the incidence of hypertension had increased to

55.8% and the incidence of diabetes mellitus had increased to 18.2%. Four low-trauma fractures occurred during the follow-up period.

In total, 101 cases had a biopsy consistent with the diagnosis and 162 cases had an ultrasound consistent with the diagnosis. In 70% of patients, the results of biopsy and ultrasound were concordant (74 positive and 192 negative), giving a kappa statistic of 0.35. In 27 patients the results were positive for biopsy but negative for ultrasound and in 88 patients the results were negative for biopsy but positive for ultrasound. The sensitivity of biopsy versus ultrasound was 39% [95% confidence interval (CI) 33% to 46%] compared with 54% (95% CI 48% to 60%) and the specificity was 100% (95% CI 97% to 100%) compared with 81% (95% CI 73% to 88%). We analysed change in the degree of ultrasound abnormality (based on halo size) during the 7-day assessment period. We found that the halo was likely to be much smaller after at least 4 days of high-dose glucocorticoid therapy than after fewer than 4 days of steroid therapy. The biopsy positivity rate also diminished significantly within 3 days of starting high doses of glucocorticoids. We also evaluated alternative strategies for combining the two tests and incorporating clinical judgement (the doctor's assessment of GCA based on the patient's characteristics and available test results). The most cost-effective strategy was to perform an ultrasound examination of all patients with suspected GCA. This strategy was more sensitive (93% vs. 91%), less specific (77% vs. 81%) and more cost-effective (incremental net monetary benefit of £485 per patient) than current standard practice, that is, a strategy involving biopsy and clinical judgement alone. The cost-effectiveness analysis accounted for the cost of the testing and the consequences of correct or incorrect diagnosis resulting in drug toxicity (e.g. fracture), as well as irreversible, potentially preventable, sight loss from anterior ischaemic optic neuropathy.

We measured inter-rater agreement for the two tests in a series of 30 cases selected from among the cohort. Among sonographers the intraclass correlation coefficient for agreement was 0.61 (95% CI 0.48 to 0.75) and for biopsy the intraclass correlation coefficient was 0.62 (95% CI 0.49 to 0.76). Agreement was strongest where the pathological findings included the presence of giant cells, but was much weaker for cases in which only minimal change was found. This suggests that the current approach of classifying test results as either positive or negative may be too simplistic because the tests are not always enough by themselves to make the diagnosis, as there can be some degree of uncertainty in interpreting the results.

Conclusions

We conclude that ultrasound, in comparison with TAB, is a more sensitive and cost-effective investigation in suspected cases of GCA. However, over one-third of the patients eventually diagnosed with GCA had neither a positive scan nor a positive biopsy, highlighting the importance of assessing the patient for clinical indicators to support the diagnosis rather than relying on test results alone. Temporal artery biopsy has a much lower sensitivity in the current study (39%) than in previously published figures (> 70%). It is tempting to speculate that this may reflect better awareness of the diagnosis of GCA and a willingness of GPs to commence treatment early in the disease course. This is supported by the relatively short time from symptom onset to diagnosis in this cohort (31 days) compared with other cohorts. The findings have potential implications for improving the management of GCA through the more effective use of available techniques to provide a clinically effective and cost-effective strategy for the diagnosis of GCA. Further research should address the issue of an independent reference diagnosis, standards for interpreting and reporting test results and the evaluation of ultrasound training and should explore the acceptability of these new diagnostic strategies in GCA. Some clinicians and patients may be uncomfortable with a strategy that does not involve a biopsy and may prefer to perform biopsies in all cases that are ultrasound negative, or in all cases of patients at medium or high risk of GCA in terms of clinical features but who have a negative ultrasound scan, to provide further evidence to rule in the disease as well as to support withdrawing therapy if both tests are negative. Although these combined strategies would be more expensive than our proposal (because a proportion of patients would require a biopsy), they remain more cost-effective than current practice (performing a biopsy in all suspected cases) and may be more acceptable to patients and clinicians.

Future work

Further research should address the issue of an independent reference diagnosis, standards for interpreting and reporting test results and the evaluation of ultrasound training and should explore the acceptability of these new diagnostic strategies in GCA.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Chapter 1 Introduction

General introduction to giant cell arteritis

Giant cell arteritis (GCA), also known as temporal arteritis, is a common form of vasculitis that affects people typically aged > 50 years.[1] GCA often progresses rapidly and, if left untreated, leads to severe pain, permanent visual loss, stroke and, in some cases, death. The incidence is approximately 220 per million per year in the UK in people aged ≥ 40 years.[2] Elsewhere, the incidence varies across the world, with published figures ranging from 150 to 250 new patients per million per year. It is more common in northern European countries, particularly in Scandinavia (313 per million per year in people aged > 70 years)[3] and in Minnesota, USA, which has a large Scandinavian-origin population (198 per million per year),[4] and it is much less common in other parts of the world such as Japan, China and Australia.

Rapid diagnosis and glucocorticoid treatment are recommended,[5] but both are problematic. Glucocorticoid treatment is usually started before a formal diagnosis is made, meaning that a proportion of patients are treated unnecessarily and are thereby exposed to side effects including weight gain, altered body habitus, hypertension, infection, osteoporosis, cataract, mood swings and thin skin. Glucocorticoid treatment also affects the accuracy of the diagnosis. The heterogeneous nature of GCA means that its diagnosis is not straightforward, but is usually based primarily on temporal artery biopsy (TAB) and supported by presenting symptoms. Glucocorticoids, by their nature, impact on inflammation; if there is a large time difference between commencement of glucocorticoids and biopsy, this reduces diagnostic accuracy. Although a positive biopsy usually (although not always) confirms GCA, the sensitivity of TAB has been estimated to vary from 39% to 91%,[6,7] resulting in a large number of false negatives in the screened population. This has led to high-dose glucocorticoid therapy being continued as a precaution (in case the patients actually have GCA), even in the absence of a positive biopsy.

Ultrasound and other forms of imaging compared with the traditional role of biopsy

An alternative to biopsy has been the development of ultrasound and other imaging techniques for the diagnosis of GCA. Imaging first emerged in the 1990s as a potential means by which to provide evidence to support a diagnosis of GCA.[8–15] High-resolution magnetic resonance imaging (MRI) of temporal arteries offers a non-invasive technique for investigating suspected GCA, but it is limited by availability and cost. Ultrasound is the most practical and widely used modality. Three meta-analyses have supported the role of ultrasound in the diagnosis of GCA.[16–18] The presence of bilateral ultrasound abnormalities (both temporal arteries involved) provides high specificity (100%) for the diagnosis of GCA, but its sensitivity was 43%.[17] Two of the meta-analyses reported concerns with the quality of the included studies[16,18] and the third did not assess the methodological quality of the included studies.[17] Currently, the use of ultrasound as a diagnostic tool for GCA is relatively limited, perhaps as a result of practical reasons relating to training to use ultrasound or equipment availability to facilitate rapid access and evaluation of patients with suspected GCA.

Ultrasound examination of temporal arteries is non-invasive and there is no ionising radiation involved. Furthermore, it can provide information about the entire length of both temporal arteries. Additional examination of the axillary arteries improves the sensitivity of ultrasound[19] because some individuals with GCA (especially those without headaches) will have isolated abnormalities in the axillary arteries but not in the temporal arteries. This may be because of the longer persistence of scan abnormalities in larger vessels than in temporal arteries, despite the use of steroids. The chief abnormality on ultrasound that suggests the diagnosis of GCA is a halo, which is defined as a dark hypoechoic area around the vessel lumen and is

thought to represent inflammatory change and oedema present in the wall and surrounding tissues of the affected blood vessel.

The role of temporal artery biopsy in the diagnosis of giant cell arteritis

A recent study of biopsy-proven GCA disease in South Australia suggested an incidence in people aged > 50 years of only 32 per million per year,[20] although the relatively low incidence may be a result of the inclusion of biopsy-confirmed cases only. Biopsy for the diagnosis of GCA has a relatively low yield.[16] The difficulty in diagnosis of GCA, which forms the main underlying question of this project, is the lack of a high-quality gold standard test. Although biopsy is reported to be the current gold standard test for diagnosis, the majority of patients in whom a diagnosis of GCA is suspected do not actually have a positive result. This may reflect the fact that there is a lower index of suspicion for diagnosis, and, therefore, more people with headaches are being evaluated for GCA; equally, it may reflect the relatively poor association between the true multivessel disease of GCA and the TAB findings to support a diagnosis of GCA.

The spectrum of different forms of giant cell arteritis

In about 50% of patients with GCA, branches of the aorta, and even the aorta itself, may be involved, suggesting that there is probably much more widespread inflammation of blood vessels (vasculitis) than previously considered.[21] In a study of 120 patients with large vessel vasculitis and 212 with more conventional cranial symptoms of GCA, but without the evidence of large vessel disease, patients with large vessel disease were significantly younger, by about 7 years, and had longer duration of symptoms prior to diagnosis (3.5 months compared with 2.2 months). There was a strong association with pre-existing polymyalgia rheumatica (PMR) in 26% of patients, compared with 15% of patients with cranial GCA, and fewer cranial symptoms (41% of patients, compared with 83% of patients with cranial GCA). Visual loss was also much less likely in large vessel GCA (4% compared with 11%). The risk of relapse of GCA features was higher in patients with large vessel disease than in patients with cranial manifestations only (4.9/10 person-years, compared with 3/10 person-years) and these patients were likely to require higher doses of steroids for longer periods of time.

Clinical presentation of giant cell arteritis

Recognising new features of GCA can be very straightforward in a patient with no previous history of headache who suddenly develops unaccustomed discomfort on the side of the head with swelling or tenderness of the temporal arteries, general systemic onset and scalp tenderness. Although symptomatic headache is very troublesome, the most feared complication of GCA is neuroischaemic damage, which can result in inflammation and occlusion of small branches of the cranial arteries (including the posterior ciliary artery and the ophthalmic artery) and ultimately in permanent visual loss. A warning symptom of ischaemia is the presence of jaw or tongue claudication, typically reported by around 50% of patients with GCA at presentation.[22] Jaw and tongue claudication refers to discomfort in the patient's masseter muscles or tongue which stops them from eating or talking. When they stop to rest their jaw or tongue, the pain resolves because it is a result of claudication of those muscles as a result of narrowing of the blood vessel supply (e.g. the facial artery and its branches). Patients who have tongue or jaw claudication are at risk of blindness because the disease can involve the posterior ciliary arteries, which supply the retina and cause unilateral, and occasionally bilateral, permanent sight loss. Sometimes the visual loss starts on one side and subsequently becomes bilateral. In an early series of 90 cases from neurology and ophthalmology clinics, up to 60% of patients presented with permanent loss of vision attributable to either ischaemic papillopathy or retinal artery occlusion;[23] other cases presenting primarily to physicians demonstrated a much lower risk of sight loss of 7.4% to 19.1%.[24–27] The risk of sight loss associated with GCA has fallen

in recent decades, from 15% of cases with ischaemic optic neuropathy between 1950 and 1979 to only 6% between 1980 and 2004.[28] There is a small but significant risk of stroke, highlighted as possibly being between 2.8% and 6% in two recent studies.[29,30] Therefore, an early diagnosis and initiation of immunosuppressive treatment with high doses of steroids is required, making the condition a medical emergency.

It is likely that the mechanisms driving the neuroischaemic complications are different from those driving the systemic inflammatory response. There is a suggestion that interleukin (IL)-12 and interferon-gamma are the main cytokines responsible for myointimal proliferation leading to vessel occlusion; by contrast, the mechanisms driving the systemic inflammatory response are likely to be IL-6 and IL-17.[31] The underlying pathological changes involve invading macrophages and lymphocytes which gain access to the blood vessels via the vasa vasorum. They generate a local inflammatory response in the blood vessel wall, starting in the adventitia, migrating through to the media and intima, with proliferation of the internal elastic lamina, intimal proliferation and swelling, and eventually resulting in vessel narrowing and complete occlusion in some cases.

Although intimal proliferation with intimal thickness and internal elastic lamina reduplication are typical features, the hallmark histological finding is the presence of multinucleated giant cells, hence the term GCA. These pathological mechanisms are the basis for the histological diagnosis of the condition, which was first recognised by Horton *et al.*[1] in 1932 who described two patients who were initially thought to have a fungal infection (actinomycosis) of the temporal arteries.

Giant cell arteritis typically affects people aged over 50 years; it is two to three times more common in women than in men. PMR is a related clinical syndrome characterised by generalised muscles aches and pains. PMR is common in patients over the age of 50 years and presents with widespread aches and pains, particularly involving proximal muscles. Criteria for classifying PMR are based on the presence of bilateral limb girdle discomfort, early-morning stiffness and an elevated inflammatory response.[32] Additional ultrasound evidence of bursitis around the hips improves the specificity of the criteria from 78% to 81% and maintains sensitivity of between 66% and 68%. PMR can be present in up to 50% of individuals with GCA and it may occur either before, during or after the manifestations of GCA appear, suggesting significant overlap between these two disease processes.[33] Therefore, our interpretation of any individual patient's diagnosis of GCA would be influenced by either pre-existing or concomitant diagnosis of PMR or might be validated by subsequent development of PMR.

Diagnosis and classification of giant cell arteritis

The 1990 American College of Rheumatology (ACR) classification criteria for GCA are based on the following:

- aged at least 50 years
- new onset of headache
- temporal artery abnormality on physical examination
- elevated erythrocyte sedimentation rate (ESR) typically $\geq$ 50 mm/hour
- abnormal TAB showing features of vasculitis.

Classification of a patient as having GCA requires at least three of these criteria to be present.[34] The classification criteria are not diagnostic tests and are limited by the technology available at the time when the criteria were being developed. As technology has improved, there are more sophisticated methods available for evaluating the temporal artery with ultrasound, MRI and computerised tomography (CT). Furthermore, it is possible to image the whole arterial tree more effectively for evidence of widespread vascular abnormality using CT angiography, magnetic resonance angiography and ^{18}F-fluorodeoxyglucose positron emission tomography CT. These techniques have revealed that some cases of GCA have much

more extensive vessel involvement than previously suspected.[21] Imaging has demonstrated that GCA can present without headaches but with other features such as constitutional symptoms and polymyalgia, which is also termed polymyalgia arteritica.[35]

The awareness of GCA has probably increased and it is likely that the concern regarding the threat of visual loss may affect a clinician's decision to pre-emptively treat any patient who might have the condition as soon as possible, in order to prevent these complications from occurring. As a result, it is likely that we are starting to see a change in the level of suspicion of symptoms at which a clinician is confident in starting treatment on the basis of a presumed diagnosis of GCA. Tests used for diagnosing GCA would now be performed in different circumstances than existed previously. We may be dealing with milder cases of the disease and/or more patients with a GCA-like symptom complex who do not actually have GCA. If these patients are given steroids, the standard test result from a TAB may be significantly influenced by the fact that the biopsy was performed on mild disease that had already been partly treated and/or was performed in patients who do not actually have GCA. Kisza *et al.*[36] assessed over 700 cases of GCA from 1994 to 2011, with 215 biopsy-positive cases, observing a peak incidence in 1996. Machado *et al.*[37] observed a reduction in the frequency of patients presenting with classical features, but no change in the likelihood of a positive biopsy from 1950 to 1985. In fact, Gonzalez-Gay *et al.*[38] found that the incidence of biopsy-proven GCA actually increased from 1981 to 2005. A more recent study[3] of 840 biopsy-positive cases of GCA in Sweden reported a reduction in incidence between 1997 and 2010, from 15.9/100,000 to 13.3/100,000, although this contrasts with an earlier Swedish study[39] that reported an increased incidence from 1976 to 1995, especially in women. In the UK, there was no evidence to suggest a change in incidence between 1990 and 2001.[2] This suggests that, although there may have been some changes to the epidemiology of GCA over time, with possibly a rise in incidence in women, there has been no significant change in the overall incidence of GCA. There is evidence of a diagnostic shift in other diseases too; hypothyroidism is now recognised as significant and is associated with increased comorbidity at lower levels of thyroid-stimulating hormone than before.[40]

The Diagnostic and Classification Criteria for Vasculitis Study

As a result of concerns about the classification and diagnosis of GCA and other forms of vasculitis, an international effort to improve criteria for the diagnosis of vasculitis has been under way since 2009.[41] The Diagnostic and Classification Criteria for Vasculitis Study (DCVAS) had, by 2015, recruited over 4000 patients with either a form of vasculitis or a comparator condition and included over 900 individuals with a clinical diagnosis of GCA. Patients are recruited if they have any clinical features that might be consistent with vasculitis. This includes patients who do not actually have vasculitis, because they are considered to be part of the comparator population for the study. Patients are either newly diagnosed with vasculitis or a comparator condition or have had a diagnosis made within 2 years of recruitment into the study. A detailed pro forma is used to report standardised information regarding symptoms, signs and test results (including blood tests, imaging and biopsy data) available at the time of diagnosis. A subsequent follow-up visit 6 months later is required, so that any change in the original diagnosis can be reported and used as the final submitting clinician's diagnosis. The DCVAS study has not yet reported results, but limited access to the DCVAS data was granted for The Role of Ultrasound Compared to Biopsy of Temporal Arteries in the Diagnosis and Treatment of GCA (TABUL) study.

Difficulty with diagnosis of giant cell arteritis based on the gold standard temporal artery biopsy

For clinicians managing patients who may have GCA, untreated disease can result in permanent visual loss (as discussed in González-Gay *et al.*[24]) and the condition is therefore considered to be a medical emergency. However, there are far more people with headache (it is an almost universal experience) than there are patients with GCA.

Toxicity of treatment versus need for urgent treatment

The other main consideration is that treatment for GCA, which involves high doses of glucocorticoids such as prednisolone over a prolonged period and which will result in rapid control of the inflammatory process and reduce the risk of ischaemic manifestations, is very toxic and results in side effects in over 80% of patients.[42] The most common side effects reported in the study by Proven *et al.*[42] included cataracts in 41% of patients, fractures in 38% of patients, infection in 31% of patients, hypertension in 22% of patients, diabetes mellitus in 9% of patients and gastrointestinal bleeding in 4% of patients.

With modern therapy, such as the prophylactic use of calcium, vitamin D and bisphosphonates to prevent fractures, some of these complications can be avoided. Further measures to reduce risk of treatment-related toxicity include better control of hypertension and diabetes mellitus, as well as prophylactic use of proton pump inhibitors to prevent gastrointestinal bleeding (which could relate to previous use of high doses of non-steroidal anti-inflammatory drugs combined with high doses of prednisolone). The risk of serious infections remains significant and has been estimated to be 55% higher than in age- and sex-matched controls.[43]

Therefore, the balance of risk versus benefit in a patient with suspected GCA rests heavily on our ability to be confident that the diagnosis is correct. A patient who is incorrectly diagnosed with GCA will be subjected to significant risk of steroid toxicity without experiencing any advantage. However, if the patient does have GCA but the diagnosis was not made and the patient was not established on high doses of steroids, then there is a significant risk of ischaemic complications, including permanent visual loss or stroke, which are the most important complications of the disease and which makes sight loss (and other acute ischaemic complications) from GCA a preventable medical emergency.

Diagnosis of giant cell arteritis relying on a gold standard of temporal artery biopsy

Since 1932 the conventional gold standard investigation for GCA has been a TAB.[1] The characteristic finding of histiocytes, epithelioid and giant cells (large multinucleated cells present in the arterial wall) at the intimal–medial junction is useful in diagnosis,[41] but not always present (e.g. giant cells were found in 75% of positive biopsies in a recent series). Other pathological features include transmural inflammation, adventitial infiltrates or localised infiltrates of inflammatory cells, especially lymphocytes in the media or intima. Reduplication of the internal elastic lamina and fragmentation of the internal elastic lamina are also described. Intimal cellularity and increased thickness can occur and, in a number of cases, the vessel lumen is narrowed to occlusion with associated thrombus formation.

Most patients have headache, which on closer questioning is localised around the temporal artery and is usually worse on one side than the other. The most symptomatic artery is usually selected for biopsy and is most likely to show evidence of pathological findings. In some centres, it has previously been a routine procedure to sample both temporal arteries in suspected cases, but the value of bilateral testing is relatively low,[44] with only one of 91 bilateral biopsies showing discordance. In a recent study of 132 cases undergoing bilateral biopsies, the diagnostic yield increased by 12.7%[45] as a result of the second simultaneous biopsy (38 patients had bilateral findings of GCA, compared with an additional 13 patients whose biopsies showed abnormalities confined to one side only).

The purpose of high doses of glucocorticoid therapy is to resolve inflammation. Therefore, the characteristic findings of cellular infiltration of the vessel wall with lymphocytes and giant cells may have disappeared by the time the biopsy is performed if there is a significant delay between starting treatment and obtaining the biopsy. Because of the 'clock ticking' as a result of glucocorticoids being administered as a precautionary measure (in case the patient really does have GCA), it is not usually helpful to perform a second biopsy of the opposite (and possibly asymptomatic) artery if the first biopsy is negative for patients in whom there is a suspicion of GCA. A biopsy from the opposite artery is feasible but is less likely to have a positive result,

unless the patient has active symptoms of GCA in the artery to be biopsied. Cellular infiltration is the most important histological finding but can potentially resolve within 7–10 days of commencing high-dose glucocorticoid therapy.[46] Therefore, in some patients, biopsy evidence for GCA is inadequate. Many of the changes seen in the intima and internal elastic lamina can also be found in older people who do not have any features of GCA.[47] A recent surgical series of 237 patients undergoing TABs reported positive findings of GCA in only 36 (15.1%) cases[48] and the result of the biopsy did not significantly contribute to the diagnosis. Changes suggestive of GCA are not consistently present throughout the course of the vessel.

The biopsy may not actually contain any arterial tissue. Nerves or veins were sampled in error in 14 of 567 consecutive biopsies (2.5%).[49] Biopsies of temporal arteries are typically sectioned transversely to provide an overall assessment of the artery. If the pathological abnormalities are present in the areas of artery that have not been cut, it is possible to miss the relevant findings. If the biopsy length is small, the characteristic histological features, which may occur sporadically along the length of the tissue obtained (skip lesions), may be missed. The biopsy is typically sectioned in cross-section and it is possible that, if the material obtained is quite small, only a few cross-sections will be available to view. If the abnormalities to be detected are not seen in these cross-sections, the interpretation would be that the biopsy was normal. However, it is possible that if a longer specimen had been obtained and more cross-sections had been viewed then the pathological changes might have been evident. Obtaining specimens that have been subjected to more sections increases the diagnostic yield slightly but leads to significantly more work and expense for the pathology laboratory.[50]

Biopsy length (after fixation) varies in different studies. Shrinkage is well recognised, with a recent study of 62 biopsies showing an average of 4.6 mm of shrinkage from the time of surgical excision to fixation.[51] A study of 966 biopsies from six different hospitals suggested that a length of at least 0.7 cm increased the diagnostic yield from 12.9% to 24.8% positive results.[52] By contrast, another study of 151 biopsies from 149 patients yielded 20 positive biopsies (13.3%), and there was no difference in the length of positive (mean 0.7 cm) compared with negative (mean 0.65 cm) biopsies.[53] The British Society for Rheumatology (BSR) guidelines recommend between a 1- and 2-cm length of artery to provide an adequate specimen, usually from only the symptomatic or most symptomatic side.[5]

The presence of inflammatory infiltrates in the vasa vasorum was reported in 6.5% of 354 biopsies considered positive in one large study of patients with clinical features of GCA.[54] However, it remains controversial whether or not these findings, as well as some of the other ‘characteristic findings’ suggesting GCA, may in fact occur in patients with other forms of vasculitis such as antineutrophil cytoplasm antibody (ANCA)-associated vasculitis.[55–58]

It has been suggested that TAB may be a useful test to diagnose other forms of vasculitis, which could mimic GCA.[59]

There is an inevitable tension between obtaining enough material to make a diagnosis and initiating therapy before disease-related complications set in. In practice, it is common for patients to start on treatment as soon as a physician suspects the diagnosis, typically based on symptoms suggesting the diagnosis of GCA and possible laboratory investigations such as an elevated C-reactive protein (CRP) level or ESR. Treatment is commonly initiated in primary care and the primary care physician would typically contact secondary care services to request confirmation of the diagnosis with a biopsy. However, the acute phase response markers are not reliable tests to diagnose GCA, although if they are elevated, the ESR and CRP level are supportive of the diagnosis but cannot be used on their own because of their lack of specificity.

Standards for diagnosis of giant cell arteritis

The BSR guidelines on managing GCA recommend that biopsy should be considered if a diagnosis of GCA is suspected and state that an early biopsy is desirable in patients with suspected cranial GCA, preferably

within 7 days of initiating high-dose steroid therapy.[5] The biopsy should be carried out by experienced surgeons to give the highest yield of positive results. Similar recommendations were made by the European League Against Rheumatism in their guidelines for the management of large vessel vasculitis.[60] Unfortunately for the NHS in England and for other health-care systems, there may be difficulty in accessing a surgical list promptly. This can result in significant delay in a biopsy being performed. Furthermore, the procedure is often performed by a relatively junior and inexperienced member of the team. The overall impact of these factors could be a reduction in the sensitivity of biopsy as a test for GCA.

Accuracy of temporal artery biopsy versus ultrasound or other imaging modalities

A meta-analysis of the use of ultrasound in GCA[16] examined 23 studies and involved 2036 patients. The weighted sensitivity and specificity of the halo sign was 69% [95% confidence interval (CI) 57% to 79%] and 82% (95% CI 75% to 87%), respectively, compared with biopsy, and 55% (95% CI 36% to 73%) and 94% (95% CI 82% to 98%), respectively, compared with ACR criteria. A study of 55 patients who underwent colour Doppler ultrasound for suspected GCA[61] reported a sensitivity and specificity of 82% and 91%, respectively, suggesting that an ultrasound scan could be a good alternative to biopsy in many patients.

However, ultrasonography of the temporal and axillary arteries is highly operator dependent and it is important to develop and maintain expertise in the technique before it can be applied. Therefore, any ultrasound study requires quality assurance of the adequate training of sonographers prior to the evaluation of patients with suspected GCA. By contrast, MRI is much less operator dependent. In a recent multicentre study, the diagnostic accuracy of MRI was investigated in 185 patients referred for suspected GCA, of whom 53% underwent TAB. The sensitivity and specificity of MRI for diagnosing GCA was 78.4% and 90.4%, respectively, and for TAB (in those patients who had biopsy), the sensitivity and specificity were 88.7% and 75%, respectively.[13] The accuracy of the imaging was high if the patients had received either no glucocorticoids or glucocorticoids for no more than 5 days, but more than 5 days of therapy resulted in a significant fall in diagnostic accuracy. A combined approach that used ultrasound to try to identify the most appropriate site for biopsy had no effect on the sensitivity of detecting histological evidence of GCA.[62]

Summary

In summary, the management of GCA requires a balance between ensuring that patients with GCA are diagnosed and treated promptly (to avoid complications such as sight loss) and avoiding the burden of unnecessary steroid treatment in people without GCA. TAB is useful in assisting with diagnosis but lacks sensitivity. Research since the 1990s on the accuracy of ultrasound suggests that ultrasound has a role as an alternative to, or in addition to, biopsy. However, within the UK, the routine use of ultrasound for GCA is restricted to only a few centres; TAB remains the standard test for the majority of patients suspected of having GCA.

Aims and objectives

The main aim of the TABUL study was to assess the relative merits of TAB and ultrasound in contributing to the diagnosis of GCA. The objectives of the TABUL study are based on two assumptions about diagnosing and treating GCA.

First, patients with suspected GCA are treated with steroids as soon as the diagnosis is suspected (in order to reduce the risk of serious vascular complications) and before any biopsy results might be available.

Therefore, the potential benefit of an ultrasound examination instead of, or in addition to, biopsy is the ability to either continue or withdraw high-dose glucocorticoid treatment appropriately owing to greater certainty of diagnosis.

Second, TAB is itself very problematic as a reference standard, because up to half or more patients with true GCA may have a negative biopsy.[6,7] This may be for a number of reasons including biopsy size, delay between onset of symptoms followed by early use of high-dose glucocorticoid therapy before biopsy and obtaining the biopsy specimen within 7–10 days of therapy commencing; furthermore, the processing and interpretation of biopsy itself can influence the outcome. A positive biopsy does confirm the diagnosis in most patients suspected of having GCA, with specificity approaching 100%. There are some exceptions because other forms of vasculitis may produce exactly the same biopsy appearances as seen in GCA. The difference for other forms of vasculitis is that patients experience clinical features in other organ systems that support that diagnosis, such as the involvement of airways or kidneys in patient with granulomatosis with polyangiitis (GPA) or eosinophilic GPA (EGPA). Ultrasound is not going to be able to achieve greater specificity than biopsy but may achieve better sensitivity if used either instead of, or in addition to, biopsy.

The first primary objective of the study was to evaluate the diagnostic performance (sensitivity and specificity) of ultrasound as an alternative to biopsy for diagnosing GCA in patients who are referred with suspected GCA and in whom a biopsy was going to be carried out.

The second primary objective was to perform a cost-effectiveness analysis to compare ultrasound as an alternative to biopsy for diagnosing GCA.

The secondary objectives in the study were to evaluate:

- interobserver agreement in the assessment of ultrasound and biopsy
- the performance (sensitivity and specificity) of alternative strategies involving ultrasound and biopsy for diagnosing GCA
- the cost-effectiveness of alternative strategies involving ultrasound and biopsy for diagnosing GCA.

Chapter 2 Methods

Summary of study design

The study used a prospective cohort design and recruited patients with suspected GCA who were undergoing a TAB, the standard diagnostic test, as part of their routine care in order to assist with establishing the diagnosis. Patients were recruited following referral from their primary care physician or a secondary care physician and consented to have an additional diagnostic test, namely an ultrasound investigation of their temporal and axillary arteries, before having their biopsy. The clinician treating the patient, as well as the patient, was blinded to the results of the ultrasound. Patients were assessed at presentation, at 2 weeks and after 6 months. The performances of TAB and ultrasound were evaluated against a reference diagnosis derived from the clinician's final diagnosis, which included any changes to the diagnosis during the follow-up period, such as the emergence of any GCA-related complications. The reference diagnosis confirmed the clinician's final diagnosis using an algorithm based on the ACR classification criteria; any unconfirmed cases (and all cases in which the ultrasound result was unblinded and seen by the clinician) were independently reviewed by a panel of experts.

Agreement between sonographers and between pathologists in their interpretation of videos and images was assessed in an inter-rater agreement exercise for a sample of recruited patients. Clinical vignettes for these patients were constructed and assessed by clinicians to see what decisions about diagnosis and treatment might have been made if ultrasound results were provided instead of biopsy results. The cost-effectiveness of the different tests and combinations of tests was assessed in an economic evaluation.

Patient and public involvement

Advice on study design was sought and obtained from patients through the registered charity Polymyalgia Rheumatica & Giant Cell Arteritis UK. Patient representatives on the Trial Steering Committee and the Data Monitoring Committee provided valuable advice and input during the study (see *Acknowledgements*).

Recruitment of sites

Sites were eligible to take part in the study if they were responsible for seeing patients with suspected GCA and used TAB as a routine test for its diagnosis. Sites were not eligible if they used ultrasound for diagnosing GCA as part of their routine practice.

Prior to study commencement, 19 hospitals in England indicated their interest in becoming study sites for potential recruitment. Sites were eligible to take part if a site principal investigator, typically a clinician (e.g. a rheumatologist or ophthalmologist) involved in the management of patients with GCA, could be identified who would have overall responsibility for the site's involvement in the study. Sites also needed to be able to identify the minimum of one pathologist who would have responsibility for assessing TABs and one sonographer with responsibility for performing and assessing ultrasound. Study sonographers needed to have some previous experience in the use of ultrasound but did not need to have specific experience in ultrasound of the temporal or axillary arteries for GCA. Sonographers could come from a variety of clinical disciplines and included rheumatologists, radiologists and radiographers. Sites also needed to provide assurance that, for any individual patient, the roles of the sonographer and the clinician managing the patient were separate. This was to prevent the managing clinician from knowing the results of the ultrasound scan, except when specifically allowed in the study protocol. It did not preclude a clinician

(e.g. a rheumatologist who carries out ultrasound) from performing either role in different patients provided that the separation of responsibilities was maintained for each participant.

All sites needed to obtain the relevant local approvals before training could be commenced. Site participation required sonographers to successfully complete a training package in ultrasound for GCA. No training was provided to the site surgeons, who were asked to perform the biopsies as part of routine care, or to pathologists, given that TAB specimen assessment is part of standard care. At some sites, additional clinicians were involved in the management of study patients and this was a requirement if the site's principal investigator was designated as the study sonographer to ensure that the ultrasound result was blinded for all patients. Research nurses at each site were responsible for co-ordinating recruitment and arranging tests to ensure that both ultrasound and biopsy procedures could be performed within 7 days of commencing high-dose glucocorticoid therapy. All these site personnel comprised the local TABUL team with responsibility for co-ordinating the study locally and completing the clinical, pathology and ultrasound data collection. The process for ultrasound training is described in the next section.

Each site was provided with study training during an initiation visit from the central TABUL study team which consisted of advice on data collection (including completion of study forms) and the process for submitting data. Specific training was provided on the completion of two measures used to assess patients: the Birmingham Vasculitis Activity Score (BVAS) and the Vasculitis Damage Index (VDI). Clinicians and research nurses were required to achieve test scores of 85% for the BVAS and 75% for the VDI (and at least 50% of all individual cases had to be correct) before they were approved for scoring the two measures. Monitoring visits were conducted as per the study standard operating procedures to ensure that the correct procedures were being followed.

Training in ultrasound for giant cell arteritis

Ultrasound assessment of temporal arteries is an established technique for the diagnosis of GCA but there is no standardised protocol in widespread use. We therefore developed a training package for performing and analysing ultrasound scans for the TABUL study. The purpose of the training package was to provide assurance that the sonographers in the study had achieved competence in scanning the temporal and axillary arteries and interpreting the results before recruiting patients to the study.

The training package included a standardised protocol for performing ultrasound in the TABUL study and an accompanying presentation. Sonographers' competence in ultrasound for GCA was assessed in three ways: (1) undertaking ultrasound assessment of 10 patients or volunteers without GCA; (2) passing an examination that tested each sonographer's competence in interpreting ultrasound videos; and (3) successfully completing a 'hot case' ultrasound assessment of a patient with active GCA. Sonographers were encouraged to attend the TABUL training day for sonographers in Oxford and/or participate in site visits from the TABUL study team. After successful completion of training, sonographers were required to submit recorded scans of recruited patients for ongoing assessment of competency in scanning and interpretation.

Sonographers were required to complete all components of the training before they were deemed eligible to assess patients recruited to the main study. An exception was made for sonographers who were already performing routine assessment for GCA; these sonographers were required to undergo part of the training protocol by scanning 10 control cases and completing their online assessment. These sonographers were exempt from completing the 'hot case' assessment on the merit of their curriculum vitae, which was assessed by the ultrasound experts for the study.

Ultrasound protocol and training requirements

The standard protocol for ultrasound and training was set out in the standard operating procedure for ultrasound and is available via the NIHR Journals Library website (www.journalslibrary.nihr.ac.uk).

The study required the use of a linear probe with a grey-scale frequency of 10 MHz or greater and a colour Doppler frequency of at least 6 MHz, using a vascular pre-set and applying colour Doppler mode as opposed to power Doppler mode. It was important to ensure that the focus was positioned around 5 mm below the skin surface for temporal artery ultrasound, in order to detect the artery. Grey-scale frequency was required to be > 10 MHz and the pulse repetition frequency was set at approximately 2–3 kHz. This was dependent on machine and vessel and would need to be altered according to the velocity of flow because this differs from artery to artery. The colour box required angle correction of at least 60º to avoid poor colour Doppler signals and inaccurate readings. The gain setting had to be adjusted to be able to just fill the lumen with colour to avoid over- or under-filling, therefore creating a potential halo or 'bleeding' over the vessel wall, which might give a false reading. We did not routinely employ a compression test to occlude the artery completely to eliminate flow; however, this is a useful test and was described to all sonographers to facilitate distinction between a true halo sign and a false one.[63]

Each site sonographer was required to register the model number and manufacturer of his or her ultrasound machine with the TABUL office to ensure that it was of sufficiently high resolution for the purposes of the study; this was also reported for the subsequent economic analysis. If the sonographer changed the machine, he or she was required to inform the central TABUL office of the change, and the TABUL office had to confirm that the machine that had been substituted was of sufficiently high quality for the study.

The protocol required each patient to lie in a recumbent or semirecumbent position on their side and pull back their hair behind their ears. Gel was applied to the area of the temporal artery and the probe was placed over the middle of the common superficial temporal artery at the level of the tragus, and the position of the probe was adjusted if necessary to locate the artery. The probe was applied in the transverse and subsequently the longitudinal plane or vice versa. After completing a sweep of the artery in one plane, the probe was rotated by 90° and a further sweep was performed in the opposite plane. The level of the bifurcation between frontal and parietal branches of temporal arteries serves as the marker point to define the start of the frontal and parietal branches, respectively. The patient was then asked to turn over to the other side so that the opposite temporal artery could be scanned. The axillary artery was examined by asking patient to remove outer clothing to expose the axilla. Gel was applied to the inner aspect of the upper arm and the ultrasound probe was placed over the midaxillary line, and swept along the expected course of the artery. The probe was applied in either the longitudinal or the transverse plane and swept along until the brachial artery branch was identified. The sweep was then repeated with the probe rotated at 90º, so that both longitudinal and transverse scans were performed. A longitudinal static image was obtained for normal cases and a transverse and longitudinal static image was obtained for abnormal cases.

The sonographers were required to sequentially scan the complete length of common superficial temporal arteries with their frontal and parietal branches in transverse and longitudinal views. The axillary arteries were also assessed in transverse and longitudinal views. The assessors were required to provide video and static images in both transverse and longitudinal planes as evidence that they had adequately scanned arteries. Each video or still image had to be labelled with the patient's study identification number, and the location of the image was defined using the standard formatting abbreviation listed in *Table 1*; for example, a video sweep image of the transverse view of the left temporal artery was labelled LTSN.

The minimum recordings consisted of a 10-second transverse sweep along the length of each of the temporal arteries up to and beyond the bifurcation of the frontal and parietal branches and a still image of each axillary artery. All images had to be scanned using colour Doppler to assess for complete filling of the vessel and accurate assessment of stenosis, and aliasing of colour within the vessel. Doppler pulse wave was used to further characterise any areas of stenosis. The sonographers were asked to report the presence or absence of any abnormalities for each of the temporal and axillary arteries on the ultrasound case report form (see *Appendix 1*) while they were scanning and to indicate the relevant section(s) for abnormalities in the temporal arteries.

TABLE 1 Abbreviations used to define ultrasound arterial sites and abnormalities found in the TABUL protocol

Site	Image	Abnormality	Left	Right
Temporal artery	Initial sweep with transverse video (for normal scans)	None	LTSN	RTSN
Axillary artery	Initial sweep with longitudinal video (for normal scans)	None	LALN	RALN
Common superficial temporal artery	Transverse video	Halo	LCTH	RCTH
	Longitudinal video	Halo	LCLH	RCLH
	Transverse video	Occlusion	LCTO	RCTO
	Longitudinal video	Occlusion	LCLO	RCLO
	Doppler pulse wave	Stenosis	LCDS	RCDS
	Longitudinal still image	Stenosis	LCLS	RCLS
Parietal ramus of superficial temporal artery	Transverse video	Halo	LPTH	RPTH
	Longitudinal video	Halo	LPLH	RPLH
	Transverse video	Occlusion	LPTO	RPTO
	Longitudinal video	Occlusion	LPLO	RPLO
	Doppler pulse wave	Stenosis	LPDS	RPDS
	Longitudinal still image	Stenosis	LPLS	RPLS
Proximal frontal ramus of superficial temporal artery	Transverse video	Halo	LPFTH	RPFTH
	Longitudinal video	Halo	LPFLH	RPFLH
	Transverse video	Occlusion	LPFTO	RPFTO
	Longitudinal video	Occlusion	LPFLO	RPFLO
	Doppler pulse wave	Stenosis	LPFDS	RPFDS
	Longitudinal still image	Stenosis	LPFLS	RPFLS
Distal frontal ramus of superficial temporal artery	Transverse video	Halo	LDFTH	RDFTH
	Longitudinal video	Halo	LDFLH	RDFLH
	Transverse video	Occlusion	LDFTO	RDFTO
	Longitudinal video	Occlusion	LDFLO	RDFLO
	Doppler pulse wave	Stenosis	LDFDS	RDFDS
	Longitudinal still image	Stenosis	LDFLS	RDFLS
Axillary artery	Transverse still image	Halo	LAFTH	RAFTH
	Longitudinal still image	Halo	LAFLH	RAFLH
	Transverse still image	Occlusion	LAFTO	RAFTO
	Longitudinal still image	Occlusion	LAFLO	RAFLO
	Doppler pulse wave	Stenosis	LAFDS	RAFDS
	Longitudinal still image	Stenosis	LAFLS	RAFLS

If any abnormality was detected, then additional information by artery and section was collected in the case report form and recordings of the abnormalities were required. For a halo, the sonographer reported the maximum thickness and length and whether or not it ran along the entire length of the section. A 3-second transverse and longitudinal video was recorded to support evidence of any reported halo, stenosis or occlusion in sections of the temporal artery. A transverse and longitudinal still image was recorded to demonstrate halo or occlusion in either axillary artery. If stenosis was reported then the velocity in and out of the stenosis (and the minimum and maximum luminal diameter for axillary arteries) was reported and a longitudinal still image and Doppler pulse wave were recorded. The presence of arteriosclerosis was reported separately as an abnormality but no images of this were required. On completion of the scanning, the sonographer was required to document whether or not the ultrasound results were consistent with a diagnosis of GCA. The completed case report forms and recordings (on compact disc) were submitted to the TABUL office.

We expected the scanning protocol to take between 20 and 45 minutes for each patient. The start time, end time and total scanning time were collected for each training case or patient. The protocol also required the sonographer to ensure that the results of the ultrasound, the case report form and the recordings were not given to, or discussed with, the clinical staff involved in treating the patient. Each site was supplied with guidance on how to perform the scans (see *Appendix 2*).

Ultrasound training programme

Although the biopsy of temporal arteries has been an established test in widespread use all over the world for decades, the use of ultrasound as a diagnostic test is much more limited. Very few of the sites involved in the study had sufficient expertise to undertake proficient vascular ultrasound scanning for GCA. We therefore developed a pragmatic training programme consisting of attendance at a training day or a site visit with hands-on training. Competence in ultrasound was assessed using a video examination to correctly identify normal or abnormal scan appearances, evidence of successfully performed scans of 10 healthy control subjects, and evidence of a successfully performed scan of at least one patient with scan findings of active GCA. Sonographers were allowed to take part in the study only once all elements had been successfully completed. In addition, we required sonographers to submit recordings of scans from all patients recruited into the study for ongoing quality control.

Ultrasound protocol training was provided during a training day in Oxford at the start of the study or at site visits by the TABUL study team. The protocol and training emphasised the importance of keeping the ultrasound result blinded from the clinician treating the patient. Sonographers were also provided with a presentation on how to scan temporal and axillary arteries to look for evidence of GCA and how to document the site and nature of the findings using standardised abbreviations (see *Table 1*). The presentation was developed with the supervision of one of the authors (WAS) who had extensive expertise in GCA ultrasound. The presentation provided information on recommended techniques and described the minimum equipment required to perform optimal scanning.

Video examination

An online assessment was developed specifically for the study and consisted of groups of ultrasound images of 20 cases representing patients with or without active GCA. The cases comprised still images and videos of approximately 10 seconds' duration from consenting patients (not part of the TABUL study), supplied by two of the authors (WAS and BD). Sonographers could view the images by accessing a secure password-protected online site designed for the study. For each case, the sonographer was required to indicate the presence or absence of hypoechoic vessel wall oedema (the 'halo'). Sonographers submitted their responses to the online system for marking; they had to achieve a minimum of 75% correct answers to pass the evaluation. Sonographers who failed to pass the test at their first attempt were required to repeat the entire test or specific questions, depending on how many errors they had made.

Scanning training cases

Sonographers' competence in performing ultrasound was assessed by their provision of satisfactory scans from 10 healthy or non-GCA training cases. All training case participants were screened and consented prior to the ultrasound scan. Training cases had to be at least 50 years old and willing to attend for an ultrasound scan of their temporal and axillary arteries. Anyone with suspected GCA or a history of diagnosed or suspected GCA was ineligible, as were patients with any inflammatory condition or anyone who had taken systemic steroids or immunosuppressants in the previous 3 months.

Scanning followed the process described in the protocol. Briefly, the sonographer was required to provide correctly labelled (and anonymised) video images of both temporal and axillary arteries from 10 individual training cases, with documentation of the findings in the case report form. The case report forms and recordings were reviewed by four expert sonographers (WAS, BD, EM, APD), who assessed the sonographers' competence and provided feedback. Sonographers were required to assess additional cases as specified by the reviewer if there were concerns over their scanning. If any of the control patients showed any evidence of an abnormality consistent with GCA then the general practitioner (GP) of the individual would be informed of the result.

Assessment of a patient with active giant cell arteritis ('hot case')

All sonographers were required to scan at least one patient who had active GCA as part of their training assessment in order to demonstrate competence in detecting and reporting the abnormal findings. The 'hot case' patient was consented to the study using NHS or local hospital consent but could not be a patient recruited to the main TABUL study. The sonographer scanned the patient, completed the case report form and submitted recordings following the ultrasound protocol. The expert reviewers assessed the submitted recordings and case report form to ensure that (1) the ultrasound features were consistent with GCA and (2) that the appropriate images had been recorded, were of suitable quality and were consistent with the case report form. If the reviewers were not satisfied then the sonographer was required to complete another 'hot case' and resubmit.

Monitoring ultrasound during the study: quality control by expert review

Once a sonographer had successfully completed and passed all three components of the training assessment, they were approved to scan patients with suspected GCA who were recruited to the study. In order to ensure that the quality of scanning was maintained, a process of ongoing quality control was developed and implemented. The ultrasound case report forms and recordings for each patient were submitted and reviewed by at least one of the four expert reviewers. Recordings were uploaded to a central ultrasound database which allowed remote access for reviewers. Reviewers assessed the quality of images collected and their agreement or otherwise with the sonographer's interpretation of the recordings. If the expert reviewers had concerns about the performance of a sonographer, then the sonographer was required to undergo additional training before being approved for scanning patients in the study.

All recruited patients had their scans reviewed unless no uploaded images were submitted. At least one expert sonographer reported their agreement, disagreement or uncertainty with the assessment made by the sonographer and, if uncertain, an indication of whether or not this was attributable to concerns over the quality of the scanned images that were submitted.

Study population, recruitment and sampling

The study aimed to recruit all eligible patients who were undergoing a TAB for suspected GCA. Patients were eligible if there was a clinical suspicion of a new diagnosis of GCA and the treating clinician had decided that the patient required an urgent TAB to help determine whether or not the diagnosis was GCA. No particular symptoms were specified, although it was expected that patients would have typical symptoms of GCA such as a new onset of headache, scalp tenderness, elevated CRP level or ESR, jaw or

tongue claudication or visual loss. Patients had to be at least 18 years of age and be willing to attend for an ultrasound scan of their temporal and axillary arteries.

Patients were not eligible for the study if they had had a previous diagnosis of GCA or if it was not possible to arrange for their ultrasound and biopsy to be performed within 7 days of starting higher doses of glucocorticoids (defined as > 20 mg of oral prednisolone or equivalent daily). Patients were also ineligible if they had prolonged use (> 1 month) of higher dose glucocorticoids (> 20 mg of prednisolone or equivalent per day at any time) within the previous 3 months for any condition other than PMR. A current or previous diagnosis of PMR or presenting symptoms of PMR were not exclusion criteria, because this group of patients would be likely to require investigations for possible associated GCA, if they presented with new features suggesting the diagnosis. No other selection criteria were used for the recruitment of patients.

All patients were required to give written informed consent. Additional consent was required to allow serum, plasma and deoxyribonucleic acid samples to be taken at the first assessment and serum and plasma to be taken at the second and third assessments for future, currently undefined studies. Patients were also invited to consent to allow their remaining tissue biopsy samples (not required for diagnosis) to be stored centrally in the Oxford Musculoskeletal Biobank for further, future currently undefined studies. All slides that were originally required for diagnostic purposes were stored in the Oxford Musculoskeletal Biobank or returned to the site pathologists, after they had been photographed. All screened patients were allocated a unique screening number and a screening case record form (CRF) was completed for each case (see *Appendix 3*). All eligible patients who consented were allocated a unique study identification number.

It was expected that the majority of patients would be recruited from referrals from general practice to secondary care (either to rheumatology and/or ophthalmology on-call teams). The clinician responsible for the patient's care obtained verbal consent from the potential patient and passed on their contact details to the local TABUL team. Following an initial telephone call the TABUL team provided the potential patients with the study invitation letter and participant/patient information sheet (see *Appendix 4*) and discussed the study with them. Alternatively, if a patient was attending the hospital, the study documents were given directly to them by the clinician or study team. The potential patient would then have sufficient time to read and understand the information and to ask any questions before providing written informed consent (see *Appendix 5*).

Study recruitment at sites was encouraged by providing study information flyers in non-patients areas of sites as an aide-memoire for research teams and clinicians. Awareness of the study was raised with rheumatologists at local, regional, national and international meetings such as the BSR, local meetings with GPs, ophthalmologists, vascular surgeons, rheumatologists and clinicians treating other forms of vasculitis. Guidance on recruitment was provided to all sites (see *Appendix 6*).

Sample size calculation

The sample size of 402 patients was calculated to provide 90% power at a 5% type I error rate to test the joint hypotheses that:

1. ultrasound has greater sensitivity than TAB (based on an assumed sensitivity of 76% for TAB and 87% for ultrasound)
2. the specificity of ultrasound is no less than 83% using the reference diagnosis.

The postulated sensitivity and specificity figures were based on a previous meta-analysis.[16] The sample size would allow estimation of a one-sided rectangular confidence region for ultrasound false- and true-positive

fractions, assuming 80% prevalence of GCA in patients having a biopsy for suspected GCA, with the sample size inflated (gamma 0.1) because of uncertainty in the ratio of cases to controls in a cohort design.[64]

In order to allow for losses to follow-up (failure to have either test done, lack of a follow-up assessment or patient withdrawal) the plan was to recruit 430 participants to the study. After monitoring actual recruitment and withdrawals during the course of the study, the target recruitment was increased to the range 435–445.

Clinical data collection

Patients who were referred with suspected GCA were screened to check their eligibility for recruitment into the study. Patients who were eligible and gave informed consent had a full clinical assessment at presentation. Appointments for ultrasound scans and then biopsy were arranged and patients returned for a follow-up clinical assessment after 2 weeks (*Figure 1*). After the 2-week assessment and after seeing the biopsy report, the clinician (who remained blinded to the ultrasound results) decided whether or not the patient had features consistent with a diagnosis of GCA.

The result of the ultrasound was unblinded only if the clinician concluded that the patient did not have features consistent with GCA and was therefore planning to withdraw steroid therapy rapidly. The procedure for doing so is described below (see *Ultrasound test results: procedure for revealing test results*). Clinicians were allowed to alter their decision to withdraw steroids rapidly following unblinding of the ultrasound result. Patients attended a final clinical assessment after 6 months.

Patient assessment at presentation

The first clinical assessment at presentation collected data on demographic information, relevant conditions and past medical history, symptoms, physical examination findings, laboratory test results and medication. Clinicians were also asked how certain they were of the diagnosis of GCA (definite, probable or possible). Patient data included the patient's age, sex, ethnicity, weight, blood pressure and smoking history. Comorbidity was assessed by reporting relevant current and previous medical history, and the assessment

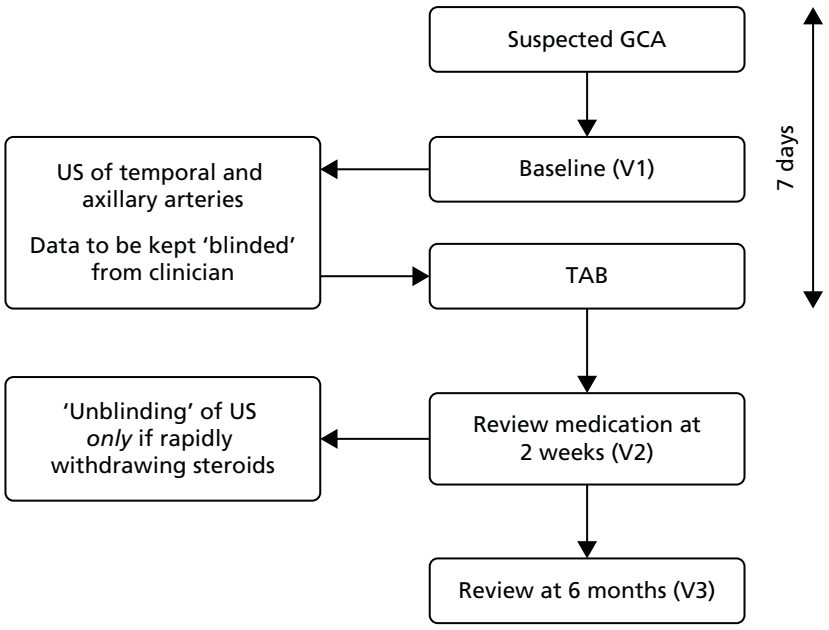


FIGURE 1 Flow of patients in the study. US, ultrasound; V, visit.

included specific questions on diabetes mellitus, hypertension, angina, myocardial infarction, heart failure, low trauma fractures and neoplasia.

Information on symptoms was collected separately for symptoms that the patient had experienced prior to commencing higher-dose glucocorticoid therapy, as well as symptoms present at the first assessment (if the patient had already started on glucocorticoid treatment). This allowed us to separately report whether or not the presenting symptoms had changed as a result of glucocorticoid therapy. The presence of the following symptoms (typically seen in GCA) was reported: anorexia, fatigue, fever/night sweats, localised pain in the head, scalp tenderness, swelling over the temporal artery, pain over the temporal artery, jaw claudication, tongue claudication, reduced or lost vision, double vision and amaurosis fugax. Symptoms of PMR (early-morning stiffness lasting longer than 1 hour, bilateral shoulder pain and bilateral hip pain) were also collected. In addition, any other symptoms that the clinicians thought were relevant could be reported manually.

Physical examination of the patient required an assessment of both temporal arteries for evidence of thickening, tenderness and reduced or absent pulsation, and of both axillary arteries for tenderness. Examination also included, if assessed, evidence of anterior or posterior ischaemic optic neuropathy, relative afferent pupillary defect, III/IV/V nerve palsy or bruits on either side and evidence of stroke, aneurysm or other features such as scalp or tongue necrosis.

The results of laboratory tests that were required for the study protocol before starting steroids and at presentation comprised ESR, CRP level and/or plasma viscosity. Additional tests included measurement of full blood count, haemoglobin, biochemistry, ANCA and urine dipstick testing if there was a clinical indication. Data were also collected on whether or not, and when, treatment with high-dose glucocorticoids for suspected GCA had been started, the route and dose and any treatment with an immunosuppressant agent. The patient was asked to complete a EuroQol-5 Dimensions (EQ-5D) 3-levels questionnaire at the assessment.[65] EQ-5D is a generic measure of health-related quality of life, necessary for the calculation of the cost-effectiveness of the two main diagnostic tests.

Patient assessment at 2 weeks and 6 months

The biopsy and ultrasound tests were completed prior to the patient assessment at 2 weeks. The results of the biopsy were provided to the clinician before the 2-week assessment but the ultrasound results were not shown. The 2-week assessment included the clinician's assessment of the biopsy report and whether or not the biopsy was consistent with GCA. It was therefore possible for the pathologist and clinician to have different opinions on whether or not the biopsy result was consistent with GCA. The patient assessments at 2 weeks and 6 months comprised changes in current conditions and symptoms, a repeat of the physical examination performed at presentation and the results of laboratory tests.

Data for two measures of disease activity and damage were also collected at 2 weeks and 6 months. The BVAS is a validated assessment questionnaire reported by the clinician in the evaluation of disease activity in systemic vasculitis.[66,67] It consists of a list of clinical features that commonly occur in patients with vasculitis together with a weighted score to provide a measure of severity of disease activity; it is widely used for clinical studies and is increasingly used in the clinical management of patients with small vessel vasculitis. It can be used to define how active disease is, to measure response to therapy or to define relapsing disease[66,68] for the purpose of clinical trials. The most current validated version of the BVAS was used.[67] The VDI is a structured assessment to evaluate damage occurring in patients diagnosed with systemic vasculitis.[69] It is a record of irreversible consequences of having a diagnosis of vasculitis. Items are reported in the VDI if they have been present for at least 3 months and have occurred since the onset of vasculitis. There is no attribution to cause and it has been used in large cohorts of patients with primary systemic small vessel vasculitis.[70] Data from the BVAS and the VDI can also be used to examine the possible presence of an alternative form of vasculitis. Data were also collected on weight, blood pressure, treatment with steroids and immunosuppressive drugs, and quality of life using the EQ-5D.

At the 2-week assessment, the clinician was required to state whether or not the patient had features consistent with GCA and, if responding yes, to indicate which of the following influenced the response: symptoms, signs, blood abnormalities, biopsy or other (to be specified). If the patient's features were not consistent with GCA then the clinician was required to give at least one alternative diagnosis. After providing the clinical diagnosis at 2 weeks, in the event that the clinician did not plan to continue high-dose glucocorticoid therapy because they did not think that the patient had GCA, they were required to contact the TABUL office for the ultrasound result. At the 6-month assessment the clinician was required to indicate if the diagnosis had changed and to indicate the influences for any patients in whom the decision was made to alter the diagnosis to GCA. At least one alternative diagnosis was required for any decision to alter the diagnosis away from GCA. The clinical CRF is shown in *Appendix 7* and guidance on completion of the CRF is shown in *Appendix 8*.

Adverse events (AEs) and any attribution to either of the diagnostic test procedures were reported on AE CRFs (see *Appendix 9*). Guidance on completion of the AE CRFs is shown in *Appendix 10*.

The standard test: temporal artery biopsy

The standard test for GCA is TAB. This normally involves a minor surgical procedure to remove a small sample of temporal artery (the BSR recommends a minimum length of 1 cm[5]) which is examined for abnormalities by a pathologist. Guidance on the collection, processing and storage of biopsy samples is shown in *Appendix 11*. Sites followed their usual practice for obtaining and processing TABs. The only changes to routine practice required by TABUL were that sites were instructed to send the actual pathological slides used to make their diagnosis to the TABUL office and that, in addition to their standard reporting of biopsy results, pathologists were required to complete a study-specific CRF (see *Appendices 12* and *13*) to report their pathological findings. We did not require any specific information from any of the surgeons undertaking the biopsy but they were all informed that the patient had been recruited to the TABUL study.

The pathologist was required to report which side or sides the biopsy had been taken from as well as the length of the biopsy (after freezing or fixation), and a note was made of whether or not it was bifurcated. They were able to add other comments on the macroscopic appearance of the sample. For each biopsy, the staining protocol was reported. The macroscopic appearance was described and a note was made of whether or not the biopsy was from the temporal artery and which sections were cut. The presence of abnormalities in the intima (arteriosclerosis or intimal hyperplasia) and the internal elastic lamina (fragmentation or reduplication) were reported. Pathologists were required to indicate if there was an inflammatory infiltrate in the sample (and the predominant site of any inflammation) and indicate if any of the following features were present: normal areas, giant cells, calcification or any other unusual features. Data were also captured on presence and causes of complete occlusion of the vessel or presence of thrombus or evidence of recanalisation in at least one section of the vessel.

The pathological diagnosis was reported as either normal or any the following: compatible with a diagnosis of GCA, compatible with another vasculitis, compatible with arteriosclerosis and compatible with any other diagnosis as specified by the pathologist. The actual pathological slides were sent to the TABUL office for image acquisition. Digital image acquisition was achieved using an Aperio Scanscope Turbo AT (Leica Biosystems, Buffalo Grove, IL, USA). Slides were loaded onto the machine's autoloader and pre-snapped to obtain a macroscopic image before proceeding with digital scanning. The macroscopic image was used to set the tissue area, focal plane, focus points, white balance, scan/slide settings and labelling description. Once the settings had been optimised the slides were scanned in fragments and digitally stitched together to form one high-resolution virtual representation of the pathology slide. These virtual slides were stored on an external physical server and a web-based database (Aperio eSlideManager V1.0, Leica Biosystems) was used to archive and store the eSlides. Slides could be viewed remotely using Aperio's web-based viewing systems (Leica Biosystems).

The biopsy result, which was the primary standard test, was defined as positive by the pathologist if the pathological diagnosis was compatible with a diagnosis of GCA. This included patients whose biopsy samples did not contain temporal artery (e.g. vein, fat, muscle or other tissue) or for whom no sample was obtained from surgery. An alternative standard test result was defined as the clinician's interpretation of the biopsy result as reported on the clinical CRF at the 2-week assessment. This was reported because we expected that the clinician might reach a different conclusion from the pathologist, based on the biopsy report.

The main analyses included patients who had no sample from surgery or a biopsy sample that did not include temporal artery; these were categorised as not compatible with a diagnosis of GCA. Additional analyses excluded the indeterminate biopsy results.

The index test: ultrasound of the temporal and axillary arteries

The index test, an ultrasound of both temporal and both axillary arteries, was performed following the protocol described earlier and is available on the NIHR Evaluation, Trials and Studies Coordinating Centre website (www.nets.nihr.ac.uk) and was subject to ongoing monitoring for quality assurance. The presence of ultrasound abnormalities (halo, occlusion, stenosis and arteriosclerosis) in different segments of the temporal arteries and in the axillary arteries (as defined in *Table 1*) was captured in the ultrasound case report form (see *Appendix 1*). The primary test result for ultrasound was defined as positive and was used for the main analyses if the sonographer responded 'yes' to the question 'In your opinion are the results consistent with a diagnosis of GCA?'. Additional analyses used alternative definitions of a positive result based on the presence or absence of a bilateral halo and on the interpretation of the ultrasound images from the expert review.

Ultrasound test results: procedure for revealing test results

The clinician treating the patient was provided with the biopsy result but did not have access to the results of ultrasound at the 2-week assessment. Study sonographers were required to keep the results of each patient's scans blinded from the managing clinician for the duration of the study. The only exception was if the managing clinician had completed the 2-week assessment and was planning to withdraw steroid treatment rapidly. In these circumstances the clinician was required to contact the TABUL office and was provided with the scan results as reported by the sonographer. The clinician then had an opportunity to reconsider their decision to withdraw steroids and alter their diagnosis. Thus, the 2-week assessment included a report of the clinician's original assessment of the diagnosis and any revision following the revealing of the ultrasound result.

The reference diagnosis

The ideal reference diagnosis for evaluating diagnostic tests is one that is independent of the tests being evaluated. No such reference diagnosis exists for GCA for evaluating the performance of biopsy and ultrasound. Criteria for classifying GCA and usual clinical practice for reaching a GCA diagnosis incorporate the results of the biopsy; therefore, they cannot be truly independent methods for defining a reference diagnosis. Furthermore, the ACR classification criteria were not intended to be used as diagnostic criteria.[34] For the purposes of the study, a partially independent approach was used, which combined elements of a clinician's final diagnosis, the ACR classification criteria (incorporating the biopsy result), the emergence of complications consistent with GCA during follow-up, the emergence of alternative vasculitis diagnoses during follow-up and expert review to determine the reference diagnosis. The process started with the clinician's final diagnosis for the patient as reported on the 6-month (or in its absence, 2-week) assessment. An algorithm was devised to determine if evidence from the biopsy and the presence or

absence of symptoms and emerging complications and diagnoses on follow-up supported the clinician's diagnosis or if expert review was required to determine the reference diagnosis.

If the clinician's final diagnosis was GCA, then a reference diagnosis of GCA was given if any of the following criteria were met:

1. a stricter version of the ACR classification criteria using either the standard or tree method was met based on the patient's symptoms and physical examination from their baseline assessment (*Table 2*)
2. the emergence of PMR during follow-up in patients with no previous history of PMR and no symptoms of PMR at presentation
3. the emergence of new or worsening jaw claudication, tongue claudication, abnormal anterior optic neuropathy, abnormal posterior optic neuropathy, or relative afferent pupillary defect during follow-up.

If the clinician's final diagnosis was not GCA, then a reference diagnosis of 'not GCA' was given. If an alternative vasculitis diagnosis was made, these included Takayasu's arteritis, large vessel vasculitis, polyarteritis nodosa, GPA, microscopic polyangiitis, EGPA, cryoglobulinemic vasculitis, IgA vasculitis

TABLE 2 Definitions and sources of items in the ACR classification criteria

Criterion	Definition	Source
(1) Age at disease onset of at least 50 years	Development of symptoms or findings beginning at ≥ 50 years of age	Baseline patient assessment: symptoms started at ≥ 50 years of age pre-steroids or at presentation
(2) New headache	New onset of or new type of localised pain in the head	Baseline patient assessment: symptoms of new onset or type of localised pain in head pre-steroids or at presentation
(3) Temporal artery abnormality	Temporal artery tenderness to palpation or decreased pulsation unrelated to arteriosclerosis of carotid arteries	Baseline patient assessment: abnormal tender temporal artery on physical examination
(4) Elevated ESR (at least 50 mm/hour)	ESR at least 50 mm/hour as assessed by the Westergren method	Baseline patient assessment: laboratory test results ESR at least 50 mm/hour pre-steroids or at presentation
(5) Abnormal artery biopsy	Biopsy specimen with artery showing vasculitis characterised by a predominance of mononuclear cell infiltration of granulomatous inflammation, usually with multinucleated giant cells	Pathology CRF: pathologist reports biopsy result as consistent with a diagnosis of GCA
(6) Claudication of jaw, tongue, or on deglutition	Development or worsening of fatigue or discomfort in muscles of mastication, tongue, or swallowing muscles while eating	Baseline patient assessment: symptoms of jaw or tongue claudication pre-steroids or at presentation
(7) Scalp tenderness or nodules	Development of tender areas or nodules over the scalp, away from the temporal artery or other cranial arteries	Baseline patient assessment: symptoms of new-onset generalised scalp tenderness pre-steroids or at presentation
Classification as GCA		
Traditional method (standard): at least three of (1) to (5) are met		
Traditional method (strict): at least four of (1) to (5) are met		
Tree method (standard): classified as GCA if (1) is met and any of (3), (5) or (6) are met. Criterion (2) replaces (5) in the absence of a TAB result; criterion (7) replaces (3) if (3) is not met		
Tree method (strict): classified as GCA if (1) is met and at least two of (3), (5) or (6) are met. Criterion (2) replaces (5) in the absence of a TAB result; criterion (7) replaces (3) if (3) is not met		

(Henoch–Schönlein purpura), or any other vasculitis to be specified. A reference diagnosis of 'not GCA' was also given if all of the following criteria were met.

1. The patient failed to meet the ACR classification criteria using either the standard or tree methods (see *Table 2*).
2. No new-onset PMR occurred during follow-up.
3. No new or worsening jaw claudication, tongue claudication, abnormal anterior optic neuropathy, abnormal posterior optic neuropathy or relative afferent pupillary defect occurred during follow-up.
4. No symptom of reduced or lost vision in either eye occurred or worsened during follow-up.
5. No evidence of abnormal III/IV/VI nerve palsy or stroke on clinical examination was observed at 2 weeks or 6 months.
6. No sudden visual loss, cerebrovascular accident or cranial nerve palsy reported on the BVAS occurred during follow-up.
7. No retinal change, optic atrophy, visual impairment/diplopia, blindness in one eye, blindness in the second eyes or cerebrovascular accident reported on the VDI occurred during follow-up.

Any patient who was not given a confirmed reference diagnosis based on the above was referred for expert review. Furthermore, any patient who had their diagnosis altered during follow-up (typically for a diagnosis altered to GCA from not GCA following unblinding of the ultrasound report) was automatically referred for expert review regardless of any confirmed reference diagnosis given above.

The expert review group comprised five rheumatologists involved in the study. Each case requiring expert review was independently assessed by three of the five rheumatologists, and no rheumatologist could review cases from their own site. A summary report for each patient was extracted from the clinical data and included information on symptoms, GCA-related complications, items from the ACR classification criteria and the clinician's diagnosis. Access to the clinical database was also given so that expert reviewers could examine all data collected as part of the study with the exception of the ultrasound results. Each expert reviewer independently reported their agreement or disagreement with the clinician's final diagnosis. The clinician's final diagnosis was supported if at least two of the experts agreed with the diagnosis. The clinician's diagnosis was altered if all three experts disagreed with the diagnosis. If two experts disagreed with the clinician's diagnosis then the patient was discussed by the relevant experts during a moderated teleconference until the three experts reached a consensus.

Inter-rater agreement data collection and analysis

The aim of the inter-rater agreement component of the study was to assess the extent of agreement between trained sonographers in their interpretation of ultrasound videos, and between experienced pathologists in their interpretation of biopsy images, for a sample of cases using data, videos and images from patients recruited to TABUL. Sonographers and pathologists assessed the same cases using a web-based exercise. Intrarater agreement was also assessed by repeating cases in the exercise. The impact of providing additional information about the patient was examined by including a brief vignette.

All pathologists and sonographers who assessed patients in the main TABUL study were asked to complete a web-based exercise. The exceptions were pathologists and sonographers who were involved in the management of TABUL or in the expert review of ultrasound for quality control (two pathologists and four sonographers.) Pathologists and sonographers who agreed were sent instructions for completing the exercise and a password to access the exercise.

The overall design involved a web exercise with 44 cases. Each case represented a patient recruited to TABUL and comprised ultrasound videos of both temporal arteries, scanned images of the biopsy slide and a brief clinical vignette describing the patient. The first five cases were defined as training/practice cases that allowed raters to familiarise themselves with the exercise. The remaining cases, the rating cases,

consisted of 30 unique cases, six repeats of unique cases (for intrarater assessment) and three reserve cases. The reserve cases were available to replace any of the 30 cases that were subsequently found to be ineligible once the exercise had started. The overall number of cases was chosen to keep the task manageable, and the aim was to have at least 10 pathologists and 10 sonographers complete the exercise. This was to allow results to be generalised to the wider populations of pathologists and sonographers.

The criteria for including a patient's videos and biopsy images as rating cases in the exercise were ultrasound videos of adequate quality of the right and left temporal arteries, biopsy slides received and scanned, inclusion of the patient in the main TABUL analyses and patient consent for the use of the images. Cases were ineligible if the biopsy specimen did not consist of artery or if the ultrasound was abnormal owing to axillary artery involvement without temporal artery abnormalities. Cases were also ineligible if the biopsy images or ultrasound videos included information that identified the patient or clinician involved or included markings indicating abnormality and this information could not be removed or hidden. Finally, cases were excluded if the quality of the ultrasound images was judged to be poor by expert review during quality control. Disagreement with the original sonographer's interpretation by expert review or difficulty in interpreting the ultrasound by expert review despite adequate quality videos were not criteria for exclusion.

Identification of cases was performed in three stages before the start of the exercise because the main TABUL database and the ultrasound and biopsy databases had not been locked at the time of initial selection and because of the work involved in ensuring that videos and images were eligible. The first stage involved identifying potentially eligible cases from the list of patients recruited to the main study who had had their ultrasound videos uploaded. This list of potentially eligible cases was ordered using random numbers generated using Stata version 13 (StataCorp LP, College Station, TX, USA). The second stage involved populating the 33 rating cases from the top of the list. Any case found to be ineligible was replaced with the next available case from the list. This process was repeated until all 33 rating cases were deemed eligible. The third stage involved pilot testing of the exercise and review of all videos and images by two pathologists (BM, KW) and two sonographers (WAS, JP) to ensure that the criteria relating to the videos and images were met. The five training cases were selected purposively starting at the bottom of the ordered list. These were selected to ensure that there were at least two abnormal and two normal cases for the biopsy images and for the ultrasound videos. A final post-exercise stage involved a further eligibility check of the rating cases against the locked database. Any of the 30 rating cases subsequently found to be ineligible were replaced with one of the three reserve rating cases for inclusion in the analyses.

A web-based exercise was designed to allow remote access to the videos and images and to capture data from the assessments made by the sonographers and pathologists. Each case in the exercise began by giving access to two video images showing left and right temporal arteries (for sonographers) or one biopsy slide image for each stain available (for pathologists). Videos could be replayed as often as required and biopsy images allowed zooming for magnification at the equivalent of up to 40 times in high resolution. Raters were asked to answer yes or no to the question 'In your opinion, do the ultrasound (or pathology) images show features of GCA?' and to answer certain or uncertain in response to the question 'How certain are you?'. They then gave their answers and confirmed that they were confident to submit their answers.

All cases were rated before and after seeing a brief clinical vignette describing the patient. The vignette was added to reflect a more realistic scenario for interpreting the videos or images. For example, the sonographer would see the patient in front of them when conducting temporal and axillary artery ultrasound. The pathologist might receive a brief description of the patient on the biopsy request form. The vignettes provided basic information on age, sex, glucocorticoid treatment, comorbidity, presenting symptoms and laboratory test results, for example '79 year old male started glucocorticoid therapy 2 days ago for suspected GCA. Patient has hypertension. Presented with new localised pain in head, jaw claudication and reduced vision. Elevated ESR and CRP'. The vignette was identical for the ultrasound and

biopsy versions of each case except for the duration of glucocorticoid therapy (which varied depending on when the test was done). For repeat cases, the core information was identical to the original case but the order of wording was altered.

Cases had to be completed in order and rating cases could not be started until all five training cases had been completed. Once a rating case had been completed it was not possible to return to that case to view the videos or images or to look at the answers given. This was because six of the cases were repeated. It was possible to return to the training cases for reference. The locations of repeated cases in the 36 rating cases were assigned before the random ordering of eligible cases. Repeated cases all made their first appearance in the first 18 cases and all made their second appearance in the final 18 cases. For each of the six repeated cases there was a minimum gap of 16 cases between its first and second appearances.

Clinical vignettes data collection and analysis

The aim of the assessment of the clinical vignettes was to determine what decision about a patient's diagnosis and treatment would have been made if there was no biopsy performed, leaving the clinician to rely on the results of the ultrasound. Two overlapping samples of cases were selected from patients recruited to the study. The first sample was the same random sample used for the assessment of inter-rater agreement. The second sample comprised all patients in the main study who had a positive biopsy and a negative ultrasound.

Clinical vignettes were structured to provide data on the patients at the times when two key decisions are made. The first is on initial presentation, when the possibility of a diagnosis of GCA is considered and a decision is taken to recommend a TAB. The second is after 2 weeks, when a decision to continue or withdraw high-dose steroids for GCA is made. Vignettes were populated with data collected during the study. Information provided at presentation comprised the patient's age, sex, relevant current conditions and medical history, symptoms, symptom onset and any laboratory test results (ESR, CRP level or ANCA) prior to starting steroids, duration and dose of steroids, new symptoms and symptoms still present at presentation, results of the physical examination at presentation and any laboratory test results (ESR, CRP level or ANCA) at presentation. Clinicians were then asked to give their indication of the likelihood of the patient having GCA (definite, probable, possible or not GCA) and indicate whether or not, in the absence of alternative tests such as ultrasound, they would recommend this patient for a TAB.

The information at 2 weeks was presented once responses to the questions had been confirmed. Information on the vignettes comprised the results of the ultrasound test and information about the patient's health after 2 weeks. The ultrasound test was reported as either consistent or not consistent with a diagnosis of GCA and included additional information on any abnormalities identified on ultrasound, for example 'consistent with GCA; halo on right temporal artery; normal left temporal artery; normal axillary arteries; no occlusion or stenosis'. Other information comprised symptoms present at 2 weeks (categorised by new, worse, no change, better and resolved), results of the physical examination at 2 weeks, results from laboratory tests and any changes in current conditions. Clinicians were then asked to give their indication of the likelihood of the patient having GCA (definite, probable, possible or not GCA) and to indicate the appropriateness of continuing to treat the patient with high-dose steroids for GCA on a nine-point scale (1, extremely inappropriate; 5, uncertain; 9, extremely appropriate).

Data on the appropriateness of continuing treatment with high-dose steroids were categorised as appropriate, inappropriate or uncertain using the method outlined in *The Rand/UCLA Appropriateness Method User's Manual*.[71] A panel median of 7–9 without disagreement is considered appropriate, a panel median of 4–6 or any median with disagreement is categorised as uncertain, and a panel median of 1–3 is categorised as inappropriate. Disagreement was determined using the interpercentile range adjusted for symmetry and the common approach of rounding up medians of 3.5 and 6.5 was applied.[71]

Statistical analysis

The statistical analyses of the diagnostic accuracy of TAB and ultrasound were specified in the statistical analysis plan (see *Appendix 14*). Sensitivities and specificities were calculated for TAB and ultrasound in comparison with the gold standard reference diagnosis. The kappa statistic was used to assess agreement between TAB and ultrasound, and McNemar's test was used to detect systematic discordance.

The inter-rater agreement between sonographers and between pathologists was evaluated using a two-way random-effects analysis of variance to estimate the intraclass correlation coefficients for agreement with 95% CIs. Both cases and raters were treated as random effects in order to generalise findings to all cases (from the sample selected) and to the potential population of trained sonographers (from the sample of sonographers doing the exercise). Intrarater agreement was evaluated by estimating kappa statistics for agreement and by examining agreement for the individual repeated cases and raters.

Statistical analysis was performed in Stata versions 12 and 13.

Pre-test probability of giant cell arteritis: definition of risk categories

The availability of data from the DCVAS study provided an opportunity to define categories of pre-test risk of a GCA diagnosis from an independent sample of patients and was used in preference to obtaining expert opinion elicited from clinical vignettes.[41] Data on 585 patients recruited to centres not participating in TABUL, and who had had a TAB, were used to derive definitions for high-, medium- and low-risk groups. The high-risk group was defined as patients with (1) claudication of the jaw or tongue and (2) elevated ESR or CRP level (ESR of at least 60 mm/hour or CRP level of at least 40 mg/l) either at pre-steroids or at presentation assessments. The low-risk group was defined as patients (1) without jaw or tongue claudication and (2) without elevated ESR or CRP level at both the pre-steroids and presentation assessments of symptoms and laboratory tests. The remaining patients were categorised as medium risk.

Changes to the study protocol

There were two substantial amendments to the study protocol. The first amendment was made in February 2011 and comprised the following key changes.

1. To alter the decision always to offer each potential participant 24 hours to consider their participation in the study. This amendment was made because there were some circumstances in which treatment may be delayed while waiting for consent, for example in an emergency (to minimise delay in normal care such as performing the biopsy) or when sites are able to provide a fast turnaround time for performing the biopsy. In these circumstances we offered the opportunity for participants to provide full written informed consent in < 24 hours from receiving information about the study.
2. To provide further clarification on the collection of additional blood and biopsy samples during the course of the study.

The second amendment was made in February 2013 and comprised the following key changes.

1. To increase the target sample size for recruitment from 430 to 435–445 (with 402 completing the primary end point).
2. To extend the recruitment period by 12 months.
3. To clarify the recruitment strategy (including the production of a poster summarising the study for use in non-patient areas).

4. To clarify the process for the managing clinician to contact the TABUL office in order to be given the results of the ultrasound result (unblind the ultrasound result).
5. To allow inclusion of patients in whom the biopsies were performed more than 7 days after starting high-dose glucocorticoids because of safety concerns about when the biopsy could be performed, for example to allow discontinuation of warfarin so that it was safe to perform the biopsy. This would be part of standard care for any patient who required a biopsy but was receiving warfarin.

Chapter 3 Site recruitment and ultrasound training

Site recruitment

The study aimed to recruit sites that routinely performed biopsy of the temporal artery as part of the care pathway in the diagnosis of GCA. Forty-four sites expressed an interest in recruiting patients to the study. Two sites already made some use of temporal artery ultrasound to assist in diagnosing GCA on a non-routine basis but were in equipoise and accepted the study requirement to keep the ultrasound result hidden from the clinician managing the patient. One site made occasional use of ultrasound to mark the area of disease for surgeons to biopsy. For the purposes of the study the site agreed to suspend this practice.

Before the study began there were 19 sites in England that had indicated an interest in taking part. Recruitment of sites in the UK began in November 2009 and the first training case, as part of the ultrasound training, was recruited in March 2010. The first approval of a site to recruit patients to the main study was in June 2010. The process of gaining the relevant research approvals [e.g. NHS research and development (R&D) approval] for sites, the availability of suitable sonography staff and staff time at each site, and the process of training sonographers delayed recruitment of sites to the study. By December 2011, six sites were approved to recruit patients (four of these had begun recruiting patients) and eight sites had started ultrasound training. The study was opened up to sites in Europe following international interest in the study to increase site recruitment, with four sites (in Germany, Ireland, Portugal and Norway) beginning the process to gain approval to recruit patients.

A total of 44 sites expressed an interest in taking part in the study, although eight sites were unable to obtain R&D approval and therefore did not progress to the ultrasound training stage. One site (the ophthalmology department at the John Radcliffe Hospital, Oxford) obtained the relevant R&D approvals but acted as a referral site for the Nuffield Orthopaedic Centre, Oxford, so did not require ultrasound training. Thus, 35 sites began the process of ultrasound training to become eligible to recruit patients to the study. Two of the sites (Stoke Mandeville Hospital, Aylesbury, and Royal Berkshire Hospital, Reading) did not have their own sonographer and instead relied on the trained sonographers from the Nuffield Orthopaedic Centre for ultrasound scanning at their sites. The two sites did not require ultrasound training but did need to complete other study requirements for site approval. Twenty of the 35 sites achieved approval to recruit patients to the study and the progress of 18 of these sites in obtaining approval is shown in *Figure 2*.

Ultrasound training

The key factor that limited progress to full-site approval to recruit patients to the study was the ultrasound training for site sonographers. There were 49 sonographers representing 35 sites who started ultrasound training and 26 sonographers representing 22 sites who passed their training. Two sites had a sonographer who passed training but the sites did not go on to recruit patients to the study. In the first site the sonographer moved to a different hospital so the site lost its approval to recruit to the study because it had no trained sonographer. The second site did not complete another component of site training (completion of BVAS and VDI training) and could not provide appropriate research nurse support to achieve approval to start the study before recruitment to the study had actually completed. Thus, there were 24 sonographers who passed training who were all located at sites that were approved and that were able to recruit patients to the study.

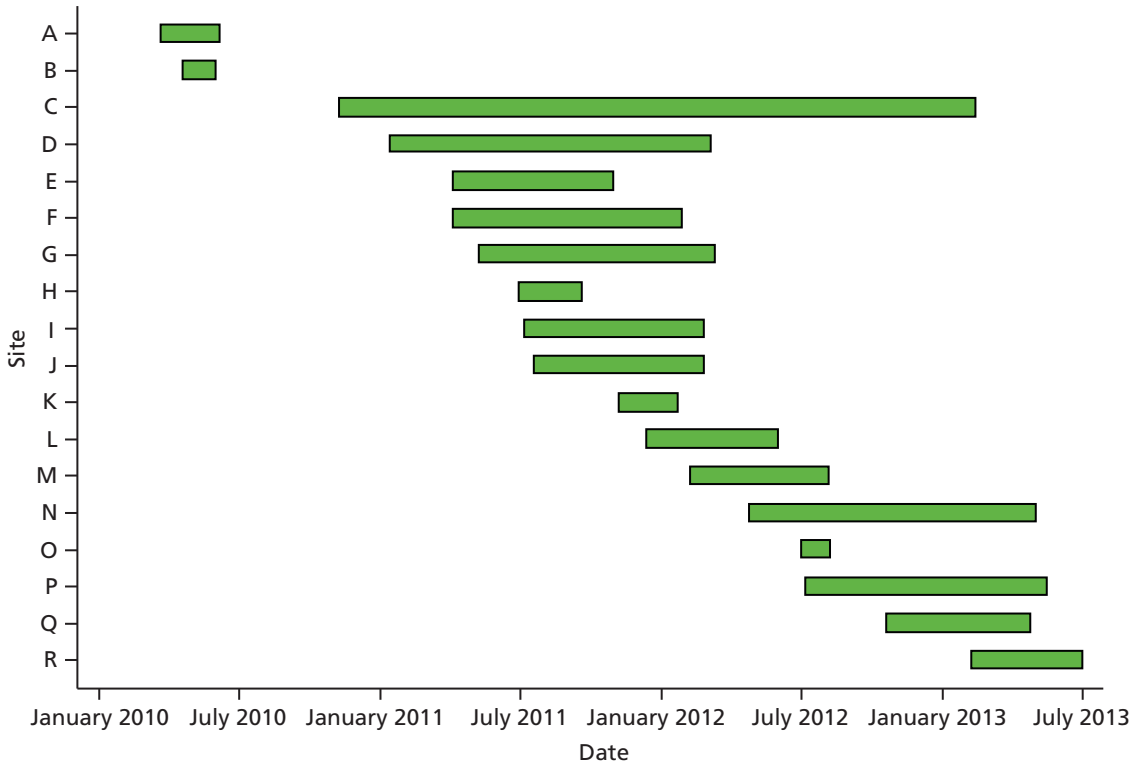


FIGURE 2 Time from starting ultrasound training to approval for patient recruitment for 18 sites (two sites not requiring ultrasound training are not shown; they received approval to recruit patients in 2011).

The main reason for not successfully passing the ultrasound training was the requirement to perform a 'hot case' assessment. Twenty-one sonographers did not provide or pass a 'hot case' assessment and 28 (58%) either passed or were exempt (*Table 3*). For the video examination component, 39 (80%) sonographers passed [although 21 (42%) needed more than one attempt at the examination] or were exempt and two experienced sonographers were exempt because they were study investigators involved in setting and marking the examination. For the training case component a total of 450 healthy volunteers or patients without any suspicion of GCA were scanned. Forty (82%) sonographers passed the component, although five were required to scan additional training cases before passing.

The 24 sonographers who scanned patients in the study were made up of 10 clinicians, six radiologists and eight professional sonographers. Four had previous experience in ultrasound for GCA and were exempt from the 'hot case' assessment and two sonographers were exempt from the video examination component as well. All 24 sonographers satisfactorily completed the training case component, as assessed by one of the study expert sonographers (WAS) not involved in scanning patients, although two sonographers needed to scan additional cases before passing. Half of the sonographers taking the video examination achieved the 75% pass mark at their first attempt with the remaining sonographers achieving this at their second or third attempt.

Seven of the 24 sonographers passed all components without requiring further attempts at any component: three professional sonographers, two radiologists and two clinicians. The 13 sonographers who did need a further attempt at one or more components comprised five professional sonographers, two radiologists and six clinicians. The four experienced sonographers (two radiologists and two clinicians) all passed those components that they took at the first attempt.

TABLE 3 Characteristics and training assessment of sonographers

Characteristic	Detail	Sonographers starting training ($N = 49$), *n* (%)	Sonographers approved at 20 recruiting sites ($N = 24$), *n* (%)
Occupation	Sonographer	15 (31)	8 (33)
	Radiologist	8 (16)	6 (25)
	Clinician	26 (53)	10 (42)
Previous experience	Yes	6 (12)	4 (17)
	No	43 (88)	20 (83)
Video examination	Pass (first attempt)	18 (37)	11 (46)
	Pass (second attempt)	11 (22)	6 (25)
	Pass (third attempt)	10 (20)	5 (21)
	Exempt (experienced)	2 (4)	2 (8)
	Not done	8 (16)	–
Ten training cases	Passed (10 cases)	35 (71)	22 (92)
	Passed (with additional cases)	5 (10)	2 (8)
	Not completed	6 (12)	–
	Not started	3 (6)	–
'Hot case' assessment	Pass (first attempt)	17 (35)	15 (63)
	Pass (second attempt)	4 (8)	4 (17)
	Pass (third attempt)	1 (2)	1 (4)
	Exempt (experienced)	6 (12)	4 (17)
	Failed	12 (24)	–
	Not done	9 (18)	–
Overall	Passed at first attempt	8 (16)	7 (29)
	Passed with further attempt(s)	13 (27)	13 (54)
	Passed with exemptions	5 (10)	4 (17)
	Not passed	23 (47)	–

Ultrasound monitoring during the study

Once a sonographer had completed training and started scanning patients recruited to the study, all their recorded scans and completed case report forms for these patients were monitored. The team of expert reviewers assessed all submitted scans and forms to monitor the quality of video and still images being recorded, as well as the sonographers' record of the scan findings, in order to identify any concerns with either the performance of the scans or the interpretation of the results. Details of the results of the expert review are reported in *Chapter 5*. The results of the expert review are reported in *Chapter 4*.

One site was suspended from scanning any further training cases because of the poor quality of scanning; the original sonographer at the site subsequently took no further part in the study. A new sonographer was identified who successfully passed the training requirements. Another site was suspended from recruiting patients to the study after the findings from the expert review disagreed with the ultrasound of

their second enrolled patient; the site was required to submit and pass an additional 'hot case' assessment before being allowed to resume recruitment. In another site, patient recruitment was suspended after expert reviewers disagreed with scans reported as ultrasound positive and recommended retraining for the sonographer. Recruitment was resumed following successful retraining.

Chapter 4 Description of the study population, recruitment and eligibility

A total of 430 patients were recruited from 20 participating centres over 42 months, as shown in *Figure 3*. The first patient was recruited in June 2010 and the last patient was recruited in December 2013. Forty-nine of these patients were excluded from the primary analyses because they did not have both an ultrasound scan and a biopsy, their biopsy was done > 10 days after starting steroid treatment or they did not have a follow-up assessment. The main study results are based on the remaining 381 patients.

Figure 3 shows that there were 730 patients originally screened for eligibility into the study, 300 of whom did not meet the inclusion criteria and were therefore not evaluated further, leaving 430 patients recruited for the baseline assessment. A further 30 patients were excluded at this stage, chiefly because it was not

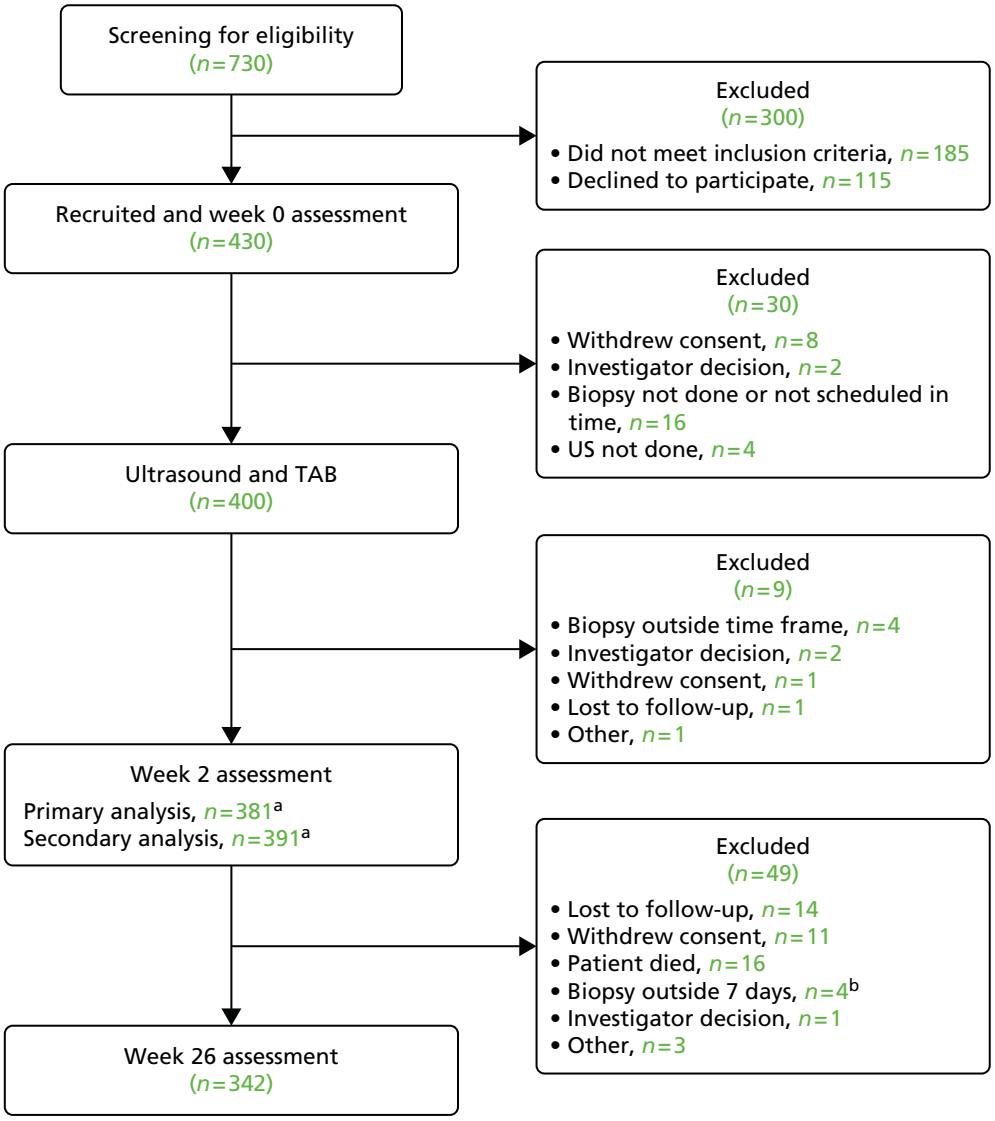


FIGURE 3 Flow of participants through the TABUL study. US, ultrasound. a, Patients included in the primary analysis underwent their biopsy within 10 days of starting steroids. Patients included in the secondary analysis underwent their biopsy at any time after starting steroids. One participant missed the 2-week visit (leaving 380 patients with a 2-week assessment) but continued in the study and completed the 26-week assessment. b, Based on original protocol exclusion criteria.

feasible to arrange a biopsy in the time frame required. Of the 400 patients who completed both assessments (the ultrasound and biopsy), a further nine were excluded, leaving 391 patients for the secondary analysis. Ten of these patients could not be included in the primary analysis because their biopsy had been performed at least 10 days after starting steroid treatment. Before the final assessment (after 6 months) another 49 patients were excluded from the analysis. The main reasons were death ($n = 16$), lost to follow-up ($n = 14$) or had withdrawn consent for the study ($n = 11$). The recruitment numbers for each centre are shown in *Figure 4*, and *Figure 5* demonstrates the cumulative recruitment over the course of the whole study. Two sites recruited the majority of patients, but eight other sites recruited > 10 cases each. Initial recruitment was slower than predicted. An extension to the recruitment period was agreed and revised planned recruitment is shown in *Figure 5*.

Summary of test results and the reference diagnosis

All 381 patients underwent ultrasound examination in accordance with the study protocol, and all patients had a TAB performed as part of the normal standard of care for investigations of patients with suspected GCA. In total, 101 patients (27%) had an abnormal biopsy that was consistent with a diagnosis of GCA. In a total of 28 patients biopsies failed; in four cases no samples were obtained and in 24 cases the biopsy sample did not contain arterial tissue. These patients are defined in the main analyses as having a diagnosis that is not consistent with GCA (i.e. they are analysed on the assumption that they do not have the disease). In total, 162 patients (43%) had an abnormal ultrasound that was compatible with a diagnosis of GCA. After expert review, the reference diagnosis of GCA was given to 257 patients, a prevalence of 66% in the study cohort. The diagnosis of GCA conventionally rests on the clinical pattern at presentation, combined with the results of laboratory tests, including ESR or CRP level, the response to steroid therapy and the TAB result. For many patients, not all of these aspects (clinical findings and symptoms, serological abnormalities or biopsy results) are entirely consistent with the diagnosis, which leads to a degree of variability in interpreting the findings. For example, patients who have symptoms suggestive of GCA,

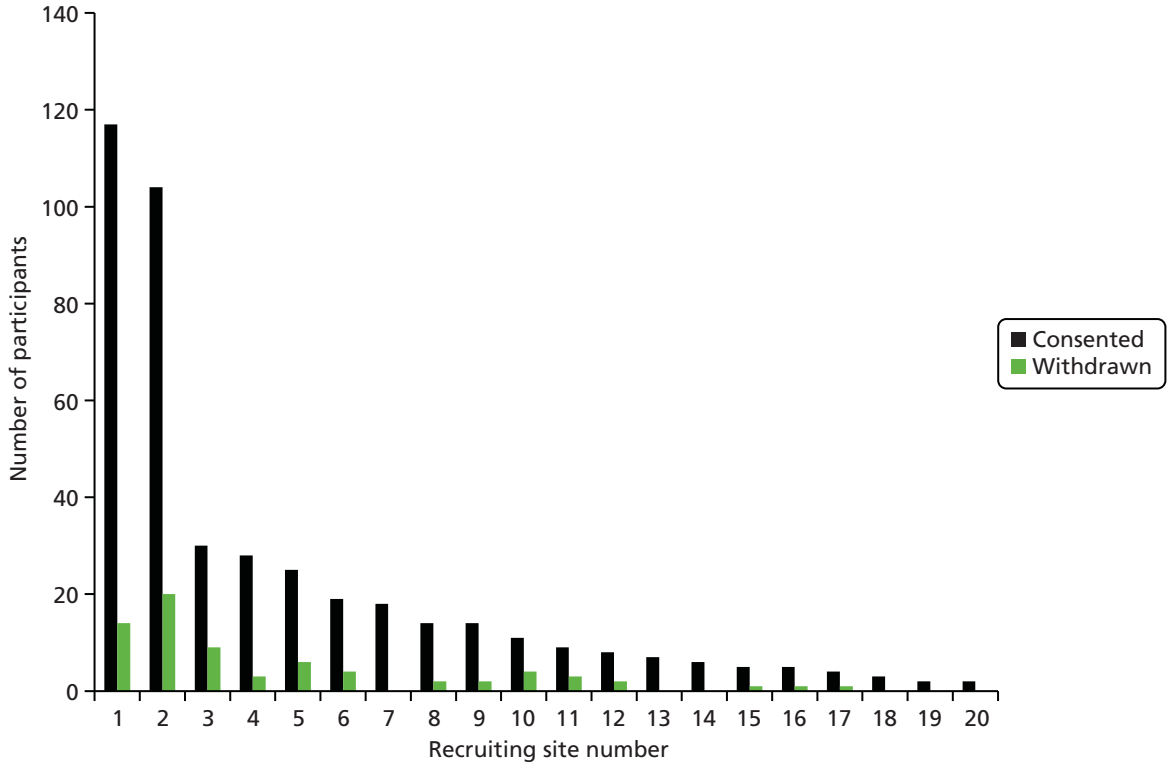


FIGURE 4 Number of patients recruited to the study per centre, separately listing consented patients and subsequent withdrawals from the study.

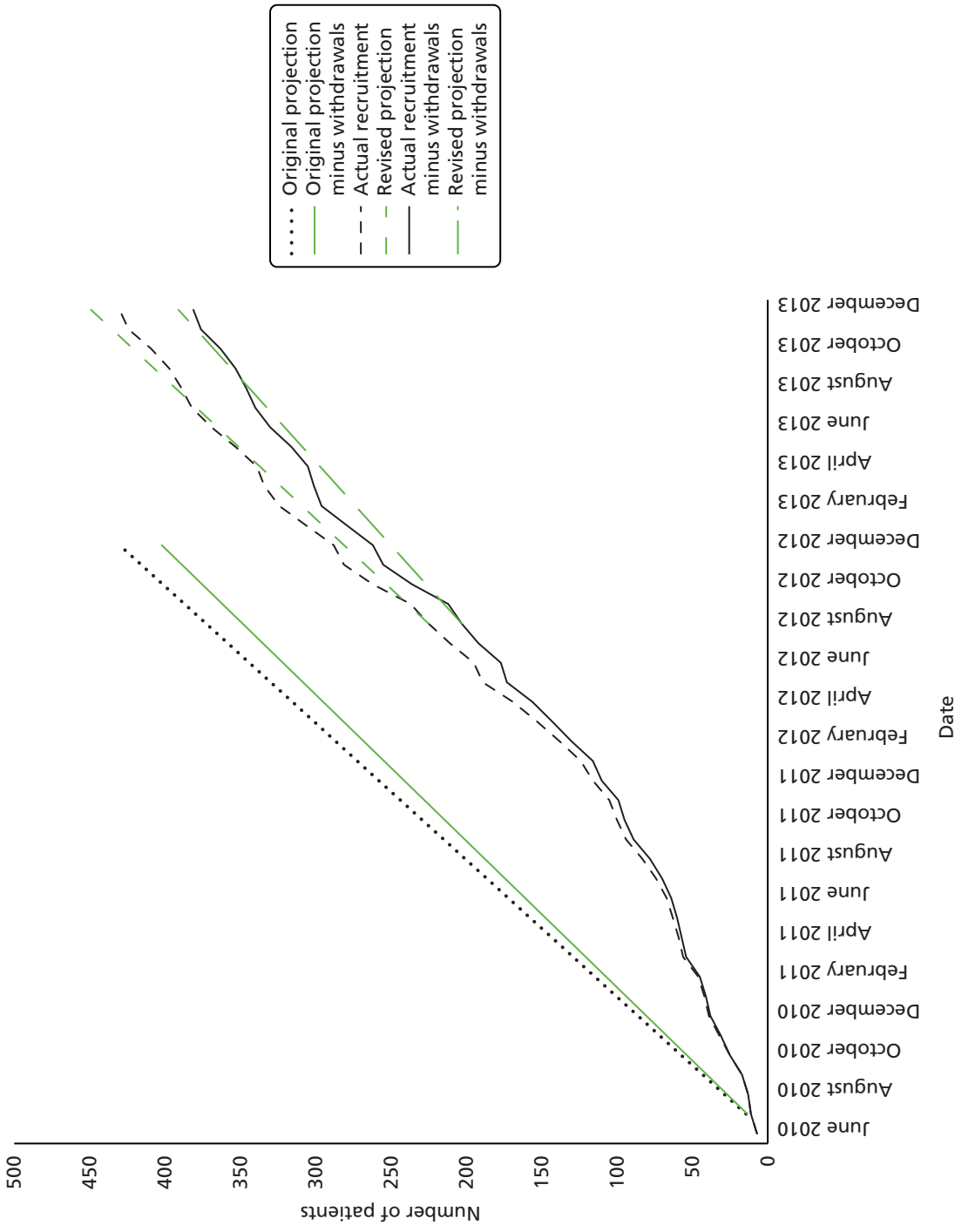


FIGURE 5 Recruitment of patients to the study, including original projected recruitment plans, revised projections and actual recruitment rate before and after withdrawals.

such as new-onset headache and jaw claudication, may actually have a normal or low ESR and or CRP level; furthermore, the biopsy result may be negative, especially if the test was performed after the patient had been treated with high doses of glucocorticoid therapy for > 7 days and/or the biopsy was small (less than 1 cm of artery). Under these circumstances it might be difficult to be absolutely certain of the diagnosis and by the time the biopsy result is provided to the clinician, it is usually too late to go back to recheck any of the tests again because, in the meantime, the patient has continued to receive high-dose glucocorticoid treatment, which is likely to significantly suppress any evidence of inflammation.

The clinical diagnosis of GCA requires the clinician to use their expertise in interpreting these different pieces of information. We therefore included an expert review in our study, so that all cases in which there was any doubt about the diagnosis of GCA were subject to expert review of the clinical and serological findings. The end result was to produce a 'reference diagnosis' of GCA based on the clinician's interpretation of all of this information. *Figure 6* summarises the different combinations of biopsy and ultrasound test results (GCA, not GCA or, in the case of biopsy, unsuccessful) and the final reference diagnosis for the 381 patients in the primary analysis group.

In 187 patients the clinician's interpretation was submitted for expert review, and in 23 patients the interpretation was altered (from an interpretation of GCA to a reference diagnosis of not GCA in 14 patients and from an interpretation of not GCA to a reference diagnosis of GCA in nine patients). *Figure 7* illustrates these interpretations and reference diagnoses with respect to the clinicians' initial assessment of GCA at presentation.

Participant characteristics

Demographics

Demographic characteristics of the cohort are shown in *Table 4*; 377 patients (99%) were aged > 50 years (one of the ACR criteria). The median age of participants was 71 years [interquartile range (IQR) 64–78 years] and 72% were female. Two recruiting centres provided the majority of patients (Nuffield Orthopaedic Centre and Southend University Hospital); 11 centres recruited fewer than 10 patients each. The majority of patients were white British (80%); most of the remainder were either white Irish or from another white background. Only 3% of patients were from a non-white background. The low numbers of non-white patients is in keeping with other data suggesting that GCA is much less common in these populations.[72]

Presenting characteristics

Current and previous medical histories at baseline are shown in *Tables 5* and *6*. The most common symptoms at baseline were localised pain in the head (88%), fatigue (65%), generalised scalp tenderness (59%) and pain over the temporal artery (51%). Although 145 and 99 patients were reported as still experiencing headache after 2 weeks and 6 months, respectively, only two patients developed new headache after the baseline visit (two new cases at 6 months). Systemic features such as fever, night sweats and anorexia affected around one-third of patients. Features suggesting accompanying PMR were reported in one-third of patients. The median time between first symptom onset and baseline was 31 days (IQR 10–93 days, $n = 377$); the median time between symptom onset and starting steroids was 33 days (IQR 13–99 days, $n = 379$). Symptoms suggesting ischaemic complications such as jaw or tongue complications were common, affecting up to 43% of patients. Baseline features of visual involvement were very common (43%), in keeping with other studies.[26] However, when we separated patients with GCA from the non-GCA patients, the frequency of visual features was only marginally higher at baseline (45% vs. 37%), 2 weeks (30% vs. 23%) and 6 months (27% vs. 22%) in the patients with GCA, as shown in *Table 7*. The frequency of ischaemic optic neuropathy on physical examination (when performed) was higher in the GCA group than in the non-GCA group at baseline (10% vs. 5%), 2 weeks (6% vs. 1%) and 6 months (4% vs. 3%). Ten new cases of reduced or lost vision in either eye were reported during follow-up: three cases (1%) were reported at 2 weeks and seven cases (2%) were reported at 6 months. Six (2%) new cases of double vision were reported at 6 months. The clinician overseeing the patient's care was responsible for reporting these

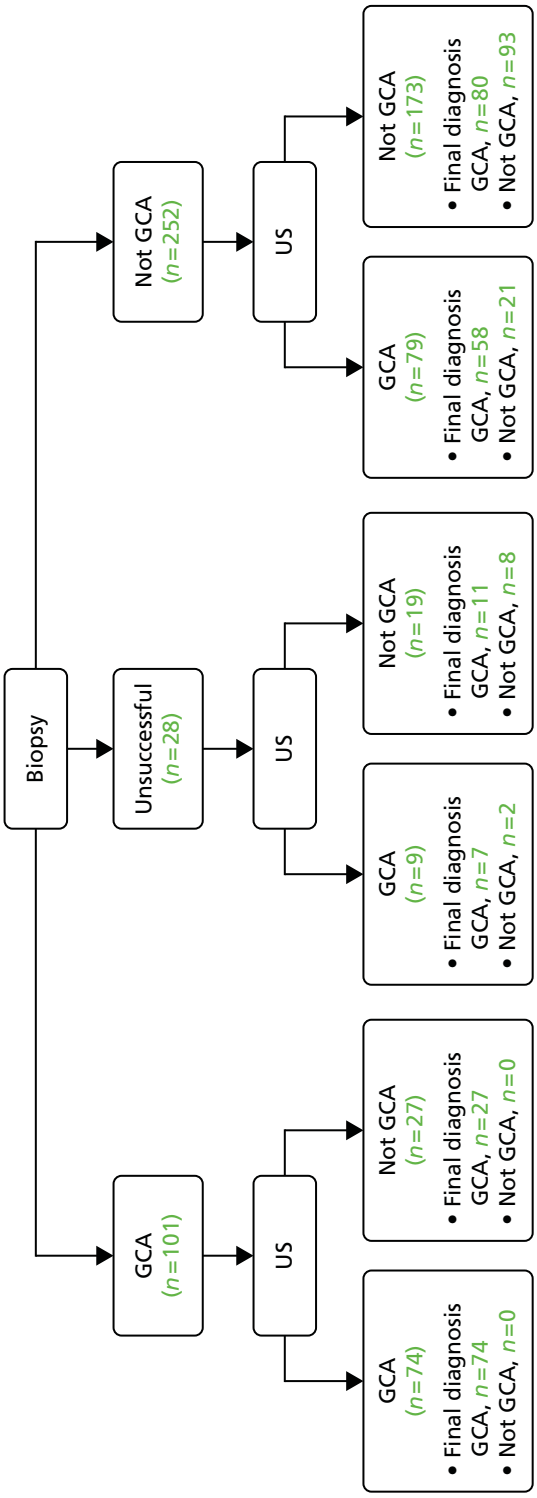


FIGURE 6 Simplified diagnosis flow for the primary analysis group ($n = 381$). US, ultrasound.

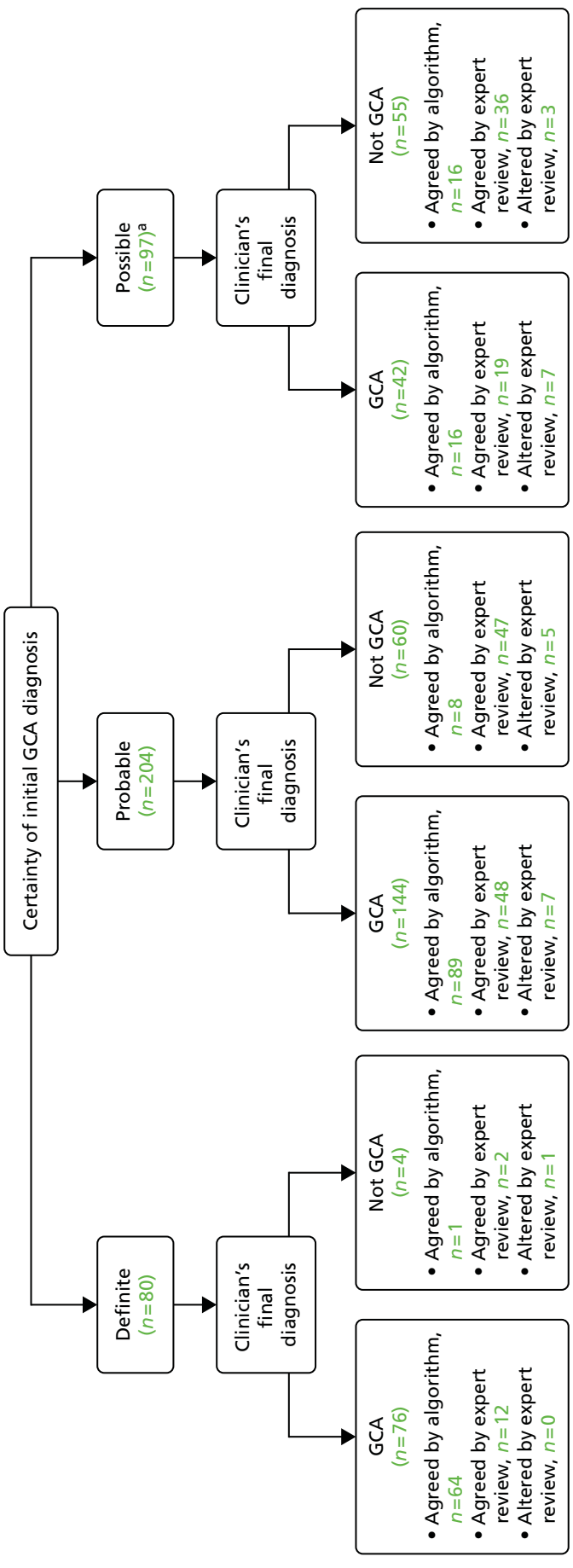


FIGURE 7 Simplified flow of clinician and reference diagnoses for the primary analysis group ($n = 381$). a, In one patient, data on the certainty of initial diagnosis were missing and, therefore, this patient has been included in the possible group.

TABLE 4 Characteristics of study participants

Characteristic	Summary ($N = 381$)
Age (years)	
Number (%) of responses	381 (100.0)
Mean (SD)	71.1 (9.8)
Median (IQR)	71.7 (64.3–77.8)
Sex,* n *(%)	
Male	108 (28.3)
Female	273 (71.7)
Site,* n *(%)	
Chapel Allerton Hospital, Leeds, UK	16 (4.2)
City Hospital, Birmingham, UK	4 (1.0)
Dudley Hospital, Dudley, UK	4 (1.0)
Gateshead Hospital, Gateshead, UK	14 (3.7)
Great Yarmouth Hospital, Great Yarmouth, UK	2 (0.5)
Hospital de Santa Maria, Lisbon, Portugal	2 (0.5)
Hospital of Southern Norway Trust, Kristiansand, Norway	25 (6.6)
Jena University Hospital, Jena, Germany	12 (3.1)
Musgrave Park, Belfast, UK	6 (1.6)
Nuffield Orthopaedic Centre, Oxford, UK	111 (29.1)
Princess Alexandra Hospital, Harlow, UK	7 (1.8)
Queen Alexandra Hospital, Portsmouth, UK	7 (1.8)
Queen's Hospital Romford, Essex, UK	8 (2.1)
Queen's Medical Centre, Nottingham, UK	22 (5.8)
Royal Berkshire Hospital, Reading, UK	4 (1.0)
Royal Derby Hospital, Derby, UK	3 (0.8)
Southend University Hospital, Southend, UK	90 (23.6)
St Vincent Hospital, Dublin, Ireland	18 (4.7)
Stoke Mandeville Hospital, Stoke, UK	20 (5.2)
Sunderland Royal Hospital, Sunderland, UK	6 (1.6)
Ethnic group,* n *(%)	
White British	303 (79.5)
Irish	22 (5.8)
Other white background	45 (11.8)
Other mixed background	1 (0.3)
Indian	5 (1.3)
Pakistani	2 (0.5)
Other Asian	1 (0.3)
Caribbean	1 (0.3)
Chinese	1 (0.3)

IQR, interquartile range; SD, standard deviation.

TABLE 5 Symptoms by visit

Symptoms	Baseline[a] (*N* = 381), *n* (%)	2 weeks (*N* = 381), *n* (%)		6 months (*N* = 335), *n* (%)	
		All	New	All	New
Localised pain in the head	337 (88.5)	145 (38.1)	0 (0.0)	99 (29.6)	2 (0.6)
Generalised scalp tenderness	223 (58.5)	83 (21.8)	6 (1.6)	49 (14.6)	4 (1.2)
Pain over temporal artery	194 (50.9)	67 (17.6)	1 (0.3)	45 (13.4)	7 (2.1)
Swelling over temporal artery	92 (24.1)	25 (6.6)	2 (0.5)	12 (3.6)	4 (1.2)
Bilateral shoulder pain	123 (32.3)	40 (10.5)	1 (0.3)	42 (12.5)	11 (3.3)
Bilateral hip stiffness or pain	68 (17.8)	17 (4.5)	3 (0.8)	20 (6.0)	7 (2.1)
Early-morning stiffness > 1 hour	75 (19.7)	22 (5.8)	3 (0.8)	26 (7.8)	11 (3.3)
Fatigue	246 (64.6)	141 (37.0)	10 (2.6)	120 (35.8)	11 (3.3)
Anorexia	140 (36.7)	45 (11.8)	3 (0.8)	27 (8.1)	6 (1.8)
Symptoms of fever or night sweats	143 (37.5)	60 (15.7)	4 (1.0)	53 (15.8)	10 (3.0)
Jaw claudication	163 (42.8)	62 (16.3)	2 (0.5)	26 (7.8)	3 (0.9)
Tongue claudication	20 (5.2)	7 (1.8)	1 (0.3)	3 (0.9)	1 (0.3)
Reduced or lost vision in either eye	133 (34.9)	95 (24.9)	3 (0.8)	80 (23.9)	7 (2.1)
Amaurosis fugax	14 (3.7)	5 (1.3)	5 (1.3)	3 (0.9)	2 (0.6)
Double vision	31 (8.1)	11 (2.9)	0 (0.0)	9 (2.7)	6 (1.8)

a Present at either pre-steroids or baseline.

TABLE 6 Medical history and conditions at baseline

Clinical feature (*N* = 381)	Current, *n* (%)	Past, *n* (%)
Medical history		
PMR	28 (7.3)	9 (2.4)
Stroke/TIA	5 (1.3)	27 (7.1)
Migraine	13 (3.4)	4 (1.0)
Headache	7 (1.8)	3 (0.8)
Shingles	1 (0.3)	6 (1.6)
Sinusitis	6 (1.6)	1 (0.3)
Conditions		
Diabetes mellitus	54 (14.2)	0 (0.0)
Hypertension	200 (52.5)	9 (2.4)
Angina	28 (7.3)	24 (6.3)
Myocardial infarction	0 (0.0)	23 (6.0)
Heart failure	19 (5.0)	8 (2.1)
Malignancy	9 (2.4)	53 (13.9)
Low trauma fracture (hip, spine, forearm, other)	1 (0.3)	56 (14.7)

TIA, transient ischaemic attack.

TABLE 7 Visual features by visit

Visual feature	Baseline		2 weeks		6 months	
	GCA (*N* = 257), *n* (%)	Not GCA (*N* = 124), *n* (%)	GCA (*N* = 257), *n* (%)	Not GCA (*N* = 124), *n* (%)	GCA (*N* = 227), *n* (%)	Not GCA (*N* = 108), *n* (%)
Symptoms and physical examination						
Any visual feature[a]	115 (44.7)	46 (37.1)	77 (30.0)	29 (23.4)	61 (26.9)	24 (22.2)
Visual loss	94 (36.6)	39 (31.5)	69 (26.8)	26 (21.0)	58 (25.6)	22 (20.4)
Anterior or posterior ischaemic optic neuropathy	25 (9.7)	6 (4.8)	16 (6.2)	1 (0.8)	9 (4.0)	3 (2.8)
BVAS						
Blurred vision			32 (12.5)	21 (16.9)	10 (4.4)	4 (3.7)
Sudden visual loss			25 (9.7)	2 (1.6)	3 (1.3)	1 (0.9)
VDI						
Blindness (no cataracts)			8 (3.1)	1 (0.8)	15 (6.6)	0 (0.0)
Blindness and cataracts			0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Optic atrophy			2 (0.8)	0 (0.0)	3 (1.3)	1 (0.9)
Visual impairment/diplopia			8 (3.1)	3 (2.4)	26 (11.5)	7 (6.5)
Combined						
Any visual features[b]	115 (44.7)	46 (37.1)	84 (32.7)	34 (27.4)	71 (31.3)	28 (25.9)
Any visual loss[c]	94 (36.6)	39 (31.5)	69 (26.8)	26 (21.0)	58 (25.6)	22 (20.4)
Optic neuropathy or atrophy	25 (9.7)	6 (4.8)	18 (7.0)	1 (0.8)	11 (4.8)	3 (2.8)

a Defined as presence of reduced or lost vision in either eye, double vision or amaurosis fugax.
b Defined as presence of reduced or lost vision in either eye, double vision or amaurosis fugax, blurred vision or sudden visual loss on the BVAS, blindness or visual impairment/diplopia on the VDI.
c Defined as presence of visual loss, sudden visual loss on the BVAS or blindness on the VDI.

data, which may or may not have been independently verified by an ophthalmologist. Ascertaining whether or not the visual features are definitely related to GCA is very difficult. We expected that there would be a tendency to report any visual features as possibly related to GCA, however unlikely this is, because the consequences of missing early ischaemic ophthalmological complications would be disastrous for the patient. If we look for more robust evidence of visual loss directly as a result of GCA, we may have to accept that reporting the number of patients with ischaemic optic neuropathy will underestimate the real risk, while accepting all reported visual loss will overestimate the real risk. The presence of ischaemic optic neuropathy or an afferent pupillary defect could be explained by a complication of a presumed diagnosis of GCA. However, non-arteritic anterior ischaemic optic neuropathy[73] can present in a similar way to GCA with visual loss but is not typically associated with headache or an elevation of the acute phase response. Non-arteritic anterior ischaemic optic neuropathy was reported in 5% of the non-GCA cases in this study at baseline.

Twenty-eight patients had a diagnosis of PMR at baseline and a further nine patients had a previous history of PMR. Levels of hypertension were high (52% of the cohort), 14% of the cohort had pre-existing diabetes mellitus at baseline, 7% were suffering from angina and 5% had heart failure. A total of 2% of the cohort had a current history of cancer but 14% had a previous history of any form of malignancy. Around 15% had previously suffered a low-trauma fracture; one of the patients had a fracture at the time of presentation. Not all patients had ESR or CRP level measured prior to starting steroids. Only 73% of patients had a CRP level tested before starting treatment and 73% had an ESR performed before steroids

were commenced. There were 75 (19.7%) patients in whom neither ESR nor CRP levels were measured before starting steroids, and in only one of these was plasma viscosity measured.

We would expect a dramatic and rapid reduction in the acute phase response as a result of glucocorticoid therapy. Laboratory results for ESR were higher prior to the use of high doses of glucocorticoid therapy [mean 46.5 mm/hour, standard deviation (SD) 33.4 mm/hour] when compared with the results at baseline (mean 37.1 mm/hour, SD 31.4 mm/hour). Similar results were found for CRP values, which were higher before glucocorticoid therapy than at baseline (mean 63.8 mg/l, SD 58.9 mg/l, compared with mean 39.0 mg/l, SD 40.4 mg/l).

Visual features over time are displayed in *Table 7*. Visual features at baseline were reported in a total of 162 (42%) participants, with a slightly higher proportion in the reference GCA group than in the group of patients whose diagnosis was not GCA. Thirty-seven per cent of patients with GCA and 31% of patients without GCA experienced visual loss at baseline; these values fell to 26% and 20%, respectively, at 6 months. If we look at reporting of visual features based on data in the BVAS and VDI assessments at 2 weeks and 6 months, respectively, blurred vision was reported as frequently in patients with GCA as in those who did not have GCA. Sudden visual loss was more often reported in patients with GCA than in patients without GCA (10% vs. 2%) at 2 weeks. The VDI reported blindness (not related to cataract) in eight patients (3%) with GCA at the 2-week assessment, and in one patient in the non-GCA group; by 6 months, 7% of patients with GCA were reported as blind. In one case this was recorded as blindness in both eyes; in all other cases blindness was recorded as occurring in one eye. The number of patients reported as having visual impairment or diplopia increased from 2 weeks to 6 months in both groups, possibly suggesting a side effect of the glucocorticoid treatment. Combining the data from the main CRF with the BVAS and VDI reporting of visual features, there were slightly more visual features (and specifically visual loss) at each visit in the GCA patients than in the non-GCA patients. More objective findings of ischaemic optic atrophy were less common in both groups but not dissimilar at baseline (10% vs. 5%), 2 weeks (7% vs. 1%) and 6 months (5% vs. 3%).

Table 8 shows findings from the physical examination at baseline; the most common symptom was tenderness of the temporal arteries (50% abnormal) that was most commonly unilateral, but in 11% was bilateral. Thickening of one or both temporal arteries was reported in 27% of patients and reduced or absent pulsation in the temporal artery was detected in 91 patients. Tenderness of either axillary artery was much less common and reported in only 34 patients. Bruits were detected in 15 individuals and could in some patients represent extracranial large vessel vasculitis, but in other patients could have been pre-existing bruits due to atherosclerosis. In fact, only 5 out of 15 patients with detectable bruits had abnormal findings on ultrasound of the axillary arteries. Of the 296 who did not have detectable bruits, 42 had abnormal axillary findings on ultrasound. In seven patients, stroke was part of the initial presentation of their GCA. Cranial nerve palsy was reported in three patients. Three patients presented with aneurysms of an artery at diagnosis. Four participants had no abnormal features reported.

Table 9 shows the physical examination findings by the length of time on steroids. Reduced or absent pulsation and thickened temporal artery appear to have been less common in those patients who had been on steroids for ≥ 3 days than in those on glucocorticoid therapy for a shorter duration. The documentation of physical examination findings was structured to elicit specific features that would be expected to occur in patients with GCA. Some of these physical examination findings would require input from other clinical staff such as ophthalmologists to confirm the presence or absence of anterior ischaemic optic neuropathy/posterior ischaemic optic neuropathy; this would explain the large number of missing values attributed to these two items. Reporting of relative afferent pupillary defect was often omitted. This feature could have been evaluated by a generalist with no specific expertise in ophthalmology, but it would require the use of a torch or an ophthalmoscope to shine in the patient's eyes. It is possible that in some centres such equipment was not available in the department while patients were being seen.

Table 9 summarises the relationships between glucocorticoid use and the presence of physical findings. It demonstrates that 92 of the patients had either not started high doses of glucocorticoids at all or had

TABLE 8 Findings from the physical examination at baseline for all patients, with findings for patients with GCA or not shown separately

	Reference GCA, *n* (%)				Reference not GCA, *n* (%)			
Feature (*N* = 381)	Unilateral	Bilateral	Normal	Missing	Unilateral	Bilateral	Normal	Missing
Tender temporal artery	97 (36.7)	28 (10.6)	132 (50.0)		53 (41.7)	14 (11.0)	56 (44.1)	1 (0.8)
Thickened temporal artery	54 (20.5)	32 (12.1)	171 (64.8)		12 (9.4)	4 (3.1)	107 (84.3)	1 (0.8)
Reduced or absent pulsation in temporal artery	57 (21.6)	22 (8.3)	178 (67.4)		10 (7.9)	2 (1.6)	111 (87.4)	1 (0.8)
Tender axillary artery	15 (5.7)	6 (2.3)	234 (88.6)	2 (0.8)	10 (7.9)	3 (2.4)	109 (85.8)	2 (1.6)
Bruits	9 (3.4)	6 (2.3)	196 (74.2)	45 (17.0)			100 (78.7)	22 (17.3)
Anterior ischaemic optic neuropathy	20 (7.6)	3 (1.1)	70 (26.5)	159 (60.2)	4 (3.1)		32 (25.2)	84 (66.1)
Posterior ischaemic optic neuropathy	3 (1.1)	1 (0.4)	72 (27.3)	175 (66.3)	3 (2.4)		25 (19.7)	96 (75.6)
Relative afferent pupillary defect	12 (4.5)	1 (0.4)	169 (64.0)	70 (26.5)	2 (1.6)		88 (69.3)	30 (23.6)
III/IV/VI nerve palsy	3 (1.1)		210 (79.5)	41 (15.5)			100 (78.7)	22 (17.3)
Feature	Present		Absent	Not assessed	Present		Absent	Not assessed
Aneurysm	3 (1.1)		211 (79.9)	43 (16.3)			109 (85.8)	15 (11.8)
Stroke	5 (1.9)		246 (93.2)	6 (2.3)	2 (1.6)		116 (91.3)	6 (4.7)

TABLE 9 Findings from physical examination at baseline by length of time on steroids

Feature	Abnormal, *n* (%)			
	Unilateral	Bilateral	Normal	Missing
***Not started steroids or started same day* (N = *92*)**				
Tender temporal artery	32 (34.8)	11 (12.0)	49 (53.3)	0 (0.0)
Thickened temporal artery	18 (19.6)	12 (13.0)	62 (67.4)	0 (0.0)
Reduced or absent pulsation in temporal artery	24 (26.1)	6 (6.5)	62 (67.4)	0 (0.0)
Tender axillary artery	6 (6.5)	0 (0.0)	85 (92.4)	1 (1.1)
Bruits	4 (4.3)	3 (3.3)	66 (71.7)	19 (20.7)
Anterior ischaemic optic neuropathy	6 (6.5)	0 (0.0)	27 (29.3)	55 (59.8)
Posterior ischaemic optic neuropathy	2 (2.2)	0 (0.0)	26 (28.3)	62 (67.4)
Relative afferent pupillary defect	4 (4.3)	0 (0.0)	51 (55.4)	35 (38.0)
III/IV/VI cranial nerve palsy	2 (2.2)	0 (0.0)	20 (21.7)	70 (76.1)
***1–2 days after starting steroids* (N = *149*)**				
Anterior ischaemic optic neuropathy	12 (8.1)	1 (0.7)	41 (27.5)	94 (63.1)
Bruits	2 (1.3)	2 (1.3)	116 (77.9)	28 (18.8)
III/IV/VI nerve palsy	1 (0.7)	0 (0.0)	25 (16.8)	122 (81.9)
Posterior ischaemic optic neuropathy	3 (2.0)	0 (0.0)	38 (25.5)	106 (71.1)
Reduced or absent pulsation in temporal artery	28 (18.8)	10 (6.7)	111 (74.5)	0 (0.0)
Relative afferent pupillary defect	5 (3.4)	0 (0.0)	106 (71.1)	36 (24.2)
Tender axillary artery	7 (4.7)	6 (4.0)	135 (90.6)	1 (0.7)
Tender temporal artery	58 (38.9)	23 (15.4)	68 (45.6)	0 (0.0)
Thickened temporal artery	30 (20.1)	15 (10.1)	104 (69.8)	0 (0.0)
***≥ 3 days after steroids* (N = *138*)**				
Anterior ischaemic optic neuropathy	6 (4.3)	2 (1.4)	33 (23.9)	93 (67.4)
Bruits	3 (2.2)	1 (0.7)	112 (81.2)	20 (14.5)
III/IV/VI nerve palsy	0 (0.0)	0 (0.0)	18 (13.0)	116 (84.1)
Posterior ischaemic optic neuropathy	1 (0.7)	1 (0.7)	32 (23.2)	102 (73.9)
Reduced or absent pulsation in temporal artery	14 (10.1)	8 (5.8)	115 (83.3)	1 (0.7)
Relative afferent pupillary defect	5 (3.6)	1 (0.7)	99 (71.7)	29 (21.0)
Tender axillary artery	12 (8.7)	3 (2.2)	121 (87.7)	2 (1.4)
Tender temporal artery	58 (42.0)	8 (5.8)	71 (51.4)	1 (0.7)
Thickened temporal artery	17 (12.3)	9 (6.5)	111 (80.4)	1 (0.7)

started them only on the same day as the initial assessment. Overall, 149 patients had received high doses of glucocorticoids for 1–2 days prior the assessment and 138 patients had been treated with glucocorticoids for at least 3 days before assessment. *Table 9* shows that clinically detectable abnormalities in the temporal arteries were less evident the longer patients had been treated with high doses of steroids, but, nevertheless, 26 patients still had detectable, thickened temporal arteries and 66 had tender temporal arteries despite 3 days of high-dose glucocorticoid therapy.

Anterior ischaemic optic neuropathy was reported in 27 patients (7% of the cohort) at baseline, which included 23 patients with a subsequent diagnosis of GCA and four patients with a subsequent diagnosis of not GCA; it was reported in six patients who had received either no steroid therapy or < 1 day of steroids; in 13 patients who had received between 1 and 2 days of steroids; and in eight patients with ≥ 3 days of treatment with high doses of steroids. The length of time on steroids may have been a reflection of the severity of the condition (i.e. with patients with visual symptoms being treated more aggressively by their primary care physician before referral to the study).

Table 10 shows that the ESR and CRP level were higher before steroids (mean 46 mm/hour, SD 33.4 mm/hour) than at baseline (mean 37.1 mm/hour, SD 31.4 mm/hour). Similar results were reported for CRP values, which were higher pre-steroids than at baseline (mean 62.6 mg/l, SD 58.5 mg/l, compared with mean 39.3 mg/l, SD 43.8 mg/l). Not all patients had their ESR or CRP level measured prior to starting steroids. CRP level was measured before starting treatment in 74% of patients and ESR was measured in 73% before steroids were commenced. The CRP level and ESR values reported in patients who were diagnosed as having GCA were higher than in those patients diagnosed as not having GCA. This is likely to be explained by the inherent bias in the diagnosis, which would have been influenced by these results.

Ultrasound results

Ultrasound examination was performed on all 381 patients. Abnormalities consistent with GCA were found in 162 (43%) of the ultrasound scans (*Table 11*). *Table 11* shows that the majority of patients with abnormal scans had changes in the temporal arteries (35%) and that 11.5% had abnormalities in the axillary and temporal arteries. Abnormalities were more likely to be bilateral than unilateral (29% vs. 20%). Halo was the most commonly cited reason for reaching a diagnosis of GCA (42.5%). Stenosis (12%) or occlusions (11%) were seen less commonly, but there was an overlap with patients also showing halo. The maximum length of halo in those patients in whom a halo was present was 20 mm (median) in the axillary arteries and 9 mm in the temporal arteries. It was, however, sometimes extremely difficult (especially in temporal arteries) to measure the length as a result of vessel tortuosity. In some patients the halo extended the entire length of the scanned artery. The maximum reported median thickness of halo was 1.1 mm (IQR 0.6–1.4 mm) in the axillary arteries and 0.6 mm (IQR 0.4–0.9 mm) in the temporal arteries.

TABLE 10 Laboratory test results at baseline

	Pre-steroids		Baseline	
Test	GCA (*N* = 257)	Not GCA (*N* = 124)	GCA (*N* = 257)	Not GCA (*N* = 124)
ESR value (mm/hour)				
Number (%) of responses	187 (72.8)	92 (74.2)	231 (89.9)	110 (88.7)
Mean (SD)	55.0 (33.5)	29.4 (26.0)	44.5 (33.0)	21.7 (20.7)
Median (IQR)	53.0 (28.0–83.0)	18.0 (8.5–49.5)	38.0 (19.0–63.0)	14.0 (6.0–33.0)
CRP value (mg/l)				
Number (%) of tests	191 (74.3)	87 (70.1)	238 (92.6)	113 (91.1)
CRP level in the normal range (no value reported), *n* (%)	35 (13.6)	38 (30.6)	63 (24.5)	75 (60.5)
Number (%) of CRP values reported	156 (60.7)	49 (39.5)	175 (68.1)	38 (30.6)
Mean (SD)	70.4 (56.6)	38.1 (58.1)	42.2 (42.3)	26.4 (49.1)
Median (IQR)	54.0 (27.0–101.5)	16.0 (7.8–36.1)	31.0 (14.0–54.0)	10.4 (3.0–24.6)

TABLE 11 Ultrasound findings in 381 patients with suspected GCA

US finding	Summary ($N = 381$)
Presence of abnormality,* n *(%)	
No	195 (51.2)
Yes	186 (48.8)
Site of abnormality,* n *(%)	
Temporal	133 (34.9)
Axillary	9 (2.4)
Both temporal and axillary	44 (11.5)
Spread of abnormality,* n *(%)	
Unilateral	75 (19.7)
Bilateral	111 (29.1)
Sonographers' opinion,* n *(%)	
Not GCA	219 (57.5)
GCA	162 (42.5)
Any halo	162 (42.5)
Any stenosis	45 (11.8)
Any occlusion	41 (10.8)
Axillary halo maximum length (mm)	
Number of measurements	15
Mean (SD)	26.1 (22.9)
Median (IQR)	20.0 (12.0–34.0)
Axillary halo maximum thickness (mm)	
Number of measurements	62
Mean (SD)	1.1 (1.0)
Median (IQR)	1.1 (0.6–1.4)
Minimum, maximum	0.1, 6.7
Temporal halo maximum length (mm)	
Number of measurements	181
Mean (SD)	12.0 (11.3)
Median (IQR)	9.0 (6.0–14.0)
Temporal halo maximum thickness (mm)	
Number of measurements	461
Mean (SD)	0.7 (0.7)
Median (IQR)	0.6 (0.4–0.9)
Minimum, maximum	0.1, 8.8
If scan abnormal, number of abnormal segments	
Number of measurements	186
Mean (SD)	3.6 (2.8)
Median (IQR)	2.5 (1.0–6.0)

US, ultrasound.

Table 12 details the artery on which the halo was identified. In total, at least one halo on ultrasound was reported in 162 patients, in the majority of whom ($n = 118$) haloes were seen only on the temporal artery (bilateral, $n = 60$; unilateral, $n = 58$). By contrast, just nine patients had a halo on the axillary artery only, with no halos seen in the temporal arteries. In the remaining 35 patients halos were observed on both temporal and axillary arteries.

In 24 patients, ultrasound showed abnormalities but the sonographer's diagnosis was not GCA. *Table 13* describes the characteristics of these patients. The majority of the abnormalities were found in the temporal arteries (18 patients) but eight patients had axillary artery abnormalities. The abnormal findings were unilateral in 14 patients and bilateral in 10 patients and halo was detected in 10 patients, stenosis in nine and occlusion in four. Of these 24 assessments, 10 were in agreemeent with the ultrasound expert reviewers, seven were in disagreement and seven were unclear. In 23 of these 24 patients, the scan findings were attributed to atherosclerosis; one abnormal case was attributed to the use of radiotherapy for breast cancer.

Having completed training (which included 10 ultrasound test cases and at least one hot case), 23 sonographers performed ultrasound scans in the TABUL study. Around half of the scans in the study were undertaken by two sonographers, who performed more than 80 scans each. *Figure 8* shows the number of ultrasound assessments undertaken by the 23 sonographers.

It is possible that the reliability of sonographers who completed fewer than 10 scans was lower than the reliability of sonographers recruiting more than 10 patients. We examined the evidence for this, which demonstrated an effect on sensitivity but not specificity (see *Chapter 5*). *Table 14* compares the

TABLE 12 Halo findings in the temporal arteries compared with the axillary arteries

Halo findings		Temporal			
Axillary	*N*	Bilateral ($N = 81$), *n* (%)	Left ($N = 34$), *n* (%)	Right ($N = 38$), *n* (%)	None ($N = 228$), *n* (%)
Bilateral	20	13 (16.0)	3 (8.8)	1 (2.6)	3 (1.3)
Left	13	6 (7.4)	0 (0.0)	4 (10.5)	3 (1.3)
Right	11	2 (2.5)	4 (11.8)	2 (5.3)	3 (1.3)
None	337	60 (74.1)	27 (79.4)	31 (81.6)	219 (96.1)

TABLE 13 Characteristics of the ultrasound assessments with abnormalities but where the sonographer's diagnosis is not GCA

US finding	Summary ($N = 24$)
Site of abnormality,* n *(%)	
Temporal	18 (75.0)
Axillary	8 (33.3)
Spread of abnormality,* n *(%)	
Unilateral	14 (58.3)
Bilateral	10 (41.7)
Any halo	10 (17.5)
Any stenosis	9 (15.8)
Any occlusion	4 (7.0)

US, ultrasound.

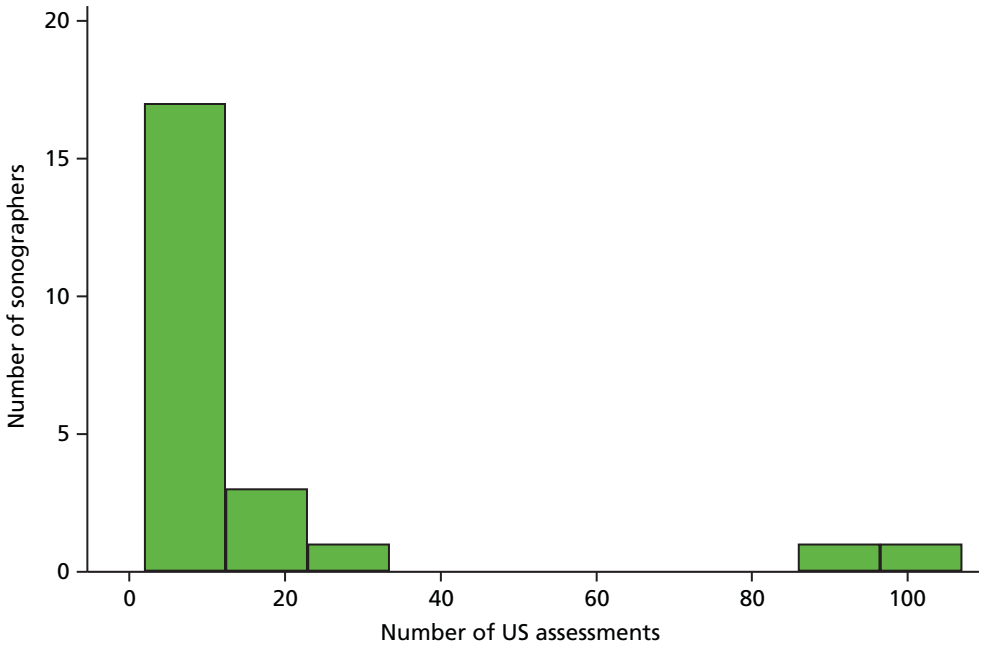


FIGURE 8 Number of ultrasound study assessments performed by the sonographers ($n = 23$). US, ultrasound.

TABLE 14 Comparison of sonographer diagnosis and ultrasound expert review

	US expert review diagnosis, *n* (%)			
Diagnosis	**GCA ($n = 109$)**	**Not GCA ($n = 212$)**	**Unclear ($n = 58$)**	**Not reviewed ($n = 2$)**
Sonographer diagnosis				
GCA	95 (87.2)	47 (22.2)	22 (37.9)	0 (0.0)
Not GCA	14 (12.8)	165 (77.8)	36 (62.1)	2 (100.0)

US, ultrasound.

sonographers' diagnoses with the ultrasound expert review. The expert reviewers agreed with the sonographers' findings in 260 out of 381 patients for whom the images were clear. In 61 patients (16%), there was a disagreement about interpretation of the scan findings, but in a further 60 patients the main reason for disagreement was on the basis of an unclear or unreviewed scan result, suggesting that technical ability to perform the scan rather than interpretation of the scan result was the main problem. The limitation of technical proficiency at scanning is an important problem to address and highlights the need to consider more training if this is the main issue for the sonographer. It is also possible that the problem is a result of patient factors; for example, the presence of very tortuous temporal arteries can make it more difficult for less experienced sonographers to adequately visualise the whole of the artery.

Figure 9 shows the time interval from starting steroids to undertaking the ultrasound scan or the biopsy as well as the number of days between performing each test. Scans were performed more quickly than biopsies (as part of the protocol, the scan had to be performed before the biopsy because if the biopsy was carried out first, then that section of the artery would no longer be available for scanning). In general, the scan was easier to obtain at very short notice, typically within a few days of starting the steroid treatment, whereas the biopsy was sometimes not possible to schedule until later in the week after commencing steroids (or even later for 10 patients). However, despite these potential difficulties, for 215 out of 391 patients (55%), the tests were performed within 2 days of each other, including 52 patients (13%) for whom the scan and biopsy were performed on the same day.

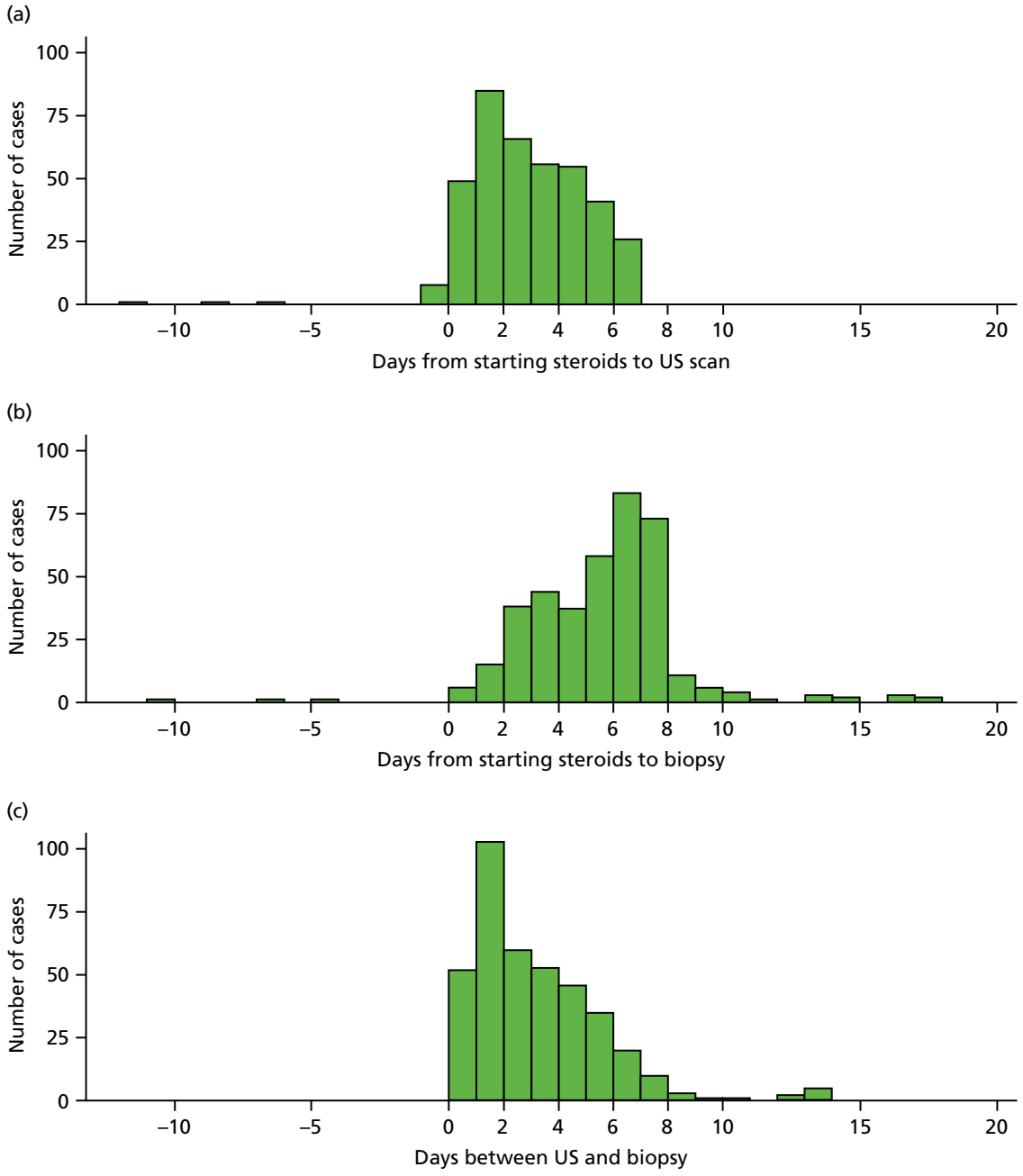


FIGURE 9 Days between starting steroids and performing ultrasound or TAB, and number of days between performing ultrasound and TAB for the 391 patients included in the secondary analysis. US, ultrasound. (a) Days from starting steroids to ultrasound scan; (b) days from starting steroids to biopsy; and (c) days between ultrasound and biopsy.

We attempted to measure the time taken to perform each scan by asking each sonographer to report the time of starting and finishing each scan. Complete data were available for 371 patients. The median time to complete a scan was 30 minutes (IQR 20–35 minutes). Looking across the centres, however, there was considerable variation. Some centres had much longer scanning times (median of 45 minutes); by contrast, the shortest scanning time was only 8.5 minutes, as shown in *Table 15*.

We could not see any relationship between the duration of the scan and when during the course of the study the scan was performed (*Figure 10*). It is likely that the scan times recorded were an estimate of the actual time taken. Ultrasound positive scans appeared to take longer than negative scans; the median time taken for positive scans was 33.5 minutes (IQR 30–40 minutes), compared with 25 minutes (IQR 20–30 minutes) for negative scans, as shown in *Figure 11*. This indicates that it takes longer to scan and document and record

TABLE 15 Reported time (minutes) taken to perform an ultrasound scan of both temporal and both axillary arteries by site

Site	*N*	Mean	SD	Median	IQR
1	18	9.3	4.0	8.5	6–12
2	6	22.5	16.0	15.0	15–20
3	2	20.0	0.0	20.0	20–20
4	7	35.4	21.2	25.0	20–40
5	109	26.0	8.0	25.0	20–30
6	21	25.8	8.7	25.0	20–29
7	4	27.5	10.4	27.5	20–35
8	4	32.5	11.9	27.5	25–40
9	19	32.1	16.7	28.0	25–35
10	8	33.8	13.1	29.0	27–40.5
11	24	26.7	8.6	30.0	20–30
12	13	33.8	10.8	30.0	25–40
13	88	33.5	6.8	30.0	30–40
14	16	32.9	14.3	33.5	25–40
15	4	36.3	12.5	37.5	27.5–45
16	7	50.7	19.9	40.0	35–60
17	12	44.6	12.1	40.0	37.5–52.5
18	3	44.3	5.1	43.0	40–50
19	6	45.0	14.1	45.0	35–60
All	371	29.9	12.3	30.0	20–35

Ranked from the fastest to the slowest median time. No data were available for scan times at one site.

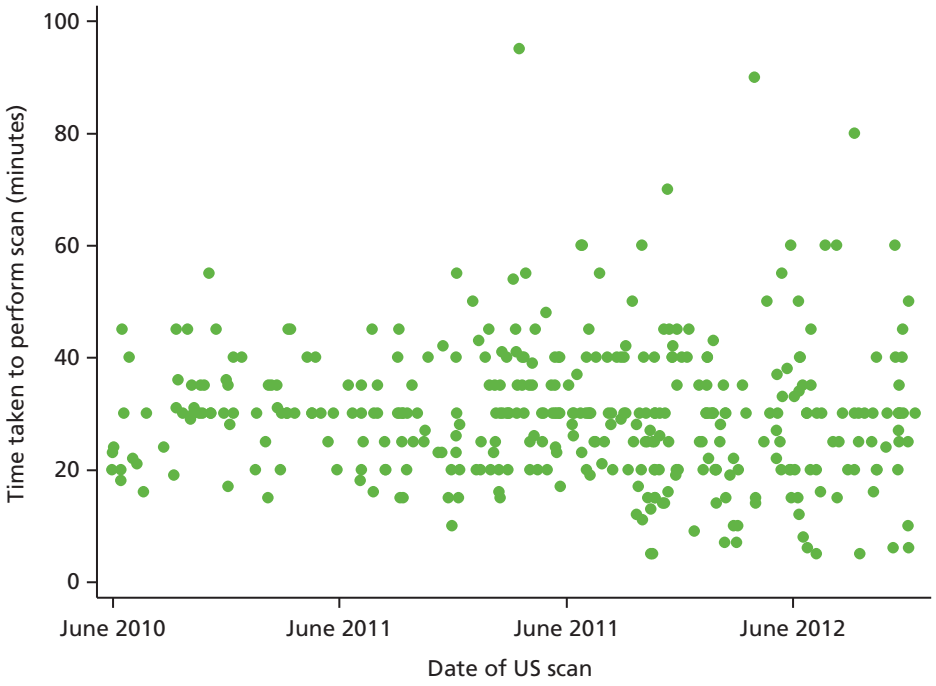


FIGURE 10 Range of times taken to complete scans during the course of the study ($n = 371$). US, ultrasound.

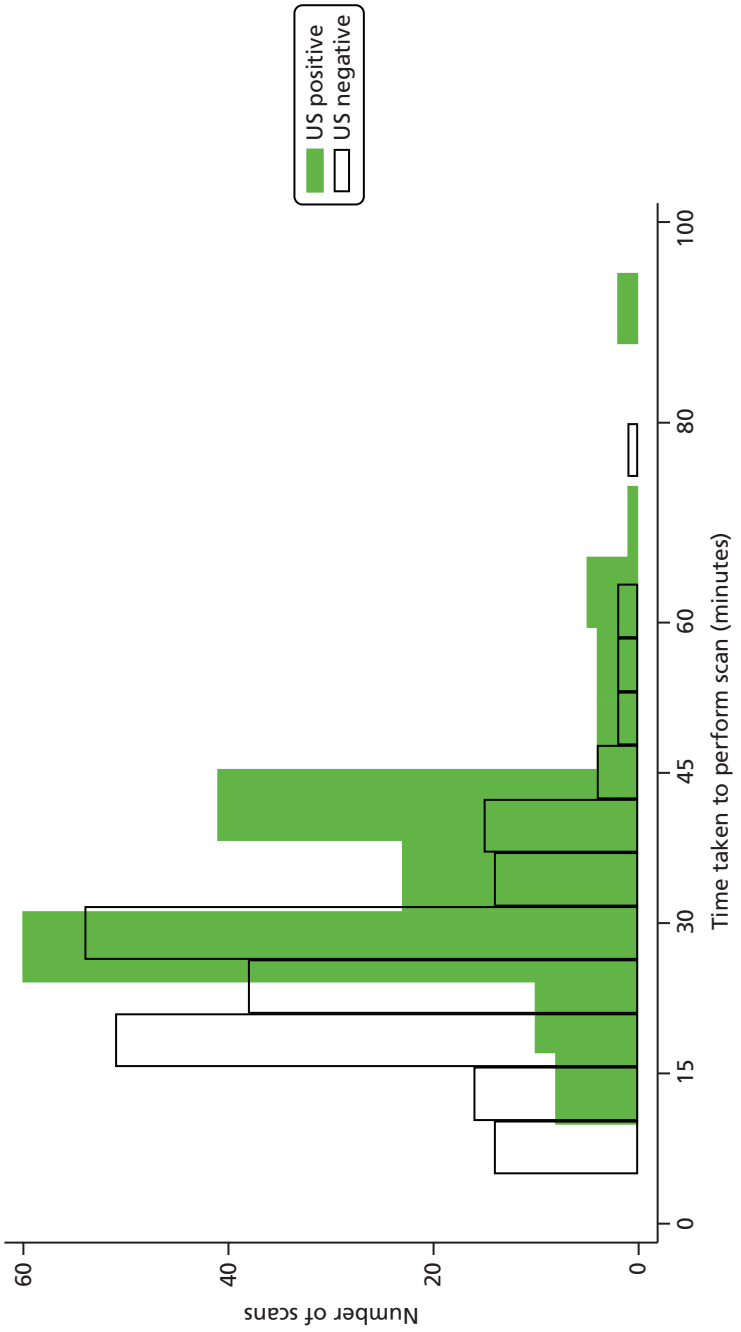


FIGURE 11 Time taken to perform ultrasound scans, comparing positive with negative scans. US, ultrasound.

areas if there are abnormalities than if normal areas only are found. There was a wider range of times taken to complete scans towards the end of the study, which might be explained by the inclusion of a larger number of sites.

Biopsy results

As part of the study protocol, all patients were scheduled to undergo a TAB within 7 days of starting high-dose glucocorticoid therapy. *Table 16* shows that a significant minority ($n = 28$, 7%) of biopsy procedures resulted in no useful tissue. The most common reason for a failed biopsy was that the surgeon took a sample that contained vein instead of artery ($n = 13$, 3.4%). Although this could reflect the difficulty in obtaining material from tortuous vessels, it could also reflect the relative inexperience of the surgeon given the task of obtaining the biopsy. The BSR guidelines recommend that a surgical biopsy with a minimum of 1 cm of temporal artery is obtained for each patient with suspected GCA; the procedure should be performed by a trained surgeon with experience in the technique. We did not mandate this, given that the study was comparing current standard practice in the NHS with the new technique of ultrasound. It seems likely that some of the biopsies were performed by less experienced surgeons, resulting in relatively poor diagnostic yield with no artery at all in 13% of patients. In addition, the length of temporal artery obtained in 43% of patients was below the BSR-recommended length of 1 cm (*Table 17*). These factors could have contributed to the relatively poor performance of biopsy as a diagnostic test in GCA. *Table 15* shows that giant cells were seen in 19% of biopsies overall, representing 71% of patients with GCA (72/101). Occlusion was reported in 25 biopsies. Of the 161 biopsies with abnormal pathology, four (1%) were compatible with another vasculitis and 35 (9%) were compatible with arteriosclerosis. *Table 16* highlights some potential issues in interpreting the biopsy results. Of the biopsy-negative patients who were ultimately diagnosed as not having GCA according to the reference diagnosis, 19% had intimal hyperplasia and 35% showed fragmentation or reduplication of the internal elastic lamina. The frequency of these changes was lower than those seen in patients with a positive biopsy, but almost identical to those seen in patients who were diagnosed as having GCA but who had a 'negative' biopsy. These findings raise further concerns about the validity of interpreting the TAB in the absence of cellular changes.

The rheumatologists interpreted the biopsy findings at 2 weeks, as well as evaluating the patient's clinical condition. In 11 patients the rheumatologist over-ruled or ignored the pathologist's conclusions, switching the diagnosis from not being consistent with GCA to being consistent with GCA (*Table 18*). There were no patients in whom the pathologists' diagnosis of GCA was over-ruled by rheumatologists, and it seems most likely that the rheumatologists would also have agreed with the pathologists' diagnosis in the case of two patients in whom information was missing.

Table 19 shows that there is no clear association between biopsy findings in those patients with a positive biopsy and presenting symptoms. Different histological features were present in patients with all three types of symptoms at baseline. We have not included a comparison between histological features and the presence of headache because this was a very common symptom at presentation.

Clinical and reference diagnoses

Any treatment decisions were independent of the study itself. The only eligibility criterion was that the clinician suspected a diagnosis of GCA and was intending to arrange a TAB to establish the likely diagnosis. High-dose steroid treatment was not an exclusion criterion, as long as the patient had not been given high-dose steroid treatment for > 7 days prior to obtaining the scan and biopsy. The physician was allowed to use any treatment, which could include methotrexate or another immunosuppressive therapy. The duration of use of therapies (apart from high-dose steroids) did not influence eligibility for the study.

TABLE 16 Characteristics of the TABs by reference diagnosis and biopsy result

Biopsy characteristics	All (*N* = 381)	Biopsy positive (*N* = 101)	Biopsy negative	
			Reference GCA (*N* = 156)	Reference not GCA (*N* = 124)
Biopsy sample,* n *(%)				
Temporal artery definitely obtained	353 (92.7)	101 (100.0)	138 (88.5)	114 (91.9)
Vein	13 (3.4)		9 (5.8)	4 (3.2)
Fat or muscle	5 (1.3)		3 (1.9)	2 (1.6)
Nerve	2 (0.5)		2 (1.3)	0 (0.0)
Fat or muscle, vein and nerve	2 (0.5)		2 (1.3)	0 (0.0)
Other	2 (0.5)		0 (0.0)	2 (1.6)
No sample obtained	4 (1.0)		2 (1.3)	2 (1.6)
Occlusion,* n *(%)				
No	336 (88.2)	77 (76.2)	143 (91.7)	116 (93.5)
Yes	25 (6.6)	24 (23.8)	1 (0.6)	0 (0.0)
Features normal areas	234 (61.4)	18 (17.8)	118 (75.6)	98 (79.0)
Features giant cells	72 (18.9)	72 (71.3)	0 (0.0)	0 (0.0)
Features calcification	44 (11.5)	18 (17.8)	19 (12.2)	7 (5.6)
Other unusual features	22 (5.8)	11 (10.9)	6 (3.8)	5 (4.0)
Normal pathology	205 (53.8)	0 (0.0)	108 (69.2)	97 (78.2)
Abnormal pathology	161 (42.3)	101 (100.0)	38 (24.4)	22 (17.7)
Compatible with a diagnosis of GCA	101 (26.5)	101 (100.0)	0 (0.0)	0 (0.0)
Compatible with a diagnosis of other vasculitis	4 (1.0)	2 (2.0)	1 (0.6)	1 (0.8)
Compatible with a diagnosis of arteriosclerosis	35 (9.2)	0 (0.0)	22 (14.1)	13 (10.5)
Compatible with another diagnosis	27 (7.1)	1 (1.0)	16 (10.3)	10 (8.1)
Intima normal	196 (51.4)	10 (9.9)	99 (63.5)	87 (70.2)
Intima abnormal,* n *(%)				
Arteriosclerosis present	39 (10.2)	14 (13.9)	16 (10.3)	9 (7.3)
Intimal hyperplasia present	149 (39.1)	88 (87.1)	37 (23.7)	24 (19.4)
Lamina normal	186 (48.8)	15 (14.9)	90 (57.7)	81 (65.3)
Lamina abnormal,* n *(%)				
Fragmentation	156 (40.9)	84 (83.2)	44 (28.2)	28 (22.6)
Reduplication	82 (21.5)	26 (25.7)	31 (19.9)	25 (20.2)
Length of sample (mm)				
Number of measurements	371	100	150	121
Mean (SD)	11.4 (7.4)	12.0 (8.9)	10.9 (7.1)	11.5 (6.5)
Median (IQR)	10.0 (7.0–15.0)	10.0 (7.0–15.0)	9.0 (6.0–15.0)	10.0 (7.0–14.0)

TABLE 17 Biopsy diagnosis by length of biopsy sample

Length	Biopsy diagnosis		
	Normal (*N* = 206), *n* (%)	Consistent with GCA (*N* = 101), *n* (%)	Other pathological diagnosis[a] (*N* = 60), *n* (%)
Biopsy length[b]			
TAB length < 1 cm	106 (51.5)	43 (42.6)	15 (25.0)
TAB length ≥ 1 cm	98 (47.6)	57 (56.4)	44 (73.3)
Missing	2 (1.0)	1 (1.0)	1 (1.7)

a 33 patients were diagnosed with arteriosclerosis, three had multiple possible diagnoses, including other forms of vasculitis, and 24 had miscellaneous other conditions.
b BSR guidelines suggest that the biopsy sample is at least 1 cm long.

TABLE 18 Comparison of pathologists and rheumatologists interpretation of biopsy

Pathologist's interpretation	*N*	Rheumatologist's interpretation		
		GCA (*N* = 110), *n* (%)	Not GCA (*N* = 256), *n* (%)	Missing (*N* = 15), *n* (%)
GCA	101	99 (90.0)	0 (0.0)	2 (13.3)
Not GCA	280	11 (10.0)	256 (100.0)	13 (86.7)

TABLE 19 Comparison of biopsy findings with symptoms present at baseline for those 101 patients whose biopsy was defined by the pathologists as consistent with GCA

Feature	*N*	Symptoms at baseline		
		Visual, *n* (%)	PMR, *n* (%)	Jaw/tongue claudication, *n* (%)
Intima	7	5 (9.8)	3 (7.1)	5 (6.8)
Internal elastic lamina	14	5 (9.8)	7 (16.7)	8 (10.8)
Media	21	8 (15.7)	6 (14.3)	14 (18.9)
Adventitia	32	9 (17.6)	14 (33.3)	18 (24.3)
Vasa vasorum	5	3 (5.9)	1 (2.4)	3 (4.1)
Transmural	38	21 (41.2)	11 (26.2)	26 (35.1)

The clinicians recorded the baseline and pre-steroid clinical features for all patients with suspected GCA. They were also given access to any available blood tests results and could request any investigation apart from a temporal artery and axillary ultrasound scan. In 10% of cases no baseline ESR result was available, and in 27% of cases no pre-steroid ESR result was available; in 8% of cases there was no baseline CRP level was available and in 27% of cases no pre-steroid CRP value was available.

Table 20 shows the initial diagnosis and treatment recommended for the patients in the study. In 21% of patients the clinicians reported definite GCA, in 54% they reported probable GCA and in 25% they reported possible GCA. The level of certainty of diagnosis of GCA is potentially biased in the data set because all patients were required to have at least the possibility of GCA in order to be eligible for inclusion in the study. It is conceivable that although the GP who referred the patient might have thought it possible that the patient had GCA, the study clinician reviewing the patient may have thought otherwise. However,

TABLE 20 Initial diagnosis and treatment

Features at diagnosis	Baseline ($N = 381$), *n* (%)
Certainty of GCA diagnosis	
Definite	80 (21.0)
Probable	204 (53.5)
Possible	96 (25.2)
Missing	1 (0.3)
Taken high-dose glucocorticoid therapy	
No	54 (14.2)
Yes	327 (85.8)
Taking immunosuppressant agents	
No	362 (95.0)
Yes	18 (4.7)
Missing	1 (0.3)

given the constraints of options available to the study clinician, they could define the patient only as having definite, probable or possible GCA. Therefore, the category of 'possible' GCA might actually contain a mixture of patients whom the study clinician might have considered did not have GCA and patients with a low likelihood of GCA. The majority of the patients were already in receipt of high doses of oral glucocorticoids (89%) at the time of the baseline visit. Very few patients (5%) were taking an immunosuppressive agent (only 18 out of the 381 patients) and, in all cases, these drugs were being given for other comorbid medical conditions rather than for suspected GCA (100% of the 18 patients).

Table 21 describes the clinical diagnosis made at 2 weeks and 6 months and shows that the majority of patients had a clinical diagnosis of GCA at both 2 weeks (67%) and 6 months (70%). In other words, most patients who were initially diagnosed as having GCA did not have any change made to their clinical diagnosis. However, in 19 patients the diagnosis was changed from not GCA to GCA following unblinding of the ultrasound results (after the 2-week visit the diagnosis has been reported). In a further 25 patients the diagnosis was changed at 6 months (this constitutes 6% of patients with available data); in 17 of these patients the diagnosis was switched from GCA to another diagnosis and in three cases the patients were reported as having a reference diagnosis of GCA. Twenty-one patients had their diagnosis changed following expert review of all the clinical data. In 13 of these patients the diagnosis was changed from GCA to another diagnosis, and in eight patients the diagnosis was changed from not GCA to GCA.

There were fewer data available at 6 months than at 2 weeks (46 fewer patients available at 6 months). Three patients were diagnosed with other forms of vasculitis at the 2-week visit and five patients had a diagnosis of another form of vasculitis at 6 months. These data highlight the potential overlapping presentation between different forms of vasculitis. In patients who did not have GCA or any form of vasculitis, non-specific headache was the most common diagnosis made (14% at 2 weeks and 12% at 6 months).

Table 22 shows the features present at 2 weeks and 6 months that were reported as influencing the clinician in making a diagnosis of GCA. There was consistent influence from the clinical symptoms (98%), signs (70%) and blood abnormalities (65%) at the 2-week visit: biopsy results influenced findings in 40% of cases. For the three cases diagnosed as GCA at 6 months but not 2 weeks, it is difficult to comment on the pattern of influence, but it looks similar to the findings at baseline.

TABLE 21 Clinical diagnosis at 2 weeks and 6 months

Clinical diagnosis	Visit, *n* (%)	
	2 weeks (*N* = 381)	6 months (*N* = 335)
GCA	257 (67.5)	234 (69.9)
Other vasculitis		
Takayasu's arteritis	1 (0.3)	1 (0.3)
EGPA	0 (0.0)	1 (0.3)
GPA	1 (0.3)	1 (0.3)
Retinal vasculitis	0 (0)	1 (0.3)
Other[a]	1 (0.3)	1 (0.3)
Other disease		
Non-specific headache	55 (14.4)	39 (11.6)
Multiple alternative diagnoses	12 (3.1)	10 (3.0)
Cervical spondylosis	7 (1.8)	6 (1.8)
Migraine	7 (1.8)	6 (1.8)
Myofascial pain	8 (2.1)	6 (1.8)
Temporomandibular dysfunction	7 (1.8)	6 (1.8)
Sinusitis	7 (1.8)	5 (1.5)
Shingles	1 (0.3)	0 (0.0)
Other	16 (4.2)	17 (5.1)

a ANCA-related vasculitis (2 weeks) or primary cerebral vasculitis (6 months).

TABLE 22 Influences on GCA diagnosis at 2 weeks and 6 months

GCA diagnosis influence	Visit, *n* (%)	
	2 weeks (*N* = 257)	6 months (*N* = 3)
Influenced by symptoms	251 (97.7)	3 (100.0)
Influenced by signs	181 (70.4)	1 (33.3)
Influenced by blood abnormalities	167 (65.0)	3 (100.0)
Influenced by biopsy report	104 (40.5)	1 (33.3)
Influenced by other factor(s)	16 (6.2)	1 (33.3)
Response to steroids	11 (4.3)	0 (0.0)

Characteristics and outcomes over time

The prevalence of diabetes mellitus increased from 14% at baseline to 18% at 6 months. By contrast, other conditions appeared to be unchanged in frequency across the visits (*Table 23*). Twenty-four participants had new-onset hypertension during the follow-up period and five participants who had documented hypertension at baseline no longer had it reported as an active condition during the follow-up period. Four fractures occurred during the 6-month follow-up. The fracture that occurred at 2 weeks was of the spine/vertebrae.

TABLE 23 Change in the prevalence of comorbid conditions over time

Condition	Visit, *n* (%)		
	Baseline (*N* = 381)	2 weeks (*N* = 381)	6 months (*N* = 335)
Diabetes mellitus	54 (14.2)	54 (14.2)	61 (18.2)
Hypertension	200 (52.5)	204 (53.5)	187 (55.8)
Angina	28 (7.3)	28 (7.3)	24 (7.2)
Heart failure	19 (5.0)	19 (5.0)	19 (5.7)
Neoplasia[a]	9 (2.4)	0 (0.0)	0 (0.0)
Low-trauma fracture (hip, spine, forearm, other)[a]	1 (0.3)	1 (0.3)	3 (0.9)

a Occurred since last visit.

Physical examination findings at each study visit are shown in *Table 24*. The number of abnormalities decreased at each study visit. The prevalence of thickened temporal artery fell from 50% at baseline to 13% at 2 weeks which would be in keeping with the expected clinical resolution of the physical findings of the disease as a result of treatment.

TABLE 24 Physical examination findings over time

Feature	Baseline (*N* = 381), *n* (%)		2 weeks (*N* = 381), *n* (%)		6 months (*N* = 335), *n* (%)	
	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal
Tender temporal artery	192 (50.4)	188 (49.3)	51 (13.4)	329 (86.4)	21 (6.3)	314 (93.7)
Thickened temporal artery	102 (26.8)	278 (73.0)	27 (7.1)	353 (92.7)	6 (1.8)	329 (98.2)
Reduced or absent pulsation in temporal artery	91 (23.9)	289 75.9)	51 (13.4)	328 (86.1)	26 (7.8)	309 (92.2)
Tender axillary artery	34 (8.9)	343 (90.0)	22 (5.8)	356 (93.4)	10 (3.0)	324 (96.7)
Bruits	15 (3.9)	296 (77.7)	7 (1.8)	298 (78.2)	7 (2.1)	271 (80.9)
Anterior ischaemic optic neuropathy	27 (7.1)	102 (26.8)	16 (4.2)	101(26.5)	12 (3.6)	82 (24.5)
Posterior ischaemic optic neuropathy	7 (1.8)	97 (25.5)	2 (0.5)	87 (22.8)	4 (1.2)	75 (22.4)
Relative afferent pupillary defect	15 (3.9)	257 (67.5)	9 (2.4)	249 (65.4)	11 (3.3)	221 (66.0)
III/IV/VI nerve palsy	3 (0.8)	310 (81.4)	3 (0.8)	304 (79.8)	1 (0.3)	274 (81.8)
	Present,* n *(%)	***Absent,* n *(%)***	***Present,* n *(%)***	***Absent,* n *(%)***	***Present,* n *(%)***	***Absent,* n *(%)***
Stroke	7 (1.8)	362 (95.0)	5 (1.3)	367 (96.3)	6 (1.8)	323 (96.4)
Aneurysm	3 (0.8)	320 (84.0)	2 (0.5)	326 (85.6)	2 (0.6)	294 (87.8)

Chapter 5 Agreement between ultrasound, biopsy and the reference diagnosis

Primary analysis

The primary outcome was the performance of ultrasound and biopsy in relation to the reference diagnosis of GCA. The reference diagnosis (defined in *Chapter 2*) for each patient was based on the 2-week and 6-month clinical diagnosis, as well as on the opinion of an expert review panel that assessed patient data (without the ultrasound result).

Ultrasound versus biopsy

The results of ultrasound and biopsy diagnosis were discordant in 115 patients (30%; *Table 25*). The two tests had fair agreement ($\kappa = 0.35$); overall, ultrasound was more likely than biopsy to find evidence consistent with a diagnosis of GCA (162 ultrasound-positive cases vs. 101 biopsy-positive cases, $p \leq 0.0001$).

Biopsy versus reference diagnosis

Temporal artery biopsy had a sensitivity of 39% (95% CI 33% to 46%) and a specificity 100% (95% CI 97% to 100%) for the reference diagnosis. All of the 101 participants whose biopsy was positive for evidence of GCA had a reference diagnosis of GCA. By contrast, 156 participants who had a reference diagnosis of GCA had a TAB that was not consistent with that diagnosis (*Table 26*).

Ultrasound versus reference diagnosis

Ultrasound examination had a sensitivity of 54% (95% CI 48% to 60%) for GCA, which is higher than that of biopsy, but had a lower specificity of 81% (95% CI 73% to 88%). Ultrasound examination showed evidence of findings consistent with GCA in 23 patients in whom GCA was not the ultimate diagnosis. By contrast, in 118 patients with a reference diagnosis of GCA, the ultrasound examination did not show features consistent with GCA (*Table 27*). When comparing the sensitivity and specificity of ultrasound and

TABLE 25 Giant cell arteritis diagnoses tabulated by biopsy and ultrasound method

	Biopsy				
US	GCA	Not GCA	Total	Kappa statistic	McNemar's test
GCA	74	88	162		
Not GCA	27	192	219		
Total	101	280	381	0.35	$p \leq 0.0001$

US, ultrasound.

TABLE 26 Giant cell arteritis diagnosis tabulated by biopsy and reference standard

	Reference diagnosis				
Biopsy	GCA	Not GCA	Total	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)
CA	101	0	101		
Not GCA	156	124	280		
Total	257	124	381	39 (33 to 46)	100 (97 to 100)

TABLE 27 Giant cell arteritis diagnosis tabulated by ultrasound and reference diagnosis

	Reference diagnosis				
US	GCA	Not GCA	Total	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)
GCA	139	23	162		
Not GCA	118	101	219		
Total	257	124	381	54 (0.48 to 0.60)	81 (73 to 88)

US, ultrasound.

biopsy, we have to bear in mind that negative and positive biopsy results would have influenced the final diagnosis (the reference standard); by contrast, the ultrasound result had no influence on either a final positive diagnosis or a final negative diagnosis. Thus, true ultrasound-positive and biopsy-negative patients may have been misclassified as non-GCA and false-positive biopsy patients whose ultrasound scan results were negative may have been misclassified as GCA. Therefore, the sensitivity and specificity of ultrasound could be a significant underestimate, whereas the sensitivity and specificity for biopsy might be falsely high.

Main results: robustness to variations in sample, biopsy diagnosis and reference diagnosis

Per protocol population: biopsy within 7 days of starting steroids

The primary analysis was based on 381 patients who had a TAB within 10 days of starting steroid treatment. We repeated this analysis, excluding 23 participants whose biopsy had been performed more than 7 days after starting steroids. The agreement between biopsy and ultrasound was marginally higher when both ultrasound and biopsy were performed within 7 days of starting steroids. Ultrasound and biopsy findings disagreed in 103 cases (28.8%, as shown in *Table 28*) and the kappa statistic was slightly larger than for the primary analysis ($\kappa = 0.37$).

The sensitivity of biopsy was very similar for the patients whose biopsies were performed within 7 days of commencing steroids, compared with the primary analysis group (sensitivity 40%, specificity 100%; *Table 29*).

Successful biopsy

Twenty-eight participants had an unsuccessful biopsy; in four participants no material was obtained at all (usually because the surgeon was unable to identify any structure resembling an artery during the procedure) and in 24 patients the sample consisted of material other than temporal artery. Repeating the primary analysis for the participants who had a successful biopsy (see *Table 29*) results in a similar sensitivity estimate for the value of biopsy compared with the population used for the primary analysis (42%, 95% CI 36% to 49%).

Table 16 summarises the biopsy findings and shows that the most common surgical error was to obtain vein instead of artery, which occurred in 13 patients. In five patients, fat or muscle was obtained,

TABLE 28 Giant cell arteritis diagnosis by ultrasound and biopsy for all patients in whom biopsy was performed within 7 days of commencing steroids

	Biopsy				
US	GCA	Not GCA	Total	Kappa statistic	McNemar's test
GCA	71	77	148		
Not GCA	26	184	210		
Total	97	261	358	0.37	$p = 0.0000$

US, ultrasound.

TABLE 29 Diagnostic accuracy for the variations in sample and biopsy diagnosis

		Reference diagnosis			
		GCA		Not GCA	
Source of diagnostic test result	*N*	Test+/true+	Sensitivity (%) (95% CI)	Test–/true–	Specificity (%) (95% CI)
Per protocol: biopsy	358	97/241	40 (34 to 47)	117/117	100 (97 to 100)
Successful biopsies only	353	101/239	42 (36 to 49)	114/114	100 (97 to 100)
Biopsy diagnosis from the rheumatologist	381	111/257	43 (37 to 49)	123/124	099 (96 to 100)
Population with 6-month data					
Biopsy	335	90/227	40 (33 to 46)	108/108	100 (97 to 100)
US	335	124/227	55 (48 to 61)	87/108	81 (72 to 88)

US, ultrasound.

in two patients nerve tissue was obtained and in four other patients the material consisted of fat or muscle, vein or nerve or other tissue.

Biopsy diagnosis from the rheumatologist

We analysed the data according to the rheumatologist's interpretation of the biopsy findings at 2 weeks. In 11 patients the rheumatologist over-ruled the pathologist's findings by switching the diagnosis from not being consistent with GCA to being consistent with GCA. The results are shown in *Table 18*; one participant was incorrectly diagnosed as having GCA. The sensitivity was slightly higher (43%, 95% CI 37% to 49%) than for the pathologist's interpretation (39%, 95% CI 33% to 46%). The disparity between pathologists' findings and the clinicians' interpretation of the biopsy result would primarily reflect confidence in the clinical diagnosis and interpretation of any comments in the biopsy that might be consistent with the diagnosis of GCA, probably influenced by how long the patient had been on high-dose glucocorticoids prior to the biopsy being obtained. For example, if the patient had a very compelling history and examination to suggest GCA, supported by a high acute-phase response, and had experienced considerable improvement with high-dose glucocorticoid therapy, the clinician might interpret minor changes in the biopsy, such as internal elastic lamina reduplication or fragmentation, as being consistent with resolving GCA.

Participants with 6-month data

Of the primary analysis set, 335 participants completed their 6-month follow-up. There is little difference in the sensitivity and specificity of ultrasound and biopsy after excluding patients without 6-month data (see *Table 29*).

Using final clinician diagnosis in place of the reference diagnosis

The reference diagnosis was based on the 2-week and 6-month clinical diagnoses, as well as the opinion of an expert review panel that assessed all of the patient data apart from the ultrasound results (see the detailed algorithm in *Chapter 2*). A sensitivity analysis was conducted by substituting the clinician's final diagnosis (which consisted of the clinician's decision on diagnosis at 6 months or at 2 weeks in the absence of 6-month data) instead of the reference diagnosis.

Twenty-one patients had a change of diagnosis following an expert review, using the original clinician's diagnosis (from 6 months or 2 weeks if no 6-month data were available) in place of the reference diagnosis. The effect was to change eight patients' results from GCA to not GCA; a further 13 patients switched from not GCA to GCA. The sensitivity and specificity of biopsy and ultrasound showed similar results to the primary analysis; the specificity of ultrasound was slightly higher (85% vs. 81%) when the clinician's final diagnosis was used in place of the reference diagnosis (*Table 30*).

TABLE 30 Diagnostic accuracy of biopsy and ultrasound with respect to clinician's final diagnosis

	Clinician's final diagnosis			
	GCA (*N* = 262)		Not GCA (*N* = 119)	
Diagnostic test (*N* = 381)	Test+	Sensitivity (%) (95% CI)	Test–	Specificity (%) (95% CI)
Biopsy	101	39 (33 to 45)	119	100 (97 to 100)
US	144	55 (49 to 61)	101	85 (77 to 91)

US, ultrasound.

Variations in ultrasound

We reviewed the variations in the interpretation of the ultrasound findings in the context of a diagnosis of GCA in order to investigate whether or not the sensitivity and specificity of ultrasound for diagnosis of GCA could be improved.

Halo with positive opinion of giant cell arteritis

The presence or absence of a halo on its own is the most important finding in considering the diagnosis of GCA. There was tight concordance between a positive halo and a positive overall ultrasound finding consistent with the diagnosis of GCA. Of 381 participants, 10 were reported to have a negative ultrasound result, even though a halo was detected. A further 10 cases had ultrasound reported as positive despite the absence of a halo. The distribution of reference diagnosis is similar in these participants, which would mean that the detection of a halo alone on the ultrasound scan was similar to that of the overall interpretation of the ultrasound scan, including other features such as stenosis or occlusion. Combining the two (i.e. presence of a halo and overall positive ultrasound diagnosis) also gives similar results to those obtained previously, as shown in *Table 31*.

Of the 10 patients reported as having features consistent with GCA on ultrasound scan and in whom no halo was seen, nine had abnormalities in the temporal arteries and one patient had abnormalities in both axillary and temporal arteries. Five patients had abnormalities at one site, two at two sites, two at three sites and one at four sites. Six patients had occlusion and six had stenosis.

Of the 10 patients in whom the ultrasound features were thought not to be consistent with GCA despite the presence of a halo, seven had a halo in one site, two had a halo at two sites and one had a halo at seven sites.

TABLE 31 Variations in interpretation of ultrasound findings in relation to supporting or not supporting a diagnosis of GCA, including the influence of expert review of the ultrasound results

	Reference diagnosis			
	GCA (*N* = 257)		Not GCA (*N* = 124)	
Change in US result (*N* = 381)	Test+	Sensitivity (%) (95% CI)	Test–	Specificity (%) (95% CI)
Original ultrasound diagnosis as reported in *Table 27*	139	54 (48 to 60)	101	81 (73 to 88)
Variations in US diagnosis				
Halo plus positive opinion	132	51 (45 to 58)	104	84 (76 to 90)
Bilateral halo plus positive opinion	84	33 (27 to 39)	115	93 (87 to 97)
US expert review opinion				
Change where disagreement is the most common	113	44 (38 to 50)	108	87 (80 to 92)
Change for reviews that are certain (i.e. no reviewer says disagree)	128	50 (44 to 56)	105	85 (77 to 91)

US, ultrasound.

Bilateral halo and positive opinion of giant cell arteritis

We investigated if the presence of halo on both sides (left and right temporal arteries or axillary arteries) affected the likelihood of interpreting the ultrasound findings as being consistent with GCA or not. This could be used as a stricter definition of positive ultrasound results by reducing false positives (increasing specificity) but potentially increasing false negatives (reducing sensitivity).

The results are shown in *Table 31*. The modified criteria resulted in 59 patients reclassified as 'not GCA' and a higher specificity (93%; 95% CI 87% to 97%). The presence of bilateral halo coupled with positive overall interpretation identified patients with GCA at a sensitivity of only 33%, because only a small proportion of patients demonstrated this feature (93 patients). This suggests that ultrasound could be used as a 'rule in' test, whereby the presence of bilateral halo indicates a positive diagnosis and thereby avoids TAB in around one-quarter of participants with few false positives, albeit with lower sensitivity and specificity than TAB alone.

Axillary involvement

A potential benefit of ultrasound is that it scans both temporal and axillary arteries. Of the abnormal ultrasound scans, 53 showed axillary involvement, which in 27 cases was bilateral. Of the 53 ultrasound assessments with axillary involvement, nine showed no temporal involvement. In three of these cases, the patient was biopsy positive and in six cases the patient was biopsy negative; seven patients were given the reference standard diagnosis of GCA and two were reported as not having GCA. Based on these data, in only a few patients would the diagnosis be changed to GCA on the basis of an ultrasound scan showing axillary involvement. Therefore, the role of ultrasound in the detection of axillary artery involvement may be important but limited because only a small number of patients are likely to have isolated axillary involvement in the absence of temporal involvement as demonstrated by ultrasound. In other words, the presence of abnormalities in the axillary arteries provides further support for the diagnosis of GCA. The low numbers may reflect the inclusion of patients predominantly presenting with cranial GCA.

Ultrasound expert review

As part of the study protocol, all ultrasound scans obtained by individual sonographers were reviewed centrally by an expert panel. This was made possible because the protocol required recording of still and video images from the scan procedure for all participants. The images were uploaded onto a secure password-protected central web-based system designed for this purpose, so that the images could be reviewed online or downloaded for review by the expert panel. The reviewers were asked to provide an assessment of the quality of the available images as either clear or unclear (the latter because of the absence of sufficient images, poor-quality images or because the reviewer was unsure for other technical reasons). If the images were clear, the expert was asked either to agree with the sonographer's interpretation or to disagree with it.

If we used data obtained from the ultrasound findings according to the expert panel, this would result in 14 out of 219 patients having their ultrasound diagnosis changed from not consistent with GCA to consistent with GCA, and 47 out of 162 patients would have their ultrasound findings changed from consistent with GCA to not consistent with GCA.

The expert reviewers provided stricter definitions of scans being consistent with GCA or not (see *Table 31*), which resulted in a lower sensitivity (44%) and higher specificity (87%) than for the original interpretation of ultrasound findings by the sonographers (sensitivity 54% and specificity 81%; see *Table 31*). One ultrasound reviewer assessed all patients (but every case was reviewed by at least two reviewers). Using this reviewer's decisions alone would result in four changes with respect to the method above: two to disagree and two to agree. All four of these patients were ultrasound positive and reference diagnosis GCA negative. Hence, the results from this reviewer are identical to the overall results.

Following the analysis of expert reviewers' opinions, we examined the effect of changing only the ultrasound interpretation findings in patients when there was consensus among the reviewers. The results of this

interpretation (see *Table 31*) were that six patients had their results changed from not being consistent with GCA to being consistent with GCA; a further 21 patients were changed from being consistent with GCA based on ultrasound to being not consistent with GCA based on ultrasound. The sensitivity and specificity are closer to the original ultrasound diagnosis because fewer patients have been changed.

Two-week diagnosis and test findings

Two-week diagnosis with biopsy finding

At the review visit, 2 weeks after baseline, the clinician was asked for their diagnosis based on observed signs, symptoms, laboratory test results and the biopsy result. *Table 32* shows that the sensitivity (91%) and specificity (81%) are high for the 2-week diagnosis compared with the reference diagnosis. There was disagreement between the 2-week diagnosis and the reference diagnosis for 46 participants (12%). The 2-week opinion of the clinician was based on the clinical presentation and subsequent findings; the biopsy result is likely to have been one of the major contributions to this opinion. This introduces some circularity to the interpretation of data because we are independently evaluating the role of biopsy in contributing to the diagnosis when, in fact, the biopsy has already contributed to the diagnosis by forming part of the clinical opinion of the clinician interpreting all the data available at the time (but not including information from the ultrasound scan that was kept confidential from the clinician managing the case, at least until they had formally reported their diagnosis).

Two-week diagnosis with biopsy and unblinded ultrasound findings

If at 2 weeks the clinician's diagnosis was not GCA and he or she was considering rapidly withdrawing steroids, the ultrasound findings were unblinded. Following this, the diagnosis changed for 19 participants (all to GCA). We analysed the sensitivity and specificity of the 2-week diagnosis compared with reference diagnosis when accounting for the unblinding of these 19 patients. The sensitivity was higher when including the results of the unblinding for the 2-week diagnosis than when we included only the results of the 2-week diagnosis without the unblinding (96% vs. 91%), but at the same time the specificity of the 2-week diagnosis for the reference diagnosis was lowered (77% vs. 81%), as shown in *Table 33*.

TABLE 32 Giant cell arteritis diagnosis tabulated by reference and 2-week diagnosis

	Reference diagnosis				
Two-week diagnosis	GCA	Not GCA	Total	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)
GCA	234	23	257		
Not GCA	23	101	124		
Total	257	124	381	91 (87 to 94)	81 (73 to 88)

TABLE 33 Giant cell arteritis diagnosis tabulated by reference and 2-week diagnosis (updated post ultrasound unblinding)

	Reference diagnosis				
Two-week diagnosis	GCA	Not GCA	Total	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)
GCA	247	29	276		
Not GCA	10	95	105		
Total	257	124	381	96 (93 to 98)	77 (68 to 84)

Ultrasound: learning effect

Ultrasonography of the temporal arteries is operator dependent; therefore, as part of the study protocol, training was given to sonographers at the beginning of the study to ensure proficiency with the technique before applying it to study participants. Some sonographers with sufficient experience were deemed exempt from the full training. The results of the training attempts are shown in *Chapter 3*. In addition to providing the training, as part of the protocol, all scans performed by each site's sonographer were recorded and the images were sent to the TABUL office in Oxford, so that they could be uploaded onto a server for assessment by the expert reviewers. Scans were reviewed during the course of the study as part of the quality control and some sonographers were retrained if necessary.

The ultrasound scans are split into two groups:

1. First 10 – this included the first 10 scans (post training) within the TABUL cohort for each sonographer who received full training, or all scans before retraining for those sonographers who subsequently received further training.
2. After 10 – included all patients after the first 10 scans for the sonographers who received full training. It also includes all scans from sonographers exempt from full training. If a sonographer was retrained during the study, it included all scans after the date of retraining.

The specificity of ultrasound was almost the same for the first 10 scans (82%) and the later scans (81%), but the sensitivity was higher for the later scans (increasing from 45% to 62%), as shown in *Table 34*. This strongly suggests that there is a learning effect, as the sonographers become more experienced at performing the scan, and that this will predominantly influence sensitivity of the test result.

We repeated the analysis, excluding the sonographers deemed to be experts and exempt from full training (leaving $n = 110$ patients), splitting the data based on the first five scans versus the rest of the scans. There appears to be an improvement in sensitivity in the scans assessed between the first five patients scanned and subsequent patients scanned (rising from 42% to 59%, respectively), with similar levels of specificity (80% vs. 82%).

Timing effect

Accuracy of tests in relation to time since starting steroids

All patients had an ultrasound test performed more rapidly than or on the same day as a biopsy. Overall, 107 patients had an ultrasound performed within 1 day of starting steroid treatment, whereas only 26 patients

TABLE 34 Diagnostic accuracy of ultrasound by sonographer training-level subgroups

		Reference diagnosis			
		GCA		Not GCA	
Learning curve subgroups	*N*	Test+/true+	Sensitivity (%) (95% CI)	Test–/true–	Specificity (%) (95% CI)
All sonographers					
Including first 10 scans	181	54/120	45 (36 to 54)	50/61	82 (70 to 91)
Excluding first 10 scans	200	85/137	62 (53 to 70)	51/63	81 (69 to 90)
Non-experts only					
Including first five scans	70	23/55	42 (29 to 56)	12/15	80 (52 to 96)
Excluding first five scans	40	17/29	59 (39 to 76)	9/11	82 (48 to 98)

had a biopsy performed within 1 day of starting steroid treatment. By comparison, 246 patients had their biopsy performed after having started steroids at least 5 days previously, compared with only 57 patients who had an ultrasound scan performed after 5 days of steroid therapy. Within the time frame of the study, the sensitivity was higher (64%) for participants whose test was up to 1 day after starting steroids than for those whose test was ≥ 2 days after starting steroids (47%); the specificity remained unchanged (*Table 35*).

Table 36 shows the potential effect of duration on high doses of glucocorticoid therapy on the interpretation of the biopsy and ultrasound test results. For those patients with a reference diagnosis of GCA, the proportion who were correctly detected by biopsy decreased with time since starting steroids. Sensitivity was highest for the biopsies that were performed within 3 days of starting steroids (48%; 95% CI 37% to 60%). Sensitivity was lowest (33%; 95% CI 22% to 46%) for biopsies that were performed ≥ 7 days after starting steroids.

The effect of delay in performing biopsy in relation to ultrasound on the agreement between two tests

We investigated whether or not the modest agreement between biopsy and ultrasound tests results was affected by the time interval between performing each test. *Table 37* shows that, contrary to expectation, the agreement between tests was similar when the biopsy was performed within 1 day of ultrasound ($\kappa = 0.33$) or when the biopsies were performed either 2 or 3 days after ultrasound ($\kappa = 0.4$), or ≥ 4 days after ultrasound ($\kappa = 0.32$).

TABLE 35 Diagnosis of biopsy and ultrasound by time since starting steroids

	Reference GCA		Reference not GCA		Total	
Number of days since starting steroids	Test GCA, *n* (%)	Test not GCA, *n* (%)	Test GCA, *n* (%)	Test not GCA, *n* (%)	Test GCA, *n* (%)	Test not GCA, *n* (%)
Days between starting steroids and TAB						
TAB performed before steroids	2 (50.0)	2 (50.0)	0 (0.0)	1 (100.0)	2 (40.0)	3 (60.0)
Same day or 1 day	6 (42.9)	8 (57.1)	0 (0.0)	7 (100.0)	6 (28.6)	15 (71.4)
2 days	11 (42.3)	15 (57.7)	0 (0.0)	12 (100.0)	11 (28.9)	27 (71.1)
3 days	20 (54.1)	17 (45.9)	0 (0.0)	7 (100.0)	20 (45.5)	24 (54.5)
4 days	13 (52.0)	12 (48.0)	0 (0.0)	12 (100.0)	13 (35.1)	24 (64.9)
5 days	13 (35.1)	24 (64.9)	0 (0.0)	21 (100.0)	13 (22.4)	45 (77.6)
6 days	18 (31.0)	40 (69.0)	0 (0.0)	25 (100.0)	18 (21.7)	65 (78.3)
7 days	14 (35.0)	26 (65.0)	0 (0.0)	33 (100.0)	14 (19.2)	59 (80.8)
≥ 8 days	7 (30.4)	16 (69.6)	0 (0.0)	9 (100.0)	7 (21.9)	25 (78.1)
Days between starting steroids and US						
US performed before steroids	6 (60.0)	4 (40.0)	0 (0.0)	3 (100.0)	6 (46.2)	7 (53.8)
Same day	27 (73.0)	10 (27.0)	4 (33.3)	8 (66.7)	31 (63.3)	18 (36.7)
1 day	34 (59.6)	23 (40.4)	4 (14.3)	24 (85.7)	38 (44.7)	47 (55.3)
2 days	27 (54.0)	23 (46.0)	2 (12.5)	14 (87.5)	29 (43.9)	37 (56.1)
3 days	13 (41.9)	18 (58.1)	5 (20.0)	20 (80.0)	18 (32.1)	38 (67.9)
4 days	14 (36.8)	24 (63.2)	3 (17.6)	14 (82.4)	17 (30.9)	38 (69.1)
5 days	14 (51.9)	13 (48.1)	1 (7.1)	13 (92.9)	15 (36.6)	26 (63.4)
6 or 7 days	8 (57.1)	6 (42.9)	4 (33.3)	8 (66.7)	12 (46.2)	14 (53.8)

US, ultrasound.

TABLE 36 Diagnostic accuracy of biopsy by time since starting steroids

Time between test and starting steroids	*N*	Reference diagnosis: GCA Test+/true+	GCA Sensitivity (%) (95% CI)	Not GCA Test–/true–	Not GCA Specificity (%) (95% CI)
Biopsy					
≤ 3 days	108	39/81	48 (37 to 60)	27/27	100 (87 to 100)
Between 4 and 6 days	178	44/120	37 (28 to 46)	58/58	100 (94 to 100)
≥ 7 days	105	21/63	33 (22 to 46)	42/42	100 (92 to 100)
US					
≤ 1 day	147	67/104	64 (54 to 74)	35/43	81 (67 to 92)
≥ 2 days	244	76/160	47 (40 to 56)	69/84	82 (72 to 90)

US, ultrasound.

TABLE 37 Agreement between ultrasound and biopsy by time between the two tests

Time between biopsy and US	*N*	Biopsy positive: US positive	Biopsy positive: US negative	Biopsy negative: US positive	Biopsy negative: US negative	Kappa statistic	McNemar's test
Biopsy same day or 1 day after	155	27	13	33	82	0.33	$p = 0.0045$
2–3 days after	113	21	8	22	62	0.4	$p = 0.0161$
≥ 4 days after	123	28	7	35	53	0.32	$p = 0.0000$

US, ultrasound.

Sequential and combined test analyses

Performing both ultrasound and biopsy may not be necessary in all patients. We would speculate that ultrasound could be useful as a 'rule-in' test to support the diagnosis of GCA. If the ultrasound result was consistent with GCA and the clinical features supported that diagnosis, the diagnosis of GCA could be made without any further testing required. If, however, ultrasound was not consistent with GCA, patients would be recommended to have a biopsy in order to help to decide whether or not they had GCA. This test strategy can be explored in the TABUL data set and lends itself to a full economic evaluation, which is provided in *Chapter 7*.

The following steps describe a potential algorithm for investigating patients with suspected GCA:

- Patients present with the clinical or laboratory features suggesting a diagnosis of GCA.
- An ultrasound scan is performed and, if the results show evidence supporting a diagnosis of GCA, a diagnosis of GCA is made.
- If the ultrasound scan does not show features consistent with GCA, the patient is scheduled to have a TAB.
- If the TAB is supportive of a diagnosis of GCA, the patient is diagnosed with GCA.
- If both TAB and ultrasound are negative, the conclusion is that the patient does not have ultrasound or histological evidence to support the diagnosis of GCA.

Table 38 illustrates the effects of applying this sequential strategy on the 381 patients in the TABUL study. Overall, 162 patients would have been diagnosed with GCA based on the ultrasound scan alone, the majority (139, 86%) correctly. The remaining 219 ultrasound-negative patients would then have a biopsy. Twenty-seven of these ultrasound-negative patients had a positive biopsy and would also have been diagnosed with GCA. The 192 patients who were both scan and biopsy negative would not have received a diagnosis of GCA despite the fact that in almost half of these patients the reference diagnosis was GCA.

Table 38 shows the number of patients who would have a positive or negative result on the test compared with the eventual reference diagnosis via not GCA or GCA.

Table 39 shows the accuracy of applying a sequential strategy to the TABUL cohort. The effect of applying a second test, the biopsy, to patients who are ultrasound negative is to improve on the sensitivity of the ultrasound-only strategy (from 54% to 65%) while maintaining its specificity at 81% (although specificity is lower than the 100% obtained for a biopsy-only strategy). If this strategy was used for the cohort, 162 (43%) patients would have avoided having a TAB. However, on the basis of this strategy, without a clinician over-riding (ignoring) the test results, 91 true cases of GCA as defined by the reference diagnosis would have been missed and 23 patients would be wrongly diagnosed as having GCA according to the reference diagnosis. It is a difficult dilemma because no single test or evaluation can be used to rule out the diagnosis, whereas any single test or evaluation could be used to rule it in, over-riding a negative result.

Pre-test probability of having giant cell arteritis or not

We investigated which (if any) subgroups of the cohort could have been diagnosed without the need for biopsy and/or ultrasound, entirely based on a pre-test clinical assessment of patients being at high, medium or low likelihood of having GCA.

There is likely to be significant bias because the clinician's opinion on the diagnosis would be strongly influenced by the factors defining the risk groups; therefore, the items to define risk groups were extracted

TABLE 38 Effect of implementing the sequential strategy of ultrasound followed by biopsy if required

	Reference diagnosis		
Test results	GCA, *n* (%)	Not GCA, *n* (%)	Total, *n* (%)
US positive	139 (85.8)	23 (14.2)	162 (100)
US negative, biopsy positive	27 (100.0)	0 (0.0)	27 (100)
US negative, biopsy negative	91 (47.4)	101 (52.6)	192 (100)

US, ultrasound.

TABLE 39 Accuracy of sequential diagnostic strategy (ultrasound first)

	Reference diagnosis				
Test diagnosis	GCA	Not GCA	Total	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)
GCA	166	23	189		
Not GCA	91	101	192		
Total	257	124	381	65 (58 to 70)	81 (73 to 88)

from DCVAS data, representing an independent cohort, giving the process external validity. We applied the same rules for DCVAS to the TABUL data to define participants as follows:

- Participants were defined as being at high risk of having GCA if they had an elevated ESR or CRP level (ESR of > 60 mm/hour or CRP level of > 40 mg/l) and jaw or tongue claudication at presentation or prior to use of steroids.
- Participants were defined as being at medium risk of having GCA if they had either elevated ESR/CRP level or jaw/tongue claudication at presentation/before steroids.
- Participants were defined as being at low risk if they had neither elevated ESR/CRP level nor jaw/tongue claudication at presentation/before steroids.

Table 40 shows the relationship between pre-test risk groups and the clinician's certainty of a diagnosis of GCA reported at baseline. The proportion of participants with 'definite' GCA is higher in the high-risk group (42%) than in the medium- and low-risk groups (20% and 9%, respectively). There was good agreement between the clinicians' certainty of diagnosis and the pre-test risk of diagnosing GCA. Among the TABUL cohort, 93% of the high-risk group had a reference diagnosis of GCA. The prevalence of GCA was lower (78%) in the medium-risk group and lower still in the low-risk group (39%).

Accuracy of test within pre-test subgroup

Table 41 shows the accuracy of biopsy and ultrasound within the pre-test probability subgroups. The sensitivity of biopsy increases as the pre-test risk increases. The sensitivity of ultrasound is slightly higher in the medium- and high-probability subgroups than in the low-probability group (57%, 57% and 44%, respectively). The specificity of ultrasound is similar across the subgroups; however, it is difficult to make a comparison owing to the small numbers of participants without GCA, in the medium- and high-probability subgroups ($n = 34$ and $n = 6$, respectively).

Diagnostic strategies

We considered the implications of introducing a test strategy dependent on the pre-test probability of a patient having or not having a diagnosis of GCA. As the prevalence of GCA is very high in the high-risk group (93%), one strategy could be not to perform either ultrasound or biopsy in this group and simply diagnose the patients as having GCA without any further testing (we have defined these patients as H0). Although this would be most economic, by avoiding either test, in clinical practice both clinicians and

TABLE 40 Relationship between pre-test risk and diagnosis

		Pre-test risk, *n* (%)		
Diagnostic certainty	Total (*N* = 381)	High (*N* = 89)	Medium (*N* = 154)	Low (*N* = 138)
Certainty of GCA diagnosis at baseline	Definite	37 (41.6)	31 (20.1)	12 (8.7)
	Probable	43 (48.3)	94 (61.0)	67 (48.6)
	Possible	8 (9.0)	29 (18.8)	59 (42.8)
Reference diagnosis	GCA	83 (93.3)	120 (77.9)	54 (39.1)
	Not GCA	6 (6.7)	34 (22.1)	84 (60.9)
Biopsy diagnosis	GCA	52 (58.4)	40 (26.0)	9 (6.5)
	Not GCA	37 (41.6)	114 (74.0)	129 (93.5)
US diagnosis	GCA	48 (53.9)	71 (46.1)	43 (31.2)
	Not GCA	41 (46.1)	83 (53.9)	95 (68.8)

US, ultrasound.

TABLE 41 Diagnostic accuracy of biopsy and ultrasound by pre-test probability group

	Reference diagnosis			
	GCA		Not GCA	
Pre-test probability of GCA per diagnostic test	Test+/true+	Sensitivity (%) (95% CI)	Test–/true–	Specificity (%) (95% CI)
***High pre-test probability* (n = *89*)**				
Biopsy	52/83	63 (51 to 73)	6/6	100 (54 to 100)
US	47/83	57 (45 to 67)	5/6	83 (36 to 100)
***Medium pre-test probability* (n = *154*)**				
Biopsy	40/120	33 (25 to 43)	34/34	100 (90 to 100)
US	68/120	57 (47 to 66)	31/34	91 (76 to 98)
***Low pre-test probability* (n = *138*)**				
Biopsy	9/54	17 (8 to 29)	84/84	100 (96 to 100)
US	24/54	44 (31 to 59)	65/84	77 (67 to 86)

US, ultrasound.

patients would find it difficult to accept the diagnosis without at least some attempt to support the diagnosis with further investigation (biopsy or scan). We would therefore also consider a strategy of performing an initial ultrasound in the high-risk group and then performing a biopsy if the scan is negative (i.e. the scan is not consistent with a diagnosis of GCA). We would define a positive result on ultrasound as consistent with a diagnosis of GCA using four possible criteria as follows.

1. The sonographer's opinion is that the ultrasound scan is consistent with a diagnosis of GCA (defined as H1).
2. Bilateral halo is present (in either the temporal or axillary arteries) (H2).
3. Either the sonographer's opinion is that the ultrasound is consistent with a diagnosis of GCA or there are abnormalities in the axillary arteries (regardless of the overall sonographer opinion) (H3).
4. Bilateral halo or any axillary involvement is present (H4).

In the medium-risk groups we considered the above four strategies in which ultrasound is performed first, followed by biopsy (M1 to M4 would be equivalent to H1 to H4).

In the low-risk groups, we considered the same four strategies as well as two further strategies.

1. Using a negative ultrasound result as a 'rule-out' test for GCA. If ultrasound is positive, then perform a biopsy and take the diagnosis from the biopsy result (L5).
2. Use the absence of any abnormal finding on the ultrasound as a 'rule-out' test for GCA. If there are any abnormalities, perform a biopsy and take the diagnosis from the biopsy result (L6).

The accuracy of the diagnostic test strategies for each subgroup is shown in *Table 42*. These strategies are combined and the accuracy of all possible combinations displayed in *Figure 12* (for full results, see *Appendix 15*) alongside point estimates for biopsy and ultrasound alone. It is apparent from the graph that TAB alone provides relatively poor performance in helping to diagnose GCA; by contrast, many of the combined strategies have better sensitivity and specificity than ultrasound alone. The two combined strategies that give the highest sensitivity are H0-M1-L1 and H0-M1-L3.These combine no test in the high-risk group, testing with ultrasound first and following with biopsy if the ultrasound result is negative in the medium- and low-risk groups, or following with biopsy if the ultrasound is negative and there is no axillary involvement in the low-risk group.

TABLE 42 Summary of potential diagnostic strategies for each pre-test risk group

Risk group	Description	Reference diagnosis, *n* (%)		
		GCA	Not GCA	
High risk		***Sensitivity* (N = *83*)**	***Specificity* (N = *6*)**	***Number of TAB required* (N = *89*)**
H0	Assume GCA positive (no diagnostic test performed)	83 (100.0)	0 (0.0)	0 (0.0)
H1	GCA if either US or TAB positive	64 (77.1)	5 (83.3)	41 (46.1)
H2	GCA if either bilateral halo on US or TAB positive	59 (71.1)	5 (83.3)	53 (59.6)
H3	GCA if either US positive, US axillary involvement or TAB positive	67 (80.7)	5 (83.3)	36 (40.4)
H4	GCA if either bilateral halo on US, US axillary involvement or TAB positive	63 (75.9)	5 (83.3)	47 (52.8)
Medium risk		***Sensitivity* (N = *120*)**	***Specificity* (N = *34*)**	***Number of TAB required* (N = *154*)**
M1	GCA if either US or TAB positive	75 (62.5)	31 (91.2)	83 (53.9)
M2	GCA if either bilateral halo on US or TAB positive	53 (44.2)	32 (94.1)	115 (74.7)
M3	GCA if either US positive, US axillary involvement or TAB positive	75 (62.5)	31 (91.2)	83 (53.9)
M4	GCA if either bilateral halo on US, US axillary involvement or TAB positive	54 (45.0)	32 (94.1)	112 (72.7)
Low risk		***Sensitivity* (N = *54*)**	***Specificity* (N = *84*)**	***Number of TAB required* (N = *138*)**
L1	GCA if either US or TAB positive	27 (50.0)	65 (77.4)	95 (68.8)
L2	GCA if either bilateral halo on US or TAB positive	17 (31.5)	77 (91.7)	117 (84.8)
L3	GCA if either US positive, US axillary involvement or TAB positive	27 (50.0)	62 (73.8)	92 (66.7)
L4	GCA if either bilateral halo on US, US axillary involvement or TAB positive	18 (33.3)	72 (85.7)	111 (80.4)
L5	GCA if both US positive and TAB positive	6 (11.1)	84 (100.0)	43 (31.2)
L6	GCA if any abnormality on US and TAB positive	8 (14.8)	84 (100.0)	51 (37.0)

US, ultrasound.

Appendix 15 contains an extensive list of combinations of different strategies that could be applied to improve accuracy in diagnosing GCA, dependent on the initial pre-test probability of the diagnosis being high, medium or low.

Exploratory findings

Birmingham Vasculitis Activity Score and Vasculitis Damage Index

The BVAS and the VDI have not been widely used either in patients with GCA nor in disease controls because they were designed for use in patients who already had a diagnosis of vasculitis;[66,69] therefore, this is an exploratory part of the study. The BVAS and the VDI could be useful in the evaluation of patients

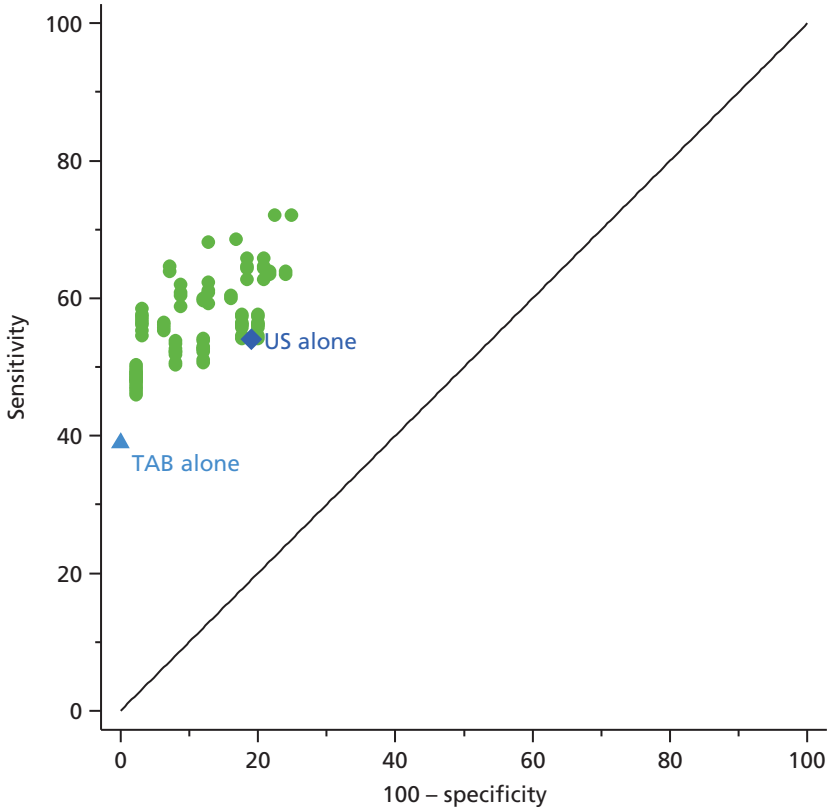


FIGURE 12 Sensitivity and specificity for the diagnostic strategy combinations. US, ultrasound.

in whom another form of vasculitis is suspected. Five patients in the study had vasculitis that was not GCA. Two of these patients had BVAS values of at least 12, indicating significant multisystem features; one of these patients had a VDI score of five at 6 months, indicating extensive damage. However, items on the BVAS and VDI forms include features relevant to GCA, and the BVAS and the VDI could be seen as further opportunities to cross-check that correct information has been recorded on the main CRF pages, particularly in relation to presence of headache, complications as a result of GCA, visual loss or stroke. We recorded the BVAS and the VDI score at 2 weeks and 6 months, but not baseline, to reduce the burden of assessments required. As the VDI scores only items that are present for at least 3 months, there would be minimal difference between the baseline and 2-week VDI scores.[69,70]

Birmingham Vasculitis Activity Score and Vasculitis Damage Index as diagnostic tools

An analysis of the VDI score and the BVAS in relation to patient diagnosis was undertaken to assess whether or not these may play a role in ruling GCA in or out. *Table 43* shows the 2-week BVAS and VDI score by the clinician-reported diagnoses at 2 weeks, as well as the eventual reference diagnosis. Neither measure appears particularly associated with diagnosis. A BVAS of ≥ 4 was observed in 45 (12%) patients with a non-GCA diagnosis at 2 weeks and 62 (16%) patients with GCA. Only seven of the non-GCA cases and 26 of the GCA cases (9% overall) actually had at least one item of VDI damage recorded at 2 weeks. After 6 months, about one-third of patients were recorded as having damage in both GCA and non-GCA groups.

There does not appear to be a difference between the proportions of participants with a 6-month VDI score of ≥ 1 if we compare the 2-week diagnosis with the reference diagnosis (see *Table 43*). There is also no difference in BVAS when comparing patients grouped according to the 2-week diagnosis and the reference diagnosis.

TABLE 43 Relationship between the BVAS/VDI and diagnosis at 2-weeks/reference diagnosis

	Diagnosis			
	Two-week, *n* (%)		Reference, *n* (%)	
Score	Not GCA (*n* = 124)	GCA (*n* = 257)	Not GCA (*n* = 124)	GCA (*n* = 257)
BVAS: 2 weeks				
0	33 (26.6)	113 (44.0)	43 (34.7)	103 (40.1)
1	29 (23.4)	34 (13.2)	26 (21.0)	37 (14.4)
2 to 3	17 (13.7)	48 (18.7)	15 (12.1)	50 (19.5)
4 to 6	29 (23.4)	37 (14.4)	26 (21.0)	40 (15.6)
≥ 7	16 (12.9)	25 (9.7)	14 (11.3)	27 (10.5)
VDI score: 2 weeks				
0	117 (94.4)	229 (89.1)	115 (92.7)	231 (89.9)
≥ 1	7 (5.6)	26 (10.1)	8 (6.5)	25 (9.7)
VDI score: 6 months				
0	65 (52.4)	144 (56.0)	72 (58.1)	137 (53.3)
≥ 1	39 (31.5)	84 (32.7)	35 (28.2)	88 (34.2)

The reliability of assessing the Birmingham Vasculitis Activity Score and Vasculitis Damage Index

Sixty-six study investigators were asked to complete 20 training cases for the BVAS and 20 for the VDI. This consisted of paper case vignettes with half a page of description for each case. The assessors were asked to complete the BVAS or the VDI for each of these cases. The pass marks were 85% agreement with the gold standard for the BVAS and 75% agreement with the gold standard for the VDI (and no case with a score of < 50% for either the BVAS or the VDI) in order to qualify each investigator for participation in the study. Sixty-one investigators completed BVAS and VDI training. The average pass mark was 89.6% for the BVAS and 86.4% for the VDI, but these included the values for investigators who failed at least one of the assessments. Twenty-two investigators were asked to repeat at least one of the BVAS cases and 18 were asked to repeat at least one of the VDI cases. Altogether, 52 investigators eventually passed the assessments. Three further investigators were exempted from the assessments (on the basis that they had already demonstrated expertise in performing the BVAS and the VDI for other studies), giving a total of 55 investigators certified to perform the BVAS and the VDI.

Recording the BVAS at the 2-week visit would include reporting all items occurring since the onset of the current condition regardless of duration and regardless of whether or not they had resolved.[66,67] In other words, if headache symptoms had been present for 2 weeks longer than at baseline, that is, because they had already been reported on the CRF at baseline, they should still have been reported on the first BVAS, which was completed at the 2-week visit, to reduce the burden of assessments required at the baseline visit. In practice, this would mean that patients may have experienced features of their current presentation (such as headache) for 2 weeks longer than they would have if evaluated at the baseline visit. In retrospect, this may have caused some confusion among assessors, as evidenced by the fact that 113 patients with GCA were reported as having no items on the BVAS at the 2-week assessment. It was not relevant to report the VDI score at the baseline visit as well as the 2-week visit; we elected to report it during the 2-week visit, because this would minimise the amount of the work required at the baseline visit.[69,70]

The VDI assessment performed at 6 months would aim to capture all damage occurring irrespective of cause, which is a principle of the VDI. Therefore, any items relating to possible disease activity as a result of GCA would not necessarily be reflected in the VDI. Equally, the VDI could report events that may have

occurred at least 3 months prior to the 6-month assessment date, for example, development of the loss of vision or stroke. However, when recording items in the VDI, the emphasis is on documenting the presence of damage occurring after the onset of vasculitis, regardless of the cause of the damage (the item could relate to disease or it could have been a complication of treatment, infection or exacerbation of or new development of an unrelated comorbidity).

Centre effect

Twenty centres participated in the TABUL study; *Table 44* shows the reference diagnosis and pre-test risk and clinical pre-test certainty of GCA by centre. Overall, there was a good spread of patients with or without GCA in centres recruiting 10 or more patients: between 50% and 100% of the patients recruited from these centres had a reference diagnosis of GCA.

The pre-test risk of likelihood of having GCA based on our external model from the DCVAS data set shows that, for centres recruiting at least 10 patients, there was a good spread of high-, medium- and low-risk patients. The clinician's pre-test certainty of diagnosis also showed a good spread across definite, probable and possible cases for all centres recruiting at least 10 patients. We conclude that the selection criteria used by different centres recruiting patients for the study were similar and allows for greater generalisability of our results.

Health-related quality of life

The primary role of the EQ-5D data is to inform the economic analysis and modelling (see *Chapter 7*), but they are also presented here as a summary of the state of health among patients within the cohort.

Table 45 shows EQ-5D over time, EQ-5D health state and thermometer health state increase by 2-week assessment, but this effect is not sustained at 6 months. There is little difference in the EQ-5D by reference diagnosis or steroid use at 6 months (*Table 46*).

Some of the patients who did not have GCA may have been treated with long-term steroids for other reasons, for example, PMR. It is conceivable that other comorbidities may have influenced the EQ-5D more strongly than GCA itself, but the tables do not suggest that there was a significant impact of having GCA on health-related quality of life as measured by the EQ-5D at any of the time points assessed; nor was there any significant change in health-quality related of life during the period of the study.

TABLE 44 Diagnosis and pre-test risk by centre

		Reference diagnosis, *n* (%)		Pre-test risk, *n* (%)			Clinician pre-test certainty, *n* (%)		
Centre	*N*	GCA	Not GCA	High	Medium	Low	Definite	Probable	Possible
Chapel Allerton Hospital, Leeds, UK	16	12 (75)	4 (25)	4 (25)	6 (38)	6 (38)	10 (63)	4 (25)	2 (13)
City Hospital, Birmingham, UK	4	4 (100)	0 (0)	1 (25)	2 (50)	1 (25)	0 (0)	4 (100)	0 (0)
Dudley Hospital, Dudley, UK	4	4 (100)	0 (0)	1 (25)	2 (50)	1 (25)	0 (0)	3 (75)	1 (25)
Gateshead Hospital, Gateshead, UK	14	10 (71)	4 (29)	2 (14)	9 (64)	3 (21)	4 (29)	9 (64)	1 (7)
Great Yarmouth Hospital, Great Yarmouth, UK	2	2 (100)	0 (0)	1 (50)	1 (50)	0 (0)	0 (0)	2 (100)	0 (0)

TABLE 44 Diagnosis and pre-test risk by centre (*continued*)

		Reference diagnosis, *n* (%)		Pre-test risk, *n* (%)			Clinician pre-test certainty, *n* (%)		
Centre	*N*	GCA	Not GCA	High	Medium	Low	Definite	Probable	Possible
Hospital de Santa Maria, Lisbon, Portugal	2	1 (50)	1 (50)	1 (50)	1 (50)	0 (0)	0 (0)	1 (50)	1 (50)
Hospital of Southern Norway Trust, Kristiansand, Norway	25	21 (84)	4 (16)	5 (20)	14 (56)	6 (24)	6 (24)	12 (48)	7 (28)
Jena University Hospital, Jena, Germany	12	11 (92)	1 (8)	2 (17)	7 (58)	3 (25)	5 (42)	7 (58)	0 (0)
Musgrave Park, Belfast, UK	6	4 (67)	2 (33)	2 (33)	3 (50)	1 (17)	0 (0)	4 (67)	2 (33)
Nuffield Orthopaedic Centre, Oxford, UK	111	60 (54)	51 (46)	16 (14)	44 (40)	51 (46)	11 (10)	66 (59)	34 (31)
Princess Alexandra Hospital, Harlow, UK	7	7 (100)	0 (0)	3 (43)	2 (29)	2 (29)	5 (71)	2 (29)	0 (0)
Queen Alexandra Hospital, Portsmouth, UK	7	6 (86)	1 (14)	3 (43)	2 (29)	2 (29)	1 (14)	5 (71)	1 (14)
Queen's Hospital Romford, Essex, UK	8	7 (88)	1 (13)	5 (63)	1 (13)	2 (25)	3 (38)	4 (50)	0 (0)
Queen's Medical Centre, Nottingham, UK	22	12 (55)	10 (45)	7 (32)	8 (36)	7 (32)	3 (14)	12 (55)	7 (32)
Royal Berkshire, Reading, UK	4	4 (100)	0 (0)	2 (50)	2 (50)	0 (0)	0 (0)	4 (100)	0 (0)
Royal Derby Hospital, Derby, UK	3	2 (67)	1 (33)	1 (33)	0 (0)	2 (67)	1 (33)	1 (33)	1 (33)
Southend University Hospital, Southend, UK	90	61 (68)	29 (32)	21 (23)	37 (41)	32 (36)	25 (28)	37 (41)	28 (31)
St Vincent Hospital, Dublin, Ireland	18	16 (89)	2 (11)	3 (17)	6 (33)	9 (50)	1 (6)	11 (61)	6 (33)
Stoke Mandeville Hospital, Stoke, UK	20	10 (50)	10 (50)	7 (35)	5 (25)	8 (40)	5 (25)	13 (65)	2 (10)
Sunderland Royal Hospital, Sunderland, UK	6	3 (50)	3 (50)	2 (33)	2 (33)	2 (33)	0 (0)	3 (50)	3 (50)

TABLE 45 EuroQol-5 Dimensions assessment by all patients in the TABUL study by visit

	Visit		
Measure	Baseline ($n = 365$)	2 weeks ($n = 369$)	6 months ($n = 328$)
EQ-5D health state			
Number (%) of responses	363 (99.5)	364 (98.6)	326 (99.4)
Mean (SD)	0.66 (0.27)	0.73 (0.26)	0.70 (0.29)
Median (IQR)	0.73 (0.62–0.80)	0.80 (0.65–1.00)	0.73 (0.62–1.00)
EQ-5D health state: change from baseline			
Number (%) of responses	–	350 (94.9)	312 (95.1)
Mean (SD)	–	0.07 (0.25)	0.02 (0.31)
Median (IQR)	–	0.00 (0.00–0.14)	0.00 (–0.11–0.20)
EQ-5D thermometer health state			
Number (%) of responses	360 (98.6)	369 (100.0)	325 (99.1)
Mean (SD)	53.8 (29.7)	58.8 (30.5)	56.8 (30.8)
Median (IQR)	60.0 (30.0–80.0)	70.0 (40.0–84.0)	65.0 (30.0–80.0)
EQ-5D thermometer health state change from baseline			
Number (%) of responses	–	351 (95.1)	309 (94.2)
Mean (SD)	–	4.9 (22.4)	1.4 (26.3)
Median (IQR)	–	1.0 (–1.0 to 10.0)	0.0 (–10.0 to 11.0)

TABLE 46 Six-month EQ-5D by reference diagnosis and steroid use at 6 months

	Reference diagnosis		Steroid usage at 6 months	
Measure ($n = 381$)	GCA ($n = 224$)	Not GCA ($n = 104$)	On steroids ($n = 251$)	Not on steroids ($n = 77$)
6-month EQ-5D health state				
Number (%) of responses	224 (100.0)	102 (98.1)	251(100.0)	75 (97.4)
Mean (SD)	0.72 (0.28)	0.65 (0.31)	0.71 (0.29)	0.67 (0.30)
Median (IQR)	0.78 (0.62–1.0)	0.69 (0.62–0.85)	0.74 (0.62–1.00)	0.73 (0.62–0.85)

Safety and adverse events

We expected that the two interventions (biopsy and ultrasound) would produce a different profile of AEs. We would expect the biopsy to result in discomfort, bruising, bleeding or infection around the biopsy site. By contrast, we would expect very little in terms of harm from the ultrasound scan. We specifically sought to document any potential harm caused by the interventions in our study.

In addition to collecting information on any adverse effects of the two main interventions, we had an option for sites to collect information on any other adverse outcomes during the observation period. All participants (100%) experienced at least one AE during the study. A total of 1229 AEs were reported during follow-up (including repeated events). *Table 47* shows expected AEs and *Table 48* shows AEs related to study tests. Fifty-seven patients experienced an AE related to the study test. Of these, 53 (6.3%) of all expected adverse events were definitely related to biopsy, 10 were possibly related to biopsy and two were definitely related to scanning. It was expected that the proportion of the AEs that would be related to the study test would be 81%.

TABLE 47 Expected AEs

Expected AEs	*n* (%)
Number of participants who experienced > 1 expected AE	170 (44.6)
Number of all expected AEs (including repeated events)	836
Severity	
Mild	660 (78.9)
Moderate	154 (18.4)
Severe	22 (2.6)
Related to scan?	
Definitely related	2 (0.2)
Not related	833 (99.6)
Unable to assess	1 (0.1)
Related to biopsy?	
Definitely related	53 (6.3)
Possibly related	6 (0.7)
Not related	777 (92.9)
Event type	
Biopsy wound problems	15 (1.8)
Post-biopsy problems	38 (4.5)
US painful	2 (0.2)
Blurred vision	47 (5.6)
Breathlessness	20 (2.4)
Return of GCA	9 (1.1)
Mood/CNS/dizziness	112 (13.4)
Infection	85 (10.2)
Skin change/bruising	50 (6.0)
Flushing/sweating	64 (7.7)
Hypertension/ischaemic heart disease	34 (4.1)
Diabetes mellitus	32 (3.8)
Weight gain/bloating/indigestion	80 (9.6)
Weakness	38 (4.5)
Other drug toxicity	14 (1.7)
Other	196 (23.4)

CNS, central nervous system; US, ultrasound.

TABLE 48 Adverse events related to study tests

AEs related to study tests	*n* (%)
Number of participants who experienced > 1 AE related to tests	57 (15.0)
Number of all related AEs (including repeated events)	75
Severity	
Mild	66 (88.0)
Moderate	9 (12.0)
Related to study tests	
Definitely related to biopsy	63 (84.0)
Probably related to biopsy	10 (13.3)
Definitely related to scan	2 (2.7)
Expected?	
No	14 (18.7)
Yes	61 (81.3)
Event type	
Biopsy wound problems	14 (18.7)
Post-biopsy problems	44 (58.7)
US painful	2 (2.7)
Mood/CNS/dizziness	4 (5.3)
Infection	1 (1.3)
Neuropathy	4 (5.3)
Other	6 (8.0)

CNS, central nervous system; US, ultrasound.

The serious AEs reported during follow-up are shown in *Table 49*; 65 participants experienced 104 serious AEs, none of which was related to either study test. *Table 47* describes the details of the expected AEs that occurred during the course of the study. This was not a mandatory part of the data collection and we suspect that this is an underestimate of events occurring during the first 6 months of disease in patients with GCA. We have based our analysis on 170 participants in whom at least one AE was reported. In total, 836 events were reported, only 3% of which were classed as severe. The majority of events were unrelated to either scan or biopsy. Most events consisted of the complications that would be expected in association with the diagnosis and treatment of GCA.

In *Table 48* we have summarised the experience of AEs that are directly related to the study investigations. Overall, 57 patients experienced 75 AEs related to the tests. A total of 63 out of the 75 of those events were either definitely or probably related to biopsy. Two events were definitely related to the scan (which consisted of pain at the time of the ultrasound examination). Several patients experienced biopsy-related wound problems or post-biopsy problems such as pain or numbness, whereas none of the patients described any of these features in relation to the ultrasound scan. Previous studies have suggested a much lower rate of complications from biopsies. In one study only two complications were reported from 412 biopsies performed on 394 patients;[74] in a smaller study of 45 cases, there were no biopsy-related complications at all.[75] Complications from biopsy can be serious, including facial nerve injury, as reported in four cases when the biopsy was attempted in the pre-auricular area.[76] An incidence of facial nerve injury of 16% was reported in a study of 75 patients undergoing biopsy, of whom only 42% recovered.[77]

TABLE 49 Serious AEs

SAEs	*n* (%)
Number of participants who experienced > 1 SAE	65 (17.1)
Number of all SAEs (including repeated events)	104
Severity	
Mild	4 (3.8)
Moderate	41 (39.4)
Severe	54 (51.9)
Missing	5 (4.8)
Related to scan?	
Related	0 (0.0)
Related to biopsy?	
Related	0 (0.0)
Expected?	47 (45.2)
Seriousness	
Hospitalisation required	74 (71.2)
Death	16 (15.4)
Life- or limb-threatening	3 (2.9)
Persistent or significant disability/incapacity	5 (4.8)
Hospitalisation prolonged	2 (1.9)
Other important medical event	4 (3.8)
Event type	
Blurred vision	1 (1.0)
Breathlessness	5 (4.8)
Return of GCA	2 (1.9)
Mood/CNS/dizziness	9 (8.7)
Infection	22 (21.2)
Skin change/bruising	2 (1.9)
Flushing/sweating	1 (1.0)
Hypertension/ischaemic heart disease	24 (23.1)
Diabetes mellitus	6 (5.8)
Weight gain/bloating/indigestion	1 (1.0)
Weakness	3 (2.9)
Admission	1 (1.0)
Cancer	5 (4.8)
Renal impairment/failure	2 (1.9)
Gastrointestinal bleed	1 (1.0)
Other	19 (18.3)

CNS, central nervous system; SAE, serious adverse event.

We suspect that previous studies may have significantly underestimated the morbidity associated with TAB. We do not think that the rate of complications that we have reported is outside the expected number seen in clinical practice.

Table 49 describes the serious AEs in the cohort, none of which was related to either investigation; they largely reflected the effects of older age, as well as of having GCA. There were 16 deaths in the study cohort, reflecting an elderly population. Seventy-four patients required hospitalisation. We conclude that the rate of AEs suggests that the study population was typical of many cohorts of patients with GCA experiencing comorbidity and the complications of their disease and its treatment.

Chapter 6 Analysis of inter-rater agreement and clinical vignettes

Participation in the agreement and vignette exercises

Twenty sonographers from 16 sites and 26 pathologists from 19 sites were asked if they wished to form the TABUL sonographers and TABUL pathologists groups that were responsible for the inter-rater agreement exercises. Twelve sonographers from 10 centres and 14 pathologists from 13 centres joined the groups and completed the exercise. Some centres had no eligible sonographers or pathologists to invite because of their involvement as expert reviewers or designers of the exercise.

Twenty clinicians with experience in managing GCA and involved in, or associated with, the TABUL study were invited to review vignettes for the clinical vignette exercise. Sixteen indicated that they were able to do the exercise and 14 completed assessments of all 30 clinical vignettes.

Selection of patients

A total of 255 initially eligible patients were identified at the first stage of selection; the first 33 patients from the randomly ordered list were selected for further screening. Twelve (36%) of these patients did not meet the inclusion criteria and were replaced with the next 12 eligible patients. Following piloting of the exercise and checking of the videos and images in the third stage, a further two patients were replaced because the ultrasound images were considered to be of inadequate quality. Three further patients were highlighted because of concerns about the ultrasound videos; two were retained because it was deemed that their videos were difficult to interpret rather than of inadequate quality; and one case was modified to include an alternative video for the same patient that better supported the original sonographer's interpretation. Finally, one of the rating cases was replaced post exercise with one of the three rated reserve cases because the patient did not complete a follow-up assessment and was excluded from the main analyses. Therefore, there were 30 unique cases, six of which were repeated, for evaluation by 12 sonographers and 14 pathologists.

Inter-rater agreement between sonographers and pathologists

Ratings based on images alone

Sonographers and pathologists rated each case as either consistent with GCA or not consistent with GCA; they were also asked to report the confidence they had in their findings, using four categories to indicate the level of certainty in their decision. The rating was done before and after seeing a brief clinical vignette describing a few key characteristics of the patient.

The distribution of the results (consistent with GCA or not consistent with GCA) by the sonographers for the 30 original cases assessed before being shown the clinical vignette is shown in *Figure 13*. The 12 sonographers unanimously agreed in 10 of the 30 cases: four as GCA positive and six as GCA negative. In half the cases there was no unanimous agreement, but no more than two of the sonographers differed from the majority view. In five cases there was greater disagreement, with three or four sonographers differing from the majority.

All 14 pathologists agreed unanimously on 11 cases, six of which were consistent with GCA and five of which were not consistent with GCA (*Figure 14*). There were 13 cases in which no more than two

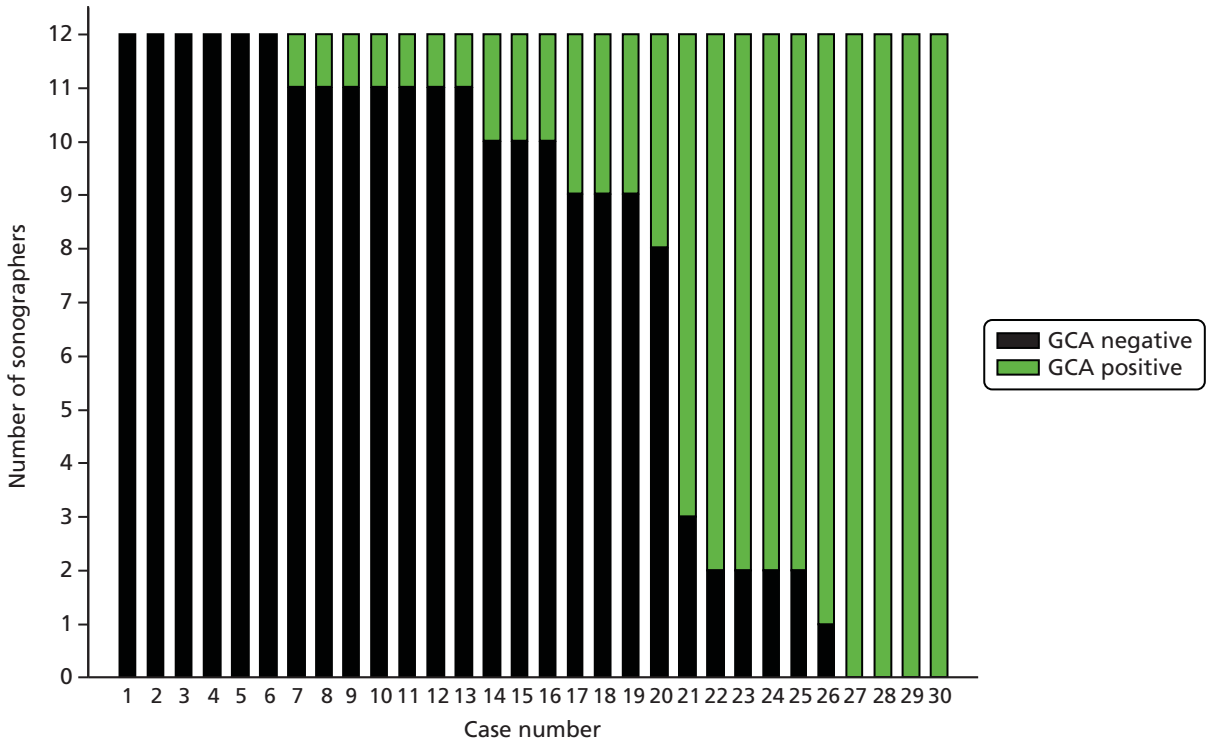


FIGURE 13 Frequency of evaluations consistent with or not consistent with GCA by 12 sonographers rating 30 cases.

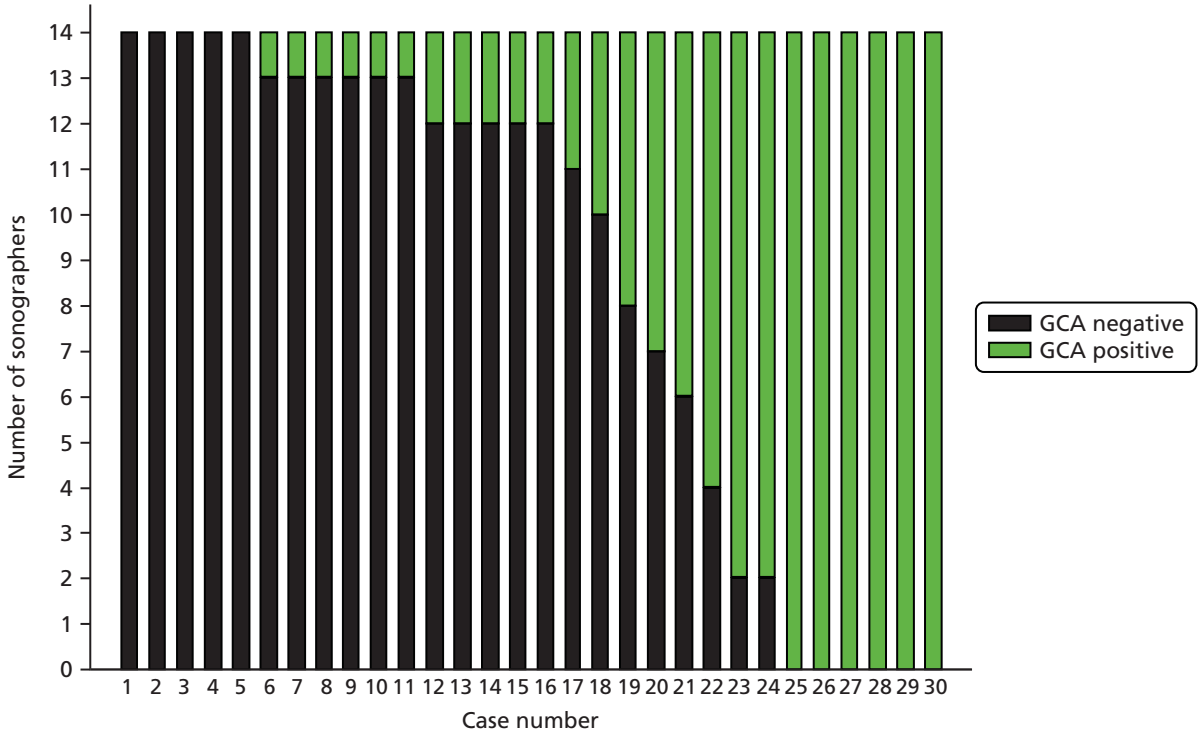


FIGURE 14 Frequency of evaluations consistent with or not consistent with GCA by 14 pathologists rating 30 cases.

pathologists differed from the majority view. In six cases, there was greater disagreement, in one of which opinion was evenly divided, with equal numbers of pathologists defining the patient as having or not having GCA.

Eight of the 30 cases involved patients who had been assessed as biopsy positive by the original reporting pathologist. All eight cases reported evidence of giant cells and these were the eight cases in the exercise that all (six cases), or all except two (two cases), of the pathologists judged to be consistent with GCA. A ninth patient was reported as biopsy negative by the original reporting pathologist but interpreted as biopsy positive by the clinician based on the abnormalities described in the biopsy report (intimal hyperplasia, fragmentation and reduplication consistent with previous GCA but no active inflammation). Two of the pathologists judged the case to be GCA positive, but most concurred with the original biopsy report and judged the case to be GCA negative.

For each GCA-positive or GCA-negative assessment, the sonographers and pathologists were asked to indicate if they were certain or uncertain in their assessments. Analysis of differences in the certainty of assessments indicated that sonographers judged fewer ratings as certain than did pathologists. For GCA-positive ratings, 69.0% were judged as certain by sonographers, whereas 79.8% were judged as certain by pathologists. For GCA-negative ratings the sonographers were certain for 54.5%, whereas the pathologists were certain for 71.0%. However, a comparison between the 14 pathologists and the 12 sonographers in the proportion of cases judged as certain did not provide strong evidence for a difference (Wilcoxon rank-sum test, $p = 0.13$). The distribution of these ratings is shown in *Figures 15* and *16*.

Ratings based on images and vignettes

The sonographers and pathologists were asked to give their assessment of each case before and after seeing a brief vignette describing the patient. The vignettes provided information on the patient's age, sex, main symptoms and blood abnormalities. The additional clinical information had little impact on the assessments made by the sonographers and pathologists; fewer than 5% of cases overall were amended following the provision of the brief vignettes (*Table 50*).

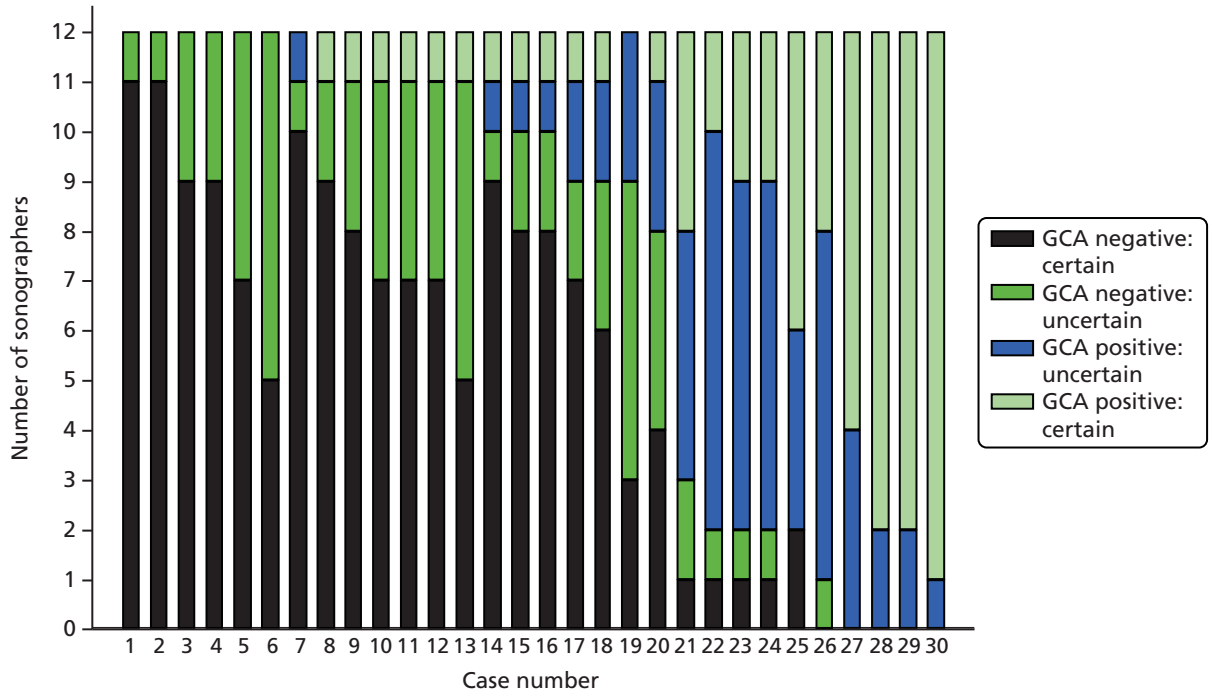


FIGURE 15 Frequency of certain and uncertain positive and negative assessments by 12 sonographers rating 30 cases.

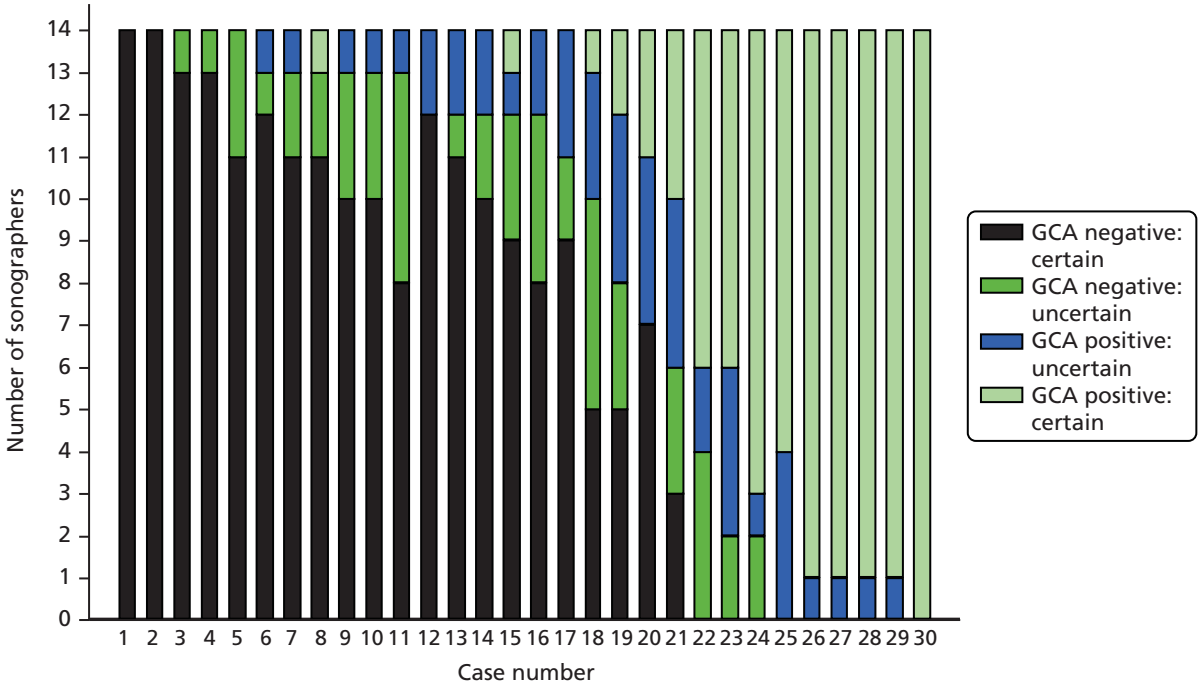


FIGURE 16 Frequency of certain and uncertain positive and negative assessments by 14 pathologists rating 30 cases.

TABLE 50 Alteration of assessments after provision of a brief clinical vignette

	Sonographers		Pathologists	
Overall	*n*	%	*n*	%
GCA positive, no change	132	36.7	159	37.9
GCA positive to GCA negative	2	0.6	3	0.7
GCA negative to GCA positive	13	3.6	11	2.6
GCA negative, no change	213	59.2	247	58.8

Figures for 30 cases assessed by 12 sonographers (360 in total) and 14 pathologists (420 in total).

The extent of agreement between the sonographers and between the pathologists was evaluated by estimating the intraclass correlation coefficient (ICC). There was little difference between the two groups when restricting the decision to a binary GCA positive or GCA negative (*Table 51*). The intraclass correlations were 0.61 (95% CI 0.48 to 0.75) for the sonographers and 0.62 (95% CI 0.49 to 0.76) for the pathologists. A small reduction in agreement was observed if the agreement was assessed from the post-vignette ratings: 0.58 (95% CI 0.44 to 0.72) for the sonographers and 0.59 (95% CI 0.45 to 0.73) for the pathologists.

There was better agreement between the pathologists if the certainty of the assessment was taken into account. The ICC for the pathologists was 0.72 (95% CI 0.60 to 0.83), whereas that for the sonographers was 0.58 (95% CI 0.44 to 0.72). In other words, sonographers and pathologists achieved similar levels of diagnostic accuracy, but sonographers were less confident in their diagnosis; perhaps this reflected their limited experience of ultrasound in GCA in comparison with pathologists' assessment of the histological features. An analysis of the cases assessed after seeing the brief vignette produced slightly lower ICCs for both sonographers and pathologists.

TABLE 51 Intraclass correlation coefficients (with 95% CIs)

GCA positive or negative	Sonographers	Pathologists
Pre-vignette cases	0.612 (0.484 to 0.748)	0.621 (0.491 to 0.756)
Post-vignette cases	0.581 (0.450 to 0.724)	0.587 (0.454 to 0.730)
GCA positive/negative with certainty		
Pre-vignette cases	0.575 (0.442 to 0.719)	0.719 (0.597 to 0.830)
Post-vignette cases	0.562 (0.424 to 0.711)	0.677 (0.548 to 0.799)

Estimated using two-way random-effects analysis of variance.

The extent of agreement between the sonographers interpreting ultrasound videos and between pathologists interpreting images of TABs was similar for the decision to categorise cases as positive or negative. However, there was still a fair degree of disagreement. There was limited impact of the brief vignettes, representing the type of information that a sonographer (while scanning a patient) or pathologist (seeing a biopsy request form) might be aware of in routine practice, on the assessments made and the extent of inter-rater agreement.

Intrarater agreement for sonographers and pathologists

Intrarater agreement was evaluated by including repeats of six of the cases during the exercise. Overall, there were 10 instances of inconsistent assessments between the original and repeated cases made by the 12 sonographers and seven made by the 14 pathologists (*Table 52*). Overall, raw agreement was 86.1% for the sonographers and 91.7% for the pathologists.

Analysis of the consistency of individual sonographers and pathologists is shown in *Table 53*. No individual assessor was inconsistent for more than two of the cases. Over half (57%) of the pathologists and one-third of the sonographers were completely consistent in their assessment of the six repeated cases. However, no statistically significant difference was observed between the 12 sonographers and the 14 pathologists in the number of cases that were inconsistent (Wilcoxon rank-sum test, $p = 0.21$).

Kappa statistics were used to estimate 'chance-corrected' intrarater agreement for the sonographers and pathologists. For the six of the cases that were repeated, the 14 pathologists achieved raw agreement of 91.7% for categorising cases as GCA or not GCA and an overall kappa statistic of 0.83 (*Table 54*).

TABLE 52 Analysis of consistency of assessments by 12 sonographers and 14 pathologists for the six repeated cases

	Sonographers				Pathologists			
Repeated case	Both negative	Differed	Both positive	Raw agreement (%)	Both negative	Differed	Both positive	Raw agreement (%)
1	12	0	0	100	7	3	4	79
2	1	4	7	67	0	0	14	100
3	11	1	0	92	9	4	1	71
4	0	2	10	83	0	0	14	100
5	12	0	0	100	13	0	1	100
6	7	3	2	75	13	0	1	100

TABLE 53 Intrarater agreement: number of inconsistent cases by rater

Number of inconsistent cases	Sonographers, *n* (%)	Pathologists, *n* (%)
0	4 (33)	8 (57)
1	6 (50)	5 (36)
2	2 (17)	1 (7)
3	0 (0)	0 (0)
≥ 4	0 (0)	0 (0)

TABLE 54 Kappa statistics for intrarater agreement for repeated cases

	Sonographers			Pathologists		
Diagnosis	Raw agreement (%)	Expected agreement (%)	Kappa statistic	Raw agreement (%)	Expected agreement (%)	Kappa statistic
GCA positive or GCA negative	86.1	55.6	0.69	91.7	50.3	0.83
GCA positive or GCA negative with level of certainty[a]	94.6	72.1	0.81	93.8	58.2	0.85

a Quadratic weights used for estimating kappa statistic.

The 12 sonographers achieved raw agreement of 86.1% and an overall kappa statistic of 0.69 for the same cases. Once ratings are reported using four categories to allow for certainty in the assessments by the sonographers and pathologists, the weighted kappa statistics for agreement are similar: 0.85 for the pathologists and 0.81 for the sonographers.

Analysis of clinical vignettes with ultrasound in the absence of biopsy

Fourteen clinicians reviewed the 30 clinical vignettes and provided responses for clinical decisions at two stages: (1) the likelihood of a diagnosis of GCA and whether or not to perform a biopsy after seeing information from a patient's initial presentation; and (2) the likelihood of a diagnosis of GCA and whether or not to continue with high-dose steroids after seeing a brief written summary of the patient's ultrasound results and clinical information after 2 weeks.

In 21 (70%) of the vignettes, the majority of the panel considered the likelihood of GCA to be probable or definite, and for two vignettes the panel was split evenly (*Table 55*). In the remaining seven vignettes, which were considered least likely to be GCA, the majority of the panel would perform a biopsy; the one exception was for vignette case 8, for which only five of the 14 panel members would recommend a biopsy.

There was some evidence of an association between certainty of GCA and recommendation for biopsy. Members of the panel were generally consistent in not recommending a biopsy for patients whom they considered not to have GCA; biopsy was indicated for 10% of the time in these cases. Biopsy was most frequently recommended (94% of the time) by panellists for vignettes judged as probable GCA. The percentage of biopsy recommendations was lower for vignettes judged as definite GCA (78% recommended for biopsy) and those judged as possible GCA (80% recommended for biopsy). These findings suggest some reluctance to recommend biopsies in patients considered to have little chance of

TABLE 55 Certainty of GCA and recommendation for a TAB at presentation for 30 clinical vignettes

	Certainty of GCA				
Case	Definite	Probable	Possible	Not GCA	Perform biopsy, *n* (%)
1	3	10	1	0	14 (100)
2	2	11	1	0	14 (100)
3	0	0	9	5	8 (57)
4	0	6	5	3	12 (86)
5	2	10	2	0	13 (93)
6	0	7	5	2	11 (79)
7	0	10	4	0	14 (100)
8	0	0	7	7	5 (36)
9	0	11	3	0	13 (93)
10	8	6	0	0	11 (79)
11	4	10	0	0	12 (86)
12	1	9	4	0	13 (93)
13	0	2	7	5	9 (64)
14	1	10	3	0	14 (100)
15	0	7	6	1	11 (79)
16	6	7	0	1	11 (79)
17	0	1	10	3	9 (64)
18	3	10	1	0	12 (86)
19	1	8	5	0	12 (86)
20	6	7	1	0	11 (79)
21	3	9	2	0	11 (79)
22	3	9	2	0	12 (86)
23	3	8	3	0	12 (86)
24	4	9	1	0	12 (86)
25	0	5	7	2	10 (71)
26	4	9	1	0	13 (93)
27	2	8	4	0	11 (79)
28	10	3	1	0	11 (79)
29	2	3	7	2	9 (64)
30	2	9	3	0	13 (93)

having GCA based on vignettes describing their symptoms and results of physical examinations and blood tests. The findings also suggest some reluctance to recommend biopsies in patients who were regarded as having very clear-cut evidence of GCA, based on their clinical presentation and the results of blood tests. However, in patients who are diagnosed as having GCA without undergoing a biopsy, there may be a concern that there is no irrefutable evidence of GCA if the diagnosis is subsequently questioned; by contrast, the clinician's interpretation is perceived as always open to reconsideration.

Table 56 describes the panellists' assessment of the diagnosis of GCA, once the ultrasound test result was revealed and information about the symptoms, blood tests and physical examination had been provided. Panel recommendations for continuing treatment for GCA with high-dose steroids were categorised as agree, disagree or uncertain. The uncertain category was used when the median rating of the 14 panellists

TABLE 56 Certainty of GCA and appropriateness of continuing high-dose steroid treatment at the 2-week assessment for 30 clinical vignettes

		Certainty of GCA				Appropriateness of continuing with high-dose steroids for GCA[a]				
Case	US result	Definite	Probable	Possible	Not GCA	1–3	4–6	7–9	Median	Appropriateness
1	+	12	2	0	0	1	0	13	9	Appropriate
2	–	1	7	4	2	4	3	7	6.5	Uncertain (D)
3	–	0	0	2	12	12	2	0	1.5	Inappropriate
4	–	0	5	3	6	7	5	2	4.5	Uncertain (D)
5	+	14	0	0	0	0	0	14	9	Appropriate
6	+	12	1	1	0	0	2	12	9	Appropriate
7	–	0	4	6	4	4	4	6	5	Uncertain (D)
8	–	0	0	1	13	14	0	0	1	Inappropriate
9	–	0	6	4	4	4	5	5	5.5	Uncertain (D)
10	+	13	1	0	0	0	0	14	9	Appropriate
11	–	1	6	5	2	3	4	7	6	Uncertain
12	–	0	6	5	3	3	5	6	6	Uncertain
13	–	0	0	6	8	9	4	1	2	Inappropriate
14	–	0	5	5	4	5	4	5	5	Uncertain (D)
15	–	0	7	2	5	5	2	7	6	Uncertain (D)
16	+	13	1	0	0	0	0	14	9	Appropriate
17	–	0	1	6	7	8	4	2	2.5	Inappropriate
18	–	1	6	3	4	4	4	6	6	Uncertain (D)
19	–	0	5	4	5	5	5	4	5.5	Uncertain (D)
20	–	4	9	1	0	1	2	11	8	Appropriate
21	+	14	0	0	0	0	0	14	9	Appropriate
22	+	14	0	0	0	0	0	14	9	Appropriate
23	+	14	0	0	0	0	0	14	9	Appropriate
24	–	2	6	4	2	3	4	7	6.5	Appropriate
25	–	0	3	4	7	8	3	3	3	Inappropriate
26	+	14	0	0	0	0	0	14	9	Appropriate
27	+	14	0	0	0	0	1	13	9	Appropriate
28	+	14	0	0	0	0	0	14	9	Appropriate
29	–	1	2	4	7	7	4	3	4	Uncertain (D)
30	–	0	8	4	2	3	3	8	7	Appropriate

D, disagreement; US, ultrasound.
a On a 9-point scale (1, extremely inappropriate; 5, uncertain; 9, extremely appropriate).

lay in the mid-range or if there was wide variation (indicated as disagreement) in the recommendations of the panellists regardless of the median. In 11 of the cases, the ultrasound test results were reported as consistent with GCA, and panel members' views were weighted strongly towards a definite diagnosis of GCA for these vignettes. The panel members were also in agreement that high-dose steroids for GCA should be continued for all 11 cases.

There were 19 ultrasound-negative vignettes and there was a reluctance to classify any of these vignettes as definite GCA. In only 4 of the 19 vignettes did a majority of the panel categorise the patient as having probable or definite GCA; in three of these four cases, the panel was in agreement that it was appropriate to continue with high-dose steroids and in the fourth case the panel was uncertain, owing to disagreement. Of the remaining 15 vignettes, the panel agreed that it was inappropriate to continue with high-dose steroids in five of the case vignettes and was uncertain in the others. There was one vignette, number 20, for which the biopsy was positive but the ultrasound was reported as negative. Despite the fact that the panel members were not aware of the positive biopsy result, they were still in agreement that it was appropriate to continue with high-dose steroids.

Chapter 7 Cost-effectiveness analysis

Introduction

The economic evaluation of the two tests needs to consider any differences in the diagnostic accuracy between them, as well as the costs and impact of the tests in terms of the development of GCA-related complications, treatments and related side effects.

The starting point for the modelling is the statistical output showing the sensitivity and specificity of the two individual tests and any diagnostic strategies which incorporate them (see *Chapter 5*). Sensitivity is the proportion of patients with true GCA who are detected by the test or strategy; the remaining proportion is made up of 'false negatives', that is, patients who test negative despite having GCA. Specificity is the proportion of patients without GCA who are classified as negative by the test or strategy; the remaining proportion is made up of 'false positives', that is, patients who test positive but who do not have GCA. A problem with false-negative and false-positive results is that patients falling into these categories may initially be managed in a different way, with potentially adverse consequences, compared with how they would have been managed had their true disease status been known earlier. The economic analysis estimates the relative cost-effectiveness of the alternative tests and strategies by quantifying and trading off the following.

- The different costs of the tests or strategies.
- The different proportions of false negatives and false positives.
- The cost and health-related quality-of-life impact of a false negative, that is, when a patient remains undetected with GCA for up to around 2 months, with the attendant risk of developing complications such as vision loss.
- The cost and health-related quality-of-life impact of a false positive, that is, initiating or continuing treatment with high-dose steroids in a patient without GCA for many months and the impact that any unnecessary treatment has on the risk of AEs such as fractures, diabetes mellitus and weight gain.

The primary objective of the economic analysis is to estimate the cost-effectiveness of ultrasound instead of biopsy for the diagnosis of GCA. The secondary objective is to estimate the cost-effectiveness of performing a biopsy following ultrasound as an alternative to TAB alone in the diagnosis of GCA. In addition, alternative diagnostic strategies have been evaluated using estimates of sensitivity and specificity from statistical modelling (as described in *Chapter 5*).

Biopsy and ultrasound are also evaluated when used in conjunction with clinical judgement, that is, the clinician's decision on the diagnosis at 2 weeks based on knowledge of the patient's symptoms, signs and available test results such as blood tests and the biopsy. This more closely reflects current clinical practice of using biopsy results to aid the clinical diagnosis rather than to define the diagnosis.

Methods

In this section, the model structure is described, followed by details of the evidence sources used for the various parameter values in the model. These include the performance of the diagnostic testing strategies; risks of GCA-related complications and glucocorticoid-related AEs; and associated costs and health-related quality-of-life effects. The costs of the tests and medications are also covered.

The development of the economic model structure was informed by evidence from published research on GCA in order to understand the main complications of the disease and steroid-related side effects. This was supplemented with evidence from previous economic and decision-analytic studies of GCA[78,79] and an analysis of outcomes and cost-effectiveness of a fast-track service for GCA.[80]

Model structure

The model structure takes the form of a combination of three submodels: first, a decision tree for the initial diagnostic testing; second, a risk submodel of the incidence of GCA-related complications and steroid-related AEs over 2 or 3 years; and third, a submodel of the lifetime effects of these incident complications and AEs. The model structure is shown in *Figure 17*.

Approach to obtaining values for parameters used in the economic model

We carried out a search for review articles in GCA and key evidence sources such as guidelines on managing GCA, prescribing steroids and steroid-related complications. We also consulted the National Institute for Health and Care Excellence (NICE) Clinical Knowledge Summary[81] for GCA. It became clear from an initial assessment of these sources that there was limited evidence on rates of complications in GCA. Visual complications were most commonly reported but there was much heterogeneity of reported outcomes and results were rarely for time periods relevant to our analysis. Furthermore, given the relatively similar test performance of biopsy and ultrasound (especially when used in conjunction with clinical judgement) and the low incidence of major comorbidity, it seemed likely that complication rates would not be a major driver of cost-effectiveness.

We used an iterative approach to the cost-effectiveness modelling. Further review of this evidence was not required once it became apparent that the results were unlikely to be sensitive to model parameters relating to complications of GCA and steroids and that the cost difference between biopsy and ultrasound was the major driver of the cost-effectiveness. Instead, our modelling focused on two aspects of test performance that would be more important than had been previously realised: the need to focus on the implications of using the test results in conjunction with clinical judgement and uncertainty around the reference diagnosis for GCA.

The main sources of evidence for the model are summarised in *Table 57*.

The specific evidence sources are provided in the detailed sections that follow.

Test accuracy was derived from an analysis of data collected in the TABUL study. For other parameters, such as the risk of complications from GCA (which were relatively infrequent in TABUL), evidence was obtained from alternative sources. The precise sources of data are described in greater detail in the following sections.

Performance of diagnostic strategies

The economic analysis considered three types of diagnostic strategy, as summarised in *Table 58*. One type relies on the use of test results alone for the diagnosis of GCA. Such strategies may be as simple as testing biopsy positive or biopsy negative, or they may involve combinations or components of tests. The second type of strategy involves the combination of test results with clinical judgement (the clinician's assessment based on the patient's characteristics and available test results) after the clinician has assessed the patient at the 2-week visit. The third type, sequential diagnostic strategies, involves applying test results in combination with characteristics of patients.

The sequential diagnostic strategies include those based around the three categories of pre-test risk defined in *Chapter 2* and reported in *Chapter 5*. The high-risk group comprised patients with tongue or jaw claudication and a high ESR or CRP level at presentation or before starting steroids. A high ESR level was defined as at least 60 mm/hour. A high CRP level was defined as at least 40 mg/l. The low-risk group comprised patients with no evidence of claudication and no evidence of a high ESR or CRP level at presentation or before starting steroids. The medium-risk group comprised the remaining patients.

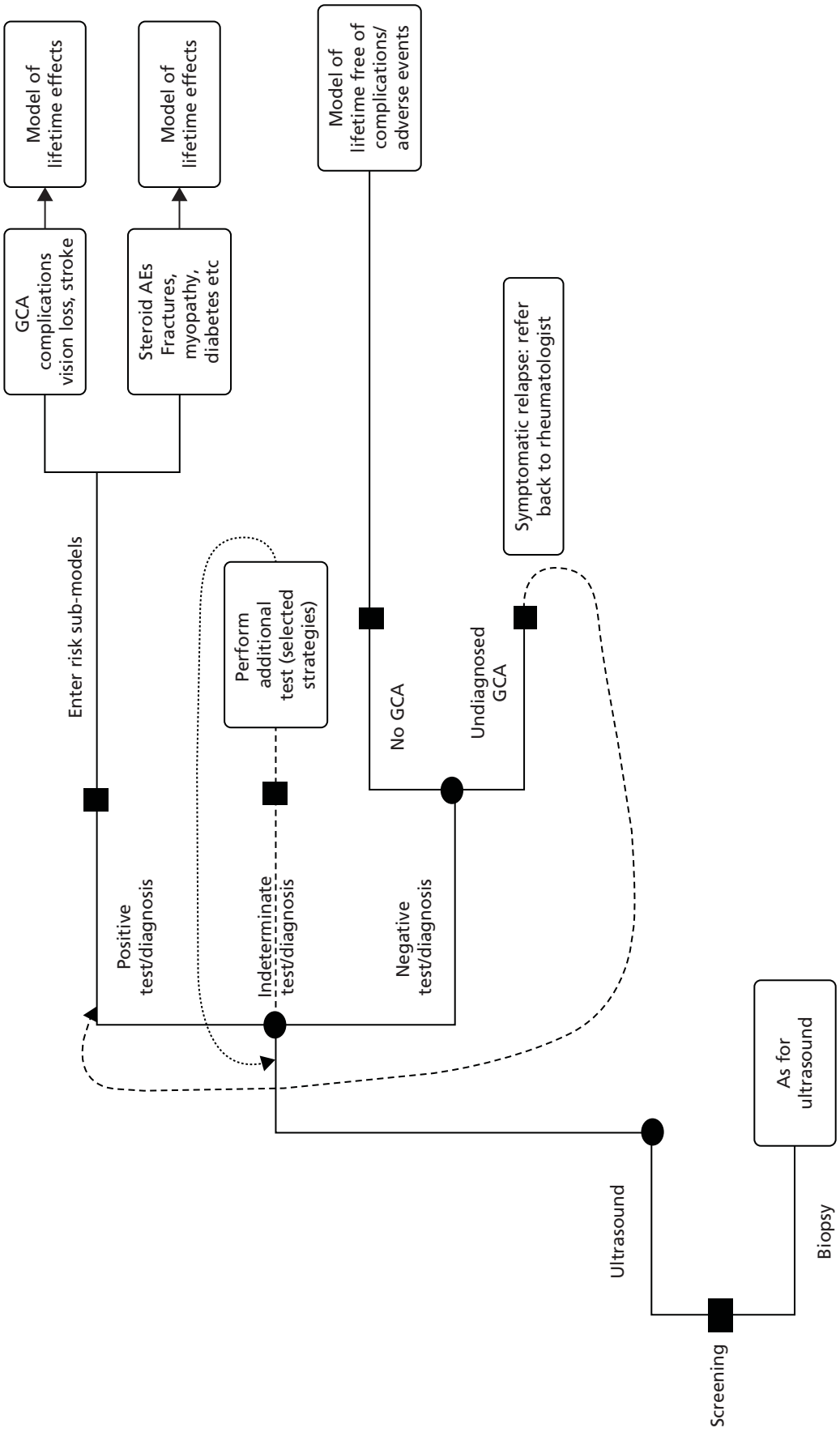


FIGURE 17 Economic evaluation model structure.

TABLE 57 Main sources of evidence for the model

Type of evidence	Source of parameter values/evidence
Accuracy of diagnostic strategies (sensitivity and specificity)	Statistical analyses of TABUL data
Risks of complications of GCA	Review articles and guidelines
	NICE Clinical Knowledge Summary[81] for GCA and key cited articles, other economic/modelling studies
Risks of AEs with steroids	Review articles and guidelines on use of steroids and key cited articles, citation searches
Costs and quality-of-life impact of GCA-related complications	Various sources
Costs and quality-of-life impact of steroid-related complications	Advice from a technology-assessment team that was reviewing the evidence for a NICE report which has now been published[82]
Cost of biopsy and US	NHS Reference Costs[83]
Steroid dosing schedule for GCA	Clinical advice and analysis of TABUL data
Cost of steroids	*British National Formulary*[84]

US, ultrasound.

TABLE 58 Types of diagnostic strategy

Type	Example
Diagnostic tests alone	Biopsy
Diagnostic tests used in conjunction with clinical judgement	Biopsy and clinical judgement
Sequential diagnostic strategies	Assume GCA if high risk, otherwise GCA if either US or biopsy positive

US, ultrasound.

Central to the cost-effectiveness of the alternative test strategies are the impacts of missing some true cases of GCA (the 'false negatives') and incorrectly categorising some patients without GCA as having the disease (the 'false positives') and, therefore, receiving unnecessary treatment. These are measured by the sensitivity and specificity of the test strategies. A strategy with high sensitivity will have few false-negative cases and a strategy with high specificity will have few false-positive cases, but, invariably, the threshold chosen will act positively on one at the expense of the other.

The performance (sensitivity and specificity) of the different test strategies was, in most cases, obtained from the data analysed from the TABUL study and reported in *Chapter 5*. The data used to determine a strategy indicated a positive or negative diagnosis of GCA each patient was obtained from the test results for biopsy and ultrasound, the clinical data collected at the baseline and 2-week assessments, and the clinician's assessment of the GCA diagnosis at the 2-week assessment. The performance of the different test strategies was evaluated against the reference diagnosis, as reported in *Chapter 4*. The only exception was for test strategies involving a combination involving ultrasound and clinical judgement.

The sensitivity and specificity of the set of diagnostic strategies within the economic evaluation are shown in *Table 59*. We included strategies specified in the protocol objectives and additional ones with the best performance from those analysed within *Chapter 5*.

TABLE 59 Sensitivity and specificity of alternative diagnostic strategies

Strategy	Sensitivity (%)	Specificity (%)	Having US (%)	Having biopsy (%)
Technology-only strategies				
Biopsy only (all patients)	39	100	0	100
US only (all patients)	54	81	100	0
Biopsy and US (both in all patients)	65	81	100	100
US followed by biopsy if US is negative	65	81	100	57
Technology followed by risk factors				
US and biopsy with additional prognostic baseline factors	67	81	100	57
Biopsy and older age and claudication with 81% specificity	68	81	0	100
Biopsy and older age and claudication with 90% specificity	59	90	0	100
Pre-test probabilities used to filter who needs a test				
Composite pre-test strategy H0M1L1	72	77	77	46
Composite pre-test strategy H0M1L3	72	75	77	46
Composite pre-test strategy H0M5L7	68	77	77	0
Technology and clinical judgement (proportion continue with steroids)				
Two-week decision: biopsy and clinical judgement	91	81	0	100
Two-week decision: US and clinical judgement	89	77	100	0
Two-week decision: biopsy and US and clinical judgement	96	73	100	100

H0, in high-risk group, assume GCA positive (no diagnostic test performed); M1, in medium-risk group, GCA if either ultrasound or biopsy positive; M5, GCA if ultrasound positive; L1, in low-risk group, GCA if either ultrasound or biopsy positive; L3, GCA if either ultrasound positive, ultrasound axillary involvement or TAB positive; L7, GCA if ultrasound positive; US, ultrasound.

Performance of ultrasound plus clinical judgement strategy

For this strategy, an additional source of diagnostic data was required because the design of the TABUL study blinded clinicians to the ultrasound result. Therefore, we were unable to determine their opinion of the diagnosis based on the ultrasound together with clinical judgement. In the study, all patients had both ultrasound and biopsy tests but only the biopsy test result was given to the clinician managing the patient. Decisions about continuing treatment and the clinician's diagnosis were therefore based on the biopsy result and a clinical assessment of the patient after 2 weeks. The ultrasound result was made available only if the clinician intended to rapidly withdraw steroids at 2 weeks based on a negative biopsy and his or her clinical assessment of the patient. The clinician could then change their treatment decision, that is, continue with steroid treatment, and alter their diagnosis after seeing the ultrasound result. TABUL data are therefore available on the treatment decisions made after 2 weeks only on the basis of the biopsy results; it is not known what treatment decisions would have been made if the ultrasound test result, but not the biopsy result, were provided to the clinician. Some assumptions are therefore required about what diagnoses and decisions about treatment would have been made. For the purposes of the economic analysis the focus is on the treatment decision to continue or withdraw treatment with high-dose steroids because it is this decision that has implications for the risk of developing GCA complications or steroid-related AEs.

An algorithm was devised that would allow an implied treatment decision to be arrived at by considering how the availability of the ultrasound rather than the biopsy would have influenced clinicians' decision-making. To do this, it is necessary to consider this separately according to what the biopsy and ultrasound test results were; in other words, there are four possible combinations of biopsy and ultrasound test results (both positive, both negative, only biopsy positive and only ultrasound positive).

A summary of the reasoning and inferred steroid treatment decision for each of the four combinations is shown in *Table 60*.

For the case in which the biopsy is negative and the ultrasound is positive, two scenarios are described. For scenario 2 (cases for which the ultrasound result was unblinded), for consistency with the TABUL study, we allowed the ultrasound result to be over-ruled by clinical judgement, which was the case for five patients.

For the final combination, a positive biopsy and a negative ultrasound, it is not possible to infer what the treatment decision would be; therefore, in the case of these 27 patients, an alternative approach based on clinical vignettes was used to elicit the treatment decisions that would have been made.

TABLE 60 Inferred outcomes for the ultrasound plus judgement strategy according to biopsy result and ultrasound result

Actual test results for US and biopsy	Information available to the clinician and rationale for inferring their decision if the US result available and biopsy result blinded	Implied treatment decision for the US plus judgement strategy
Biopsy positive AND US positive	It is assumed that treatment would be continued in all cases with a positive US result. Similarly, a positive biopsy would almost certainly result in a clinical diagnosis of GCA and continuation of treatment. In TABUL this was the case for all positive biopsies so it is assumed that the decision would be the same	Same as actual treatment decision with biopsy, that is, continue steroid treatment
Biopsy negative AND US negative	The clinical diagnosis and treatment decision relies on other factors, for example signs, symptoms, blood tests, response to treatment, in the presence of a negative test result. It is assumed that the same decision would have been reached regardless of which negative test result was provided to the clinician	Same as actual treatment decision with biopsy (either continue with or withdraw steroids)
Biopsy negative AND US positive	Scenario 1: no unblinding of the actual US result happened in the study. In this situation the clinician's decision in the study was to continue with steroid treatment because other factors such as patient symptoms strongly suggested GCA. It is assumed that a positive US result would have supported this and so would not have altered this decision	Same as actual treatment decision with biopsy, that is, continue steroid treatment
	Scenario 2: the clinician planned to withdraw steroids in the study because neither the biopsy nor signs and symptoms suggested GCA, so the US result was unblinded. In this situation the actual decision of the clinician after unblinding the US result is known and has effectively taken the biopsy, US and patient symptoms, etc. into account. For our implied treatment decision (for which the clinician would not know the biopsy result) it is assumed that knowledge of the negative biopsy result in the study did not ultimately have any influence on the decision made in the light of the US result and patient symptoms, etc.	Same as actual treatment decision from TABUL following unblinding of US (either continue with or withdraw steroids)
Biopsy positive AND US negative	In the study the decision was to continue treatment for all these patients. However, it is not possible to know if the same decision would have been made on the basis of other factors (symptoms, etc.) alone. Therefore, it cannot be assumed that the decision would be the same based on the US result and clinical judgement	Cannot be inferred from study data and outcomes. Need to obtain decisions about treatment from a separate clinical vignettes exercise

All of these 27 patients were included in the clinical vignette exercise as part of the original random sample (as reported in *Chapter 6*) or in an additional sample for the economic analysis. The panel members rating the vignettes reported their assessment of the diagnosis (definite, probable, possible or not GCA) and the appropriateness of continuing treatment with high-dose steroids (on a scale from 1 = extremely inappropriate to 9 = extremely appropriate) using data collected at presentation and at 2 weeks plus the result of the ultrasound. The economic analysis used the available results from the clinical vignettes, from the first 12 clinicians who completed the exercise.

To dichotomise the continuation of high-dose steroids ratings into a yes/no outcome, a score of 5 or higher was used to indicate a decision to continue treatment. Scores of 4 or lower would indicate a decision not to continue. This threshold resulted in 63% of vignettes being categorised as 'possible GCA' by panel members falling into the 'continue treatment' group. Alternative thresholds of 3, 4 or 6 would have resulted in 100%, 87% and 18% of 'possible GCA' vignettes being categorised as 'continue treatment', respectively.

A simulation was then run to model the diagnosis and treatment decisions if treatment decision had been made by a single clinician for each patient, as was the case in the TABUL study. Decisions were randomly sampled using the ratings from all 12 clinicians on the panel. The simulation was repeated for each vignette 100 times in order to give equal weight to the ratings from all clinicians. By comparing the sampled results with the reference standard, the expected (average) numbers of true positives and false negatives were obtained; there were no false positives or true negatives because all 27 were biopsy positive.

Application of the simulated results from the vignettes to the test strategy that combined ultrasound with clinical judgement produced a sensitivity of 89.1% and a specificity of 76.6% (see *Table 59*). These figures were slightly lower than the equivalent figures for the strategy involving biopsy and clinical judgement.

Risks of complications of giant cell arteritis

Visual complications

Visual complications represent the greatest burden of complications of GCA, with about 25% of cases resulting in sight loss if left untreated.[85] The major presenting symptoms are amaurosis fugax (a transient shade, dimming, fogging, blurring or monocular blindness), transient diplopia (double vision) or unilateral or bilateral partial or complete vision loss.

For the economic model, we needed to identify the risk of onset of visual complications after patients had presented to their GP, because an estimated 92% of visual complications arise prior to the initiation of high-dose steroid treatment and therefore would not be affected by the diagnostic strategies considered in TABUL. To do this, we created a submodel of visual complications, combining and modelling data from various sources, as shown in *Figure 18*.

Blindness in both eyes is rare in GCA[86] because steroid treatment is usually started when sight loss occurs in one eye and should reduce the risk of sight loss occurring in the other eye. It is therefore assumed that there will be no cases of bilateral sight loss and that steroids will have been started in all cases of unilateral sight loss. The stages during the diagnostic and treatment pathway during which visual loss arises are illustrated in *Figure 17*, based on 30% of patients experiencing visual complications,[87] 15% experiencing permanent visual loss[87] and 92% of visual complications arising before treatment as initiated.[88] Of this 92%, one-fifth of complications are assumed to be attributable to an initial false-negative diagnosis. Eight per cent are estimated to arise in true positives after steroid treatment has started. The required estimates of incidence rates of new visual loss among true positives and false negatives are shown by the solid arrows.

In order to assign costs of treatment and the quality-of-life impact of visual complications, we required an assessment of severity, based on a previously reported analysis[89] (*Table 61*).

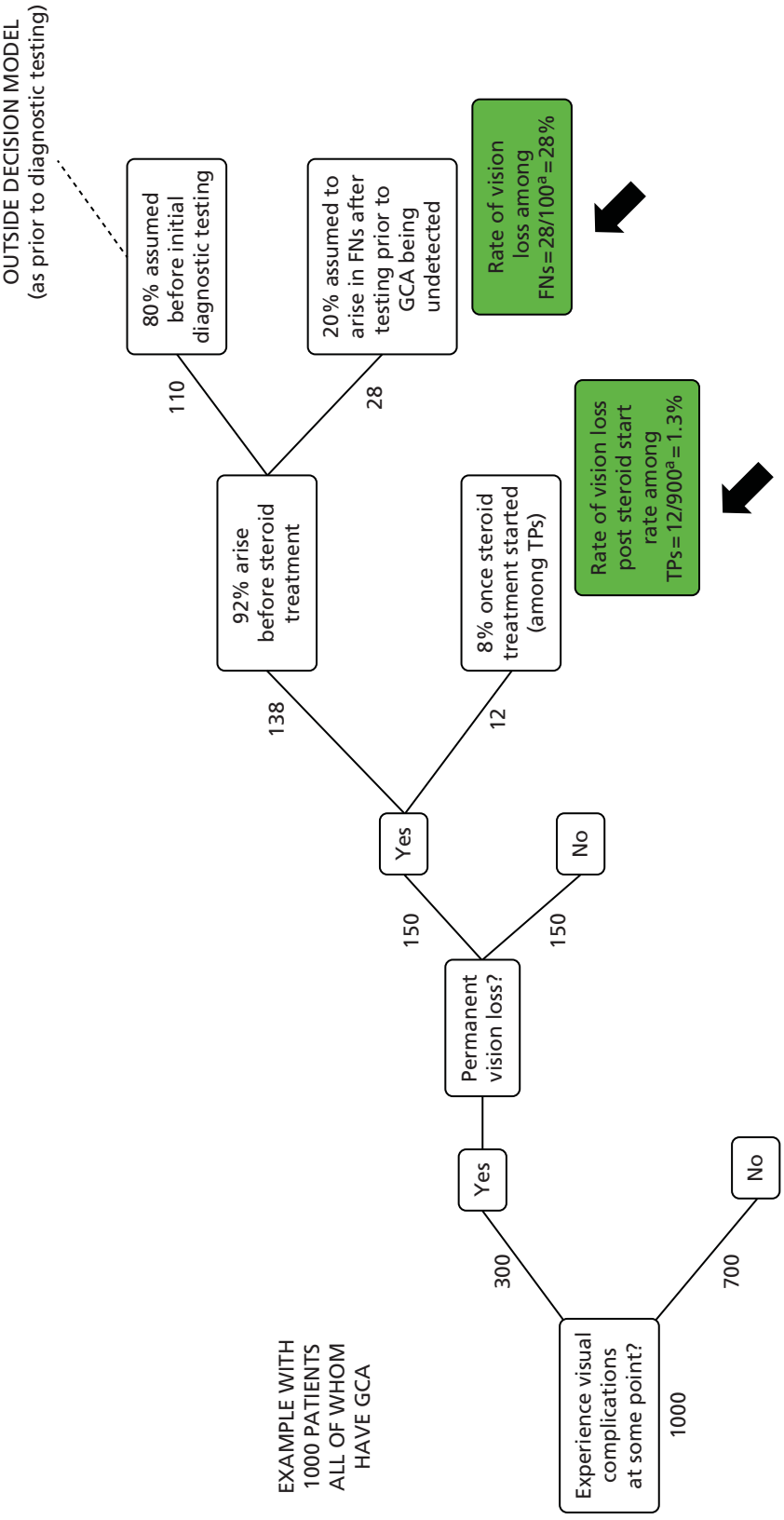


FIGURE 18 Logic modelling of evidence to obtain incidence rates of new onset of visual complications. a, Relative proportion of TPs : FNs assumed to be 90 : 10 after initial diagnostic test; so, of the 1000 GCA cases, 900 would be TPs and 100 would be FNs. FN, false negative; TP, true positive.

TABLE 61 Analysis of the severity of visual loss by initial visual acuity in one eye

Visual acuity	Oral therapy or intravenous therapy, *n* (%)
20/50–20/70	12 (13)
20/80–20/100	4 (4)
20/200–20/400	5 (6)
Counting fingers	19 (21)
Hand motion	15 (17)
Light perception	13 (15)
No light perception	21 (24)

Rates calculated from data in Hayreh *et al.*[89]

Although visual acuity is the primary criterion for determining vision loss, other types of vision loss (e.g. peripheral vision loss or contrast sensitivity loss) are recognised as disabilities even if central visual acuity is 20/20. Partial sight loss in the centre of vision is different to partial sight loss in the periphery, but we have no information on the nature of GCA-induced visual loss.

Stroke

For the incidence of GCA-related stroke, the models assume that 2.64% of cases of GCA result in a stroke, as per Amiri *et al.*,[90] and further assume that strokes arise after presentation to the patient's GP. It is also assumed that stroke occurring as a result of GCA has the same severity and likelihood of fatality as stroke unrelated to GCA. Sixty per cent of strokes were assumed to be minor; case fatality in major strokes was assumed to be 50%.

Mortality from giant cell arteritis

There have been numerous studies reporting an increased risk of mortality in the years following a diagnosis of GCA. However, we decided that it was not necessary to include this in the model because there is no evidence to suggest that a delay in the diagnosis of several weeks (as a result of an initial false-negative test result) has an impact on this mortality risk. Hence, it is unlikely to have any impact on the relative cost-effectiveness of different test strategies.

Use of steroids and risk of complications

Oral corticosteroids have potent systemic effects, including numerous side effects. Evidence on complications arising from treatment with steroids is based on studies relating to oral corticosteroids; almost all patients in TABUL were treated with oral high-dose glucocorticoid therapy. The dose schedule for individuals with GCA is shown in *Table 62*. The second column describes the typical dose schedule for a true positive, that is, a patient with GCA with ongoing treatment. The third column describes a shorter duration of therapy for false-negative cases; this was adopted on the basis that steroid doses are likely to be tapered more quickly in the absence of ongoing features of the disease. The data from the TABUL study placed some doubt on this assumption; therefore, we performed a sensitivity analysis to include a dose schedule for false positives that was the same as that for true positives.

The list of all possible side-effects of steroids is long, but they vary in severity and burden to the patient and the NHS. Even treatment with low-dose steroids is associated with weight gain, hyperglycaemia, diabetes mellitus, increased blood pressure and hypertension, decreased bone mineral density with increased risk of fracture, cognitive dysfunction, increased risk of infection and cataracts.[91] The economic analysis focused on those AEs that were reported to have a high-cost impact or a detrimental effect on quality of life and that were clearly attributable to the use of steroids (as opposed to possibly arising, at least in part, as a result of

TABLE 62 High-dose oral glucocorticoid regimen typically used for treating GCA, with tapering over time

Month	Dose (mg): true positives	Dose (mg): false positives	Month	Dose (mg): true positives	Dose (mg): false positives
1	60	60	13	11.5	3
2	52.5	53	14	10.5	2
3	45	44	15	9.5	1
4	37.5	36	16	8.5	0
5	30	28	17	7.5	0
6	22.5	19	18	6.5	0
7	17.5	13	19	5.5	0
8	16.5	10	20	4.5	0
9	15.5	9	21	3.5	0
10	14.5	7	22	2.5	0
11	13.5	5	23	1.5	0
12	12.5	4	24	0	0

having GCA). The AEs included in the model were fractures, diabetes mellitus and hyperglycaemia, symptomatic steroid myopathy and steroid psychosis. Hypertension was not included because data from the TABUL study showed little change in the use of antihypertensive medication. As rates of AEs in TABUL were only for a 6-month period, we sought evidence from other studies for the rates to be used in the economic model.

Fractures

The model includes vertebral body compression fractures, fractures of the hip/femoral neck, wrist/forearm and proximal humerus (shoulder). The approach to modelling incidence of fractures in a GCA cohort is to start with risks in the general population, then to apply uplift (hazard ratio) for the impact of steroid treatment, and then to apply a relative risk for the effect of bone-protection therapy (*Table 63*).

The model used the fracture risks per annum shown in *Table 63*. These are specific to the 70–74 years age group of the general population,[92] the average age in TABUL being 71 years, and are prior to adjustment for the effect of steroids.

We also obtained the hazard ratios for the increased risks because of the use of steroids with a dose exceeding 7.5 mg daily from the same source.[92] These are 5.2 for vertebral fracture, 2.35 for hip fracture and 1.79 for osteoporotic fracture, which we used for fractures of the wrist/forearm and humerus. Although uncertain, the evidence and clinical opinion suggest that the excess risk of fractures disappears within 1 year of stopping steroid therapy.

TABLE 63 Fracture risks per annum in the general population

Sex	Vertebral fracture	Hip fracture	Wrist fracture	Proximal humerus
Men	0.299	0.213	0.161	0.120
Women	0.533	0.379	0.699	0.246
Average	0.416	0.296	0.430	0.183

Prevention of fractures

We assumed that all patients treated with high-dose steroids were classed as being at high risk of fractures and so received bone protection therapy. There are various therapies available but, for simplicity, we assumed that treatment was with a combination of a bisphosphonate, vitamin D and calcium, the standard dose and costs[84] for which are shown in *Table 64*. We assumed that the relative risks for fracture following bone-protection therapy were 0.57 for vertebral fractures and 0.61 for fractures of the hip, forearm or humerus.[92]

Diabetes mellitus and hyperglycaemia

In Niederkohr and Levin[78] the combined overall incidence of hyperglycaemia and diabetes mellitus was 4.8%, the majority of which was likely to be hyperglycaemia below the threshold for diabetes mellitus. Duru *et al.*[93] reported the incidence of diabetes mellitus alone to be in the range 0–3%. For the model, we used 1.5% as an estimate of the incidence of GCA-related diabetes mellitus. It was assumed that 80% of these cases might be reversible (i.e temporary hyperglycaemia). It was assumed that episodes of temporarily raised glucose would not be given a permanent label of diabetes mellitus (such a label would result in a significant burden to the individual and resource use). For the remaining 20% of patients, in whom it was assumed the incident diabetes mellitus was permanent, a proportion of these were likely to have had non-diabetic hyperglycaemia before starting steroid treatment for suspected GCA. The impact of starting steroids meant that the diagnosis of diabetes mellitus may have emerged earlier than it would otherwise have done, that is, these patients would have eventually developed diabetes mellitus at some point in the future regardless of their steroid therapy. Although it is therefore difficult to attribute a proportion of the burden of such accelerated diagnoses to the use of steroids, we judged that it would be reasonable to assume that the costs of managing diabetes mellitus would be incurred 5 years earlier than they would otherwise have been without steroid treatment, that is, the impact of steroids accelerates the occurrence of diabetes mellitus by 5 years.

Other adverse events

We assumed that the annual incidence of symptomatic steroid myopathy was 3.4% and the annual incidence of steroid psychosis was 7.6% based on a GCA study by Niederkohr and Levin.[78] For the many other common and mild AEs, for example moon face (round, puffy-shaped swollen face), there is likely to be a very small cost burden to the NHS. However, collectively there is a significant impact on quality of life; therefore, an overall adjustment to quality of life was applied (see *Chapter 7*, *Health utilities*). For the impact of diabetes mellitus on utility, based on Brown *et al.*,[94] we assumed a multiplier of 0.88, which leads to a decrement in quality of life because of diabetes mellitus of 0.09. This is assumed to persist indefinitely because diabetes mellitus is a progressive condition and individuals with a longer duration of diagnosed diabetes mellitus can be expected to have a greater prevalence of complications and associated loss of quality of life.

TABLE 64 Bone protection therapy

Medication	Dose	Dose cost
Sodium alendronate (non-proprietary) alone[a]	10 mg daily	28-tablet pack = £1.64
Vitamin D (cholecalciferol) with calcium carbonate	10 µg per day of cholecalciferol	Accrete D3® net price 60-tablet pack (10 µg) = £2.95
Risedronate sodium, calcium carbonate and cholecalciferol (Actonel® Combi, Warner Chilcott)	Weekly cycle of 1 Actonel Once a Week® (Risedronate sodium) tablet on the first day followed by one calcium and cholecalciferol sachet daily for 6 days	24-sachet plus four-tablet pack = £19.12

a We assumed that GPs would prescribe alendronate on its own rather than in combination with colecalciferol. This is because it would be much cheaper to prescribe alendronate and to separately prescribe vitamin D and calcium because they are much cheaper. Bisphosphonates have been estimated to reduce the risk of fractures by 43%.[92]

Unit costs of tests, medications and treatments

The evidence sources for the unit costs are described below. All costs are then adjusted for inflation to bring them to 2014/15 levels.

Biopsy and ultrasound

Biopsy is estimated to cost £493 based on NHS Reference Costs for 2011/12[83] (for lymph node biopsy/salivary gland biopsy). This is assumed to include theatre cost, surgeon time, pathologist time, sample processing, camera, microscope and other pathology equipment and administration cost. It has been pointed out that some 'biopsy costs' shown in NHS Reference Costs[83] may be understated, as they include relatively minor procedures such as the removal of warts. However, we used a specific procedure code, lymph node biopsy/salivary gland biopsy, which we expect to be robust in this case.

In the TABUL study, the typical time taken to perform ultrasound of both temporal and axillary arteries was 30 minutes, although there was considerable variation (scans took between 20 and 60 minutes, depending on the experience of the sonographer and the extent of the abnormalities to be defined). The cost of a 'direct access' (as opposed to outpatient) ultrasound scan taking 20 minutes or more is £57 based on the NHS Reference Costs for 2013/14.[95] This is assumed to include equipment cost, equipment maintenance and calibration, sonographer time, radiology space/room cost, radiologist interpretation cost, administration cost and a contribution for hospital overheads. Training costs for a hospital to set up a new GCA sonography service are classed as 'implementation costs' so, in line with NICE convention, they are excluded from the cost-effectiveness analysis. With uplifts for inflation, the costs for biopsy and ultrasound are £514 and £58, respectively.

Giant cell arteritis-related complications

The costs of vision loss shown in *Table 65* are applied to the visual acuity states in the model that are worse than 6/60 m (20/200 feet), that is, those meeting the legal definition of blindness, in line with the ranibizumab and pegaptanib sodium HTA assessment.[96]

The costs of registration of blindness, provision of low-vision aids and low-vision rehabilitation are one-off rather than recurrent costs. Community care costs were estimated as the annual cost for a local authority home care worker and residential care costs were based on the annual cost of private residential care (taking into account that approximately 30% of residents pay themselves). Using the estimated annual costs in *Table 65* gives a cost of £5090 for the first year of blindness and £4903 for each subsequent year.

The 5-year cost of a non-fatal stroke was estimated to be £29,400 in a NICE report.[97]

TABLE 65 Costs of vision loss below best corrected visual acuity of 6/60 in the better-seeing eye

Service	Receiving services (%)	Unit cost (£)	Annual cost (£)
Blind registration	95	115	109
Low-vision aids	33	150	50
Low-vision rehabilitation	11	259	28
Community care	6	6,552	393
Residential care	30	13,577	4073
Depression	39	431	168
Hip replacement	5	5379	269

Adapted from *Table 42* in Colquitt *et al.*[96] This monograph may be freely reproduced for the purposes of private research and study and may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising.

Steroid-related adverse events

The unit costs of AEs were obtained from published studies[98–101] and are shown in *Table 66*.

Inflation

All unit costs were inflated to 2014/15 values using the Hospital and Community Health Services index.[102]

Health utilities

Utilities are valuations of health-related quality-of-life on a scale from 0 to 1, with 0 being equivalent to dead and 1 being equivalent to perfect health. A loss of quality-of-life attributable to a complication such as vision loss or an AE such as a fracture is called a utility decrement. The utility decrements used in the model are shown in *Table 67.* The baseline utility for someone of 71 years of age is 0.716, based on an age-related annual decrease in utility of 0.004.[107]

For visual loss, we obtained the required utility decrement by combining data on substates of visual loss. The quality of life of various visual states was studied in Brown *et al.*,[94] showing a wide range of utilities associated with different levels of vision within the range of legal blindness (visual acuity < 20/200). We used the reported time trade-off (TTO) values rather than standard gamble (*Table 68*), as these are consistent with the EQ-5D quality-of-life instrument preferred by NICE. Multiplying these TTO values by the proportional occurrence of visual loss by severity in *Table 61* gives a weighted value of 0.524 (on a scale of 0 to 1). As the TTO values are on a scale of 0 to 1, this was used as a multiplier to the age-specific utility in the model, giving an overall utility value for vision loss of 0.375, which represents a decrement of 0.34 compared with the baseline utility of 0.716.

Utilities associated with vision loss tend to be higher after the first year, which we speculate is because of a degree of adjustment made to the condition.

Model time horizon

Cost-effectiveness analyses need to capture all significant costs and utility effects that are relevant to the intervention and condition of interest. As steroid treatment causes fractures and diabetes mellitus in a small minority of patients, and because these have lifetime cost and/or quality-of-life impacts, it is necessary for the model to take a long-term perspective. The model horizon is, therefore, 40 years, which is effectively a lifetime perspective for a cohort with a baseline age of 71 years.

Mortality

As the model has a lifetime perspective, it is necessary to include both the mortality rate for the general population and any excess mortality arising from GCA or steroid-related side effects. General population mortality rates were obtained from standard Office for National Statistics tables.[108] Stroke mortality is modelled explicitly. For fractures, excess mortality was applied when vertebral or hip fractures occurred, leading to an absolute estimated 1-year mortality of 4.4% and 6.0%, respectively (estimates for patients aged 71 years, rates derived from van Staa *et al.*[109]).

TABLE 66 Unit costs of steroid-related AEs

Event type	Cost (£)	Source
Vertebral body compression fracture	1152	Gutiérrez *et al.*[98]
Hip fracture	4222	Gutiérrez *et al.*[99]
Forearm fracture	690	Gutiérrez *et al.*[98]
Proximal humerus fracture	690	Gutiérrez *et al.*[98]
Symptomatic steroid myopathy	2079	[a]Bernatsky *et al.*[100]
Diabetes mellitus	2520	Manson *et al.*[101]

a CAD 4099 converted to GBP using 2008 conversion rate of 1.9717.

TABLE 67 Utility decrement values for complications and AEs

Health state	Multiplier (when applicable)	Utility value	Utility decrement versus baseline value	Source
Baseline utility		0.716		See *Health Utilities*
Major stroke		0.260	–0.46	Post *et al.*[103]
Minor stroke		0.550	–0.17	Post *et al.*[103]
Vision loss	0.524	0.375	–0.34	See *Health Utilities*
Vertebral body compression fracture				
Year 1	0.570	0.408	–0.31	ScHARR[104]
≥ Year 2	0.660	0.473	–0.24	
Hip fracture				
Year 1	0.690	0.494	–0.22	ScHARR[104]
≥ Year 2	0.850	0.609	–0.11	
Proximal humerus fracture				
Year 1	0.860	0.616	–0.10	ScHARR[104]
≥ Year 2	1.000	0.716	0.00	
Forearm fracture				
Year 1	0.880	0.630	–0.09	ScHARR[104]
≥ Year 2	0.980	0.702	–0.01	
Diabetes mellitus	0.880	0.630	–0.09	Brown *et al.*[105]
Symptomatic steroid myopathy		0.707	–0.01	Roberts *et al.*[106]
Steroid-induced psychosis		0.665	–0.05	Roberts *et al.*[106]
General decrement for steroid users			–0.03	Niederkohr and Levin[78]

ScHARR, School of Health and Related Research.
The utility decrement for weight gain is included within the general utility decrement from general symptoms of GCA (not AEs).

TABLE 68 Utility values of alternative visual states

Visual state	Mean utility TTO method[94] except when specified[a]	Comments
20/50–20/70	0.88	Assumed equally spaced between perfect health and 20/200
20/80–20/100	0.77	
20/200–20/400	0.65	
Light perception to counting fingers[b]	0.47	
No light perception in one eye	0.37	Assumed in between LP and NLP each eye in Brown *et al.*[94]
No light perception in each eye	0.26	

a TTO (on a scale of 0 to 1, with 1 being perfect vision).
b The acuity state 'Hand Motion' falls within this category.

Discount rates and perspective

Discount rates of 3.5% per annum are applied for both costs and health benefits as measured in quality-adjusted life-years (QALYs) in line with NICE guidance.[110] Discounting is undertaken to ensure that both the overall costs and overall benefits are reported in comparable terms, in their present value. A sensitivity analysis is undertaken with alternative rates of 0% for benefits (QALYs), as long-term benefits are heavily discounted when a rate of 3.5% is applied. In line with NICE guidance, the model takes a health and social care perspective. Wider societal impacts, such as time off work and private care home costs, are excluded (except for specific sensitivity analyses).

Sensitivity analysis

The values described so far for the main analysis are referred to as the 'base-case' values. However, model parameters have some uncertainty around their 'true' value, either because of sample sizes (as evidenced by reported 95% CIs) or because there are multiple heterogeneous studies from which is it difficult to obtain an unequivocal single 'best estimate'. It is therefore standard practice to carry out sensitivity analyses. Here the term 'sensitivity' refers to how much the economic outcomes change according to changes in model parameters from their 'base-case' values.

Sensitivity analyses were undertaken for the following strategies:

- biopsy alone
- ultrasound alone
- biopsy in combination with clinical judgement (current routine care)
- ultrasound in combination with clinical judgement.

Uncertainty around the various parameters works both ways, so, for example, if the base-case estimate of the cost of ultrasound is £57, we could test out what happens if the cost were 20% higher or 20% lower. Given that initial analyses indicated that ultrasound is likely to be more cost-effective than biopsy, values for the sensitivity analyses have been chosen, as shown in *Table 69,* in the direction which is likely to make the cost-effectiveness of ultrasound and biopsy closer than in the base case.

Alternative reference diagnosis of giant cell arteritis

For GCA there is currently a lack of a universally accepted reference or gold standard definition for the diagnosis of GCA. As a result, the performance (sensitivity and specificity) of each test or composite screening strategy is inevitably influenced by the choice of reference standard. In TABUL, clinical judgement played a major part in the reference standard, as well as the biopsy and ultrasound results. However, there are alternative, more narrowly defined, reference standards that could be used for the purpose of sensitivity analysis, such as the ACR criteria or combinations of the tests and ACR criteria/risk factors. The concern is that if we vary the reference standard diagnosis, this will influence the relative cost-effectiveness of the potential screening strategies.

We have therefore tested the impact of three alternative reference standards, which are defined such that there are fewer 'true' GCA cases (*Table 70*). This is an exploratory analysis and the alternative reference standards are merely to explore whether or not fewer true GCA cases might alter the base-case conclusions and, having not been comprehensively evaluated, do not purport to have applicability to clinical practice.

The outcomes of the following subset of strategies were compared against the alternative reference standard diagnoses:

- biopsy alone: as per protocol
- ultrasound alone: as per protocol
- a composite strategy (H0M5L7) in which high-risk cases are treated as GCA and others are treated as GCA only if the ultrasound is positive
- biopsy and clinical judgement.

TABLE 69 Sensitivity analyses to be undertaken

Number	Parameter	Values
Risks of GCA-related complications		
1	Baseline risk of GCA-related complications: reduce the risks to reflect introduction of a fast-track pathway across the UK The fast-track pathway in Patil *et al.*[80] involved raising awareness of the fast-track pathway in general practice (including publicity to patients) and providing training to GPs to enable them to spot the symptoms of GCA, with reminders every 3 months. Patients with features of GCA and ischaemic symptoms were referred to A&E for assessment, receiving advice from both ophthalmology and rheumatology specialties There was an overall reduction in inpatient costs and cost of re-admissions, the savings being partially offset by the training costs. There was little or no difference in costs of medication, GP appointments, investigations or outpatient appointments[111] The UK Department of Health working group is now evaluating 'rollout' of a fast-track pathway across the UK[112]	Apply a hazard ratio of 0.41 for GCA-related events vs. conventional pathway (0.41 = 9%/22% as per fast-track study)[80]
2	Ratio needed for the calculation of rates of new permanent visual loss post presentation: ratio of true positives to false negatives over past few decades (using biopsy and clinical judgement). TABUL suggests that a ratio of 90 : 10 may be the most up-to-date estimate	Historically this may have been lower, at around 80/20, allowing for more recent improved recognition of signs and symptoms of GCA
3	Split of cases of visual loss that arise before treatment into those arising before presentation vs. cases in false negatives	70/30 (i.e. 7 times more before presentation; assumed to be 80/20 in base case)
Test performance/costs		
4	Higher test sensitivity for biopsy and clinical judgement than suggested by mean in TABUL	94% (the upper 95% CI) vs. 91% base case
5	Higher test specificity for biopsy and clinical judgement than suggested by mean in TABUL	88% (the upper 95% CI) vs. 81% base case
6	Cost of US: £57 in base case per NHS Reference Costs[95]	£144 per TABUL reimbursement costing
Risks of steroid-related AEs		
7	Persistence of raised fracture risk after cessation of steroids: duration over which the risk gradually tapers off from the level at steroid cessation to zero	3 years after cessation (assumed to be 1 year in base case)
8	False negatives: shorter time to detection of GCA following tapering off steroid treatment after initial test	1 month instead of > 2 months (base case)
9	Longer time to withdrawal of steroids in false positives than in base case	Assume same steroid schedule as for true positives
Cost and utility/quality of life		
10	Overall cost and quality-of-life burden of AEs attributable to steroids	100% higher than base case
11	Unit cost of vision loss (defined by visual acuity worse than 20/200)	Reduced by 20% (from £5090 to £4072 in year 1)
12	Utility (quality-of-life) multiplier for visual loss (on 0–1 scale, in which 1 = perfect health, 0 = equivalent to death)	Increased from 0.762 to 0.800
13	Alternative discount rate for QALYs	0% for QALYs

TABLE 69 Sensitivity analyses to be undertaken (*continued*)

Number	Parameter	Values
Willingness-to-pay threshold		
14	Base case used £20,000/QALY	£30,000 per QALY
No over-ruling of US result		
15	As described earlier in relation to *Table 60*, in TABUL there were five individuals for whom the US result was over-ridden by clinical judgement A sensitivity analysis will examine the effect of allowing no over-ruling of the US result when calculating the implied treatment decision for ultrasound and judgement	The resulting sensitivity and specificity become 93.9% and 72.6%, respectively (compared with 93.1% and 72.6% for the base case)

US, ultrasound.

TABLE 70 Definitions of the alternative reference standards

Number	Alternative reference standard	Number of GCA cases	Total GCA cases as % of TABUL reference standard
1	As per reference standard diagnosis EXCEPT change to 'NOT GCA' where: • reference diagnosis = GCA positive AND • low risk AND • negative TAB AND • negative US	234	91
2	As per reference standard diagnosis EXCEPT change to 'NOT GCA' where: • reference diagnosis = GCA positive AND • negative TAB AND • negative US AND • either low risk or (medium risk AND no optic neuropathy AND normal temporal artery)	215	84
3	As per reference standard diagnosis EXCEPT change to 'NOT GCA' where: • reference diagnosis = GCA positive AND • negative TAB AND • negative US AND • no tongue/jaw claudication AND • normal temporal artery AND • no optic neuropathy AND • ESR < 80	244	95

Results

In this section, results are presented for the base case, then for the various sensitivity analyses, and the budget impact, all based around the diagnostic reference standard applicable in the TABUL study. We then investigate how varying the reference standard changes the results.

The two main measures of cost-effectiveness are the incremental cost-effectiveness ratio (ICER) and net monetary benefit (NMB). Both of these are all-encompassing measures that trade off additional costs of diagnosis, medication and treatment of complications against benefits in terms of improved life expectancy

and quality of life (e.g. through reduced incidence of blindness or reduced incidence of fractures). Central to these measures is:

1. The QALY; for example, 2 years spent with a utility of 0.6 gives 1.2 QALYs.
2. The value placed on 1 QALY gained [often referred to as the willingness-to-pay (WTP) threshold], which, in the UK, is stated by NICE to be typically in the range £20,000 to £30,000 per QALY. We will use a threshold of £20,000 per QALY for our analysis because this is more usual for groups that are not disadvantaged.

The preferred measure is the ICER. This shows how cost-effective one strategy is compared with another by dividing the incremental costs by the incremental QALYs, but this can become complex to present when there are many strategies. We shall therefore report ICERs to compare a small number of strategies, but we shall use the NMB to compare the cost-effectiveness across all strategies. The NMB is the overall monetary value of a screening/treatment strategy taking account of both costs and health benefits, with the health benefits valued at £20,000 per QALY. The higher the NMB, the more cost-effective a strategy is; this allows easy comparison across multiple strategies.

To calculate the NMB of a strategy, the steps are:

1. Calculate the total costs incurred (including the cost of the tests, medications and treatment of complications and AEs).
2. Calculate the total QALYS over the model time horizon, in this case 40 years.
3. Multiply the total QALYs by the WTP threshold of £20,000 per QALY.
4. Deduct the costs calculated in (1) from the value in (3) to obtain the NMB.

Base-case results

In *Table 71*, results are shown for various alternative diagnostic strategies, all assuming base-case model parameters.

Columns 2 and 3 show the performance of each screening strategy. Columns 4 and 5 show the proportion of patients who would undergo each test. Columns 6–10 are the economic outcomes. Column 8 is the NMB measure of cost-effectiveness. The NMB figures for each strategy appear to be of roughly the same magnitude, and, although this might suggest that they are all almost the same, this would be an incorrect interpretation. The higher the incremental net benefit in column 9 of a given strategy compared with the biopsy-only strategy, the more cost-effective that strategy is. It should be remembered that these monetary differences are per patient. The budgetary impact of selected strategies is explored later. Column 10 shows the ranking of each diagnostic strategy in terms of cost-effectiveness (based on the NMB); the lower the ranking the more cost-effective the strategy. The last two columns show two clinical outcome measures.

It may be easier to understand how the results compare visually on a cost-effectiveness plane, as shown in *Figure 19*. The most cost-effective strategy is indicated by bold font, that is, '2-week decision: ultrasound and judgement'. The green dotted line is known as a cost-effectiveness threshold, and it represents a line along which any point would have the same cost-effectiveness (any point has a cost-effectiveness ratio of £20,000 per QALY relative to this strategy). Any points below the line have a more favourable ratio of additional costs to additional benefits and would be a more cost-effective option (if there were any). Any points above the line are not cost-effective. In the case of the strategy '2-week decision: combined biopsy and ultrasound and judgement' connected by a blue dashed line, the gradient is clearly much steeper than the green dotted line, indicating that the marginally higher QALY gains are not achieved in a cost-effective way. Numerically, the additional 0.0018 (7.6482 – 7.6464) QALYs cost an extra £485 (£1406 – £921), giving an ICER of £271,864, which far exceeds the acceptable threshold of £20,000 per QALY. For all other strategies, both the costs and QALYs are inferior (higher cost and fewer QALYs) compared with the

TABLE 71 Results for alternative screening strategies

Strategy	Sensitivity (%)	Specificity (%)	Having US (%)	Having TAB (%)	Total costs per patient (£)	Total QALYs per patient	NMB at £20,000/QALY per patient (£)	Incremental NMB per patient (£)[a]	Cost-effectiveness rank[b]	New onset irreversible vision loss (%)[c]	Fractures over 2 years among cohort (%)
Technology-only strategies											
Biopsy only (all patients)	39	100	0	100	1965	7.5958	149,950	–	13	0.60	5.17
US only (all patients)	54	81	100	0	1371	7.6036	150,701	751	10	0.51	5.49
Biopsy and US (both in all patients)	65	81	100	100	1757	7.6162	150,567	617	11	0.43	5.49
US followed by biopsy if US negative	65	81	100	57	1538	7.6162	150,786	836	8	0.43	5.49
Technology followed by risk factors											
US and biopsy with additional prognostic baseline factors	67	81	100	57	1512	7.6185	150,857	907	7	0.42	5.49
Biopsy and age and claudication with 81% specificity	68	81	0	100	1664	7.6196	150,728	778	9	0.41	5.49
Biopsy and age and claudication with 90% specificity	59	90	0	100	1757	7.6132	150,507	557	12	0.47	5.34

continued

TABLE 71 Results for alternative screening strategies (*continued*)

Strategy	Sensitivity (%)	Specificity (%)	Having US (%)	Having TAB (%)	Total costs per patient (£)	Total QALYs per patient	NMB at £20,000/QALY per patient (£)	Incremental NMB per patient (£)[a]	Cost-effectiveness rank[b]	New onset irreversible vision loss (%)[c]	Fractures over 2 years among cohort (%)
Pre-test probabilities used to filter who needs a test[d]											
Composite pre-test strategy H0M1L1	72	77	77	46	1389	7.6225	151,060	1110	5	0.38	5.56
Composite pre-test strategy H0M1L3	72	75	77	46	1392	7.6216	151,041	1091	6	0.38	5.59
Composite pre-test strategy H0M5L7	68	77	77	0	1200	7.6179	151,159	1209	4	0.41	5.56
Technology and clinical judgement (proportion continue with steroids)											
Two-week decision: biopsy and judgement	91	81	0	100	1396	7.6459	151,523	1573	3	0.26	5.49
Two-week decision: US and judgement	93	77	100	0	921	7.6464	152,008	2058	1	0.24	5.57
Two-week decision: biopsy and US and judgement	96	73	100	100	1406	7.6482	151,558	1608	2	0.22	5.63

US, ultrasound.
a Versus 'biopsy only' strategy, measured per patient (£).
b Rank in order of cost-effectiveness.
c Defined as < 20/200 and reported as percentage of true GCA cases.
d H0, in high-risk group, assume GCA positive (no diagnostic test performed); M1, in medium risk group, GCA if either ultrasound or biopsy positive; M5, GCA if ultrasound positive; L1, in low risk group, GCA if either ultrasound or biopsy positive; L3, GCA if either ultrasound positive, ultrasound axillary involvement or TAB positive; L7, GCA if ultrasound positive.

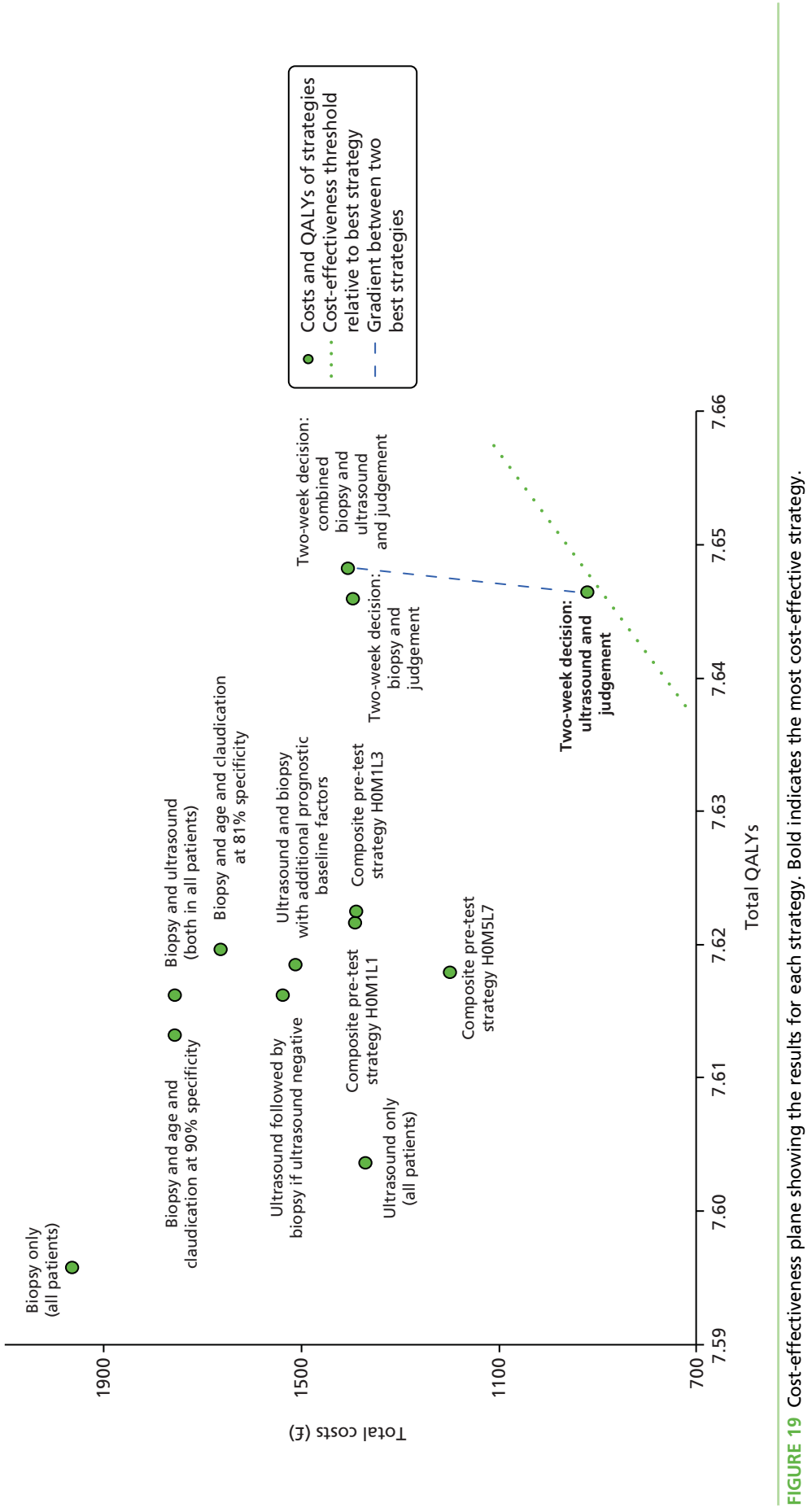


FIGURE 19 Cost-effectiveness plane showing the results for each strategy. Bold indicates the most cost-effective strategy.

optimal 'Two-week decision: ultrasound and judgement' strategy, which is thereby said to dominate these strategies (including 'Biopsy and judgement').

In light of these results, we undertook some further refinement of the ultrasound and clinical judgement strategy as shown in *Table 72.*

The results lead to the following findings:

1. The most cost-effective strategies are those that include an element of clinical judgement.
2. Ultrasound and clinical judgement is the most cost-effective strategy, with the highest incremental NMB. This is largely because of the difference in the cost of the tests (*Table 73*).
3. For the strategy in (2) above, the estimated cost saving is £475 patient and there is a very small QALY gain of 0.0005 compared with biopsy and judgement. Rather than calculating an ICER, ultrasound is said to dominate biopsy in this case as ultrasound results in both cost savings and QALY gains.
4. Ultrasound alone is more cost-effective than biopsy alone.
5. The three sequential diagnostic strategies that incorporate pre-test probabilities (those ranked 4, 5 and 6) offer a level of cost-effectiveness between those involving clinical judgement and those (ranked 7 to 13) that include neither clinical judgement nor pre-test probabilities.

A further finding from the additional analyses in *Table 72* is that the ultrasound and judgement strategy may be improved slightly by undertaking a biopsy in cases in which the pre-test risk is high and the ultrasound and judgement decision would be not to treat. It should be noted that only 2% of individuals in TABUL were referred for biopsy under such a strategy, so there is some uncertainty around the benefit of a biopsy in such circumstances. It would also require the timing of the decision to perform a biopsy to be made after the outcome of the ultrasound plus judgement strategy is known. This is likely to mean that the biopsy is delayed (so may be less accurate than in our model because of the change in histology since patient presentation). Alternatively, an earlier biopsy would be possible if an ultrasound plus judgement outcome was obtained before 2 weeks. However, this would mean that there is less information available to the clinician on the patient's symptoms and response to steroid treatment which, in turn, may lead to a less accurate outcome as a result of a more rapid ultrasound and clinical judgement strategy.

Detailed analysis of results for ultrasound plus judgement versus biopsy plus judgement

It is useful to break down the cost and QALY differences further, as shown in *Table 73*, to understand how they arise. As previously stated, the cost difference is largely because of the difference in cost of the tests. In terms of QALYs, compared with biopsy and judgement, ultrasound plus judgement leads to fewer false negatives and so lower loss of health due to complications of GCA (difference = 0.0023). However, approximately 75% of this QALY gain is offset by loss of health through prescribing steroids to a greater number of false-positive cases (0.0017).

Sensitivity analyses

Table 74 shows the NMB (based on a £20,000/QALY acceptability threshold) under various alternative model assumptions. The results relate to the biopsy plus clinical judgement strategy compared with the ultrasound plus clinical judgement. The base-case difference was £485 in favour of ultrasound plus clinical judgement.

The results from the sensitivity analyses indicate that the improved cost-effectiveness of ultrasound and judgement (compared with biopsy and judgement) is not sensitive to alternative assumptions, with only alternative cost or test sensitivity assumptions reducing the incremental NMB result below £400 (from £485 base-case result). This is because the difference in the cost of the tests, in particular, is a very strong driver of the cost-effectiveness. Even doubling the cost and quality-of-life burden from steroid-related AEs did not change the outcome much.

TABLE 72 Further exploratory analyses of the ultrasound plus judgement strategy

Strategy	Sensitivity (%)	Specificity (%)	Having US (%)	Having TAB (%)	Total costs per patient (£)	Total QALYs per patient	NMB at £20,000/QALY per patient (£)	Cost-effectiveness rank[a]	New onset irreversible vision loss (%)[b]	Fractures over 2 years among cohort (%)
Further exploratory analyses: when US and judgement decision is not GCA refer for biopsy in some cases										
Refer for biopsy if high risk	94	77	100	2	920	7.6478	152,035	1	0.24	5.57
Refer for biopsy if medium or high risk	95	77	100	13	965	7.6487	152,009	2	0.23	5.57

US, ultrasound.
a Rank in descending order of cost-effectiveness.
b Defined as < 20/200 and reported as percentage of true GCA cases.

TABLE 73 Differences in costs and QALYs

Cost or QALY element[a]	Difference[b]
Lower test cost of US	–£456
Lower cost of treating complications of GCA in false negatives	–£27
Higher cost of steroids and treating AEs in (mainly because of difference in false positives)	£8
Total cost difference	–£475
Lower QALY loss from GCA complications in false negatives	0.0023
Greater QALY loss from overtreatment of false positives	–0.0017
Other difference	–0.0001
Total QALY difference	0.0005
Monetary value of QALY difference at £20,000 per QALY	£10
Incremental NMB (–475 to 10)[c]	–£485

US, ultrasound.
a Lower or higher stated for ultrasound relative to biopsy.
b Ultrasound minus biopsy.
c The £485 reconciles to the difference between the incremental NMB figures of £1680 and £2173 in *Table 71*.

TABLE 74 Results from sensitivity analyses

Number	Parameter	Details	Biopsy and clinical judgement (£)	US and clinical judgement (£)	Difference (US minus biopsy) (£)
Base case (for reference)			151,523	152,008	485
Risks of GCA-related complications					
1	Baseline risk of GCA-related complications. Reduce the risks to reflect introduction of a fast-track pathway across the UK. The UK Department of Health working group is now evaluating 'rollout' of a fast-track pathway across the UK[112]	Apply a hazard ratio of 0.41 for GCA-related events vs. conventional pathway (0.41 = 9%/22% as per fast-track led by Patil *et al.*[80])	152,179	152,625	446
2	Ratio needed for the calculation of rates of new permanent visual loss post presentation: ratio of true positives to false negatives over past few decades (using biopsy and clinical judgement). TABUL suggests that a ratio of 90 : 10 may be the most up-to-date estimate (see *Figure 17*)	Historically this may have been lower, around 80 : 20, allowing for more recent improved recognition of signs and symptoms of GCA	151,545	152,001	456
3	Split of cases of visual loss that arise before treatment into those arising before presentation vs. cases in false negatives	70 : 30 (i.e. 7 times more before presentation; assumed to be 80 : 20 in base case)	151,487	151,980	493

TABLE 74 Results from sensitivity analyses (*continued*)

Number	Parameter	Details	Biopsy and clinical judgement (£)	US and clinical judgement (£)	Difference (US minus biopsy) (£)
Test performance/costs					
4	Higher test sensitivity for biopsy and clinical judgement than suggested by mean in TABUL	94% (the upper 95% CI) vs. 91% base case	151,626	152,008	381
5	Higher test specificity for biopsy and clinical judgement than suggested by mean in TABUL	88% (the upper 95% CI) vs. 81% base case	151,592	152,008	415
6	Cost of US: £57 in base case per NHS Reference Costs[95]	£144 per TABUL reimbursement costing	151,523	151,919	397
Risks of steroid-related AEs					
7	Persistence of raised fracture risk after cessation of steroids: duration over which the risk gradually tapers off from the level at steroid cessation to zero. In base case, assuming the risk tapers off to zero after 1 year	Assume risk gradually tails off from the level at steroid cessation to zero: 3 years after cessation (assumed to be 1 year in base case)	150,595	151,067	472
8	False negatives: shorter time to detection of GCA following tapering off steroid treatment after initial test	1 month instead of > 2 months (base case)	151,626	152,087	461
9	Longer time to withdrawal of steroids in false positives than in the base case	Assume same steroid schedule as for true positives	151,471	151,944	473
Cost and utility/quality of life					
10	Overall cost and quality-of-life burden of AEs attributable to steroids	100% higher than base case	149,568	150,026	458
11	Unit cost of vision loss (defined by visual acuity worse than 20/200)	Reduced by 20% (from 5090 to 4072 in year 1)	151,611	152,092	480
12	Utility (quality-of-life) multiplier for visual loss (on scale 0 to 1, where 1 = perfect health, 0 = equivalent to death)	Increased from 0.764 to 0.800	151,885	152,350	466
13	Alternative discount rate for QALYs	0% for QALYs	205,407	205,896	489
Willingness-to-pay threshold					
14	Base case used £20,000/QALY	£30,000 per QALY	228,011	228,463	452
No over-ruling of US result					
15	As described in relation to *Table 60*, there were five patients whose US results were over-ridden by clinical judgement A sensitivity analysis will examine the effect of allowing no over-ruling of the US result when calculating the implied treatment decision for ultrasound and judgement	Sensitivity and specificity become 94% and 73%, respectively (compared with 93% and 73% for the base case)	151,523	151,995	472

US, ultrasound.

Results based around an alternative reference standard

All of the results presented so far have been based on the reference diagnosis defined for the TABUL study. In this section, we show the impact of alternative reference diagnoses that involve fewer true GCA cases by removing some cases that rely solely on clinical judgement.

The results and their interpretation are best shown graphically (*Figure 20*). The *x*-axis shows four reference standards: the one used in the TABUL study and then the three alternatives described earlier, with increasing proportions of cases for which results might be considered more borderline. The *y*-axis shows the incremental NMB of the four selected alternative diagnostic strategies compared with biopsy alone.

The results show that, for all alternative reference standards tested, ultrasound plus clinical judgement remains the most cost-effective strategy. It is only by adopting a reference standard with a significant reduction in cases of GCA (16% fewer GCA cases than in the TABUL cohort) that a diagnostic strategy based on pre-test risks and ultrasound might potentially become as cost-effective as ultrasound combined with clinical judgement.

Budget impact

In the UK population, the annual incidence of GCA in those aged over 40 years is about 1 per 4500 people (or 22 per 100,000),[113] giving an annual incidence of about 7000 cases.

The cost savings arising at the point of testing through use of ultrasound instead of biopsy (both alongside clinical judgement) would be £456 (which represents the difference between £514 for a biopsy and £58 for ultrasound) per case or around £4,735,000 annually for the UK. Taking account of higher treatment costs for biopsy (because of slightly lower sensitivity), the cost savings would be £475 per case or around £4,933,615 annually for the UK.

If we use the strategy of ultrasound combined with clinical judgement but refer for biopsy cases that were judged to be 'not GCA' if they had a high pre-test probability of GCA, the cost saving would be £477 per case, or around £4,950,000 annually for the UK.

Discussion

Statement of principal findings

The results indicate that ultrasound alone is more cost-effective than biopsy alone, largely because of its much lower cost and, to a lesser extent, its higher sensitivity.

In practice, patients are stratified for the risk of having GCA or not, based on demographic factors such as age and sex, the clinical presentation and, in particular, the presence of more specific GCA-related symptoms such as jaw claudication and/or visual loss combined with the evidence of an acute phase response (elevated CRP level or ESR). Therefore, the biopsy test or ultrasound test are never used in isolation and should be regarded as supplementary to the rest of the clinical evaluation in such patients; this combination increases the sensitivity of the tests considerably. This is reflected in the main set (base-case) results, which show that the most cost-effective strategies are based on a test in conjunction with clinical judgement. Current clinical practice involves biopsy with clinical judgement. The results indicate that ultrasound plus clinical judgement is more cost-effective than biopsy plus clinical judgement, with a relative cost saving of £475 per patient and a larger QALY gain of 0.0005; thus, ultrasound is said to dominate biopsy in this case (both in terms of cost savings and QALY gains). This is a very small difference in QALYs, however, which is equivalent to < 1 day of full health on average across presenting patients. Ultrasound plus judgement is also estimated to result in a marginally lower incidence of vision loss (owing to its slightly higher sensitivity) than biopsy and judgement.

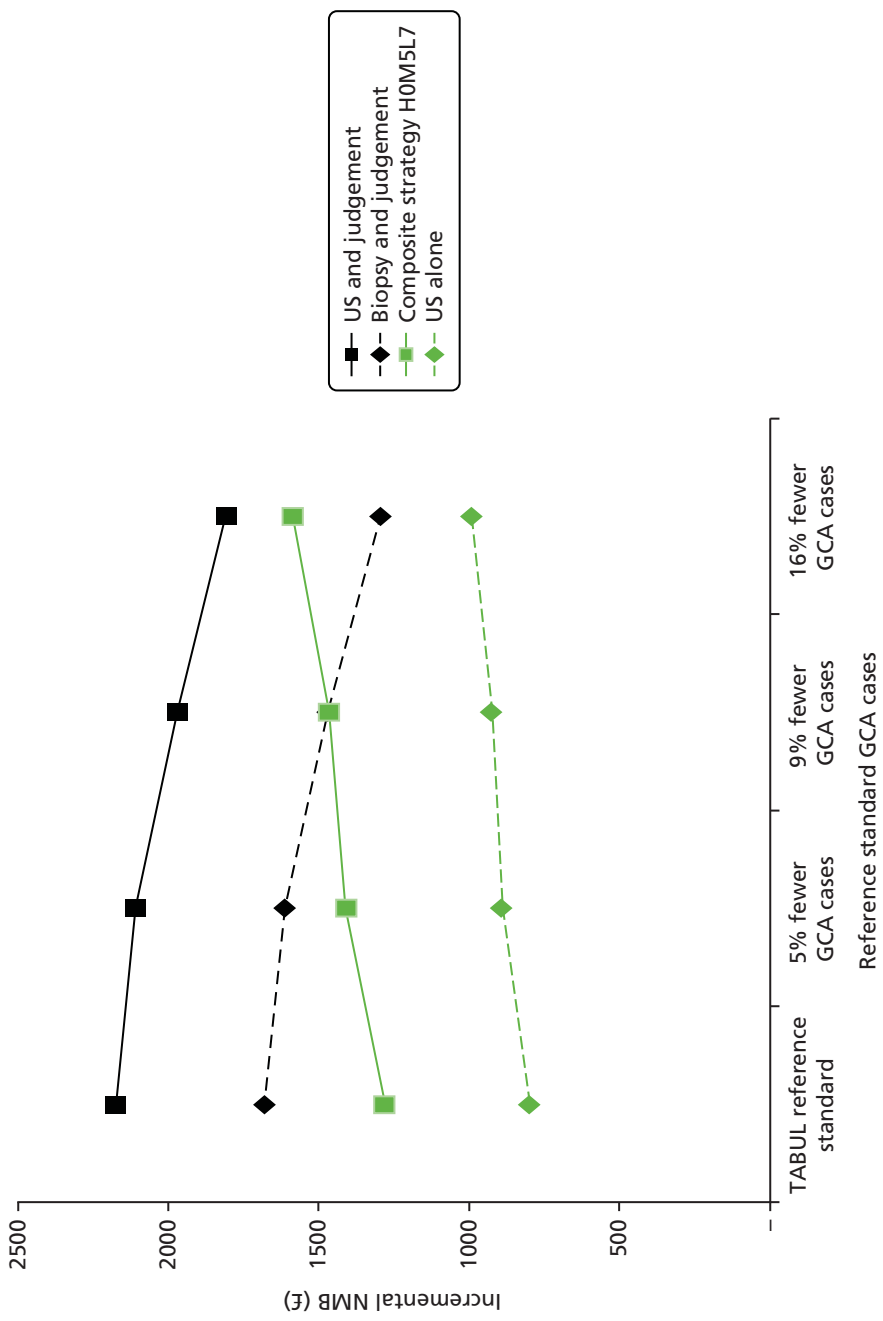


FIGURE 20 Incremental NMB for four diagnostic strategies compared with biopsy alone according to reference standard adopted. H0M5L7, treat all high-risk patients, diagnosis for low-/medium-risk patients as per ultrasound result; US, ultrasound.

One-way sensitivity analyses show that these findings are highly insensitive to changes in nearly all model parameters. The only parameters having any sizeable effect, in terms of partly reducing the difference in cost-effectiveness, are the cost of ultrasound and uncertainty around the sensitivity of ultrasound and biopsy.

In conjunction with clinical judgement, performing both a biopsy and an ultrasound test in all patients is less cost-effective than ultrasound alone because the additional costs of testing are not justified by the small reduction in treatment cost and increase in QALYs.

When we explored the impact of alternative diagnostic reference standards with up to 16% fewer GCA cases, ultrasound plus clinical judgement remained the most cost-effective strategy.

Drivers of cost-effectiveness

By far the most dominant driver is the cost of TAB because it is estimated to be almost nine times the cost of an ultrasound (£514 compared with £58). It is this that makes ultrasound plus clinical judgement more cost-effective than biopsy plus clinical judgement. When comparing strategies involving clinical judgement with equivalent strategies without clinical judgement (e.g. biopsy plus judgement compared with biopsy alone), the different sensitivities to GCA are the main driver of the results.

Strengths and limitations

This is, to our knowledge, the first published economic evaluation of ultrasound compared with biopsy. The evaluation not only includes costs incurred at the point of diagnostic testing, but also the costs and QALY implications of different rates of false-positive and false-negative cases. We also carried out additional analyses to allow for the fact that there is not a single universally accepted gold standard for diagnosing GCA (see *Chapter 8* for further discussion on the lack of a gold standard).

No evidence source was found for the cost of a biopsy of the temporal artery. It was therefore necessary to use the cost of a procedure similar in terms of complexity and therefore resource use, a lymph node/salivary gland biopsy. Ideally, a micro-costing study could have been undertaken to arrive at an estimate specific to TAB. However, sensitivity analysis around the difference in cost between ultrasound and biopsy showed that this only had a small impact on reducing the favourable cost-effectiveness of ultrasound.

The diagnostic outcomes for ultrasound plus clinical judgement were not a formal outcome of the TABUL study so we had to use an algorithm (see *Methods*). Although this approach, and specifically the use of a vignette exercise to obtain the outcome for 27 patients, introduces some uncertainty around the sensitivity and specificity of ultrasound and clinical judgement, this is very unlikely to be large enough to have a material effect on the economic findings. This can be seen in the sensitivity analysis that varied the sensitivity and specificity of biopsy.

Owing to the complexity involved, our model was not sophisticated enough to include the impact of a quicker turnaround of results with ultrasound and any benefits arising from being able to lower the steroid dose sooner for cases with a negative diagnosis so there might be some further benefit to ultrasound-based strategies not accounted for in the modelling.

Any general limitations of the TABUL study, as discussed in *Chapter 8*, that pertain to the observed diagnostic yields (test sensitivity and specificity) apply to the economic analysis too. However, we carried out uncertainty (sensitivity) analysis around these parameters and this did not affect the conclusions.

Implications

The results indicate that ultrasound plus clinical judgement is the most cost-effective strategy. Such use of ultrasound rather than biopsy would result in significant reductions in costs as a result of the much lower cost of the test (£514 vs. £58). Frequently, the upfront cost can be a barrier to uptake of cost-effective technologies for which the economic benefits only materialise over the long term. This is not the case here, however, with estimated savings to the UK of £4,735,000 annually based on annual incidence of 7000 cases.

Unanswered questions and further research

We were unable to identify a study that would enable us to calculate dose-specific risks of fractures for each fracture type. Studies generally reported hazard ratios by category of average steroid dose, for example > 7.5 mg per day, rather than for specific and varying doses over time. This is a specific example of the difficulty of synthesising the range of heterogeneous evidence available on risks of steroid therapy in terms of study duration, starting dose, tapering schedule and set of AEs reported. Sensitivity analysis indicates that our results are very insensitive to uncertainty around the burden of steroid-related AEs. However, in a different context to this evaluation, for example, with a less dominant difference between the costs of the tests, the difficulties in synthesising such evidence could be a far greater limitation.

Chapter 8 Discussion and conclusions

We have undertaken a large multicentre evaluation of two diagnostic tests in patients with newly suspected GCA. We performed both tests (ultrasound of temporal and axillary arteries and biopsy of the temporal artery) in all cases. We kept the results of the scan blinded from the clinicians until after the primary end point had been achieved (the clinicians' diagnosis was recorded 2 weeks after initial assessment). In order to conduct the study we needed to establish a new training programme for ultrasound of temporal and axillary arteries and to measure the quality of all scans being performed. We will discuss our main study findings, based on the original hypothesis examining the sensitivity and specificity of both tests as well as an economic analysis of the tests alone or in different combinations. We will describe and summarise the patient cohort and discuss the potential advantages and disadvantages of the reference standard diagnosis used to compare the outcome of the two tests. We summarise the ultrasound training programme and the scan results during the course of the study. We comment on the biopsy findings in the cohort, and on the clinical diagnoses. We have subjected the clinical data to scrutiny by an expert panel and have provided interobserver comparisons of the ultrasound and biopsy data. We assess the changes in diagnosis or test result following expert review. We discuss the value of combined strategies and the added role of clinical judgement or clinical risk stratification on either or both tests. We look at the generalisability and implications of this study in routine practice.

Main findings

We conducted a prospective multicentre study to compare the relative value of ultrasound assessment of both temporal and both axillary arteries with TAB in 381 patients with newly suspected GCA. In order to ensure proficiency in performing ultrasound scans, we created an extensive training programme, which was then compared with the established standard procedure of TAB, usually from the most symptomatic side, in this patient population. No training was provided for performance of the biopsy within the study. All patients in the study underwent both tests in sequence (ultrasound first followed by biopsy) and our analysis included those who underwent both tests within 10 days of commencing high doses of glucocorticoids. Usual care was given to the patients by their clinicians. The ultrasound result was not revealed to the clinician caring for the patient, unless they specifically requested the result (because they were planning rapidly to reduce and withdraw glucocorticoid therapy) after they made their clinical diagnosis at 2 weeks' follow-up, which was the main primary outcome in the study. A final follow-up assessment was conducted at 6 months in case the diagnosis had changed.

The main objectives of the study were to compare the diagnostic performance (sensitivity and specificity) of ultrasound as an alternative to biopsy for diagnosing GCA in patients who are referred with suspected GCA and in whom a biopsy was going to be carried out and to perform a cost-effectiveness analysis to compare different potential investigation strategies for diagnosing GCA, incorporating either or both ultrasound and biopsy. The original hypothesis was that ultrasound would be a more sensitive test than biopsy and would have a specificity of at least 83%.

Early studies suggested that biopsy had 95% sensitivity and 100% specificity for GCA;[114] later studies reported somewhat lower results of around 68–69% sensitivity but very high specificity.[11,115] Patients who had a positive biopsy but who did not have GCA were reported to have other forms of vasculitis.[59]

We wanted to compare the performance of ultrasound, which we predicted would provide 87% sensitivity and 83% specificity or higher. Among 381 patients who had ultrasound and TAB for suspected GCA, 101 (27%) had a TAB consistent with GCA and 162 (43%) had an ultrasound result compatible with GCA. The sensitivity of biopsy for diagnosis GCA was 39%, much lower than previously published; the specificity was 100%. By contrast, the sensitivity of ultrasound was 54% with a specificity of 81%. Therefore, we failed to find evidence to support our primary hypothesis because, although ultrasound was

more sensitive than biopsy, it did not achieve specificity greater than 83%. Nevertheless, we demonstrated that the current sensitivity of biopsy is much lower than previously published and that, in comparison, the sensitivity of ultrasound is superior (14% higher). The specificity of ultrasound was 81%, which is lower than expected from our original hypothesis. We cannot conclude that ultrasound can replace biopsy, based on these findings. However, the data support a significant challenge to the role of biopsy as a 'gold standard' test for diagnosing GCA. A combination strategy using both tests in sequence, with all patients undergoing an ultrasound scan, but only scan-negative cases undergoing a biopsy, has 65% sensitivity and 81% specificity for the reference standard diagnosis of GCA. The addition of risk stratification based on initial clinical features and measures of ESR or CRP levels can further increases the sensitivity to 77.1% and specificity to 91.2%.

The cost-effectiveness analysis indicates that ultrasound alone is more cost-effective than biopsy alone largely because of its much lower cost (£58 vs. £514) and higher sensitivity (54% vs. 39%). The use of ultrasound combined with clinical judgement is not only more cost-effective than biopsy plus clinical judgement but is estimated to result in both cost savings (largely owing to the lower cost of ultrasound) and a very small QALY gain.

Patient details

A total of 730 patients were screened for the study: 430 participants were recruited from 20 sites in five countries (England, Ireland, Norway, Germany and Portugal); and 300 patients either did not meet the inclusion criteria or declined to participate. From the 430 patients included, there were 39 withdrawals prior to the primary analysis being performed (at the 2-week assessment) and a further 49 withdrawals after the primary analysis was performed. Of the remaining 391 patients, 10 were excluded from the primary analysis; hence, the primary analysis was performed on 381 patients in total. The average age of the cohort was 71.1 years and 72% of patients were female. The majority of patients (80%) were of white British ethnicity, and the remainder were from either a white Irish or other white background or a non-white background (3%). The majority of patients (88%) had significant new headache at presentation: fatigue was reported in 65%, generalised scalp tenderness in 59% and 51% had pain over one or both temporal arteries. PMR was present in 7%. Visual symptoms were frequently present at baseline (reduced or lost vision reported in 133, amaurosis fugax in 14 and double vision in 31 patients). By 2 weeks, three, five and zero patients experienced new loss of vision, amaurosis fugax and double vision, respectively; by 6 months, new reports of these visual features occurred in seven, two and six patients, respectively. Anterior ischaemic optic neuropathy was reported in 27 patients (7%) at baseline; posterior ischaemic optic neuropathy was reported in seven patients (2%); by 2 weeks these findings were present in 4% and 0.5% of patients, and by 6 months in 4% and 1% of patients, respectively. The results of inflammatory markers (ESR and CRP level) were not always available (10% of patients had no baseline ESR result and 8% had no baseline CRP level result). The median ESR at baseline was 43 (IQR 60–70) and the median CRP level was 46 (IQR 90–91). Many patients had hypertension at baseline (52%); 7% had angina and 14% had a previous history of cancer. There was a small increase in the occurrence of diabetes mellitus during the course of the study from 14% at baseline to 18% at 6 months. The main physical findings were of tenderness (50%) or thickening (27%) of one or both of the temporal arteries, which were less likely to be detected if the patients had received even a few days of steroid treatment.

Use of the reference diagnosis

There was no absolute gold standard that we could apply in this study to decide whether or not the patient definitely had GCA. Use of the ACR classification criteria for GCA[34] included using the result of the biopsy; this would bias the interpretation of the clinician's opinion in favour of stronger agreement with the biopsy test when it was positive, and perhaps bias it against that diagnosis if it was negative. We attempted to address this by including additional aspects of the patient's condition that would be compatible with the

clinical diagnosis of GCA, such as the presence or development of visual loss attributed to GCA, the presence of stroke or PMR. Other features such as jaw or tongue claudication or significant elevation of the ESR or CRP level (above 60 and above 40, respectively) would have contributed to the clinician's assessment and the likelihood of diagnosing GCA. However, the interpretation of all the clinical features, laboratory findings and results from the specific investigations would have to be considered individually on a case-by-case basis. This would mean that the clinician could over-ride/ignore any individual results in favour of or against the diagnosis of GCA. This is a clear limitation of the current study. However, including the ability to adjust the reference standard diagnosis in light of the development of changes to the clinical state in the 6 months following initial assessment (e.g. the development of features consistent with the diagnosis of GCA or, equally, the development of features consistent with another diagnosis) strengthens the argument for using this reference standard diagnosis as the gold standard, albeit a less than perfect one.

The use of presenting features to predict the likelihood of a diagnosis was suggested by Gabriel *et al.*[116] In a review of > 500 patients, the likelihood of a negative TAB was increased substantially in the absence of claudication and the absence of significant elevation of the ESR. In the current study we used the reference diagnosis rather than the biopsy as the standard for the model, but with similar findings. A limitation of this study is the lack of a robust unequivocal standard for diagnosis against which each test could be compared. In the absence of this, diagnostic criteria are being developed in the DCVAS study which might provide a better surrogate gold standard than currently exists. The clinical evaluation of the patient at baseline and after 2 weeks has the strongest influence on the diagnosis at 6 months. Part of the difficulty is the concern of clinicians that if any of the clinical features, combined with measurement of the acute phase response, are suggestive of GCA, despite negative further testing (biopsy or imaging), there is a clearly demonstrated unwillingness to dismiss the diagnosis. Rather, the tests (biopsy or imaging) are being used to provide further enhancement of the clinical opinion.

Ultrasound training

Ultrasound has not yet superseded TAB as a diagnostic test. This may reflect the poor consistency of the scanning technique as a result of the lack of a standardised scanning protocol. We developed a standardised protocol that was implemented in 439 healthy controls and subsequently in patients with suspected GCA. We assessed each patient for evidence of typical ultrasound features of GCA: the presence of a halo surrounding the vessel wall, stenosis or occlusion of the vessel. A detailed scanning protocol was developed for all patients and controls. We reported the presence or absence of ultrasound features of GCA in each segment of each temporal artery (common, parietal, frontal proximal and frontal distal) and both axillary arteries. Sonographers were asked to acquire video and static images for each patient to ensure accuracy of findings. The sonographer measured and documented halo diameter (based on a normal range of up to 0.5 mm for the temporal artery and up to 1.0 mm in the axillary artery) and length; pulse Doppler measurements prior to and within a stenosis (confirmed if the highest maximum systolic velocity was over twice the lowest maximum systolic velocity); and arterial occlusion. Each study site sonographer was required to be proficient in the protocol by scanning at least 10 healthy controls, passing an online test showing normal and abnormal scans (pass mark > 75%) and scanning a patient with ultrasound evidence of active GCA. The scanning protocol was started by 33 sites, with only 22 sites completing the training in 6.7 months (range 0.2–16.4 months). A total of 439 controls were scanned across 31 sites (one sonographer covered three sites). The online test was passed by 39 sonographers (multiple sonographers at some sites) with an average of two attempts (range 1–4); 22 sonographers successfully scanned an active GCA patient, as validated by the expert panel. The longest delay in completing the training was a result of difficulty in recruiting a patient with active GCA, which was necessary prior to commencement of the main study. Common issues encountered were a lack of time away from clinical duties and locating a new suspected GCA case for the hot case assessment. We have created a bank of 857 sets of consistently recorded images of temporal and axillary arteries from patients with suspected GCA and from healthy controls. Expert review of the scans confirmed that the overall rate of disagreement was 16%. Quality and accuracy are imperative for the clinical use of ultrasound data

in diagnosis. We have developed an effective protocol, including training, which ensures consistency and proficiency in scanning. The methodology can be adapted and extended to allow for additional artery assessment, including carotid, vertebral and subclavian, extending the value of a structured approach. We recommend the current study scanning protocol as the standard approach for diagnosis of GCA using ultrasound.

How could we improve on the ultrasound training programme in practice?

We developed a novel training programme as part of the current study. The programme was based on published evidence of performing ultrasound examination of GCA; most of the publications were from experts within the study investigator group. The basic elements of the training programme consisted of (1) a tutorial/lecture [which could be provided as a recording or annotated Microsoft PowerPoint® (version 97–2003; Microsoft Corporation, Redmond, WA, USA) presentation], (2) hands-on training for novice sonographers (which would not be required by more experienced sonographers), (3) evidence of recorded images to show proficiency in performing scans on healthy individuals to demonstrate non-diseased temporal and axillary arteries (primarily done remotely) and (4) evidence provided by sonographers of recorded images to show their proficiency at performing scans in at least one individual with active GCA, to demonstrate diseased temporal or axillary arteries (primarily done remotely). We implemented the training requirements for the purpose of this study, which was deliberately based in non-academic as well as academic centres, in order to test the practicality of establishing this new technique of ultrasound in large numbers of local hospitals, where resources might be limited. We discovered significant variation in the uptake of the training, primarily driven by local factors such as the availability of sonographers and ultrasound machines; as a result, only half of the centres that originally attempted the training programme actually preceded with the study. Given the nature of the condition (i.e. presentation with acute-onset symptoms and the need to undertake scanning within a short time of starting steroid therapy), there are minimum basic requirements in any individual centre to ensure that the technique is performed to the correct standard and can be undertaken in a timely fashion. Furthermore, a minimum number of cases scanned per annum would be advisable to ensure ongoing quality control; we found that scanning reliability was higher for sonographers who had scanned at least five cases during the study compared with those who had scanned fewer than this number. In practice, therefore, we may need to explore other ways in which to deliver the training material and to develop a programme to maintain proficiency in training. We speculate that some of the training elements could be provided as courses, whereas other elements are bespoke to individual centres and would require clear demonstration of the sonographers' abilities to scan and to be able to clearly distinguish cases from non-cases. The nature of the training programme itself could be adapted depending on the expertise of the sonographer, for example, shortening it for more expert centres, while still maintaining minimum standards. Targeted training would need to be more intense for novice sonographers (similar to the full training programme in this study), and less intense for more expert centres (requiring the sonographers to provide evidence that they are competent at performing the scans, by providing evidence that they have been regularly scanning cases, as well as being able to submit the scans of an active case as proof that they can adequately recognise an abnormal case. For centres with some experience, but that have not been performing scans regularly, we could ask their sonographers to undertake the online quiz, to make sure that they can recognise normal and abnormal scans, as well as to provide scans from an active case that they have recently seen. Implementation of training programmes would be facilitated by their certification through Royal Colleges or national bodies, such as the BSR. This would encourage accredited training and it would be feasible to apply for this training activity to be recognised as continuing professional development.

Ultrasound findings

Ultrasound abnormalities consistent with GCA were found in 162 patients, predominantly in the temporal arteries, but in 31% of patients, the axillary arteries were also involved and in a small number of patients

(2.4%) they were exclusively involved. The predominant abnormalities on ultrasound that were considered to be consistent with GCA were the presence of a halo in 162 patients, stenosis in 45 patients and occlusion in 41 patients. The median halo size was 0.6 mm (range 0.4–0.9 mm) as measured in temporal arteries. In patients with abnormal ultrasound scans, the median number of segments of artery involved was 2.5 (range 1–6).

We measured differences in the size of the halo around the arteries depending on the duration of steroid therapy prior to scanning; we correlated halo size with ischaemic symptoms of GCA. We analysed data from 301 out of 415 patients with clinically defined definite or probable GCA at baseline using linear and logistic regression models to determine the relationship between halo size and days of steroid treatment and also with ischaemic symptoms of GCA (jaw and tongue claudication, amaurosis fugax and reduced, lost or double vision). Fifty per cent of patients were scanned on or before receiving 2 days of high-dose steroid treatment. Forty-three per cent (131) of patients had a halo in one or more temporal segments, 49% (146) of patients had bilateral temporal artery halos and 13% (38) of patients had axillary involvement. The linear regression model showed a consistently smaller halo size in temporal arteries during the 7 days of steroid treatment ($p < 0.005$). The likelihood of finding a halo diminished with time, until day 4 of steroid treatment ($p < 0.005$). Jaw claudication occurred more frequently in patients with a halo ($p < 0.05$). Temporal artery symptoms correlated with ipsilateral ultrasound findings ($p < 0.05$). The findings suggest that, in newly diagnosed GCA, ultrasound halo size decreases rapidly with steroid treatment and correlates with the presence of ischaemic symptoms, supporting its early use as a diagnostic and potentially prognostic marker.

Biopsy findings

Only 353 out of 381 biopsies performed actually contained a sample of temporal artery; the remainder either consisted of another tissue (such as vein or nerve) or no sample was obtained at all. The median length of artery biopsied was 10 mm (range 7–15 mm). In 161 patients the TABs were defined as abnormal and in 101 patients (27% overall) this was compatible with the diagnosis of GCA. In four patients the biopsy was compatible with another form of vasculitis. In a further 35 patients, arteriosclerosis was the dominant finding; 27 patients had a variety of other diagnoses (not GCA or vasculitis). Fragmentation in the internal elastic lamina was reported in 156 biopsies, and reduplication in 82 patients. Thirty-nine per cent of patients had intimal hyperplasia and 10% had arteriosclerosis in the intima. Of the 101 biopsies consistent with GCA (27% overall and 39% of the patients diagnosed with GCA), giant cells were present in 72 biopsies (representing 19% of the overall cohort, but 71% of biopsies of patients with GCA). In 99% of biopsy-positive cases, inflammatory infiltrates were present, which were transmural in 42% and adventitial in 18% as the predominant sites of inflammation. Furthermore, seven patients had evidence of recanalisation in at least one section of the biopsy.

Histological features in biopsy-positive patients with GCA were not confined to one form of inflammation. The most common finding was transmural inflammation. The relatively low number of positive biopsies may reflect the low index of suspicion of GCA in the cohort, technical difficulties in obtaining an adequate sample, skip lesions or the effects of glucocorticoid therapy in changing the biopsy result. These findings highlight the need for a better diagnostic strategy for patients with suspected temporal arteritis.

Change in diagnosis after expert review

Following expert review of the clinical cases, 21 patients had a change in diagnosis: in 13 of these, the diagnosis changed from GCA to not GCA; and in eight patients the diagnosis changed from not GCA to GCA. The diagnoses were predominantly made on the basis of symptoms and signs, blood abnormalities and, to a lesser extent, the biopsy report. The most common diagnosis in patients who did not have GCA was non-specific headache, myofascial pain, migraine, temporomandibular dysfunction and sinusitis.

Five patients in total were diagnosed with another form of vasculitis including Takayasu's arteritis, EGPA, GPA and other undefined forms of vasculitis.

The confidence in the clinical diagnosis of GCA at the baseline was > 75% in favour of probable or definite GCA; 86% of the patients, regardless of the confidence in diagnosis, were being treated with high doses of steroids at baseline. Most patients did not have any change in their clinical diagnosis by the observing clinicians from the 2-week assessment to the 6-month assessment. However, 19 patients had their diagnosis changed from not GCA to GCA at the 2-week assessment after unblinding of the ultrasound result. In 25 patients the diagnosis was changed at 6 months (6% of all patients); in 17 of these patients the diagnosis was changed from GCA to not GCA and in three patients a diagnosis of GCA was made. In the remaining five patients the diagnosis changed (but not from or to GCA).

Ultrasound compared with biopsy results

There was a significant association between the biopsy and ultrasound results ($\kappa = 0.35$), but more scans were positive than biopsies, so that ultrasound was more likely to be used to diagnose GCA than biopsy (162 positive ultrasound cases compared with 101 biopsy cases). Eighty-eight patients who had ultrasound evidence consistent with GCA had a negative biopsy and 27 patients with biopsy evidence of GCA had a negative ultrasound. There was a small number of patients (23) to whom steroids were given for longer than 7 days prior to the biopsy being performed. If we excluded those patients from the analysis, the agreement between ultrasound and biopsy increases slightly with a kappa of 0.37. The finding of a halo appeared to be the most useful aspect of the ultrasound result to support a diagnosis of GCA. Combining halo assessment with other aspects of ultrasound, namely stenosis or occlusion, did not increase the overall interpretation of the ultrasound scan as being positive or negative. Ninety-three patients had bilateral halo and a clinical diagnosis of GCA. Axillary involvement on ultrasound was present in 53 patients, nine of whom did not have temporal artery involvement; three were biopsy positive and, in total, seven were given a reference of diagnosis of GCA, suggesting that ultrasound of the axillary arteries can provide further support for the diagnosis of GCA in the absence of either temporal artery ultrasound or biopsy evidence to suggest GCA.

The effect of training and expert review of scan results on diagnosis

Expert review of the ultrasound images was part of the protocol and was undertaken for ongoing quality control purposes during the study. In 16% of scans the expert reviewers' interpretation of the scans differed from the sonographer's interpretation; 14 patients were interpreted as GCA by the reviewers but not GCA by the sonographer, and a further 47 were interpreted as not GCA by the reviewers but as GCA by the sonographer. The overall impact of using the reviewers' interpretation in place of the sonographer's interpretation was to increase the specificity of ultrasound from 81% to 87% but to reduce sensitivity from 54% to 44%. One potential explanation for the lower sensitivity using the reviewers' interpretations is that the recorded ultrasound images and videos that the reviewers saw did not capture the abnormalities seen by the sonographer during a patient's scan. A second potential explanation is that a sonographer's interpretation may have been influenced by seeing the patient, for example, by observing a tender or thickened artery during the scan, something the reviewers would not be aware of. For the majority of discordant interpretations (those interpreted as not GCA by the reviewers) it is unclear if the difference indicates problems with the sonographer's interpretation or, as the reduction in sensitivity might suggest, merely difficulties in capturing abnormalities in scan recordings. The discordant interpretations that the reviewers interpreted as GCA were fewer in number, but may indicate issues with a sonographer's interpretation of the scans. For two sonographers in the study, retraining was required before they resumed scanning patients.

In 19 patients, the 2-week diagnosis based on the clinical findings and biopsy was not of GCA, but the unblinding of the ultrasound result suggested that there were findings compatible with GCA. Unblinding improved the sensitivity but reduced the specificity of the 2-week assessment compared with the reference diagnosis (sensitivity of 0.96 and specificity of 0.77). We observed a training effect among the sonographers. There was no significant change in the specificity of ultrasound for GCA by sonographers when comparing their early (first 10) scans with their subsequent scans, but the sensitivity improved from 45% to 62%, strongly suggesting an improvement in the ability to detect the presence of halo. Such an effect suggests that it is possible to achieve improved accuracy with ultrasound as sonographers gain experience in scanning. It also raises the question of whether or not more extensive training and/or supervision should be provided in addition to the training protocol developed for this study.

The effect of delay in testing and the effect of steroids

The accuracy of biopsy was likely to be greatest if performed within 3 days of starting steroids (sensitivity of 48% at this stage, compared with 33% for biopsies performed from ≥ 7 days after the commencement of steroid treatment). For ultrasound, the accuracy was highest for patients seen on no more than one dose of steroids, but was still maintained up to 7 days. The effect of delay between the scan and the biopsy being performed did not appear to influence the probable agreement between the tests.

Combination strategies and pre-test probability of having giant cell arteritis

We derived a risk of having GCA based on data obtained from an independent cohort of patients (based on the DCVAS study). We divided patients with GCA in the DCVAS cohort into three risk groups: those with an ESR > 60 mm/hour or a CRP level > 40 mg/l combined with the presence of jaw or tongue claudication were in the highest-risk group for having GCA; the lowest-risk patients had none of these features; medium-risk patients had only either an elevated ESR or CRP level or symptoms of jaw or tongue claudication. There was a significant relationship between the assignment of patients to one of these risk groups and the certainty of diagnosis of GCA reported at baseline in the TABUL study cohort, the reference diagnosis given and the biopsy findings. Although there was a trend for the ultrasound result, it was not as consistent. In other words, the patients in the lowest-risk group still had a 31% likelihood of a positive ultrasound compared with only 7% having a positive biopsy. Within the highest-risk group, the sensitivity of biopsy was 63% and for ultrasound it was 57%, with specificities of 100% and 80%, respectively. However, in the medium-risk group, biopsy had only 33% sensitivity, with a specificity of 100%, whereas ultrasound had 57% sensitivity and 91% specificity. Furthermore, in the low-risk group, biopsy had the least sensitivity of 17%, with a specificity of 100%, whereas ultrasound had a sensitivity of 44% and a specificity of 77%. One potential option is not to do either test (biopsy or ultrasound) if patients are in the high-risk group, because there is 93% prevalence of likelihood of diagnosis GCA according to the reference diagnosis in these patients. Looking at the potential combination of strategies, the risk group (high, medium or low) would affect the sensitivity and specificity of diagnosing GCA by performing either ultrasound and/or TAB (depending on the results of the ultrasound). In every instance, combination strategies produced better receiver operating characteristic curves than biopsy alone, supporting the role of ultrasound in supplementing or, in some patients, replacing biopsy as the diagnostic test for GCA.

Assessment using vasculitis activity and damage scores and quality of life

We used standardised generic scores of vasculitis activity and damage (the BVAS and the VDI score) in this cohort of patients, primarily to screen for the possibility that some patients had a more widespread form of

a different type of systemic vasculitis (and this was actually true in five patients). As an exploratory outcome, we found that in 257 patients with GCA, disease activity scores were not significantly different from patients who did not have GCA. This shows that the disease activity score is not discriminatory between GCA and non-GCA (it was never designed for this purpose). However, the BVAS was more likely to be lower at 6 months than at 2 weeks. We have to bear in mind that it is likely that the scores were under-reporting disease activity at 2 weeks, because a significant number reported no abnormalities in the GCA group. There did not appear to be any discriminatory effect of measuring the VDI score at 2 weeks or at 6 months between patients and controls (the VDI was not designed to discriminate), but there was an increase in the number of patients and controls with at least one item of damage reported after 6 months compared with the 2-week assessment. Quality of life, as measured by the EQ-5D, did not differ significantly between patients and controls and neither did it change significantly after 6 months.

Adverse events

In total, 1229 AEs were reported during the study; every patient suffered at least one event, the majority of which were related to steroid therapy. When looking at events related to the study tests, 63 patients had an AE definitely related to biopsy, 10 had events possibly related to biopsy and two had events definitely related to the scan. There were 104 serious AEs among 55 participants, but none of them was related to the study test. The serious events included 16 deaths and 74 hospitalisations; all of these characteristics would be expected in a population of patients with suspected GCA and in whom high doses of steroids have been used.[42]

Inter-rater agreement

We undertook inter-rater testing to evaluate agreement between pathologists and between sonographers in their assessment of images biopsy and ultrasound images. We selected the ultrasound scan recordings and histology slides from 33 patients in the study (a mixed group chosen at random, some of whom had a reference diagnosis of GCA and some of whom did not). We performed an inter-rater exercise separately for 14 pathologists and 12 sonographers. Agreement among 14 pathologists based on ICC was 0.62; among the sonographers it was 0.61. This would suggest that the level of certainty for interpretation of either test is variable, and it is perhaps more variable for pathologists than previously appreciated. The agreements between observers for both tests were similar.

Strengths and weaknesses of the study

We recruited a large cohort of patients mostly from primary care practices in the UK to a large number of centres, including academic and non-academic centres, to establish the generalisability of our findings. We developed an ultrasound training module as part of the study to ensure proficiency of testing. We did not offer any training in biopsy techniques or in biopsy processing and interpretation, as these are standard and, as such, should not be required by participating sites. We were able to compare the effects of ultrasound and biopsy independently on the diagnosis; however, the classification criteria for GCA include the results of biopsy, introducing an inherent bias in the diagnosis of GCA which would be likely to be given more or less weight depending on whether or not the biopsy was positive or negative. Despite this bias we were still able to demonstrate that ultrasound was an effective strategy for diagnosis in a significant proportion of patients. Nevertheless, neither of the tests is perfect, and we do not have a true gold standard to compare the effectiveness of each test. Unblinding of the ultrasound result at 2 weeks could have biased the results, but, in fact, a sensitivity analysis suggested that it had only a marginal effect on the outcome of the study.

Evolution in the presentation and suspicion of giant cell arteritis

Greater awareness of GCA may prompt primary care physicians to initiate treatment at a very early stage, which might affect the likelihood of obtaining a positive test result. Studies of pathological specimens obtained in other forms of vasculitis suggests that, whereas previously a biopsy showed clear evidence of abnormality, if awareness of the disease and clinical suspicion of the diagnosis lead to earlier investigation and treatment, we might actually be changing the natural history of the disease such that we do not see the characteristic features of the disease as previously described on biopsy. For example, nasal tissue biopsies have been reported to provide diagnostic appearances in 24–53% in patients with GPA.[117,118] It is possible that the level of suspicion for the diagnosis of GCA may have changed in line with the suspicion of the diagnosis of the other form of vasculitis.[119]

Generalisability of current findings

One of the potential criticisms of the project is that we were introducing a specialist form of ultrasound imaging to NHS hospitals and comparing this with established practice. The specialist techniques of ultrasound imaging of temporal and axillary arteries might be perceived as being feasible to implement only in specialised centres where more time and resources might be available to perform these scans and that it might require more specialised equipment. However, we deliberately chose to recruit participants from non-academic centres, as well as academic centres, in order to test whether or not our technique was generalisable and could be applied, with suitable training in ultrasound performance and interpretation.

Although there is a difference between sonographers in terms of experience, as demonstrated by our evaluation of performance for centres recruiting fewer than 10 patients or more than 10 patients, this in itself is not an issue of whether the centre is an expert academic centre or a non-academic centre. This is to do with the volume of patients evaluated. Given the relative frequency of GCA in the general population and the likelihood that patients who have suspected GCA are referred to hospital for assessment, there is an opportunity for all centres to increase the number of patients evaluated to improve the sensitivity and specificity of ultrasound as a diagnostic test for GCA.

What are the implications of the study findings?

Our data suggest that TAB is less effective as a diagnostic test for GCA than was previously appreciated. Although it retains a high specificity, the sensitivity is only 39%. This could be because patients who are being evaluated with this test have low pre-test probability of the diagnosis. However, patients were selected for inclusion in the study on the basis that they have at least a possibility of GCA; in 53.5% of patients there was probable diagnosis of GCA and 21% of patients were reported as having a definite diagnosis of GCA at presentation as reported by the clinician. In comparison to other cohorts of patients undergoing TAB, the biopsy yield was actually higher than the 15.1% reported previously.[48] Difficulty in interpreting the biopsy result is undoubtedly made worse by not obtaining any arterial tissue at all, which occurred in 28 patients in the cohort. In a previous cohort of 567 consecutive biopsies, 2.5% had no arterial tissue,[49] suggesting limitations to the technique. The biopsy length obtained was an average of 1 cm in the current study, which is the minimum recommended by the BSR guidelines.[5] However, other studies have suggested that 0.7 cm is an adequate length;[52] in fact, even smaller biopsies might be adequate, with no evidence of a difference in positive biopsies for samples < 0.65 cm compared with those longer than 0.7 cm.[53] The biopsy length referred to in the current study is the measurement taken by the pathologists once a specimen arrives in the laboratory. It is known that shrinkage occurs once the specimen has been excised; we did not measure the length of the specimen obtained by surgeons at the time of sampling.

Pathologists are usually expected to provide an opinion on the diagnosis based on the interpretation of the biopsies. Our data suggest that the variation in agreement between observers can be considerable, especially for less clear-cut cases. If the specimen did not contain characteristic features of GCA, the interpretation of changes consistent with GCA, such as reduplication of the internal elastic lamina or intimal thickening or proliferation, could be that of early features of GCA, or of healing GCA, but, equally, these findings can occur in patients who have arteriosclerosis or age-related changes in their temporal arteries specimen and do not have any features of GCA at all.[47] We should give consideration to encouraging pathologists to report on the uncertainty of interpreting the findings rather than forcing them to make a clear-cut distinction between GCA and not GCA on the basis of the histology alone if there is insufficient information to make such a distinction with confidence.

We did not provide any training specifically to either the surgeons undertaking the biopsy or to the pathologists preparing and interpreting the sample results. We did not provide any reference standards to compare abnormal results or require any evidence of proficiency by the pathologists in the interpreting biopsies. The effect of training or use of reference standards may have improved our biopsy results.

We have developed an ultrasound training protocol that was effective in allowing 20 different sites with variable experience of use of vascular ultrasound (in some cases none at all) to undertake and interpret images of the temporal and axillary arteries to a standard acceptable by an expert panel in over 90% of patients undergoing a scan. Using this methodology, we have demonstrated that we can improve on sensitivity of biopsy by using ultrasound. However, there is a lower specificity and neither technique alone provides a high rate of confidence in the diagnosis of GCA, without the interpretation of the clinical features. We have shown that ultrasound is cost-effective compared with biopsy.

However, in a significant number of patients, both tests will be negative and yet the clinician will still diagnose GCA because the patient has clinical features that strongly suggest the diagnosis (such as jaw or tongue claudication or the development of ischaemic events compatible with the clinical syndrome of GCA). Until we have a more robust measure as a diagnostic test of GCA, these two tests (biopsy and ultrasound) could be used in combination to improve early diagnosis and treatment of GCA. It is feasible that other imaging techniques could have a higher yield than ultrasound (e.g. MRI). In 64 patients who underwent MRI (and a proportion who also underwent TAB), the sensitivity and specificity of MRI was reported as 80.6% and 97.0%, respectively, compared with histology, which had 77.8% sensitivity and 100% specificity.[120] A comparison study between ultrasound and magnetic resonance showed almost identical positive and negative predictive values.[121] Unfortunately, magnetic resonance changes resolve within a few days of starting glucocorticoid therapy and access to magnetic resonance is likely to be a limiting factor, whereas access to ultrasound is much more rapid.[13] The effects of steroids on image appearances for both magnetic resonance and ultrasound have been compared in 59 patients undergoing both tests, as well as in a proportion undergoing TAB. Whereas the sensitivity of ultrasound and magnetic resonance were 92% and 90%, respectively, up to 1 day following steroid therapy, this is reduced to 50% and 80% with > 4 days of steroid therapy.[122]

It is conceivable that clinicians may feel some discomfort over having to rely on a clinical diagnosis of GCA supported by an imaging test such as ultrasound, but not confirmed by histological examination of the artery on biopsy. The concern would be that they are potentially overtreating a patient, who does not have a true diagnosis of GCA. However, the current study demonstrates that videos and images can be stored and reviewed later. Expert reviews of stored imaging tests were as reliable as expert reviews of stored biopsy specimens. There is increasing use of ultrasound as a diagnostic test in GCA in some centres for which confidence in the technical proficiency is high[123] as more scans are performed. The methods in this study will enable naive centres to gain proficiency and improve sensitivity and specificity of the tests. We have shown that it is practical and achievable to become proficient at vascular ultrasound, but that it does require specific training. Trained sonographers could initially perform scans in suspected cases that also undergo biopsies, until adequate sensitivity and specificity for ultrasound are achieved (in the current study there was an improvement in specificity after 10 scans). A recent retrospective review of 43 patients diagnosed with GCA based on ultrasound findings allows further characterisation of patients into those

who have isolated cranial vessel involvement and those who have extracranial features.[124] Patients with extracranial disease on axillary or subclavian artery ultrasound have a lower risk of permanent blindness, but a slightly higher risk of relapse and greater steroid requirement.[19,21,36,124]

Problems with interpreting tests for giant cell arteritis

The inter-rater analysis for both tests (ultrasound and biopsy) revealed that the agreement between assessors is more variable than perhaps appreciated. The variability is significantly influenced by the degree of abnormality, as is to be expected with any test result. Borderline findings are likely to be subject to more dispute by different assessors than results showing either clear-cut abnormal appearances or clear-cut highly abnormal appearances. As demonstrated in the graphs of the inter-rater agreement (*Figures 13–16*), this problem appears to be present for both interpretations of ultrasound images as well as the evaluation of histological samples.

Biopsy has been regarded as a gold standard in diagnostic testing for many conditions including vasculitis, but when there is more uncertainty about the test results, our expectations of the pathologist or sonographer should perhaps be lowered. When we originally designed the study we were expecting a positive or negative outcome from each of the tests, so that we could compare the differences. What we have discovered is that in up to one-third of cases there is insufficient information available in the sample to determine confidently whether the diagnosis should be ruled in or ruled out. For some conditions, such as thyroid cancer, pathologists recognised that indeterminate histology was a significant problem in around 10% of cases discussed in a recent analysis of 14 studies comprising > 60,000 samples.[125] In these patients, a repeat sample was obtained and in 57% of patients the repeat biopsy was sufficient to make a definitive diagnosis. However, interestingly, in 42% of patients, a second opinion from an independent pathology review of the original sample resulted in a definitive diagnosis. Histological analysis of other conditions such as ulcerative colitis can be challenging in the presence of atypical histological features, which can lead to variations in the interpretation of diagnosis or severity of the condition.[126]

Issues with the choice of reference diagnosis for giant cell arteritis

Evaluations of diagnostic tests rely on a 'gold standard' reference diagnosis in order to determine the accuracy of the test(s) being evaluated. A reference diagnosis should ideally be independent of the test(s) being evaluated and the timing of its measurement should coincide with the timing of the test(s). No reference diagnosis exists for GCA that meets these standards. ACR classification criteria exist but these are not diagnostic criteria and they use the results of biopsy. The design of the study therefore sought an approach to determining the reference diagnosis that balanced these different limitations; neither a clinical diagnosis nor the ACR classification criteria alone were considered suitable.

We used an algorithm that took the clinician's diagnosis at 2 weeks as the starting point and this decision inevitably took account of the clinician's knowledge of the result of the biopsy. This allowed clinicians to use their judgement based on their knowledge of the patient and many biopsy-negative patients were judged to have GCA. As expected, all biopsy-positive patients were judged to have GCA. The clinician's diagnosis was confirmed as the reference diagnosis depending on consistency with the ACR classification criteria for GCA and the presence or absence of specific GCA-related symptoms or complications during follow-up. In around half of patients, a reference diagnosis was not confirmed this way. We used expert review of these patients to determine the reference diagnosis and for 23 (6%) patients the expert review confirmed a reference diagnosis that differed from the clinician's diagnosis.

Our finding that interobserver agreement in interpreting biopsy images is moderate undermines the assumption that a positive biopsy should be regarded as confirming a GCA diagnosis. The reference diagnosis is not independent of the biopsy result because it is incorporated in the clinician's judgement

and is part of the ACR classification criteria used to confirm the reference diagnosis. One implication is that the 100% specificity (and also the sensitivity) of biopsy may be overestimated and that false-positive biopsy results have not been identified. A second implication is that the performance of biopsy compared with ultrasound may also be overestimated in favour of biopsy.

The lack of independence of the reference diagnosis is also an issue for interpreting testing strategies that combine test results with clinical judgement. Clinical judgement, that is, the clinician's diagnosis at 2 weeks, is part of the test strategy, but is also the starting point for determining the reference diagnosis. Clinical judgement may also draw on patients' symptoms at presentation that also feature in the ACR classification criteria and that, in turn, may confirm the reference diagnosis. This lack of independence may therefore overestimate the performance of strategies incorporating clinical judgement. The use of expert review and GCA-related symptoms and complications during longer-term follow-up (albeit only 6 months) for confirming reference diagnoses provides some protection against this lack of independence. However, the use of emerging symptoms or complications and expert review raises the issue of timing and the possibility that the reference diagnosis is capturing newly incident GCA that was not present at the times at which biopsy and ultrasound were done. Both this timing effect and the potential for expert review to incorrectly classify a patient's true diagnosis may mean that the performance of tests and testing strategies is underestimated.

The economic modelling included additional analyses based around alternative reference standards constructed for the purpose of testing whether or not the findings could be sensitive to the reference standard criteria. Under the alternative reference standards evaluated, ultrasound in combination with clinical judgement remained a more cost-effective strategy than biopsy plus clinical judgement.

Could the results of the study be used to improve the existing service for diagnosis of suspected giant cell arteritis?

The low sensitivity of biopsy for diagnosis of GCA was one of the most surprising findings from the study. There are likely to have been several factors leading to this outcome. We could speculate on how the sensitivity of biopsy could be improved. The selection of patients could be based on a higher pre-test probability of having GCA, with careful clinical evaluation of each individual case. Patients would need to be seen promptly, either before or very shortly after commencing high-dose glucocorticoid therapy, which would require a fast-track service for these patients. The biopsy procedure should be performed by senior surgeons with expertise in the procedure. The samples should be processed and evaluated by experienced pathologists with the potential for central review of the histology. The interpretation of the biopsy should include the possibility that the result is non-diagnostic or non-specific, in order to provide more detailed results to enable the clinician to weigh up the likelihood of diagnosis in the presence of intermediate or indeterminate results. This could be achieved without recourse to ultrasound, in order to enhance the current service provision for patients with suspected GCA.

Fast-track service in giant cell arteritis

The potential window of opportunity to diagnose GCA is small once the patient has been commenced on high doses of glucocorticoid therapy. To optimise either test (ultrasound or biopsy) the important first step is to develop a rapid-access service for patients with suspected GCA. The current study provides evidence for the rapid decline in diagnostic performance with time and the economic analysis supports the introduction of ultrasound as a cost-effective means of achieving the diagnosis more effectively at much lower cost than the existing standard of care. Furthermore, fast-track services for GCA, which incorporate the use of ultrasound, have been shown to reduce the incidence of sight loss in this population, further justifying their role in the management of suspected GCA.[111,127]

Summary of findings

Giant cell arteritis or temporal arteritis remains a diagnostic and therapeutic challenge. Unfortunately, the treatment options available are relatively limited and patients usually require a very high dose of steroids for prolonged periods of time, which results in significant toxicity in > 80% of patients. If, however, the diagnosis is missed and the patient is not treated with a high dose of glucocorticoid therapy, there is a significant risk of permanent visual loss or other ischaemic complications.

The current study was performed in an attempt to explore the value of ultrasound as a diagnostic tool in assisting the management of patients with suspected GCA. Ultrasound is a readily accessible investigation in most hospitals, whereas obtaining a TAB remains problematic in the NHS. Furthermore, the diagnostic value of TAB has been questioned owing to some studies reporting low sensitivity.

Ultrasound examination of temporal arteries is a relatively specialist procedure; we wanted to explore the generalisability of diagnostic testing in GCA within a NHS setting. We therefore had to design a training programme that was effective enough and applicable enough to be generalised to clinicians and sonographers working in a variety of centres throughout the UK. We deliberately chose a mixture of district general hospitals and teaching hospitals to explore this generalisability. We trained sonographers in the technique of ultrasound examination of temporal and axillary arteries by developing a training programme based on established expertise.

The effect of the training programme was tested thoroughly by an expert review panel established specifically to view all images obtained from the main study for quality control. This ensured that the images acquired and interpreted by site sonographers were of a sufficiently high standard to be comparable to those that would have been obtained by experts.

For the main study, we needed to test the value of ultrasound as a diagnostic tool without interfering with the normal diagnostic process. We therefore designed the study so that patients underwent the normal diagnostic process if they were suspected of having GCA. This meant that they underwent a clinical assessment followed by a TAB in every case. We undertook a blinded ultrasound test before biopsy was performed, but the results of the ultrasound tests were not given to the clinician managing the patient. However, the results of the biopsy test were given to the clinician as would occur in normal practice. The results of the biopsy test, together with the clinical condition of the patient when re-evaluated 2 weeks after initial assessment, were used by the clinician to make a diagnosis. If the clinician had made a diagnosis that was not GCA and was planning to bring the patient off high doses of glucocorticoid therapy or was not planning to start high doses of steroids, we built in a safety mechanism whereby the clinicians were asked to contact the TABUL office to be given the results of the ultrasound scans just in case there was a disparity between the scan result and the clinical decision. It was then up to the clinician managing the patients to decide whether or not to alter their diagnosis and management plan, but this decision was not used as the basis for the primary outcome, although it was reported.

We asked for a 6-month follow-up visit to determine whether or not any new features consistent with the diagnosis of GCA had emerged or, indeed, whether or not any features consistent with other diagnoses had emerged and whether or not the clinician had an opportunity to change the diagnosis in 6 months. We felt that this study design was realistic and represented usual practice, but with the addition of the ultrasound scan.

Our results showed that the sensitivity of biopsy was only 39%, which was lower than in previously published studies. The sensitivity of ultrasound was 54%. The specificity of biopsy was always going to be high and in this study was 100% compared with 81% for ultrasound. As both of the tests had been performed in all patients, we were able to hypothesise on a potential sequence of tests that could have been performed to try to improve sensitivity and specificity and also to look at the cost implications of these strategies. We therefore analysed the data according to a potential strategy of performing both tests

in different combinations. However, the tests were never performed in isolation from the clinical evaluation of patients; we therefore introduced two methods to define the likelihood of the patient having or not having GCA, based on clinical features and blood test results, before looking at the results of either test. One method was simply to ask the clinicians to state their opinion of the likelihood of GCA (definite, probable or possible); the other method was to use an external data set obtained from DCVAS to try to define patients as being at high, medium or low risk of having GCA. We used the presence of jaw or tongue claudication and elevated inflammatory response (ESR > 60 mm/hour or CRP level > 40 mg/l) as parameters that would define a patient as being at a high likelihood of having GCA. Patients defined as having a low likelihood of GCA did not have any of these parameters. Patients were defined as being tat intermediate risk of having GCA if they had either claudication (of jaw or tongue) or an elevated acute-phase response (ESR or CRP level) but not both of these. Using this strategy would make clinical sense. The clinicians would normally assess the patient and decide whether or not it was worthwhile to investigate a patient further for the possibility of GCA, and therefore defining who should or should not have a test such as TAB or ultrasound.

Using this strategy-based approach we demonstrated that without a clinical evaluation, a combined approach of scanning all patients and performing a biopsy only on those for whom the scan was negative would achieve a sensitivity of 65% and specificity of 81%. If we took a risk strategy approach, by only investigating those patients with the high likelihood of GCA based on clinical presentation, the sensitivity and specificity increases to 77% and 91%, respectively. However, for patients at moderate or low risk of having GCA based on the clinical presentation, the sensitivity and specificity were lower and this would inevitably result in a cohort of patients for whom there was still a clinical suspicion of diagnosis of GCA, but for whom both ultrasound and biopsy were negative. We have demonstrated that in the TABUL study there is a significant cohort of such patients (about one-quarter of all patients defined as having GCA).

In terms of cost-effectiveness, biopsy is more expensive than ultrasound (an almost ninefold difference) and this higher cost was the key factor in the greater cost-effectiveness of strategies using ultrasound. The similarity in the diagnostic performances of the two tests (when combined with clinical judgement), the estimated impact of GCA-related complications from false-negative results, and the estimated impact of steroid toxicity from false-positive results was insufficient to alter the results.

In a parallel study, we used data from the TABUL project to measure the reliability of the interpretation of ultrasound or biopsy findings. In order to do this we produced a series of 30 patients from the TABUL cohort (containing a mixture of patients with positive and negative ultrasound and biopsy results). We prepared the ultrasound scans and the histological slides of those patients and showed a brief clinical vignette together with either the scans or histology slides to a group of sonographers and pathologists, respectively, to determine inter-rater reliability of these two tests. We found that there were similar levels of agreement, with kappa values of 0.61 for sonographers and 0.62 for pathologists. The areas of disagreement among the pathologists occurred when the histological results were less clear cut, particularly when no giant cells were found. These findings suggest that the pathologist's interpretation of biopsy material should be qualified according to the level of severity of the findings. If there are very obvious features of GCA (such as transmural inflammation or giant cells), this should be stated by pathologists, but if there are much less obvious features that might be consistent with GCA, but that equally could be consistent with normal ageing, then it is important that the pathologists are able to express this diversity of possible diagnosis rather than having to state that the biopsy is consistent with GCA but not declare that the biopsy is also consistent with normal ageing findings. The clinicians managing patients may feel less comfortable with the fact that the pathologists are not giving a clear-cut interpretation of the biopsy, but this should improve the management of patients if we avoid making conclusions based on insufficient evidence. Although no clear pattern emerged from an evaluation of cases disputed by sonographers, similar remarks would apply in interpreting ultrasound findings. Until we have a more effective diagnostic tool, the clinical evaluation of the patient remains paramount in the decision-making process.

We have challenged the place of TAB as the gold standard for the diagnosis of GCA. We have demonstrated that in 381 patients with newly suspected GCA, the application of clinical risk stratification (based on the presence of ischaemic symptoms of tongue or jaw claudication and/or an elevated acute phase response), combined with either ultrasound of temporal and axillary arteries or biopsy, will result in a high sensitivity and specificity in the diagnosis of GCA. In order to achieve this, we have created a training programme to ensure the proficiency of sonographers in performing the scans. We compared the results obtained from these scans with the traditional factors used in making a diagnosis of GCA, namely the application of clinical judgement (strongly influenced by ACR classification criteria for GCA). Despite the inherent bias of using a reference diagnosis that incorporates the results of the biopsy, ultrasound examination was more sensitive but less specific than biopsy as a diagnostic test. We tested the reliability of both tests, by asking a number of pathologists and sonographers to respectively review biopsy and scan findings from an anonymous sample of patients drawn from the cohort. We showed that the reliability of both techniques was similar (ICC of 0.61–0.62), revealing that both tests have some fallibility. Further analyses of the diagnostic strategies to combine clinical risk stratification with one or both tests in appropriate cases can be used effectively to significantly improve the diagnostic accuracy of patients with newly suspected GCA. The economic evaluation of the test strategies used in this cohort of patients has shown that an ultrasound-based approach is more cost-effective than a biopsy-based strategy if used in conjunction with clinical risk assessment.

Conclusions

Implications for health care

The inclusion of ultrasound scanning of temporal and axillary arteries can be a clinically effective and cost-effective addition to the current strategy of tests to aid in the diagnosis of GCA among individuals referred from the community to hospital. We have shown that it is practical to introduce an ultrasound training module to ensure minimum standards of proficiency in scanning temporal and axillary arteries for evidence of GCA. Ultrasound is more sensitive but less specific than biopsy. It would be possible to introduce a clinical pathway that involves scanning all patients with suspected GCA without performing biopsies. Such a strategy is clinically effective as well as cost-effective (an incremental NMB of £485 per case compared with standard current practice of biopsy and clinical judgement) and avoids an invasive biopsy procedure. However, the strategy will be successful only if patients have rapid access to the diagnostic pathway while the scan abnormalities are still present (and not affected by the effects of glucocorticoid therapy).

It will be important to define the acceptability of any new diagnostic strategies in the management of GCA both for patients and for clinicians. If we follow the most cost-effective strategy, we would rely on ultrasound and clinical judgement alone as a means of diagnosis. Some clinicians and patients may be uncomfortable with this strategy and may prefer that biopsies are performed in all cases that are ultrasound negative or in all cases with medium or high risk of GCA on clinical features but in which there has been a negative ultrasound scan. The reason for the additional use of biopsy, despite a negative scan, would be to provide further evidence to rule in the disease as well as to support withdrawing therapy in the event that both tests are negative. Although these combined strategies would be more expensive, they remain more cost-effective than current practice (performing a biopsy in all suspected cases) and may be more acceptable to patients and clinicians.

Recommendations for research

The current study has challenged the previously secure place of biopsy as the gold standard in the diagnosis of GCA. Although ultrasound may not be the perfect replacement for biopsy, it has significant advantages over biopsy, as well as some limitations, as discussed in this report. The following areas would merit further exploration:

- What should be the gold standard for diagnosis of GCA? Can our assessment of the pathological findings be improved, removing the previously used dichotomous decision on normal or abnormal findings, to generate a grade of likelihood of diagnosis, especially for those patients whose biopsies do

not contain giant cells? Do we need to develop a training programme for pathologists to maintain standardisation in the reporting of findings in GCA? Do we need to re-examine surgical training in performing biopsies for patients with suspected GCA? Do we need to re-evaluate the histological characteristics that define the presence or absence of GCA? How can we better account for the influences that alter the histological findings in the temporal artery, such as the presence of arteriosclerosis, the effect of glucocorticoid therapy and the effects of ageing? Should a hierarchical approach to diagnosis be developed, with clinical features, laboratory features, ultrasound and biopsy evaluated to develop an algorithmic approach to standardise the investigation and evaluation of patients with suspected GCA?

- Are biomarkers available to improve the diagnostic certainty in GCA? Many groups have attempted to introduce alternative tests to increase the diagnostic yield in GCA. Assessments of the ESR, circulating levels of CRP, vascular endothelial growth factor or pentraxin 3 have been tested and were found to lack sensitivity and specificity for the diagnosis of GCA; however, could they add value in the diagnosis of GCA if combined with ultrasound? Are there any new biomarkers to be tested in suspected GCA?
- Can ultrasound examination of temporal arteries be used to guide responses to therapy? If ultrasound becomes more widely used than biopsy, this provides a new opportunity to assess ultrasound as a biomarker to measure the response of the scan findings to the effects of therapies. This could mean allowing more rapid reduction of glucocorticoid therapy for those cases showing a fast resolution, with rapid reintroduction for cases in which the scan abnormality is returning, either in the context of a clinical relapse or in patients who are asymptomatic. How often must a scan be repeated to evaluate this risk? In addition, adjunctive therapies (steroid-sparing agents) could be tested for their role in resolving and maintaining a normal ultrasound appearance of the arteries. The rate of ultrasound response to treatment might be a guide to future risk of relapse.
- How can we improve the standardisation of ultrasound assessment of suspected GCA? We have developed and introduced a novel training protocol for the ultrasound of temporal and axillary arteries. We applied the training protocol to all centres included in the study, most of which had never performed vascular ultrasound before. With the benefit of this training protocol, we observed that 86% of recorded images from the patients recruited into the study were technically satisfactory. It is possible that the training methods that we devised could have been improved by being tailored to the expertise of the sonographer. As technology advances, it is possible that the amount of training required to adequately prepare a sonographer to examine these arteries may decrease. In addition, as ultrasound becomes more widely used, some sites may become more familiar with the techniques, and their training requirements will be reduced. As more patients are scanned, more experience can be gained to maintain standards. Testing new training methods should be considered, as well as the development of methods to maintain expertise.
- How should we explore the acceptability of introducing new combined diagnostic strategies into clinical practice? How will clinicians respond to the idea that they should no longer be requesting a biopsy in the majority of cases to rule in or rule out GCA? Will they have confidence in the clinical features plus scan evidence of having GCA? Will they or their patients be willing to accept a diagnosis without a histopathological test to verify the diagnosis? By contrast, how acceptable will a negative scan be in ruling out the diagnosis? We shall need to consider the impact of these changes to a well-established diagnostic pathway, especially in centres that do not have any experience of using ultrasound in the assessment of GCA.

Acknowledgements

Contributions of authors

Raashid Luqmani (Chief Investigator) conceived the study design, contributed to the management and conduct of the study, acted as a reviewer for reference diagnosis and drafted and revised the report.

Ellen Lee planned and conducted the statistical analysis, drafted and revised *Chapters 4* and *5* and reviewed a draft of the report.

Surjeet Singh (Trial Manager) contributed to the design, conduct and management of the study and contributed to drafting and revising the report.

Mike Gillett planned and conducted the cost-effectiveness analysis, drafted and revised *Chapter 7* and reviewed a draft of the report.

Wolfgang A Schmidt contributed to the design and conduct of the study, was involved in ultrasound training and expert review and reviewed a draft of the report.

Mike Bradburn contributed to the design and conduct of the study, planned and supervised the statistical analysis and reviewed a draft of the report.

Bhaskar Dasgupta contributed to the design and conduct of the study was involved in ultrasound training and expert review and reviewed a draft of the report.

Andreas P Diamantopoulos contributed to the conduct of the study, was involved in ultrasound expert review and reviewed a draft of the report.

Wulf Forrester-Barker contributed to the acquisition of images, contributed to the design of the expert review, was involved in agreement and clinical vignette components and reviewed a draft of the report.

William Hamilton contributed to the design and conduct of the study, provided primary care expertise and reviewed a draft of the report.

Shauna Masters contributed to the management and conduct of the study, contributed to the design of the agreement and clinical vignette components and contributed to drafting and revising report.

Brendan McDonald contributed to the design and conduct of the study, contributed to the pathology review and agreement component and reviewed a draft of the report.

Eugene McNally contributed to the design and conduct of the study; was involved in ultrasound training and expert review and reviewed a draft of the report.

Colin Pease contributed to the design and conduct of the study, was a reviewer for reference diagnosis and reviewed a draft of the report.

Jennifer Piper contributed to conduct of study, contributed to the ultrasound training, review, image acquisition and agreement component and contributed to drafting and revising the report.

John Salmon contributed to the design and conduct of the study, provided ophthalmology expertise and reviewed a draft of the report.

Allan Wailoo contributed to the design and conduct of the study, planned and supervised the cost-effectiveness analysis and reviewed a draft of the report.

Konrad Wolfe contributed to the design and conduct of the study, contributed to the pathology review and agreement component and reviewed a draft of the report.

Andrew Hutchings conceived the study design, contributed to the management and conduct of the study, planned the statistical analysis and conducted the agreement analysis and drafted and revised the report.

Independent members of the Trial Steering Committee

Professor Michael Ehrenstein (Chairperson), Department of Medicine, UCL, London, UK.

Professor Bleddyn Davies (Patient Representative).

Professor Karim Raza, Division of Immunity and Infection, University of Birmingham, Birmingham, UK.

Professor David Mant, Department of Primary Health Care, University of Oxford, Oxford, UK.

Independent members of the Data Monitoring Committee

Dr Lyn Williamson (Chairperson), Rheumatology Department, Great Western Hospital, Great Western Hospitals NHS Foundation Trust, Swindon, UK.

Dr Kate Gilbert (Patient Representative), Polymyalgia Rheumatica & Giant Cell Arteritis UK, c/o Birmingham Arthritis Resource Centre, Central Library, Birmingham, UK.

Dr Simon Travis, Gastroenterology Unit, John Radcliffe Hospital, Oxford, UK.

Professor Jonathan Sterne, University of Bristol, Bristol, UK.

Data management, Sheffield

Amanda Loban and Christopher Ellis.

Study co-ordination, Oxford

Jana Vaskova, Keri Fathers, Leo Marcus-Wan, Nicola Farrar, Vanshika Sharma, Varun Manhas, Connor Scott, Nicky Sullivan, Denise Brown, Gareth Bicknell, Karolina Kliskey and Wulf Forrester-Barker.

Data entry, Oxford

David Gray, Samiya Mahmood, Ann-Marie Morgan and Ifzal Ahmed.

TABUL sonographers group

Paulo Batista, Andrew Beech, Emmanuel Grimley, Eric Heffernan, Laura Horton, Rainer Klocke, Thomas Neumann, Janaki Praharaju, Borsha Sarker, Ernest Wong, Steven Young Min, Ahmed Zayat, Bhaskar Dasgupta, Andreas Diamantopoulos, Eugene McNally, Jennifer Piper and Wolfgang Schmidt.

TABUL pathologists group

Salam Al-Sam, Rolf Bie, Aruna Chakrabarty, Diane Hemming, Margaret Jeffrey, Susan Kennedy, Tong Lioe, James Lowe, Madhavi Maheshwari, Fawaz Musa, Simon Payne, Ian Scott, Gerhard van Schalkwyk, Amgad Youssef, Brendan McDonald and Konrad Wolfe.

Review group for reference diagnosis

Ruth Geraldes, Raashid Luqmani, Sarah Mackie, Lorraine O'Neill and Colin Pease.

Clinical vignettes panel members

Kuljeet Bhamra, Andreas Diamantopoulos, Ruth Geraldes, Raashid Luqmani, Sarah Mackie, Damodar Makkuni, Eamonn Molloy, Geirmund Myklebust, Lorraine O'Neill, Colin Pease, Cristina Ponte, Wolfgang Schmidt, Richard Watts and Vadivelu Saravanan.

Staff at participating hospitals

Chapel Allerton Hospital, Leeds, UK

Colin Pease (Principal Investigator).

Sarah Mackie and Ann Morgan (Clinicians).

Oliver Wordsworth (Research Nurse).

Ahmed Zayat and Laura Horton (Sonographers).

Aruna Chakrabarty (Pathologist).

City Hospital, Birmingham, UK

David Carruthers (Principal Investigator).

Biruk Asfaw and Adam Slater (Research Nurses).

Andrew Filer (Sonographer).

Madhavi Maheshwari and Syi Yum Chan (Pathologists).

Dudley Hospital, Dudley, UK

Rainer Klocke (Principal Investigator and Sonographer).

George Hirsch (Clinician).

Chitra Ramful (Research Nurse).

Amgad Youssef and Sixto Batitang (Pathologists).

Gateshead Hospital, Gateshead, UK

Vadivelu Saravanan (Principal Investigator).

Susan Pugmire (Research Nurse).

Borsha Sarker and Ian Pearson (Sonographers).

Diane Hemmingway (Pathologist).

Great Yarmouth Hospital, Great Yarmouth, UK

Damodar Makkuni (Principal Investigator).

Celia Whitehouse (Research Nurse).

Srinivas Boddu (Sonographer).

Rajesh Logansundaram (Pathologist).

Hospital de Santa Maria, Lisbon, Portugal

Ruth Geraldes (Principal Investigator).

Paulo Batista and Vanessa Almeida (Sonographers).

Artur Costa Silva (Pathologist).

Ana Lopes (other).

Hospital of Southern Norway, Kristiansand, Norway

Andreas Diamantopoulos (Principal Investigator and Sonographer).

Geirmund Myklebust (Clinician).

Hanne Vestaby (Research Nurse).

Helene Hetland (Sonographer).

Rolf Bie (Pathologist).

Jena University Hospital, Jena, Germany

Thomas Neumann (Principal Investigator and Sonographer).

Anne Welzel and Peter Oelzner (Clinicians).

Iver Petersen (Pathologist).

Kristin Knoll (other).

Musgrave Park Hospital, Belfast, UK

Adrian Pendleton (Principal Investigator).

Georgina Sterrett and Rebecca Denham (Research Nurses).

Emmanuel Grimley (Sonographer).

Tong Lioe (Pathologist).

Nuffield Orthopaedic Centre, Oxford, UK

Raashid Luqmani (Principal Investigator).

Shauna Masters, Joanna Burchall, Elizabeth Nugent, Pam Lovegrove and Adedoyin Adetunji (Research Nurses).

Eugene McNally, Jennifer Piper (Sonographers).

Brendan McDonald (Pathologist).

Nicky Sullivan (other).

John Salmon (Principal Investigator, John Radcliffe Hospital).

Alexina Fantato (Research Nurse, John Radcliffe Hospital).

Princess Alexandra Hospital, Harlow, UK

Khalid Ahmed (Principal Investigator).

Nikki White, Lucy Brown and Carol Keel (Research Nurses).

Janaki Praharaju (Sonographer).

Salam Al-Sam and Evdokia Arkoumani (Pathologists).

Queen Alexandra Hospital, Portsmouth, UK

Richard Hull (Principal Investigator).

Ernest Wong and Steven Young Min (Clinicians and Sonographers).

Paula White and Julie Williams (Research Nurses).

Margaret Jeffrey (Pathologist).

Holly Pickering and Didem Agdiran (other).

Queen's Hospital, Romford, UK

Kuntal Chakravarty (Principal Investigator).

Laith Al-Sweedan and Niamh Quillinan (Clinicians).

Doris Butawan and Caron Baldwin (Research Nurses).

Faisal Alyas (Sonographer).

Ibitsam Saeed (Pathologist).

Queen's Medical Centre, Nottingham, UK

Peter Lanyon (Principal Investigator).

Fiona Pearce and Frances Rees (Clinicians).

Alan Thomas (Research Nurse).

Andrew Beech (Sonographer).

Keith Robson, Ian Scott and James Lowe (Pathologists).

Amanda Butler (other).

Royal Berkshire Hospital, Reading, UK

Antoni Chan (Principal Investigator).

Tinashe Samakomva, Julie Foxton, Linda Jones and Jennie King (Research Nurses).

Jennifer Piper (Sonographer).

Fawaz Musa (Pathologist).

Natalie Mears (other).

Royal Derby Hospital, Derby, UK

Nicholas Raj (Principal Investigator).

Louisa Badcock and Chris Deighton (Clinicians).

Alison Booth (Research Nurse).

Nicholas Raj (Sonographer).

Gerhard van Schalkwyk (Pathologist).

Jo Sherman (other).

Southend University Hospital, Southend, UK

Frances Borg (Principal Investigator).

Dimitrios Christidis, Pravin Patil, Tochi Adizie, Win Win Maw, Katerina Achilleos, Mark Williams, Mirjana Dojcinovska, Nada Hassan and Sunhoury Elsideeg (Clinicians).

Dawn Gayford and Daniela Boteanu (Research Nurses).

Bhaskar Dasgupta (Sonographer).

Konrad Wolfe, John Davies, Simon Payne and Vasudev Sharma (Pathologists).

Robert Toplis (other).

St Vincent's University Hospital, Dublin, Ireland

Eamonn Molloy (Principal Investigator).

Lorraine O'Neill (Clinician).

Phil Gallagher (Research Nurse).

Eric Heffernan (Sonographer).

Susan Kennedy (Pathologist).

Stoke Mandeville Hospital, Stoke, UK

Malgorzata Magliano (Principal Investigator).

Anna Mistry, Azeem Ahmed and Jasroop Chana (Clinicians).

Ursula Perks (Research Nurse).

Jennifer Piper (Sonographer).

Heman Manikkapurath and Vidhi Bhargava (Pathologists).

Sunderland Royal Hospital, Sunderland, UK

David Wright (Principal Investigator).

David Coady (Clinician).

Tracey Robson (Research Nurse).

Simon England (Sonographer).

Deborah Milne (Pathologist).

Data sharing statement

Data and samples can be obtained from the corresponding author; these will be provided for only those participants who have provided explicit consent for this purpose. Use of such data and samples would need to be justified by a formal application to the corresponding author for ethically approved studies.

References

1. Horton BT, Magath TB, Brown GE. An undescribed form of arteritis of the temporal vessels. *Proc Staff Meet Mayo Clin* 1932;**7**:700–1.
2. Smeeth L, Cook C, Hall AJ. Incidence of diagnosed polymyalgia rheumatica and temporal arteritis in the United Kingdom, 1990-2001. *Ann Rheum Dis* 2006;**65**:1093–8. http://dx.doi.org/10.1136/ard.2005.046912
3. Mohammad AJ, Nilsson JÅ, Jacobsson LT, Merkel PA, Turesson C. Incidence and mortality rates of biopsy-proven giant cell arteritis in southern Sweden. *Ann Rheum Dis* 2015;**74**:993–7. http://dx.doi.org/10.1136/annrheumdis-2013-204652
4. Chandran AK, Udayakumar PD, Crowson CS, Warrington KJ, Matteson EL. The incidence of giant cell arteritis in Olmsted County, Minnesota, over a 60-year period 1950–2009. *Scand J Rheumatol* 2015;**44**:215–18. http://dx.doi.org/10.3109/03009742.2014.982701
5. Dasgupta B, Borg FA, Hassan N, Alexander L, Barraclough K, Bourke B, *et al.* BSR and BHPR guidelines for the management of giant cell arteritis. *Rheumatology* 2010;**49**:1594–7. http://dx.doi.org/10.1093/rheumatology/keq039a
6. Schmidt WA, Gromnica-Ihle E. Incidence of temporal arteritis in patients with polymyalgia rheumatica: a prospective study using colour Doppler ultrasonography of the temporal arteries. *Rheumatology* 2002;**41**:46–52. http://dx.doi.org/10.1093/rheumatology/41.1.46
7. Ashton-Key MR, Gallagher PJ. False-negative temporal artery biopsy. *Am J Surg Pathol* 1992;**16**:634–5. http://dx.doi.org/10.1097/00000478-199206000-00014
8. Schmidt WA, Kraft HE, Völker L, Vorpahl K, Gromnica-Ihle EJ. Colour Doppler sonography to diagnose temporal arteritis. *Lancet* 1995;**345**:866. http://dx.doi.org/10.1016/S0140-6736(95)93005-1
9. Schmidt WA, Kraft HE, Vorpahl K, Völker L, Gromnica-Ihle EJ. Color duplex ultrasonography in the diagnosis of temporal arteritis. *N Engl J Med* 1997;**337**:1336–42. http://dx.doi.org/10.1056/NEJM199711063371902
10. Bley TA, Wieben O, Vaith P, Schmidt D, Ghanem NA, Langer M. Magnetic resonance imaging depicts mural inflammation of the temporal artery in giant cell arteritis. *Arthritis Rheum* 2004;**51**:1062–3. http://dx.doi.org/10.1002/art.20840
11. Diamantopoulos AP, Haugeberg G, Hetland H, Soldal DM, Bie R, Myklebust G. Diagnostic value of color Doppler ultrasonography of temporal arteries and large vessels in giant cell arteritis: a consecutive case series. *Arthritis Care Res* 2014;**66**:113–19. http://dx.doi.org/10.1002/acr.22178
12. Siemonsen S, Brekenfeld C, Holst B, Kaufmann-Buehler AK, Fiehler J, Bley TA. 3T MRI reveals extra- and intracranial involvement in giant cell arteritis. *AJNR Am J Neuroradiol* 2015;**36**:91–7. http://dx.doi.org/10.3174/ajnr.A4086
13. Klink T, Geiger J, Both M, Ness T, Heinzelmann S, Reinhard M, *et al.* Giant cell arteritis: diagnostic accuracy of MR imaging of superficial cranial arteries in initial diagnosis-results from a multicenter trial. *Radiology* 2014;**273**:844–52. http://dx.doi.org/10.1148/radiol.14140056
14. Puppo C, Massollo M, Paparo F, Camellino D, Piccardo A, Shoushtari Zadeh Naseri M, *et al.* Giant cell arteritis: a systematic review of the qualitative and semiquantitative methods to assess vasculitis with 18F-fluorodeoxyglucose positron emission tomography. *Biomed Res Int* 2014;**2014**:574248. http://dx.doi.org/10.1155/2014/574248

15. Prieto-González S, Arguis P, García-Martínez A, Espígol-Frigolé G, Tavera-Bahillo I, Butjosa M, *et al.* Large vessel involvement in biopsy-proven giant cell arteritis: prospective study in 40 newly diagnosed patients using CT angiography. *Ann Rheum Dis* 2012;**71**:1170–6. http://dx.doi.org/10.1136/annrheumdis-2011-200865

16. Karassa FB, Matsagas MI, Schmidt WA, Ioannidis JP. Meta-analysis: test performance of ultrasonography for giant-cell arteritis. *Ann Intern Med* 2005;**142**:359–69. http://dx.doi.org/10.7326/0003-4819-142-5-200503010-00011

17. Arida A, Kyprianou M, Kanakis M, Sfikakis PP. The diagnostic value of ultrasonography-derived edema of the temporal artery wall in giant cell arteritis: a second meta-analysis. *BMC Musculoskelet Disord* 2010;**11**:44. http://dx.doi.org/10.1186/1471-2474-11-44

18. Ball EL, Walsh SR, Tang TY, Gohil R, Clarke JM. Role of ultrasonography in the diagnosis of temporal arteritis. *Br J Surg* 2010;**97**:1765–71. http://dx.doi.org/10.1002/bjs.7252

19. Schmidt WA, Seifert A, Gromnica-Ihle E, Krause A, Natusch A. Ultrasound of proximal upper extremity arteries to increase the diagnostic yield in large-vessel giant cell arteritis. *Rheumatology* 2008;**47**:96–101. http://dx.doi.org/10.1093/rheumatology/kem322

20. Dunstan E, Lester SL, Rischmueller M, Dodd T, Black R, Ahern M, *et al.* Epidemiology of biopsy-proven giant cell arteritis in South Australia. *Intern Med J* 2014;**44**:32–9. http://dx.doi.org/10.1111/imj.12293

21. Muratore F, Kermani TA, Crowson CS, Green AB, Salvarani C, Matteson EL, Warrington KJ. Large-vessel giant cell arteritis: a cohort study. *Rheumatology* 2015;**54**:463–70. http://dx.doi.org/10.1093/rheumatology/keu329

22. Souza AW, Okamoto KY, Abrantes F, Schau B, Bacchiega AB, Shinjo SK. Giant cell arteritis: a multicenter observational study in Brazil. *Clinics* 2013;**68**:317–22. http://dx.doi.org/10.6061/clinics/2013(03)OA06

23. Graham E, Holland A, Avery A, Russell RW. Prognosis in giant-cell arteritis. *Br Med J* 1981;**282**:269–71. http://dx.doi.org/10.1136/bmj.282.6260.269

24. González-Gay MA, Blanco R, Rodríguez-Valverde V, Martínez-Taboada VM, Delgado-Rodriguez M, Figueroa M, Uriarte E. Permanent visual loss and cerebrovascular accidents in giant cell arteritis: predictors and response to treatment. *Arthritis Rheum* 1998;**41**:1497–504. http://dx.doi.org/10.1002/1529-0131(199808)41:8<1497::AID-ART22>3.0.CO;2-Z

25. Liozon E, Herrmann F, Ly K, Robert PY, Loustaud V, Soria P, Vidal E. Risk factors for visual loss in giant cell (temporal) arteritis: a prospective study of 174 patients. *Am J Med* 2001;**111**:211–17. http://dx.doi.org/10.1016/S0002-9343(01)00770-7

26. Salvarani C, Cimino L, Macchioni P, Consonni D, Cantini F, Bajocchi G, *et al.* Risk factors for visual loss in an Italian population-based cohort of patients with giant cell arteritis. *Arthritis Rheum* 2005;**53**:293–7. http://dx.doi.org/10.1002/art.21075

27. Chatelain D, Duhaut P, Schmidt J, Loire R, Bosshard S, Guernou M, *et al.* Pathological features of temporal arteries in patients with giant cell arteritis presenting with permanent visual loss. *Ann Rheum Dis* 2009;**68**:84–8. http://dx.doi.org/10.1136/ard.2007.084947

28. Singh AG, Kermani TA, Crowson CS, Weyand CM, Matteson EL, Warrington KJ. Visual manifestations in giant cell arteritis: trend over 5 decades in a population-based cohort. *J Rheumatol* 2015;**42**:309–15. http://dx.doi.org/10.3899/jrheum.140188

29. González-Gay MA, Vazquez-Rodriguez TR, Gomez-Acebo I, Pego-Reigosa R, Lopez-Diaz MJ, Vazquez-Triñanes MC, *et al.* Strokes at time of disease diagnosis in a series of 287 patients with biopsy-proven giant cell arteritis. *Medicine* 2009;**88**:227–35. http://dx.doi.org/10.1097/MD.0b013e3181af4518

30. Zenone T, Puget M. Characteristics of cerebrovascular accidents at time of diagnosis in a series of 98 patients with giant cell arteritis. *Rheumatol Int* 2013;**33**:3017–23. http://dx.doi.org/10.1007/s00296-013-2814-0

31. Weyand CM, Goronzy JJ. Immune mechanisms in medium and large-vessel vasculitis. *Nat Rev Rheumatol* 2013;**9**:731–40. http://dx.doi.org/10.1038/nrrheum.2013.161

32. Dasgupta B, Cimmino MA, Kremers HM, Schmidt WA, Schirmer M, Salvarani C, *et al.* 2012 Provisional classification criteria for polymyalgia rheumatica: a European League Against Rheumatism/American College of Rheumatology collaborative initiative. *Arthritis Rheum* 2012;**64**:943–54. http://dx.doi.org/10.1002/art.34356

33. González-Gay MA, Pina T. Giant cell arteritis and polymyalgia rheumatica: an update. *Curr Rheumatol Rep* 2015;**17**:6. http://dx.doi.org/10.1007/s11926-014-0480-1

34. Hunder GG, Bloch DA, Michel BA, Stevens MB, Arend WP, Calabrese LH, *et al.* The American College of Rheumatology 1990 criteria for the classification of giant cell arteritis. *Arthritis Rheum* 1990;**33**:1122–8. http://dx.doi.org/10.1002/art.1780330810

35. Anon. Polymyalgia arteritica. *Br Med J* 1977;**1**:1046–7. http://dx.doi.org/10.1136/bmj.1.6068.1046

36. Kisza K, Murchison AP, Dai Y, Bilyk JR, Eagle RC, Sergott R, Savino PJ. Giant cell arteritis incidence: analysis by season and year in mid-Atlantic United States. *Clin Experiment Ophthalmol* 2013;**41**:577–81. http://dx.doi.org/10.1111/ceo.12069

37. Machado EB, Michet CJ, Ballard DJ, Hunder GG, Beard CM, Chu CP, O'Fallon WM. Trends in incidence and clinical presentation of temporal arteritis in Olmsted County, Minnesota, 1950–1985. *Arthritis Rheum* 1988;**31**:745–9. http://dx.doi.org/10.1002/art.1780310607

38. González-Gay MA, Miranda-Filloy JA, Lopez-Diaz MJ, Perez-Alvarez R, Gonzalez-Juanatey C, Sanchez-Andrade A, *et al.* Giant cell arteritis in northwestern Spain: a 25-year epidemiologic study. *Medicine* 2007;**86**:61–8. http://dx.doi.org/10.1097/md.0b013e31803d1764

39. Petursdottir V, Johansson H, Nordborg E, Nordborg C. The epidemiology of biopsy-positive giant cell arteritis: special reference to cyclic fluctuations. *Rheumatology* 1999;**38**:1208–12. http://dx.doi.org/10.1093/rheumatology/38.12.1208

40. Rodondi N, den Elzen WP, Bauer DC, Cappola AR, Razvi S, Walsh JP, *et al.* Subclinical hypothyroidism and the risk of coronary heart disease and mortality. *JAMA* 2010;**304**:1365–74. http://dx.doi.org/10.1001/jama.2010.1361

41. Craven A, Robson J, Ponte C, Grayson PC, Suppiah R, Judge A, *et al.* ACR/EULAR-endorsed study to develop Diagnostic and Classification Criteria for Vasculitis (DCVAS). *Clin Exp Nephrol* 2013;**17**:619–21. http://dx.doi.org/10.1007/s10157-013-0854-0

42. 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. http://dx.doi.org/10.1002/art.11388

43. Durand M, Thomas SL. Incidence of infections in patients with giant cell arteritis: a cohort study. *Arthritis Care Res* 2012;**64**:581–8. http://dx.doi.org/10.1002/acr.21569

44. Danesh-Meyer HV, Savino PJ, Eagle RC, Kubis KC, Sergott RC. Low diagnostic yield with second biopsies in suspected giant cell arteritis. *J Neuroophthalmol* 2000;**20**:213–15. http://dx.doi.org/10.1097/00041327-200020030-00011

45. Breuer GS, Nesher G, Nesher R. Rate of discordant findings in bilateral temporal artery biopsy to diagnose giant cell arteritis. *J Rheumatol* 2009;**36**:794–6. http://dx.doi.org/10.3899/jrheum.080792

46. Allison MC, Gallagher PJ. Temporal artery biopsy and corticosteroid treatment. *Ann Rheum Dis* 1984;**43**:416–17. http://dx.doi.org/10.1136/ard.43.3.416

47. Cox M, Gilks B. Healed or quiescent temporal arteritis versus senescent changes in temporal artery biopsy specimens. *Pathology* 2001;**33**:163–6. http://dx.doi.org/10.1080/00313020120038764

48. Le K, Bools LM, Lynn AB, Clancy TV, Hooks WB, Hope WW. The effect of temporal artery biopsy on the treatment of temporal arteritis. *Am J Surg* 2015;**209**:338–41. http://dx.doi.org/10.1016/j.amjsurg.2014.07.007

49. Guffey Johnson J, Grossniklaus HE, Margo CE, Foulis P. Frequency of unintended vein and peripheral nerve biopsy with temporal artery biopsy. *Arch Ophthalmol* 2009;**127**:703. http://dx.doi.org/10.1001/archophthalmol.2009.77

50. Chakrabarty A, Franks AJ. Temporal artery biopsy: is there any value in examining biopsies at multiple levels? *J Clin Pathol* 2000;**53**:131–6. http://dx.doi.org/10.1136/jcp.53.2.131

51. Murchison AP, Bilyk JR, Eagle RC, Savino PJ. Shrinkage revisited: how long is long enough? *Ophthal Plast Reconstr Surg* 2012;**28**:261–3. http://dx.doi.org/10.1097/IOP.0b013e31824ee720

52. Ypsilantis E, Courtney ED, Chopra N, Karthikesalingam A, Eltayab M, Katsoulas N, *et al.* Importance of specimen length during temporal artery biopsy. *Br J Surg* 2011;**98**:1556–60. http://dx.doi.org/10.1002/bjs.7595

53. Kaptanis S, Perera JK, Halkias C, Caton N, Alarcon L, Vig S. Temporal artery biopsy size does not matter. *Vascular* 2014;**22**:406–10. http://dx.doi.org/10.1177/1708538113516322

54. Cavazza A, Muratore F, Boiardi L, Restuccia G, Pipitone N, Pazzola G, *et al.* Inflamed temporal artery: histologic findings in 354 biopsies, with clinical correlations. *Am J Surg Pathol* 2014;**38**:1360–70. http://dx.doi.org/10.1097/PAS.0000000000000244

55. Disdier P, Pellissier JF, Harle JR, Figarella-Branger D, Bolla G, Weiller PJ. Significance of isolated vasculitis of the vasa vasorum on temporal artery biopsy. *J Rheumatol* 1994;**21**:258–60.

56. Esteban MJ, Font C, Hernandez-Rodriguez J, Valls-Sole J, Sanmarti R, Cardellach F, *et al.* Small-vessel vasculitis surrounding a spared temporal artery: clinical and pathological findings in a series of twenty-eight patients. *Arthritis Rheum* 2001;**44**:1387–95. http://dx.doi.org/10.1002/1529-0131(200106)44:6<1387::AID-ART232>3.0.CO;2-B

57. Hamidou MA, Moreau A, Toquet C, El Kouri D, de Faucal P, Grolleau JY. Temporal arteritis associated with systemic necrotizing vasculitis. *J Rheumatol* 2003;**30**:2165–9.

58. Restuccia G, Cavazza A, Boiardi L, Pipitone N, Macchioni P, Bajocchi G, *et al.* Small-vessel vasculitis surrounding an uninflamed temporal artery and isolated vasa vasorum vasculitis of the temporal artery: two subsets of giant cell arteritis. *Arthritis Rheum* 2012;**64**:549–56. http://dx.doi.org/10.1002/art.33362

59. Genereau T, Lortholary O, Pottier MA, Michon-Pasturel U, Ponge T, de Wazieres B, *et al.* Temporal artery biopsy: a diagnostic tool for systemic necrotizing vasculitis. French Vasculitis Study Group. *Arthritis Rheum* 1999;**42**:2674–81. http://dx.doi.org/10.1002/1529-0131(199912)42:12<2674::AID-ANR25>3.0.CO;2-A

60. Mukhtyar C, Guillevin L, Cid MC, Dasgupta B, de Groot K, Gross W, *et al.* EULAR recommendations for the management of large vessel vasculitis. *Ann Rheum Dis* 2009;**68**:318–23. http://dx.doi.org/10.1136/ard.2008.088351

61. Karahaliou M, Vaiopoulos G, Papaspyrou S, Kanakis MA, Revenas K, Sfikakis PP. Colour duplex sonography of temporal arteries before decision for biopsy: a prospective study in 55 patients with suspected giant cell arteritis. *Arthritis Res Ther* 2006;**8**:R116. http://dx.doi.org/10.1186/ar2003

62. Germanò G, Muratore F, Cimino L, Lo Gullo A, Possemato N, Macchioni P, *et al.* Is colour duplex sonography-guided temporal artery biopsy useful in the diagnosis of giant cell arteritis? A randomized study. *Rheumatology* 2015;**54**:400–4. http://dx.doi.org/10.1093/rheumatology/keu241

63. Aschwanden M, Daikeler T, Kesten F, Baldi T, Benz D, Tyndall A, *et al.* Temporal artery compression sign – a novel ultrasound finding for the diagnosis of giant cell arteritis. *Ultraschall Med* 2013;**34**:47–50. http://dx.doi.org/10.1055/s-0032-1312821

64. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford: Oxford University Press; 2003.

65. EuroQol Group. EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy* 1998;**16**:199–208.

66. Luqmani RA, Bacon PA, Moots RJ, Janssen BA, Pall A, Emery P, *et al.* Birmingham Vasculitis Activity Score (BVAS) in systemic necrotizing vasculitis. *QJM* 1994;**87**:671–8.

67. Mukhtyar C, Lee R, Brown D, Carruthers D, Dasgupta B, Dubey S, *et al.* Modification and validation of the Birmingham Vasculitis Activity Score (version 3). *Ann Rheum Dis* 2009;**68**:1827–32. http://dx.doi.org/10.1136/ard.2008.101279

68. Luqmani R. Maintenance of clinical remission in ANCA-associated vasculitis. *Nat Rev Rheumatol* 2013;**9**:127–32. http://dx.doi.org/10.1038/nrrheum.2012.188

69. Exley AR, Bacon PA, Luqmani RA, Kitas GD, Gordon C, Savage CO, Adu D. Development and initial validation of the Vasculitis Damage Index for the standardized clinical assessment of damage in the systemic vasculitides. *Arthritis Rheum* 1997;**40**:371–80. http://dx.doi.org/10.1002/art.1780400222

70. Suppiah R, Flossman O, Mukhtyar C, Alberici F, Baslund B, Brown D, *et al.* Measurement of damage in systemic vasculitis: a comparison of the Vasculitis Damage Index with the Combined Damage Assessment Index. *Ann Rheum Dis* 2011;**70**:80–5. http://dx.doi.org/10.1136/ard.2009.122952

71. Fitch K. *The Rand/UCLA Appropriateness Method User's Manual*. Santa Monica, CA: Rand; 2001.

72. Piram M, Maldini C, Mahr A. Effect of race/ethnicity on risk, presentation and course of connective tissue diseases and primary systemic vasculitides. *Curr Opin Rheumatol* 2012;**24**:193–200. http://dx.doi.org/10.1097/BOR.0b013e32835059e5

73. Boghen DR, Glaser JS. Ischaemic optic neuropathy. The clinical profile and history. *Brain* 1975;**98**:689–708. http://dx.doi.org/10.1093/brain/98.4.689

74. Ikard RW. Clinical efficacy of temporal artery biopsy in Nashville, Tennessee. *South Med J* 1988;**81**:1222–4. http://dx.doi.org/10.1097/00007611-198810000-00005

75. Albertini JG, Ramsey ML, Marks VJ. Temporal artery biopsy in a dermatologic surgery practice. *Dermatol Surg* 1999;**25**:501–8. http://dx.doi.org/10.1046/j.1524-4725.1999.08296.x

76. Yoon MK, Horton JC, McCulley TJ. Facial nerve injury: a complication of superficial temporal artery biopsy. *Am J Ophthalmol* 2011;**152**:251–5.e1. http://dx.doi.org/10.1016/j.ajo.2011.02.003

77. Murchison AP, Bilyk JR. Brow ptosis after temporal artery biopsy: incidence and associations. *Ophthalmology* 2012;**119**:2637–42. http://dx.doi.org/10.1016/j.ophtha.2012.07.020

78. Niederkohr RD, Levin LA. Management of the patient with suspected temporal arteritis a decision-analytic approach. *Ophthalmology* 2005;**112**:744–56. http://dx.doi.org/10.1016/j.ophtha.2005.01.031

79. Elliot DL, Watts WJ, Reuler JB. Management of suspected temporal arteritis. A decision analysis. *Med Decis Making* 1983;**3**:63–8. http://dx.doi.org/10.1177/0272989X8300300112

80. Patil P, Achilleos K, Williams M, Maw W, Dejaco C, Borg F, *et al.* Outcomes and cost effectiveness analysis of fast track pathway in giant cell arteritis. *Rheumatology* 2014;**53**(Suppl. 2):i5–6. http://dx.doi.org/10.1093/rheumatology/keu191

81. NICE. *Clinical Knowledge Summary for Giant Cell Arteritis*. London: NICE; 2014. URL: http://cks.nice.org.uk/giant-cell-arteritis (accessed 28 October 2016).

82. Davis S, Martyn-St James M, Sanderson J, Stevens J, Goka E, Rawdin A, *et al. Bisphosphonates for Preventing Osteoporotic Fragility Fractures (Including a Partial Update of NICE Technology Appraisal Guidance 160 and 161). Technology Assessment Report: Final report to the National Institute for Health and Care Excellence*. Sheffield: School of Health and Related Research (ScHARR), University of Sheffield; 2015.

83. Department of Health. *NHS Reference Costs: Financial Year 2011 to 2012*. London: Department of Health; 2012.

84. Joint Formulary Committee. *British National Formulary* (online). London: BMJ Group and Pharmaceutical Press. URL: www.medicinescomplete.com (accessed 1 June 2015).

85. Royal National Institute of Blind People. *Giant Cell or Temporal Arteritis*. Royal National Institute of Blind People. 2015. URL: www.rnib.org.uk/eye-health-eye-conditions-z-eye-conditions/giant-cell-or-temporal-arteritis (accessed 19 March 2015).

86. Tidy C. *Giant Cell Arteritis*. Patient. 2013. URL: www.patient.co.uk/doctor/giant-cell-arteritis (accessed 19 March 2015).

87. Borchers AT, Gershwin ME. Giant cell arteritis: a review of classification, pathophysiology, geoepidemiology and treatment. *Autoimmun Rev* 2012;**11**:A544–54. http://dx.doi.org/10.1016/j.autrev.2012.01.003

88. Guida A, Tufano A, Perna P, Moscato P, De Donato MT, Finelli R, *et al.* The thromboembolic risk in giant cell arteritis: a critical review of the literature. *Int J Rheumatol* 2014;**2014**:806402. http://dx.doi.org/10.1155/2014/806402

89. Hayreh SS, Zimmerman B, Kardon RH. Visual improvement with corticosteroid therapy in giant cell arteritis. Report of a large study and review of literature. *Acta Ophthalmol Scand* 2002;**80**:355–67. http://dx.doi.org/10.1034/j.1600-0420.2002.800403.x

90. Amiri N, De Vera M, Choi HK, Sayre EC, Avina-Zubieta JA. Increased risk of cardiovascular disease in giant cell arteritis: a general population-based study. *Rheumatology*. 2016;**55**:33–40. http://dx.doi.org/10.1093/rheumatology/kev262

91. Kavanaugh A, Wells AF. Benefits and risks of low-dose glucocorticoid treatment in the patient with rheumatoid arthritis. *Rheumatology* 2014;**53**:1742–51. http://dx.doi.org/10.1093/rheumatology/keu135

92. Kanis JA, Stevenson M, McCloskey EV, Davis S, Lloyd-Jones M. Glucocorticoid-induced osteoporosis: a systematic review and cost-utility analysis. *Health Technol Assess* 2007;**11**(7). http://dx.doi.org/10.3310/hta11070

93. Duru N, van der Goes MC, Jacobs JW, Andrews T, Boers M, Buttgereit F, *et al.* EULAR evidence-based and consensus-based recommendations on the management of medium to high-dose glucocorticoid therapy in rheumatic diseases. *Ann Rheum Dis* 2013;**72**:1905–13. http://dx.doi.org/10.1136/annrheumdis-2013-203249

94. Brown MM, Brown GC, Sharma S, Kistler J, Brown H. Utility values associated with blindness in an adult population. *Br J Ophthalmol* 2001;**85**:327–31. http://dx.doi.org/10.1136/bjo.85.3.327

95. Department of Health. *NHS Reference Costs 2013 to 2014*. London: Department of Health; 2014.

96. Colquitt JL, Jones J, Tan SC, Takeda A, Clegg AJ, Price A. Ranibizumab and pegaptanib for the treatment of age-related macular degeneration: a systematic review and economic evaluation. *Health Technol Assess* 2008;**12**(16). http://dx.doi.org/10.3310/hta12160

97. NICE. *Stroke: Diagnosis and Initial Management of Acute Stroke and TIA*. 2008. URL: www.nice.org.uk/guidance/cg68/resources/stroke-costing-report2 (accessed 1 June 2015).

98. Gutiérrez L, Roskell N, Castellsague J, Beard S, Rycroft C, Abeysinghe S, *et al.* Clinical burden and incremental cost of fractures in postmenopausal women in the United Kingdom. *Bone* 2012;**51**:324–31. http://dx.doi.org/10.1016/j.bone.2012.05.020

99. Gutiérrez L, Roskell N, Castellsague J, Beard S, Rycroft C, Abeysinghe S, *et al.* Study of the incremental cost and clinical burden of hip fractures in postmenopausal women in the United Kingdom. *J Med Econ* 2011;**14**:99–107. http://dx.doi.org/10.3111/13696998.2010.547967

100. Bernatsky S, Panopalis P, Pineau CA, Hudson M, St Pierre Y, Clarke AE. Healthcare costs of inflammatory myopathies. *J Rheumatol* 2011;**38**:885–8. http://dx.doi.org/10.3899/jrheum.101083

101. Manson SC, Brown RE, Cerulli A, Vidaurre CF. The cumulative burden of oral corticosteroid side effects and the economic implications of steroid use. *Respir Med* 2009;**103**:975–94. http://dx.doi.org/10.1016/j.rmed.2009.01.003

102. Curtis L. *Unit Costs of Health and Social Care 2014*. Canterbury: PSSRU, University of Kent; 2014.

103. Post PN, Stiggelbout AM, Wakker PP. The utility of health states after stroke: a systematic review of the literature. *Stroke* 2001;**32**:1425–9. http://dx.doi.org/10.1161/01.STR.32.6.1425

104. School of Health and Related Research (ScHARR). *Bisphosphonates for Preventing Osteoporotic Fragility Fractures (Including a Partial Update of NICE Technology Appraisal Guidance 160 and 161): Assessment Report*. Sheffield: ScHARR, University of Sheffield; 2015. URL: http://www.nice.org.uk/guidance/gid-tag462/documents/osteoporosis-prevention-bisphosphonates-inc-part-rev-ta160-ta161-id782-assessment-report2 (accessed 3 April 2016).

105. Brown GC, Brown MM, Sharma S, Brown H, Gozum M, Denton P. Quality of life associated with diabetes mellitus in an adult population. *J Diabetes Complicat* 2000;**14**:18–24. http://dx.doi.org/10.1016/S1056-8727(00)00061-1

106. Roberts J, Lenton P, Keetharuth AD, Brazier J. Quality of life impact of mental health conditions in England: results from the adult psychiatric morbidity surveys. *Health Qual Life Outcomes* 2014;**12**:6. http://dx.doi.org/10.1186/1477-7525-12-6

107. Ward S, Lloyd Jones M, Pandor A, Holmes M, Ara R, Ryan A, *et al.* A systematic review and economic evaluation of statins for the prevention of coronary events. *Health Technol Assess* 2007;**11**(14). http://dx.doi.org/10.3310/hta11140

108. Office for National Statistics (ONS). *National Life Tables, 2011–2013 Release*. London: ONS; 2014.

109. van Staa TP, Kanis JA, Geusens P, Boonen A, Leufkens HG, Cooper C. The cost-effectiveness of bisphosphonates in postmenopausal women based on individual long-term fracture risks. *Value Health* 2007;**10**:348–57. http://dx.doi.org/10.1111/j.1524-4733.2007.00188.x

110. National Institute for Health and Care Excellence. Chapter 7: Incorporating Economic Evaluation. In *Developing NICE Guidelines: The Manual*. London: NICE; 2014. URL: www.nice.org.uk/process/pmg20/chapter/7-incorporating-economic-evaluation (accessed 4 June 2015).

111. Patil P, Williams M, Maw WW, Achilleos K, Elsideeg S, Dejaco C, *et al.* Fast track pathway reduces sight loss in giant cell arteritis: results of a longitudinal observational cohort study. *Clin Exp Rheumatol* 2015;**33**(Suppl. 2):103–6.

112. Patil P, Karia N, Jain S, Dasgupta B. Giant cell arteritis: a review. *Eye Brain* 2013;**5**:23–33.

113. NHS Choices. *Giant Cell Arteritis (Temporal Arteritis)*. London: NHS Choices; 2015. URL: www.nhs.uk/Conditions/giant-cell-arteritis/Pages/Introduction.aspx (accessed 4 June 2015).

114. Hedges TR, Gieger GL, Albert DM. The clinical value of negative temporal artery biopsy specimens. *Arch Ophthalmol* 1983;**101**:1251–4. http://dx.doi.org/10.1001/archopht.1983.01040020253019

115. Chmelewski WL, McKnight KM, Agudelo CA, Wise CM. Presenting features and outcomes in patients undergoing temporal artery biopsy. A review of 98 patients. *Arch Intern Med* 1992;**152**:1690–5. http://dx.doi.org/10.1001/archinte.1992.00400200120022

116. Gabriel SE, O'Fallon WM, Achkar AA, Lie JT, Hunder GG. The use of clinical characteristics to predict the results of temporal artery biopsy among patients with suspected giant cell arteritis. *J Rheumatol* 1995;**22**:93–6.

117. Maguchi S, Fukuda S, Takizawa M. Histological findings in biopsies from patients with cytoplasmic-antineutrophil cytoplasmic antibody (cANCA)-positive Wegener's granulomatosis. *Auris Nasus Larynx* 2001;**28**:S53–8. http://dx.doi.org/10.1016/S0385-8146(01)00072-4

118. Del Buono EA, Flint A. Diagnostic usefulness of nasal biopsy in Wegener's granulomatosis. *Hum Pathol* 1991;**22**:107–10. http://dx.doi.org/10.1016/0046-8177(91)90030-S

119. Dwolatzky T, Sonnenblick M, Nesher G. Giant cell arteritis and polymyalgia rheumatica: clues to early diagnosis. *Geriatrics* 1997;**52**:38–40, 43–4.

120. Bley TA, Uhl M, Carew J, Markl M, Schmidt D, Peter HH, *et al.* Diagnostic value of high-resolution MR imaging in giant cell arteritis. *AJNR Am J Neuroradiol* 2007;**28**:1722–7. http://dx.doi.org/10.3174/ajnr.A0638

121. Bley TA, Reinhard M, Hauenstein C, Markl M, Warnatz K, Hetzel A, *et al.* Comparison of duplex sonography and high-resolution magnetic resonance imaging in the diagnosis of giant cell (temporal) arteritis. *Arthritis Rheum* 2008;**58**:2574–8. http://dx.doi.org/10.1002/art.23699

122. Hauenstein C, Reinhard M, Geiger J, Markl M, Hetzel A, Treszl A, *et al.* Effects of early corticosteroid treatment on magnetic resonance imaging and ultrasonography findings in giant cell arteritis. *Rheumatology* 2012;**51**:1999–2003. http://dx.doi.org/10.1093/rheumatology/kes153

123. Schmidt WA. Role of ultrasound in the understanding and management of vasculitis. *Ther Adv Musculoskelet Dis* 2014;**6**:39–47. http://dx.doi.org/10.1177/1759720X13512256

124. Czihal M, Piller A, Schroettle A, Kuhlencordt P, Bernau C, Schulze-Koops H, Hoffmann U. Impact of cranial and axillary/subclavian artery involvement by color duplex sonography on response to treatment in giant cell arteritis. *J Vasc Surg* 2015;**61**:1285–91. http://dx.doi.org/10.1016/j.jvs.2014.12.045

125. Gerhard R, Boerner SL. Evaluation of indeterminate thyroid cytology by second-opinion diagnosis or repeat fine-needle aspiration: which is the best approach? *Acta Cytol* 2015;**59**:43–50. http://dx.doi.org/10.1159/000369332

126. Farmer M, Petras RE, Hunt LE, Janosky JE, Galandiuk S. The importance of diagnostic accuracy in colonic inflammatory bowel disease. *Am J Gastroenterol* 2000;**95**:3184–8. http://dx.doi.org/10.1111/j.1572-0241.2000.03199.x

127. Diamantopoulos AP, Haugeberg G, Lindland A, Myklebust G. The fast-track ultrasound clinic for early diagnosis of giant cell arteritis significantly reduces permanent visual impairment: towards a more effective strategy to improve clinical outcome in giant cell arteritis? *Rheumatology* 2016;**55**:66–70. http://dx.doi.org/10.1093/rheumatology/kev289

Appendix 1 Ultrasound case report

Ultrasound

Ultrasound equipment

Have you used the machine that you registered with us? ☐ Yes ☐ No

If no:

Manufacturer _____ Model _____ Probe ☐☐☐ MHz

Scan

Date of scan D D M M Y Y Y Y Initials of sonographer ☐☐☐

Was the scan performed in a single session without any interruptions? ☐ Yes ☐ No

If yes:

Start time H H : M M Stop time H H : M M

If no:

Total time spent on scan H H : M M

Right side temporal ☐ Normal* ☐ Abnormal**

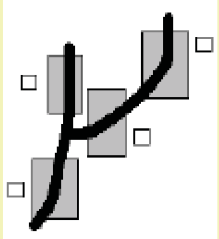
If normal:

Please indicate that the appropriate recording was made:

Video: transverse sweep from common superficial temporal artery to bifurcation showing both the frontal and parietal branch ☐ RTSN

If abnormal:

Please indicate which areas were abnormal:



Left side temporal ☐ Normal* ☐ Abnormal**

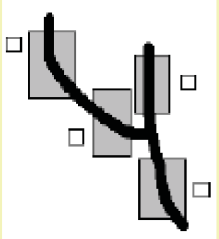
If normal:

Please indicate that the appropriate recording was made:

Video: transverse sweep from common superficial temporal artery to bifurcation showing both the frontal and parietal branch ☐ LTSN

If abnormal:

Please indicate which areas were abnormal:



Right side axillary ☐ Normal* ☐ Abnormal**

If normal:

Please indicate that the appropriate recording was made:

Still: longitudinal image of axilla (grey scale or colour Doppler) at the level of the circumflexa humeri artery. ☐ RALN

If abnormal:

Please complete the relevant section on page 4

Left side axillary ☐ Normal* ☐ Abnormal**

If normal:

Please indicate that the appropriate recording was made:

Still: longitudinal image of axilla (grey scale or colour Doppler) at the level of the circumflexa humeri artery. ☐ LALN

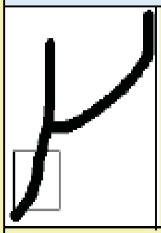
If abnormal:

Please complete the relevant section on page 4

*If all are normal, please sign page 4.

**For any areas indicated as abnormal please complete the relevant sections in pages 2 to 4.

Right common superficial temporal artery



Halo ☐ Yes ☐ No If yes:

Halo maximum thickness ☐☐.☐ mm

Halo maximum length ☐☐ mm

Does the halo run along the entire length of the section? ☐ Yes ☐ No

Occlusion ☐ Yes ☐ No

Arteriosclerosis ☐ Yes ☐ No

Stenosis ☐ Yes ☐ No If yes:

Velocity in stenosis ☐☐☐ cm/s

Velocity out of stenosis ☐☐☐ cm/s

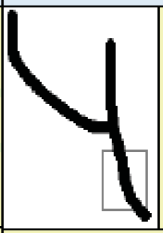
Please indicate which recordings you have taken:

Halo:[1] Transverse ☐ RCTH Longitudinal ☐ RCLH

Occlusion:[1] Transverse ☐ RCTO Longitudinal ☐ RCLO

Stenosis:[2] Longitudinal ☐ RCLS Doppler[3] pulse wave ☐ RCDS

Left common superficial temporal artery



Halo ☐ Yes ☐ No If yes:

Halo maximum thickness ☐☐.☐ mm

Halo maximum length ☐☐ mm

Does the halo run along the entire length of the section? ☐ Yes ☐ No

Occlusion ☐ Yes ☐ No

Arteriosclerosis ☐ Yes ☐ No

Stenosis ☐ Yes ☐ No If yes:

Velocity in stenosis ☐☐☐ cm/s

Velocity out of stenosis ☐☐☐ cm/s

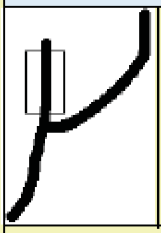
Please indicate which recordings you have taken:

Halo:[1] Transverse ☐ LCTH Longitudinal ☐ LCLH

Occlusion:[1] Transverse ☐ LCTO Longitudinal ☐ LCLO

Stenosis:[2] Longitudinal ☐ LCLS Doppler[3] pulse wave ☐ LCDS

Right parietal ramus



Halo ☐ Yes ☐ No If yes:

Halo maximum thickness ☐☐.☐ mm

Halo maximum length ☐☐ mm

Does the halo run along the entire length of the section? ☐ Yes ☐ No

Occlusion ☐ Yes ☐ No

Arteriosclerosis ☐ Yes ☐ No

Stenosis ☐ Yes ☐ No If yes:

Velocity in stenosis ☐☐☐ cm/s

Velocity out of stenosis ☐☐☐ cm/s

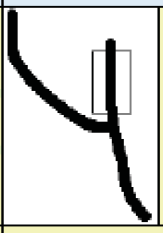
Please indicate which recordings you have taken:

Halo:[1] Transverse ☐ RPTH Longitudinal ☐ RPLH

Occlusion:[1] Transverse ☐ RPTO Longitudinal ☐ RPLO

Stenosis:[2] Longitudinal ☐ RPLS Doppler[3] pulse wave ☐ RPDS

Left parietal ramus



Halo ☐ Yes ☐ No If yes:

Halo maximum thickness ☐☐.☐ mm

Halo maximum length ☐☐ mm

Does the halo run along the entire length of the section? ☐ Yes ☐ No

Occlusion ☐ Yes ☐ No

Arteriosclerosis ☐ Yes ☐ No

Stenosis ☐ Yes ☐ No If yes:

Velocity in stenosis ☐☐☐ cm/s

Velocity out of stenosis ☐☐☐ cm/s

Please indicate which recordings you have taken:

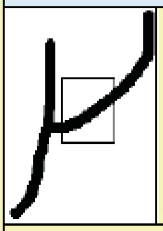
Halo:[1] Transverse ☐ LPTH Longitudinal ☐ LPLH

Occlusion:[1] Transverse ☐ LPTO Longitudinal ☐ LPLO

Stenosis:[2] Longitudinal ☐ LPLS Doppler[3] pulse wave ☐ LPDS

[1] Video 3 seconds
[2] Still image
[3] Doppler pulse wave should show doppler curves demonstrating low and high flow systolic velocities

Right proximal frontal ramus (≤ 2cm)



Halo ☐ Yes ☐ No If yes:

Halo maximum thickness ☐☐.☐ mm

Halo maximum length ☐☐ mm

Does the halo run along the entire length of the section? ☐ Yes ☐ No

Occlusion ☐ Yes ☐ No

Arteriosclerosis ☐ Yes ☐ No

Stenosis ☐ Yes ☐ No If yes:

Velocity in stenosis ☐☐☐ cm/s

Velocity out of stenosis ☐☐☐ cm/s

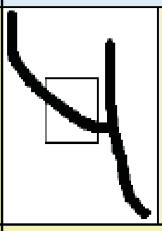
Please indicate which recordings you have taken:

Halo:[1] Transverse ☐ RPFTH Longitudinal ☐ RPFLH

Occlusion:[1] Transverse ☐ RPFTO Longitudinal ☐ RPFLO

Stenosis:[2] Longitudinal ☐ RPFLS Doppler[3] pulse wave ☐ RPFDS

Left proximal frontal ramus (≤ 2cm)



Halo ☐ Yes ☐ No If yes:

Halo maximum thickness ☐☐.☐ mm

Halo maximum length ☐☐ mm

Does the halo run along the entire length of the section? ☐ Yes ☐ No

Occlusion ☐ Yes ☐ No

Arteriosclerosis ☐ Yes ☐ No

Stenosis ☐ Yes ☐ No If yes:

Velocity in stenosis ☐☐☐ cm/s

Velocity out of stenosis ☐☐☐ cm/s

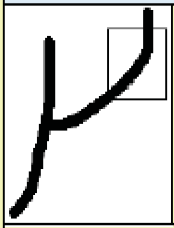
Please indicate which recordings you have taken:

Halo:[1] Transverse ☐ LPFTH Longitudinal ☐ LPFLH

Occlusion:[1] Transverse ☐ LPFTO Longitudinal ☐ LPFLO

Stenosis:[2] Longitudinal ☐ LPFLS Doppler[3] pulse wave ☐ LPFDS

Right distal frontal ramus (> 2cm)



Halo ☐ Yes ☐ No If yes:

Halo maximum thickness ☐☐.☐ mm

Halo maximum length ☐☐ mm

Does the halo run along the entire length of the section? ☐ Yes ☐ No

Occlusion ☐ Yes ☐ No

Arteriosclerosis ☐ Yes ☐ No

Stenosis ☐ Yes ☐ No If yes:

Velocity in stenosis ☐☐☐ cm/s

Velocity out of stenosis ☐☐☐ cm/s

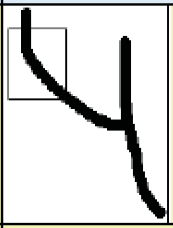
Please indicate which recordings you have taken:

Halo:[1] Transverse ☐ RDFTH Longitudinal ☐ RDFLH

Occlusion:[1] Transverse ☐ RDFTO Longitudinal ☐ RDFLO

Stenosis:[2] Longitudinal ☐ RDFLS Doppler[3] pulse wave ☐ RDFDS

Left distal frontal ramus (> 2cm)



Halo ☐ Yes ☐ No If yes:

Halo maximum thickness ☐☐.☐ mm

Halo maximum length ☐☐ mm

Does the halo run along the entire length of the section? ☐ Yes ☐ No

Occlusion ☐ Yes ☐ No

Arteriosclerosis ☐ Yes ☐ No

Stenosis ☐ Yes ☐ No If yes:

Velocity in stenosis ☐☐☐ cm/s

Velocity out of stenosis ☐☐☐ cm/s

Please indicate which recordings you have taken:

Halo:[1] Transverse ☐ LDFTH Longitudinal ☐ LDFLH

Occlusion:[1] Transverse ☐ LDFTO Longitudinal ☐ LDFLO

Stenosis:[2] Longitudinal ☐ LDFLS Doppler[3] pulse wave ☐ LDFDS

[1] Video 3 seconds
[2] Still image
[3] Doppler pulse wave should show doppler curves demonstrating low and high flow systolic velocities

Right axillary artery	Left axillary artery
Halo ☐ Yes ☐ No **If yes:**	Halo ☐ Yes ☐ No **If yes:**
Halo maximum thickness ☐☐.☐ mm	Halo maximum thickness ☐☐.☐ mm
Halo maximum length ☐☐☐ mm	Halo maximum length ☐☐☐ mm
Does the halo run along the entire length of the section? ☐ Yes ☐ No	Does the halo run along the entire length of the section? ☐ Yes ☐ No
Occlusion ☐ Yes ☐ No	Occlusion ☐ Yes ☐ No
Arteriosclerosis ☐ Yes ☐ No	Arteriosclerosis ☐ Yes ☐ No
Stenosis ☐ Yes ☐ No **If yes:**	Stenosis ☐ Yes ☐ No **If yes:**
Velocity in stenosis ☐☐☐ cm/s	Velocity in stenosis ☐☐☐ cm/s
Velocity out of stenosis ☐☐☐ cm/s	Velocity out of stenosis ☐☐☐ cm/s
Luminal minimum diameter ☐.☐ mm	Luminal minimum diameter ☐.☐ mm
Luminal maximum diameter ☐.☐ mm	Luminal maximum diameter ☐.☐ mm
Please indicate which recordings you have taken:	Please indicate which recordings you have taken:
Halo:[1] Transverse ☐ RATH Longitudinal ☐ RALH	Halo:[1] Transverse ☐ LATH Longitudinal ☐ LALH
Occlusion:[1] Transverse ☐ RATO Longitudinal ☐ RALO	Occlusion:[1] Transverse ☐ LATO Longitudinal ☐ LALO
Stenosis:[1] Longitudinal ☐ RALS Doppler[2] pulse wave ☐ RADS	Stenosis:[1] Longitudinal ☐ LALS Doppler[2] pulse wave ☐ LADS

[1] Still image
[2] Doppler pulse wave should show doppler curves demonstrating low and high flow systolic velocities

Checklist

In your opinion are the results consistent with a diagnosis of GCA? ☐ Yes ☐ No

If no, specify: ☐

Has the ultrasound scan been transferred to the TABUL central office? ☐ Yes ☐ No

If no: ☐ Technical problem ☐ Other, specify ☐

Have you reminded the participant to report any serious adverse events which occur after the ultrasound? ☐ Yes ☐ No

I certify that the ultrasound data are complete and accurate. (To be signed and dated by the sonographer)

Signature ☐ Date D D M M Y Y Y Y

Print name ☐

Appendix 2 Completion of the ultrasound case report form

The standard operating procedure for completion of the ultrasound case report form can be accessed via the following link: http://ora.ox.ac.uk/objects/uuid:7990dde3-0714-4414-b590-3e0aa1b7d761 (accessed 27 May 2016).

Appendix 3 Screening case report form

Clinical

Visit date D D M M Y Y Y Y

Screening number S P - - -

Gender ☐ Male ☐ Female Age years

Inclusion criteria	Yes	No
1 Is there a clinical suspicion of a new diagnosis of GCA, e.g. candidate has a new onset of headache, scalp tenderness, with or without elevated CRP or ESR, tongue or jaw claudication with or without visual loss?	☐	☐
2 Has there been a clinical decision that the candidate requires an urgent temporal artery biopsy to determine whether or not the diagnosis is GCA?	☐	☐
3 Has the candidate agreed and given NHS consent to undergo a temporal artery biopsy as part of standard care?	☐	☐
4 Is the candidate willing to attend for an ultrasound scan of their temporal and axillary arteries?	☐	☐
5 Has the informed consent been obtained? Participants must be willing to give informed written consent or if they are unable, because of physical disabilities (e.g. sudden onset of blindness/ vision loss which can be caused by GCA), they must be willing to give permission for a nominated friend or relative to provide written informed assent.	☐	☐
6 Is the candidate age 18 years or over?	☐	☐

Exclusion criteria	Yes	No
1 Has there been a previous diagnosis of GCA?	☐	☐
2 Has the candidate been on a course of high-dose glucocorticoids (>20mg prednisolone per day) for more than seven days prior to the dates of ultrasound and biopsy?	☐	☐
3 Has the candidate been on a long term course (>1 month) of high-dose steroids (>20mg per day at any time) for conditions other than PMR within 3 months prior to study?	☐	☐
4 Is the candidate unable to give informed written or verbal consent or no witness is available?	☐	☐
5 Is the candidate unable to undergo an ultrasound scan of the temporal and axillary arteries?	☐	☐
6 Does the candidate have any condition which could preclude a temporal artery biopsy?	☐	☐
7 Is the candidate unable to undergo an ultrasound scan and a temporal artery biopsy within 7 days of starting high dose glucocorticoids for suspected GCA?	☐	☐

I certify that the data contained on this page are complete and accurate.
(To be signed and dated by the investigator or authorised member of the investigator's staff)

Signature Date D D M M Y Y Y Y

Print name

Appendix 4 Patient information sheet

PATIENT INFORMATION SHEET

The role of ultrasound compared to biopsy of temporal arteries in the diagnosis and management of giant cell arteritis (TABUL)

We would like you to consider this research study and then decide whether or not you wish to take part. Before you decide whether to participate or not it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and to decide whether or not you wish to take part.

1. What is the purpose of this study?

Our study will examine the role of ultrasound in helping to diagnose Giant Cell Arteritis (GCA). It causes narrowing and blockage of some of these blood vessels; it can cause severe headache and in some cases may affect eyesight. It is important that a prompt diagnosis of GCA can be made in order to start treatment with steroid tablets or injections. Currently there is no test that is 100% accurate for diagnosing GCA.

To help to confirm a diagnosis of GCA, the patient will usually have a biopsy of a temporal artery (a minor surgical procedure performed under local anaesthetic to remove a 1 to 3 centimetre sample of one of the arteries to the scalp). The examination of the biopsy sample usually confirms that the patient has GCA and steroid treatment can be continued. However, some patients with GCA will have a normal biopsy result. For these patients with a normal biopsy result it is difficult to confirm if they do or don't have GCA and whether or not steroid treatment should be continued.

The main study

It is important to find better ways of diagnosing GCA to ensure that more patients are treated appropriately. Another test that may help in diagnosing GCA is examination of ultrasound scans of the arteries in the side of the head and under the arms. Ultrasound does not involve surgery; it is a simple test which can be performed in a radiology department. Gel is applied to both sides of the head and under each arm. A sound probe is placed over the artery at each of these areas to produce the scan for expert examination

The sub-studies

We are asking you to take part in the main study described above, but in addition, there are a number of separate sub-studies that you can also choose to participate in, to look at immune abnormalities in GCA; make an educational website for doctors to use and to store ultrasound images, samples and other data in a Biobank for future studies.

2. Why have I been invited?

You have been chosen because you have been suffering a new onset of headache which is suspected as being giant cell arteritis and the doctors looking after you have decided that you need a biopsy of your temporal artery to clarify the diagnosis.

3. Do I have to take part?

It is up to you to decide whether or not to take part. You are free to withdraw at any time and without giving a reason. If you decide to take part we will ask you to sign a consent form indicating your willingness to participate in the study. Any current or future healthcare that you receive will not be affected by deciding whether or not to take part in the study. Taking part in the study is voluntary.

4. What will happen if I take part?

If you take part in the study, you will need to attend for an ultrasound scan of your scalp and armpit arteries before the planned biopsy. The scan will take about 30 minutes to complete. At the first study visit, you will be assessed by the study nurse or doctor and asked to complete questionnaires about your health; you will be reviewed two weeks (study visit 2) and six months (final study visit) later where the study nurse or doctor will repeat the assessment and ask you to complete the same health questionnaires. Your ultrasound scan result will not be disclosed to you or to the doctors and nurses looking after you. If you agree to take part in any sub-studies, you will also need to give blood samples on each of the 3 study visits.

Participating in the TABUL study will not affect the care you receive from the National Health Service (NHS).

5. What do I have to do?

The main study

By the time you read this, you will have been referred to hospital for a possible diagnosis of GCA. The doctors looking after you will have told you about the study and asked your

permission to contact the study nurse or study doctor. We have given you this information sheet about the study. The study nurse or doctor will get in touch with you; they will either see you on the day of your scheduled hospital visit, or telephone you so that they can give you more information about the study. The research nurse or doctor will arrange to see you after you have had an opportunity to read the information and ask any questions. At this time you can ask further questions about the study, if you decide to take part in the study, at this point you will be asked to sign a consent form (study visit 1). A friend or relative can sign the consent form on your behalf if needed. If you agree to participate in the study, you will be assessed by the study nurse or doctor and asked to complete a questionnaire. An appointment will be made for you to have the ultrasound scan before your biopsy. We will make sure that your biopsy has been arranged. You would usually be attending hospital regularly after the biopsy to check on your health and adjust your treatment. You will be asked to attend for 2 further follow up visits after 2 weeks (study visit 2) and after 6 months (final study visit). On each study visit, the study nurse or doctor will assess you. We will ask you questions about your condition and how it affects your daily life, using standard questionnaires. We expect the first study visit to last about 60-90 minutes, and for each of the 2 subsequent study visits to last between 45-60 minutes. The doctor or nurse in the clinic will complete specialised clinical questionnaires, to assess your diagnosis, but we will also ask you to complete health questionnaire on your ability to carry out normal activities of daily living and your quality of life. We anticipate that it will take about 5-10 minutes for you to complete the questionnaires.

You will have an ultrasound scan of your scalp arteries (temporal arteries) and armpit arteries (axillary arteries) in addition to your routine biopsy of one temporal artery. The ultrasound scan is entirely painless. It will involve applying gel to the both sides of your head, and both armpits. The person performing the scan will place a sound probe over these 4 sites in order to identify the arteries. The probe will be used to record the images, and test each artery for signs of swelling, blockage and narrowing. The probe will be moved over the arteries and different settings will be applied using the dials on the machine to get the best picture. We hope to test the value of performing an ultrasound scan of these 4 arteries as an alternative to a biopsy of one temporal artery. We expect the scan to take about 30 minutes to complete. For training purposes, and if you are willing to do so, more than one scan may be performed.

We will store the biopsy samples and images of the ultrasound scans in one centre (Oxford), so that we can use the material to make sure that all the different hospitals taking part in the study are performing and interpreting the biopsy and scan results according to a high standard of accuracy

Sub studies

We would like to perform some further tests on your biopsy and blood samples we collect from you so that we can gain a better understanding of the disease. In addition to the main study, we are planning to undertake a number of sub studies, making use of the biopsy sample that will normally be taken, as well as the video images of the scans and extra blood samples which we will ask you to provide on each of the 3 study visits. These sub studies are optional. You can take part in the main study and choose not to take part in the sub studies. The sub studies will look at the immune abnormalities in the biopsy sample and blood samples to try to determine what type of cells and inflammatory chemicals are responsible for the condition. This will involve processing the biopsy sample and blood samples to extract the cells, measure the inflammatory chemicals in the blood and look at the biopsy in detail, using special attaining techniques. If you agree to the sub-studies, we will take the additional blood samples (maximum total of 85 ml):

1. **Study Visit 1** - 6 tubes of blood (approx. 35 ml total, equivalent to about 7 teaspoonfuls).
2. **Study Visit 2 (2 weeks)** - 5 tubes of blood (approx. 25 ml total, equivalent to about 5 teaspoonfuls).
3. **Study Visit 3 (6 months)** - 5 tubes of blood (approx. 25 ml total, equivalent to about 5 teaspoonfuls).

Please note that these will be in addition to your routine blood samples at each visit.

We would like to make use of the video images of the scans and of the images of the biopsy samples in order to develop and test the usefulness of a training website to teach other doctors about the best way to diagnose temporal arteritis. No personal details will be used on the website.

We would like your permission to store your samples of blood and your biopsy in a Biobank for future related studies

We will remove your personal details from all research samples so that they are all anonymous and your personal details will remain confidential. However, it will be possible to link the clinical and laboratory details through a unique laboratory code to enable us compare your clinical state with the laboratory findings. Your blood samples will be stored in a laboratory freezer until used.

We would like to store some of your DNA extracted from the blood samples to form part of our Biobank that we can use in future genetic studies. This is purely for research purposes

and you will not be told the results of the tests on your samples. The anonymous genetic information may be shared with other research groups conducting similar investigations.

6. Expenses and payments

If you incur travel expenses in order to attend especially for the ultrasound scan this extra appointment will be reimbursed on request.

7. What are the possible benefits of taking part?

Your condition of suspected giant cell arteritis will be treated by the doctors in hospital in the normal way, using widely recognised means. You will not directly benefit from taking part in this study but the information we get from this study will help improve future treatment of people with suspected giant cell arteritis.

8. What are the possible risks of taking part?

Ultrasound is a safe technique. The temporal artery biopsy procedure would be part of routine care so you would be undergoing this procedure whether you decide to take part in the study or not. Complications following a temporal artery biopsy are rare, but include bleeding, swelling over an artery due to formation of a blood clot (haematoma), damage to branches of the facial nerve, failure to identify the artery, development of infection at the biopsy site, wound breakdown and very rarely scalp necrosis.

9. Will my taking part be kept confidential?

All patient information is stored on password protected computer databases and in locked filing cabinets and will only be accessible to the TABUL research team and regulatory authorities for auditing and monitoring purposes. You will be allocated a unique study number and staff not directly involved with you will know you only by this number. When the results of the study are reported, individuals who have taken part will not be identified in any way. Again we must emphasise that none of your samples will identify you in any way as we will use your unique study number when storing them. Responsible members of the University of Oxford or the local Hospital NHS Trust may be given access to data for monitoring or audit of the study to ensure we are complying with regulations.

10. What if I change my mind about taking part?

If you decide to withdraw from the study, your standard of care will not be affected. You will **still** be asked to attend the routine follow-up clinics required by your doctor and hospital as part of your standard care. These follow up clinics will not be part of the study.

If you withdraw from the study, all samples and clinical information that we have obtained up to the point of you coming out of the study will continue to be used for the purpose of the study.

11. What if there is a problem?

If you wish to complain about any aspect of the way in which you have been approached or treated during the course of this study, you should contact Dr Raashid Luqmani on XXXX or you may contact the University of Oxford Clinical Trials and Research Governance (CTRG) office on XXXX or the head of CTRG, email XXXX.

The University has arrangements in place to provide for harm arising from participation in the study for which the University is the Research Sponsor. NHS indemnity operates in respect of any clinical treatment with which you are provided.

12. Will my GP be informed of my involvement in the study?

Yes. We will send your GP a brief letter informing them of your participation in the study.

13. What will happen to any samples I give?

The video images of your scans, tissue and blood samples collected for the TABUL study will be sent to the Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Science, in Oxford, for analysis. Samples of blood will be stored for future studies, including DNA studies to learn more about the nature of giant cell arteritis. None of these will identify you in any way as we will use your unique study number when storing them. However, if your local hospital requires the biopsy sample to be returned for clinical care, we will be able to trace the sample from the unique study number, supplied by your hospital, so that the sample can be returned. We will use the video images of the scans and the biopsy samples to train the investigators in the study to make sure that we maintain a high standard of quality control. In order to do this, at least 2 experts trained in ultrasound from the study panel will independently review all the video images of the scans; 1 or 2 expert pathologists from the study panel will independently review all the biopsy slides. If there is a disagreement with the findings of the local investigator, this will be discussed.

If you agree to take part in the sub-studies, we will store your biopsy sample and blood samples in a Biobank at oxford. All specimens will be carefully catalogued and maintained in a facility which is fully compliant with the requirements of the Human Tissue Act. This will allow us to safely and securely keep your samples in a freezer, for use in future studies,

which will be reviewed and approved by a Research Ethics Committee, before we make use of these specimens.

14. How will the information I provide be used?

We plan to publish the results in a health journal so others can read about and learn from the results of the study.

15. Who is organising and funding the research?

This nationwide trial is being funded through the Health Technology Assessment (HTA) Programme, which is part of the Department of Health. You can access information about them on the HTA website (www.hta.nhs.uk).

The Nuffield Department of Orthopaedic, Rheumatology & Musculoskeletal Sciences (www.ndorms.ox.ac.uk) a department of the University of Oxford, in Oxford will undertake the day to day running of the trial, under the supervision of Dr Raashid Luqmani. The University of Oxford will act as a sponsor for the study and will be responsible for the governance of the trial. The Sheffield Clinical Trials Unit will be responsible for collecting and monitoring the information generated.

16. Who has reviewed this study?

The Berkshire Research Ethics Committee have reviewed the study and given it a favourable opinion. In addition your local NHS hospital Trust have and your local rheumatologist or ophthalmologist have approved the study.

17. Further Information

If you require more information about this study please call one of the telephone numbers provided to speak to a clinical member of the research team or, alternatively look at the clinical trials website:

http://clinicaltrials.gov/ct2/show/NCT00974883.

Thank you for reading this.

If you have any questions or would like any more information please contact the TABUL Office by phone:

XXXX (XXXX) or XXXX (XXXX)

Or email XXXX

Please keep this information sheet for your records.

If you agree to enter the study, please sign the attached consent form and we will return a copy to you

Appendix 5 Patient consent form

Patient Consent Form

Chief Investigator: Dr Raashid Luqmani

Local Investigator: *("Please add per site")*

Address and telephone number of local investigator: *("Please add per site")*

SITE NUMBER: | | |

PARTICIPANT STUDY NUMBER:

E	**P**	**-**			**-**			

Please initial in the boxes if you agree.

PART A		
1.	I confirm that I have read and understood the information sheet dated ……. (Version …….), and have had the opportunity to ask questions and have had these answered satisfactorily.	
2.	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without my medical care or my legal rights being affected.	
3.	I agree that the video images from my ultrasound scan can be used for quality control for the study	
4.	I agree to the use of my temporal artery biopsy sample for quality control for the study	
5.	I agree for my personal information to be stored confidentially by the TABUL research team so that they can contact me in the future to invite me to participate in any future related research studies. I understand that my participation in any future related study will be entirely voluntary and I can decide not to participate	

6.	I understand that responsible members of the TABUL research team may look at sections of my medical notes where it is relevant to my taking part in research. I give permission for these individuals to have access to my records	
7.	I agree to my GP being informed of my participation in this study	
8.	I understand that my data may be accessed by responsible members of the University of Oxford and the University of Sheffield for the purpose of monitoring or audit	
9.	I agree that my non identifiable data can be stored on a password encrypted data base for the purpose of this study and undefined future related studies	
10.	I understand that the samples collected will be considered a gift to the University of Oxford	
11.	I agree that if I change my mind and withdraw consent from this study at a later date, any clinical information, samples or images obtained that have been donated by me until the time that I withdraw from the study will continue to be used for the study.	
	PART B	
	PLEASE INITIAL YES OR NO TO EACH OF THE FOLLOWING STATEMENTS	
1.	I agree that the video images from my ultrasound scans and photographic images of my temporal artery biopsy can be stored together with my anonymised clinical details and used in this study and future GCA related ethically approved studies to help improve the training in use of ultrasound and biopsy in the diagnosis of GCA	YES NO
2.	agree to give an additional 85ml of my blood during the study	YES NO
3.	I agree to the use and storage of my blood samples in a bio bank, including DNA and temporal artery biopsy, for further ethically approved studies to examine the role of inflammatory cells and mediators in the condition	YES

	NO

I agree to take part in all or part of this study as clearly outlined in the questions that I have initialled and answered **Yes** or **No** to.	

Signed…….D.O.B…….

…….

Full

name……...Date……

…

Name of

Researcher……Date……

Signed……

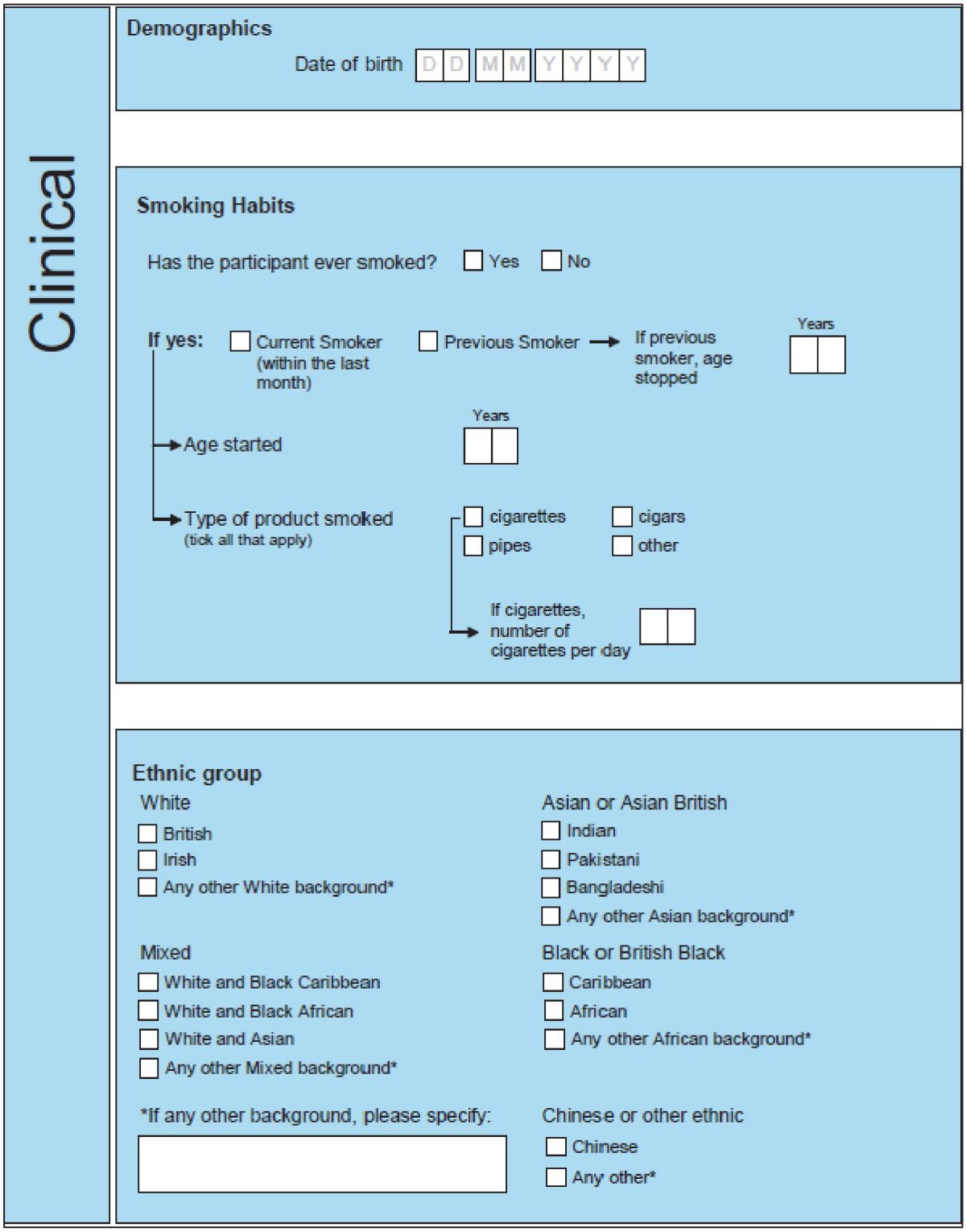
…….

(Copies: Top copy for Study Office, 1 for patient, 1 for hospital notes)

Appendix 6 Recruiting and consenting participants

The standard operating procedure for recruiting and consenting participants can be accessed via the following link: http://ora.ox.ac.uk/objects/uuid:2603e653-8498-4b1a-854a-be889f1d9c38 (accessed 27 May 2016).

Appendix 7 Clinical case report form



Conditions

Please indicate if the participant has (current), or has ever had (past), any of the following conditions. If so please provide the onset date (and resolution date for those 'past' conditions) on the conditions log page 25.

Condition	Current	Past	Never	Condition	Current	Past	Never
Diabetes Mellitus If current, controlled by: ☐ diet ☐ tablet ☐ insulin	☐	☐	☐	Low trauma fracture - of hip	☐	☐	☐
				- of spine	☐	☐	☐
				- of forearm	☐	☐	☐
Hypertension If current: ☐ on treatment ☐ not on treatment	☐	☐	☐	Other low trauma fracture:	☐	☐	☐
Angina	☐	☐	☐	Neoplasia, specify:	☐	☐	☐
Myocardial infarction	☐	☐	☐				
Heart failure	☐	☐	☐				

Previous and other current medical history ☐ Yes ☐ No

If Yes, Record any relevant medical history in the systems listed below, use the codes* to relate to the associated system. (Don't record the conditions listed in the table above).

Code*	Details	Onset date	Current?
		M M Y Y Y Y	☐ Yes ☐ No
		M M Y Y Y Y	☐ Yes ☐ No
		M M Y Y Y Y	☐ Yes ☐ No
		M M Y Y Y Y	☐ Yes ☐ No
		M M Y Y Y Y	☐ Yes ☐ No

*Codes

CV: Cardiovascular R: Respiratory HB: Hepato-biliary GI: Gastro-intestinal GU: Genito-urinary End: Endocrine K: Renal	Hm: Haematological LM: Locomotor Neo: Neoplasia Nrl: Neurological Psy: Psychological Imm: Immunological Drm: Dermatological	All: Drug allergies ENT: Eyes, ear, nose, throat IA: Inflammatory arthritis SLE: Systemic lupus erythematosus IBD: Inflammatory bowel disease	OID: Other inflammatory disease GCA: Giant cell arteritis PMR: Polymyalgia rheumatica SSV: Any form of vasculitis Oth: Other

Please tick here if medical history is continued on another page (download additional pages from the TABUL website: https://weblearn.ox.ac.uk) ☐

Presenting symptoms - pre steroids

Has the participant started a course of high dose steroids within the last 7 days? ☐ Yes ☐ No If no, please go to page 4

If yes, steroid start date D D M M Y Y Y Y

If yes, please indicate if each group of symptoms was absent or present before starting high dose steroids, by selecting yes (if present) or no (if absent). If present please tick all symptoms that were present and complete the symptoms log on page 26.

Symptom			
General?	☐ Yes	☐ No	**If yes, tick all that apply:**
☐ Development of symptoms or findings beginning at age 50 or older*			
☐ Anorexia			
☐ Fatigue			
☐ Symptoms of fever or night sweats			
☐ New onset of bilateral shoulder pain			
☐ New onset of early morning stiffness > 1 hour			
☐ New onset of bilateral hip stiffness or pain			
Pain in or around the head?	☐ Yes	☐ No	**If yes, tick all that apply:**
☐ New onset or new type of localised pain in the head*			
☐ New onset of generalised scalp tenderness			
☐ Swelling over temporal artery			
☐ Pain over temporal artery			
☐ Jaw claudication			
☐ Tongue claudication			
Visual?	☐ Yes	☐ No	**If yes, tick all that apply:**
☐ New symptom of reduced or lost vision in either eye			
☐ Double vision			
☐ Amaurosis fugax			
Any others?	☐ Yes	☐ No	**If yes, please specify:**
e.g. TIA, stroke			

*ACR Criterion for classification of GCA

Presenting symptoms - current

Have the presenting symptoms changed since the pre-steroid assessment? ☐ Yes ☐ No ☐ Not applicable (participant has not started steroids)

If no, go to page 5.

If yes or not applicable; please indicate if each group of symptoms is absent or present at the time of the visit, by selecting yes (if present) or no (if absent). If present please tick all symptoms that are present and complete the symptoms log on page 26. (If these current symptoms were also present before starting high dose steroids don't duplicate this information in the symptoms log. However, do provide the resolution dates for any symptoms which have resolved since the commencement of steroids on this log).

Symptom			
General?	☐ Yes	☐ No	**If yes, tick all that apply:**
☐ Development of symptoms or findings beginning at age 50 or older*			
☐ Anorexia			
☐ Fatigue			
☐ Symptoms of fever or night sweats			
☐ New onset of bilateral shoulder pain			
☐ New onset of early morning stiffness > 1 hour			
☐ New onset of bilateral hip stiffness or pain			
Pain in or around the head?	☐ Yes	☐ No	**If yes, tick all that apply:**
☐ New onset or new type of localised pain in the head*			
☐ New onset of generalised scalp tenderness			
☐ Swelling over temporal artery			
☐ Pain over temporal artery			
☐ Jaw claudication			
☐ Tongue claudication			
Visual?	☐ Yes	☐ No	**If yes, tick all that apply:**
☐ New symptom of reduced or lost vision in either eye			
☐ Double vision			
☐ Amaurosis fugax			
Any others?	☐ Yes	☐ No	**If yes, please specify:**
e.g. TIA, stroke			

*ACR Criterion for classification of GCA

Vital signs

Pulse rate ☐☐☐ bpm

Blood pressure ☐☐☐ / ☐☐☐ mm/Hg

Recorded weight ☐☐☐ Kg or ☐☐ st ☐☐ lbs

Specific physical examination

Right side				Left side		
Abnormal	Normal	Not Assessed	Feature, please remember to complete left and right	Abnormal	Normal	Not Assessed
☐	☐		Thickened Temporal Artery	☐	☐	
☐	☐		Tender Temporal Artery*	☐	☐	
☐	☐		Reduced or absent pulsation in temporal artery*	☐	☐	
☐	☐		Tender Axillary Artery	☐	☐	
☐	☐	☐	Anterior ischaemic optic neuropathy	☐	☐	☐
☐	☐	☐	Posterior ischaemic optic neuropathy	☐	☐	☐
☐	☐	☐	Relative afferent pupillary defect	☐	☐	☐
☐	☐	☐	III/IV/VI nerve palsy	☐	☐	☐
☐	☐	☐	Bruits	☐	☐	☐

Other features	Present	Absent	Not Assessed	
Stroke	☐	☐	☐	If present please specify:
Aneurysm	☐	☐	☐	If present please specify site of aneurysm:
Other, e.g. scalp necrosis tongue necrosis	☐	☐	☐	If present please specify:

*ACR Criterion for classification of GCA

Pre Steroid Results: ESR / CRP/ Plasma viscosity ☐ not available, or:

Date of test D D M M Y Y Y Y

ESR ☐☐☐ mm/hr or > ☐☐☐ mm/hr

Plasma viscosity ☐.☐☐ mPa.s

CRP ☐ in the normal range, or:

☐☐☐.☐☐ or > ☐☐☐.☐☐ ☐ mg/L ☐ mg/dL ☐ mmol/L

Baseline Results: ESR / CRP / Plasma viscosity

ESR ☐☐☐ mm/hr or > ☐☐☐ mm/hr

Plasma viscosity ☐.☐☐ mPa.s

CRP ☐ in the normal range, or:

☐☐☐.☐☐ or > ☐☐☐.☐☐ ☐ mg/L ☐ mg/dL ☐ mmol/L

Haematology ☐ not done ☐ pre steroid results ☐ baseline results

Haemoglobin ☐☐.☐ g/dL or ☐.☐☐ g/L or ☐☐.☐ mmol/L

Platelets ☐☐☐☐ x10^9/L or ☐☐☐☐ x10^3/µL

Total WBC ☐☐☐.☐ x10^9/L or ☐☐☐.☐ x10^3/µL

Neutrophils ☐☐☐.☐ x10^9/L or ☐☐☐.☐ x10^3/µL

ANCA ☐ not done ☐ pre steroid results ☐ baseline results

Immunofluorescence Negative | P | C | Indeterminate (Please circle the result) titre if known 1/ ☐☐☐☐

ELISA MPO ☐☐☐.☐ IU/ml or > ☐☐☐.☐ IU/ml

ELISA PR3 ☐☐☐.☐ IU/ml or > ☐☐☐.☐ IU/ml

Urine dipstick ☐ not done ☐ pre steroid results ☐ baseline results

Blood 0 | Trace | + | ++ | +++ (Please circle the result)

Protein 0 | Trace | + | ++ | +++

Diagnosis

How certain are you of the diagnosis of GCA? ☐ definite ☐ probable ☐ possible

Steroids

Has the participant taken any steroids? ☐ Yes ☐ No

If yes, are these being given for: (tick all that apply) ☐ Suspected GCA ☐ Other condition

Please record details below:

Name of Steroid	Route†	Dose (mg)*
Current steroids:		
Previous steroid preparations:		

Immunosuppressants

Is the participant currently taking any immunosuppressants? ☐ Yes ☐ No

If yes, are these being given for: (tick all that apply) ☐ Suspected GCA ☐ Other condition

If yes, please record details below:

Name of current Immunosuppressant	Route†	Total daily dose - unit	Start date
		-	D D M M Y Y
		-	D D M M Y Y
		-	D D M M Y Y
		-	D D M M Y Y
		-	D D M M Y Y

*Total daily dose in mg †e.g. PO=Oral, IV=intravenous, IM=intramuscular; IA=intraarticular

End of visit checklist

	Yes	No
Is the participant taking any other concomitant medications? If yes please complete the concomitant medications form on page 27.	☐	☐
Has the ultrasound appointment been made? The ultrasound appointment must take place before the biopsy.	☐	☐
Has the biopsy appointment been made? The biopsy appointment must take place after the ultrasound.	☐	☐
Has the participant completed the EQ5D? If yes please store the completed EQ5D in the participant's study file.	☐	☐
Will the participant continue? If no please complete the discontinuation form on page 28.	☐	☐
Has an appointment been made for visit two? Usually two weeks after baseline visit.	☐	☐

I certify that the data contained in the baseline CRF are complete and accurate.
(To be signed and dated by the investigator or authorised member of the investigator's staff)

Signature _____ Date D D M M Y Y Y Y

Print name _____

Visit date D D M M Y Y Y Y

If not two weeks please explain:

Biopsy

Has the biopsy been done? ☐ Yes ☐ No*

If yes:

Has the biopsy site been defined in the surgical records? ☐ Yes ☐ No

Did the sample consist of artery? ☐ Yes ☐ No

If yes, please specify (tick all that apply):

	Right	Left	Side not defined
common superficial temporal artery	☐	☐	☐
parietal ramus	☐	☐	☐
proximal frontal ramus (< 2cm)	☐	☐	☐
distal frontal ramus (> 2cm)	☐	☐	☐
section of artery not defined	☐	☐	

Is the biopsy report available to you? ☐ Yes ☐ No

If no report available, reason:

☐ not reported yet

☐ reported but result not made available

☐ other, specify:

Are the results consistent with GCA? ☐ Yes ☐ No

***If the biopsy has not been done, reason:**

☐ participant did not attend

☐ procedure cancelled

☐ participant refused to have biopsy

☐ participant medically unfit for biopsy

☐ other, specify:

Current conditions

For those conditions that were not current at the previous visit and have not occurred since, please tick 'not occurred since last visit'. No further details are required.
For those conditions that were not current at the previous visit but have occurred since the last visit, please tick 'occurred since last visit'.
NB it is possible that the condition has occurred and resolved since the last visit, in which case also tick 'resolved'.
For those conditions that were current at the prior visit please tick if they have resolved.
If they have not resolved please tick if they are better, worse or no change.

If more than one other type of low trauma fracture or neoplasia is recorded at the prior visit, please provide details for each separately.

Condition	If absent at prior visit: Not Occurred since last visit	If absent at prior visit: Occurred since last visit	Resolved	If present at prior visit and not resolved: Better*	Worse**	No change
Diabetes Mellitus If present, now controlled by: ☐ diet ☐ tablet ☐ insulin	☐	☐	☐	☐	☐	☐
Hypertension If present, now: ☐ on treatment ☐ not on treatment	☐	☐	☐	☐	☐	☐
Angina	☐	☐	☐	☐	☐	☐
Heart failure	☐	☐	☐	☐	☐	☐
Myocardial infarction	☐	☐				
Low trauma fracture - of hip	☐	☐				
- of spine	☐	☐				
- of forearm	☐	☐				
Other low trauma fracture: _____ _____	☐	☐ ☐				
Neoplasia, specify: _____ _____	☐	☐ ☐				

* No deterioration since last visit, the condition has improved.
**General deterioration since last visit.

For any new conditions please document the onset date on the conditions log (page 25). For any resolved conditions please document the resolution date next to the corresponding onset date on the conditions log (page 25).

Current symptoms **Do Not leave any blank rows.**

For those conditions that have not occurred at the previous visit and have not occurred since, please tick 'not occurred since last visit'. No further details are required.
For those symptoms that were absent at the previous visit but have occurred since the last visit, please tick 'occurred since last visit'.
NB it is possible that the symptom has occurred and resolved since the last visit, in which case also tick 'resolved'. For those symptoms that were present at the prior visit please tick if they have resolved
If they have not resolved please tick if they are worse, better or no change.

Symptom	If absent at prior visit: Not Occurred since last visit	If absent at prior visit: Occurred since last visit	Resolved	If present at prior visit and not resolved: Better*	If present at prior visit and not resolved: Worse**	If present at prior visit and not resolved: No change
Anorexia	☐	☐	☐	☐	☐	☐
Fatigue	☐	☐	☐	☐	☐	☐
Symptoms of fever or night sweats	☐	☐	☐	☐	☐	☐
Bilateral shoulder pain	☐	☐	☐	☐	☐	☐
Early morning stiffness > 1 hour	☐	☐	☐	☐	☐	☐
Bilateral hip stiffness or pain	☐	☐	☐	☐	☐	☐
Localised pain in the head	☐	☐	☐	☐	☐	☐
Generalised scalp tenderness	☐	☐	☐	☐	☐	☐
Swelling over temporal artery	☐	☐	☐	☐	☐	☐
Pain over temporal artery	☐	☐	☐	☐	☐	☐
Jaw claudication	☐	☐	☐	☐	☐	☐
Tongue claudication	☐	☐	☐	☐	☐	☐
Reduced or lost vision in either eye	☐	☐	☐	☐	☐	☐
Double vision	☐	☐	☐	☐	☐	☐
Amaurosis fugax	☐	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐	☐

* No deterioration since last visit, the condition has improved.
**General deterioration since last visit.

For any new symptoms please document the onset date on the symptoms log. For any resolved symptoms please document the resolution date next to the corresponding onset date on the symptoms log (page 26).

Vital signs

Pulse rate ☐☐☐ bpm

Blood pressure ☐☐☐ / ☐☐☐ mm/Hg

Recorded weight ☐☐☐ Kg or ☐☐ st ☐☐ lbs

Specific physical examination

Right side				Left side		
Abnormal	Normal	Not Assessed	Feature, please remember to complete left and right	Abnormal	Normal	Not Assessed
☐	☐		Thickened Temporal Artery	☐	☐	
☐	☐		Tender Temporal Artery*	☐	☐	
☐	☐		Reduced or absent pulsation in temporal artery*	☐	☐	
☐	☐		Tender Axillary Artery	☐	☐	
☐	☐	☐	Anterior ischaemic optic neuropathy	☐	☐	☐
☐	☐	☐	Posterior ischaemic optic neuropathy	☐	☐	☐
☐	☐	☐	Relative afferent pupillary defect	☐	☐	☐
☐	☐	☐	III/IV/VI nerve palsy	☐	☐	☐
☐	☐	☐	Bruits	☐	☐	☐

Other features	Present	Absent	Not Assessed	
Stroke	☐	☐	☐	If present please specify:
Aneurysm	☐	☐	☐	If present please specify site of aneurysm:
Other, e.g. scalp necrosis tongue necrosis	☐	☐	☐	If present please specify:

*ACR Criterion for classification of GCA

ESR / CRP / Plasma viscosity

ESR ☐☐☐ mm/hr or > ☐☐☐ mm/hr

Plasma viscosity ☐.☐☐ mPa.s

CRP ☐ in the normal range, or:

☐☐☐.☐☐ or > ☐☐☐.☐☐ ☐ mg/L ☐ mg/dL ☐ mmol/L

Haematology ☐ not done

Haemoglobin ☐☐.☐ g/dL or ☐.☐☐ g/L or ☐☐.☐ mmol/L

Platelets ☐☐☐☐ x10^9/L or ☐☐☐☐ x10^3/µL

Total WBC ☐☐☐.☐ x10^9/L or ☐☐☐.☐ x10^3/µL

Neutrophils ☐☐☐.☐ x10^9/L or ☐☐☐.☐ x10^3/µL

ANCA ☐ not done

Please circle the result

Immunofluorescence | Negative | P | C | Indeterminate | titre if known 1/ ☐☐☐☐

ELISA MPO ☐☐☐.☐ IU/ml or > ☐☐☐.☐ IU/ml

ELISA PR3 ☐☐☐.☐ IU/ml or > ☐☐☐.☐ IU/ml

Urine dipstick ☐ not done

Blood | 0 | Trace | + | ++ | +++

Please circle the result

Protein | 0 | Trace | + | ++ | +++

Birmingham Vasculitis Activity Score (BVAS)

Tick an item only if attributable to active suspected GCA. If there are no abnormalities in a section, please tick 'No' for that organ-system. If there are abnormalities, tick yes and tick all items attributable to active suspected GCA.

General? ☐ Yes ☐ No

If yes:

- ☐ Myalgia
- ☐ Arthralgia / arthritis
- ☐ Fever ≥ 38°C
- ☐ Weight loss ≥ 2kg

Cutaneous? ☐ Yes ☐ No

If yes:

- ☐ Infarct
- ☐ Purpura
- ☐ Ulcer
- ☐ Gangrene
- ☐ Other skin vasculitis

Mucous membranes / eyes? ☐ Yes ☐ No

If yes:

- ☐ Mouth ulcers
- ☐ Genital ulcers
- ☐ Adnexal inflammation
- ☐ Significant proptosis
- ☐ Scleritis / Episcleritis
- ☐ Conjunctivitis / Blepharitis / Keratitis
- ☐ Blurred vision
- ☐ Sudden visual loss
- ☐ Uveitis
- ☐ Retinal changes (vasculitis / thrombosis / exudate / haemorrhage)

ENT? ☐ Yes ☐ No

If yes:

- ☐ Bloody nasal discharge / crusts / ulcers / granulomata
- ☐ Paranasal sinus involvement
- ☐ Subglottic stenosis
- ☐ Conductive hearing loss
- ☐ Sensorineural hearing loss

Chest? ☐ Yes ☐ No

If yes:

- ☐ Wheeze
- ☐ Nodules or cavities
- ☐ Pleural effusion / pleurisy
- ☐ Infiltrate
- ☐ Endobronchial involvement
- ☐ Massive haemoptysis / alveolar haemorrhage
- ☐ Respiratory failure

Cardiovascular? ☐ Yes ☐ No

If yes:

- ☐ Loss of pulses
- ☐ Vascular heart disease
- ☐ Pericarditis
- ☐ Ischaemic cardiac pain
- ☐ Cardiomyopathy
- ☐ Congestive cardiac failure

Abdominal? ☐ Yes ☐ No

If yes:

- ☐ Peritonitis
- ☐ Bloody diarrhoea
- ☐ Ischaemic abdominal pain

Renal? ☐ Yes ☐ No

If yes:

- ☐ Hypertension
- ☐ Proteinuria > 1+
- ☐ Haematuria ≥ 10 RBCs/hpf
- ☐ Serum creatinine 125-249 µmol/L*
- ☐ Serum creatinine 250-499 µmol/L*
- ☐ Serum creatinine ≥500 µmol/L*
- ☐ Rise in serum creatinine >30% or fall in creatinine clearance >25%

* Can only be scored on the first assessment

Nervous system? ☐ Yes ☐ No

If yes:

- ☐ Headache
- ☐ Meningitis
- ☐ Organic confusion
- ☐ Seizures (not hypertensive)
- ☐ Cerebrovascular accident
- ☐ Spinal cord lesion
- ☐ Cranial nerve palsy
- ☐ Sensory peripheral neuropathy
- ☐ Mononeuritis multiplex

Other? ☐ Yes ☐ No

If yes, specify:

PERSISTENT DISEASE ONLY*

☐ Tick if ALL abnormalities are due to persistent disease.

*Active suspected GCA which is not new / worse in the prior 4 weeks.

VASCULITIS DAMAGE INDEX (VDI)

This is for recording organ damage that has occurred in patients *since the onset of suspected GCA*
Patients often have co-morbidity before onset of suspected GCA, **which must not be scored**
Record features of active disease using the Birmingham Vasculitis Activity Score (BVAS)
A new patient should **usually have a VDI score of zero**, unless:
(a) they have had suspected GCA for more than three months **and**
(b) the damage has developed or become worse since the onset of suspected GCA

Musculoskeletal? ☐ Yes ☐ No

If yes:

- ☐ Significant muscle atrophy or weakness
- ☐ Deforming/erosive arthritis
- ☐ Osteoporosis/vertebral collapse
- ☐ Avascular necrosis
- ☐ Osteomyelitis

Skin/Mucous membranes? ☐ Yes ☐ No

If yes:

- ☐ Alopecia
- ☐ Cutaneous ulcers
- ☐ Mouth ulcers

Ocular? ☐ Yes ☐ No

If yes:

- ☐ Cataract
- ☐ Retinal change
- ☐ Optic atrophy
- ☐ Visual impairment/diplopia
- ☐ Blindness in one eye
- ☐ Blindness in second eye
- ☐ Orbital wall destruction

ENT? ☐ Yes ☐ No

If yes:

- ☐ Hearing loss
- ☐ Nasal blockage/chronic discharge/crusting
- ☐ Nasal bridge collapse/septal perforation
- ☐ Chronic sinusitis/radiological damage
- ☐ Subglottic stenosis (no surgery)
- ☐ Subglottic stenosis (with surgery)

Pulmonary? ☐ Yes ☐ No

If yes:

- ☐ Pulmonary hypertension
- ☐ Pulmonary fibrosis
- ☐ Pulmonary infarction
- ☐ Pleural fibrosis
- ☐ Chronic asthma
- ☐ Chronic breathlessness
- ☐ Impaired lung function

Cardiovascular? ☐ Yes ☐ No

If yes:

- ☐ Angina/angioplasty
- ☐ Myocardial infarction
- ☐ Subsequent myocardial infarction
- ☐ Cardiomyopathy
- ☐ Valvular disease
- ☐ Pericarditis ≥ 3 mths or pericardectomy
- ☐ Diastolic BP ≥ 95 or requiring antihypertensives

Peripheral vascular disease? ☐ Yes ☐ No

If yes:

- ☐ Absent pulses in one limb
- ☐ 2nd episode of absent pulses in one limb
- ☐ Major vessel stenosis
- ☐ Claudication >3 mths
- ☐ Minor tissue loss
- ☐ Major tissue loss
- ☐ Subsequent major tissue loss
- ☐ Complicated venous thrombosis

Gastrointestinal? ☐ Yes ☐ No

If yes:

- ☐ Gut infarction/resection
- ☐ Mesenteric insufficiency / pancreatitis
- ☐ Chronic peritonitis
- ☐ Oesophageal stricture/surgery

Renal? ☐ Yes ☐ No

If yes:

- ☐ Estimated/measured GFR < 50%
- ☐ Proteinuria > 0.5g/24hr
- ☐ End stage renal disease

Neuropsychiatric? ☐ Yes ☐ No

If yes:

- ☐ Cognitive impairment
- ☐ Major psychosis
- ☐ Seizures
- ☐ Cerebrovascular accident
- ☐ 2nd cerebrovascular accident
- ☐ Cranial nerve lesion
- ☐ Peripheral neuropathy
- ☐ Transverse myelitis

Other? ☐ Yes ☐ No

If yes:

- ☐ Gonadal failure
- ☐ Chemical cystitis
- ☐ Marrow failure
- ☐ Malignancy
- ☐ Diabetes
- ☐ Other

Total VDI Score*

☐ Record the number of positive items (1 point for each).

*The VDI score can either increase or remain the same over time. Remember to carry forward any previous items of damage

Diagnosis

Does the participant have features consistent with a diagnosis of GCA? ☐ Yes ☐ No

If yes, which of the following influenced your decision (tick all that apply):

☐ symptoms ☐ signs ☐ blood abnormalities

☐ biopsy report ☐ other, specify:

If no*, please give at least one alternative diagnosis:

- ☐ non specific headache
- ☐ migraine
- ☐ myofascial pain
- ☐ temperomandibular dysfunction
- ☐ cervical spondylosis
- ☐ fibromyalgia
- ☐ sinusitis
- ☐ orbital cellulitis
- ☐ shingles
- ☐ orbital pseudotumour
- ☐ metastatic disease (cancer)
- ☐ lymphoma
- ☐ Paget's disease
- ☐ Takayasu's arteritis
- ☐ large vessel vasculitis
- ☐ polyarteritis nodosa
- ☐ Granulomatosis with polyangiitis (GPA)
- ☐ microscopic polyangiitis
- ☐ Churg-Strauss syndrome
- ☐ cryoglobulinemic vasculitis
- ☐ Henoch-Schonlein purpura
- ☐ other vasculitis, specify:
- ☐ other, specify:

***WARNING, are you considering rapidly withdrawing steroids because the participant does not have features consistent with a clinical diagnosis of GCA?** ☐ Yes** ☐ No

****IF YES: please contact 01865 737221 or 01865 227326.**

Weekends only, please contact 07905 211359

If no, please specify why you are not considering withdrawing steroids given that you do not suspect GCA.

After contacting the team are you still going to rapidly withdraw steroids? ☐ Yes ☐ No

If no, have you changed your diagnosis to:

☐ GCA? ☐ other, specify:

Steroids

Has the participant taken any steroids? ☐ Yes ☐ No

If yes, are these being given for: (tick all that apply) ☐ Suspected GCA ☐ Other condition

Please record details below:

Name of Steroid	Route†	Dose (mg)*
Current steroids:		
Previous steroid preparations:		

Immunosuppressants

Is the participant currently taking any immunosuppressants? ☐ Yes ☐ No

If yes, are these being given for: (tick all that apply) ☐ Suspected GCA ☐ Other condition

If yes, please record details below:

Name of current Immunosuppressant	Route†	Total daily dose - unit	Start date
		-	D D M M Y Y
		-	D D M M Y Y
		-	D D M M Y Y
		-	D D M M Y Y
		-	D D M M Y Y

*Total daily dose in mg †e.g. PO=Oral, IV=intravenous, IM=intramuscular; IA=intraarticular

End of visit checklist

	Yes	No
Has the participant taken any other concomitant medications? If yes please complete the concomitant medications form (page 27).	☐	☐
Has the participant had any adverse events? If yes please complete the adverse event form (separate pad).	☐	☐
Has the participant had any serious adverse events? If yes please complete the adverse event form (separate pad).	☐	☐
Has the participant completed the EQ5D? If yes please store the completed EQ5D in the participant's study file.	☐	☐
Will the participant continue? If no please complete the discontinuation form (page 28).	☐	☐
Has an appointment been made for visit three? Usually six months after baseline.	☐	☐

I certify that the data contained in the visit two CRF are complete and accurate.
(To be signed and dated by the investigator or authorised member of the investigator's staff)

Signature _____ Date D D M M Y Y Y Y

Print name _____

Visit date D D M M Y Y Y Y

Current conditions

For those conditions that were not current at the previous visit and have not occurred since, please tick 'not occurred since last visit'. No further details are required.
For those conditions that were not current at the previous visit but have occurred since the last visit, please tick 'occurred since last visit'.
NB it is possible that the condition has occurred and resolved since the last visit, in which case also tick 'resolved'.
For those conditions that were current at the prior visit please tick if they have resolved.
If they have not resolved please tick if they are better, worse or no change.

If more than one other type of low trauma fracture or neoplasia is recorded at the prior visit, please provide details for each separately.

Condition	If absent at prior visit: Not Occurred since last visit	If absent at prior visit: Occurred since last visit	Resolved	If present at prior visit and not resolved: Better*	Worse**	No change
Diabetes Mellitus If present, now controlled by: ☐ diet ☐ tablet ☐ insulin	☐	☐	☐	☐	☐	☐
Hypertension ☐ on treatment ☐ not on treatment	☐	☐	☐	☐	☐	☐
Angina	☐	☐	☐	☐	☐	☐
Heart failure	☐	☐	☐	☐	☐	☐
Myocardial infarction	☐	☐				
Low trauma fracture - of hip	☐	☐				
- of spine	☐	☐				
- of forearm	☐	☐				
Other low trauma fracture:	☐					
		☐				
		☐				
Neoplasia, specify:	☐					
		☐				
		☐				

* No deterioration since last visit, the condition has improved.
**General deterioration since last visit.

For any new conditions please document the onset date on the conditions log (page 25). For any resolved conditions please document the resolution date next to the corresponding onset date on the conditions log (page 25).

Current symptoms **Do Not leave any blank rows.**

For those conditions that have not occurred at the previous visit and have not occurred since, please tick 'not occurred since last visit'. No further details are required.
For those symptoms that were absent at the previous visit but have occurred since the last visit, please tick 'occurred since last visit'.
NB it is possible that the symptom has occurred and resolved since the last visit, in which case also tick 'resolved'. For those symptoms that were present at the prior visit please tick if they have resolved
If they have not resolved please tick if they are worse, better or no change.

Symptom	If absent at prior visit:		Resolved	If present at prior visit and not resolved:		
	Not Occurred since last visit	Occurred since last visit		Better*	Worse**	No change
Anorexia	☐	☐	☐	☐	☐	☐
Fatigue	☐	☐	☐	☐	☐	☐
Symptoms of fever or night sweats	☐	☐	☐	☐	☐	☐
Bilateral shoulder pain	☐	☐	☐	☐	☐	☐
Early morning stiffness > 1 hour	☐	☐	☐	☐	☐	☐
Bilateral hip stiffness or pain	☐	☐	☐	☐	☐	☐
Localised pain in the head	☐	☐	☐	☐	☐	☐
Generalised scalp tenderness	☐	☐	☐	☐	☐	☐
Swelling over temporal artery	☐	☐	☐	☐	☐	☐
Pain over temporal artery	☐	☐	☐	☐	☐	☐
Jaw claudication	☐	☐	☐	☐	☐	☐
Tongue claudication	☐	☐	☐	☐	☐	☐
Reduced or lost vision in either eye	☐	☐	☐	☐	☐	☐
Double vision	☐	☐	☐	☐	☐	☐
Amaurosis fugax	☐	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐	☐

* No deterioration since last visit, the condition has improved.
**General deterioration since last visit.

For any new symptoms please document the onset date on the symptoms log. For any resolved symptoms please document the resolution date next to the corresponding onset date on the symptoms log (page 26).

Vital signs

Pulse rate ☐☐☐ bpm

Blood pressure ☐☐☐ / ☐☐☐ mm/Hg

Recorded weight ☐☐☐ Kg or ☐☐ st ☐☐ lbs

Specific physical examination

Right side				Left side		
Abnormal	Normal	Not Assessed	Feature, please remember to complete left and right	Abnormal	Normal	Not Assessed
☐	☐		Thickened Temporal Artery	☐	☐	
☐	☐		Tender Temporal Artery*	☐	☐	
☐	☐		Reduced or absent pulsation in temporal artery*	☐	☐	
☐	☐		Tender Axillary Artery	☐	☐	
☐	☐	☐	Anterior ischaemic optic neuropathy	☐	☐	☐
☐	☐	☐	Posterior ischaemic optic neuropathy	☐	☐	☐
☐	☐	☐	Relative afferent pupillary defect	☐	☐	☐
☐	☐	☐	III/IV/VI nerve palsy	☐	☐	☐
☐	☐	☐	Bruits	☐	☐	☐

Other features	Present	Absent	Not Assessed	
Stroke	☐	☐	☐	If present please specify:
Aneurysm	☐	☐	☐	If present please specify site of aneurysm:
Other, e.g. scalp necrosis tongue necrosis	☐	☐	☐	If present please specify:

*ACR Criterion for classification of GCA

ESR / CRP / Plasma viscosity

ESR ☐☐☐ mm/hr or > ☐☐☐ mm/hr

Plasma viscosity ☐.☐☐ mPa.s

CRP ☐ in the normal range, or:

☐☐☐.☐☐ or > ☐☐☐.☐☐ ☐ mg/L ☐ mg/dL ☐ mmol/L

Haematology ☐ not done

Haemoglobin ☐☐.☐ g/dL or ☐.☐☐ g/L or ☐☐.☐ mmol/L

Platelets ☐☐☐☐ x10^9/L or ☐☐☐☐ x10^3/µL

Total WBC ☐☐☐.☐ x10^9/L or ☐☐☐.☐ x10^3/µL

Neutrophils ☐☐☐.☐ x10^9/L or ☐☐☐.☐ x10^3/µL

ANCA ☐ not done

Please circle the result

Immunofluorescence | Negative | P | C | Indeterminate | titre if known 1/☐☐☐☐

ELISA MPO ☐☐☐.☐ IU/ml or > ☐☐☐.☐ IU/ml

ELISA PR3 ☐☐☐.☐ IU/ml or > ☐☐☐.☐ IU/ml

Urine dipstick ☐ not done

Blood | 0 | Trace | + | ++ | +++

Please circle the result

Protein | 0 | Trace | + | ++ | +++

Birmingham Vasculitis Activity Score (BVAS)

Tick an item only if attributable to active suspected GCA. If there are no abnormalities in a section, please tick 'No' for that organ-system. If there are abnormalities, tick yes and tick all items attributable to active suspected GCA.

General? ☐ Yes ☐ No

If yes:

- ☐ Myalgia
- ☐ Arthralgia / arthritis
- ☐ Fever ≥ 38°C
- ☐ Weight loss ≥ 2kg

Cutaneous? ☐ Yes ☐ No

If yes:

- ☐ Infarct
- ☐ Purpura
- ☐ Ulcer
- ☐ Gangrene
- ☐ Other skin vasculitis

Mucous membranes / eyes? ☐ Yes ☐ No

If yes:

- ☐ Mouth ulcers
- ☐ Genital ulcers
- ☐ Adnexal inflammation
- ☐ Significant proptosis
- ☐ Scleritis / Episcleritis
- ☐ Conjunctivitis / Blepharitis / Keratitis
- ☐ Blurred vision
- ☐ Sudden visual loss
- ☐ Uveitis
- ☐ Retinal changes (vasculitis / thrombosis / exudate / haemorrhage)

ENT? ☐ Yes ☐ No

If yes:

- ☐ Bloody nasal discharge / crusts / ulcers / granulomata
- ☐ Paranasal sinus involvement
- ☐ Subglottic stenosis
- ☐ Conductive hearing loss
- ☐ Sensorineural hearing loss

Chest? ☐ Yes ☐ No

If yes:

- ☐ Wheeze
- ☐ Nodules or cavities
- ☐ Pleural effusion / pleurisy
- ☐ Infiltrate
- ☐ Endobronchial involvement
- ☐ Massive haemoptysis / alveolar haemorrhage
- ☐ Respiratory failure

Cardiovascular? ☐ Yes ☐ No

If yes:

- ☐ Loss of pulses
- ☐ Vascular heart disease
- ☐ Pericarditis
- ☐ Ischaemic cardiac pain
- ☐ Cardiomyopathy
- ☐ Congestive cardiac failure

Abdominal? ☐ Yes ☐ No

If yes:

- ☐ Peritonitis
- ☐ Bloody diarrhoea
- ☐ Ischaemic abdominal pain

Renal? ☐ Yes ☐ No

If yes:

- ☐ Hypertension
- ☐ Proteinuria > 1+
- ☐ Haematuria ≥ 10 RBCs/hpf
- ☐ Serum creatinine 125-249 µmol/L*
- ☐ Serum creatinine 250-499 µmol/L*
- ☐ Serum creatinine ≥500 µmol/L*
- ☐ Rise in serum creatinine >30% or fall in creatinine clearance >25%

* Can only be scored on the first assessment

Nervous system? ☐ Yes ☐ No

If yes:

- ☐ Headache
- ☐ Meningitis
- ☐ Organic confusion
- ☐ Seizures (not hypertensive)
- ☐ Cerebrovascular accident
- ☐ Spinal cord lesion
- ☐ Cranial nerve palsy
- ☐ Sensory peripheral neuropathy
- ☐ Mononeuritis multiplex

Other? ☐ Yes ☐ No

If yes, specify:

PERSISTENT DISEASE ONLY*

☐ Tick if ALL abnormalities are due to persistent disease.

*Active suspected GCA which is not new / worse in the prior 4 weeks.

VASCULITIS DAMAGE INDEX (VDI)

This is for recording organ damage that has occurred in patients *since the onset of suspected GCA*
Patients often have co-morbidity before onset of suspected GCA, **which must not be scored**
Record features of active disease using the Birmingham Vasculitis Activity Score (BVAS)
A new patient should **usually have a VDI score of zero**, unless:
(a) they have had suspected GCA for more than three months **and**
(b) the damage has developed or become worse since the onset of suspected GCA

Musculoskeletal? ☐ Yes ☐ No

If yes:
- ☐ Significant muscle atrophy or weakness
- ☐ Deforming/erosive arthritis
- ☐ Osteoporosis/vertebral collapse
- ☐ Avascular necrosis
- ☐ Osteomyelitis

Skin/Mucous membranes? ☐ Yes ☐ No

If yes:
- ☐ Alopecia
- ☐ Cutaneous ulcers
- ☐ Mouth ulcers

Ocular? ☐ Yes ☐ No

If yes:
- ☐ Cataract
- ☐ Retinal change
- ☐ Optic atrophy
- ☐ Visual impairment/diplopia
- ☐ Blindness in one eye
- ☐ Blindness in second eye
- ☐ Orbital wall destruction

ENT? ☐ Yes ☐ No

If yes:
- ☐ Hearing loss
- ☐ Nasal blockage/chronic discharge/crusting
- ☐ Nasal bridge collapse/septal perforation
- ☐ Chronic sinusitis/radiological damage
- ☐ Subglottic stenosis (no surgery)
- ☐ Subglottic stenosis (with surgery)

Pulmonary? ☐ Yes ☐ No

If yes:
- ☐ Pulmonary hypertension
- ☐ Pulmonary fibrosis
- ☐ Pulmonary infarction
- ☐ Pleural fibrosis
- ☐ Chronic asthma
- ☐ Chronic breathlessness
- ☐ Impaired lung function

Cardiovascular? ☐ Yes ☐ No

If yes:
- ☐ Angina/angioplasty
- ☐ Myocardial infarction
- ☐ Subsequent myocardial infarction
- ☐ Cardiomyopathy
- ☐ Valvular disease
- ☐ Pericarditis ≥ 3 mths or pericardectomy
- ☐ Diastolic BP ≥ 95 or requiring antihypertensives

Peripheral vascular disease? ☐ Yes ☐ No

If yes:
- ☐ Absent pulses in one limb
- ☐ 2nd episode of absent pulses in one limb
- ☐ Major vessel stenosis
- ☐ Claudication >3 mths
- ☐ Minor tissue loss
- ☐ Major tissue loss
- ☐ Subsequent major tissue loss
- ☐ Complicated venous thrombosis

Gastrointestinal? ☐ Yes ☐ No

If yes:
- ☐ Gut infarction/resection
- ☐ Mesenteric insufficiency / pancreatitis
- ☐ Chronic peritonitis
- ☐ Oesophageal stricture/surgery

Renal? ☐ Yes ☐ No

If yes:
- ☐ Estimated/measured GFR < 50%
- ☐ Proteinuria > 0.5g/24hr
- ☐ End stage renal disease

Neuropsychiatric? ☐ Yes ☐ No

If yes:
- ☐ Cognitive impairment
- ☐ Major psychosis
- ☐ Seizures
- ☐ Cerebrovascular accident
- ☐ 2nd cerebrovascular accident
- ☐ Cranial nerve lesion
- ☐ Peripheral neuropathy
- ☐ Transverse myelitis

Other? ☐ Yes ☐ No

If yes:
- ☐ Gonadal failure ☐ Chemical cystitis
- ☐ Marrow failure ☐ Malignancy
- ☐ Diabetes ☐ Other

Total VDI Score*

☐ Record the number of positive items (1 point for each).

*The VDI score can either increase or remain the same over time. Remember to carry forward any previous items of damage

Diagnosis

Has the clinical diagnosis changed compared to visit 2? ☐ Yes ☐ No

If no, no further details are required on this page.

If yes:

Does the participant have features consistent with a diagnosis of GCA? ☐ Yes ☐ No

If yes, which of the following influenced your decision (tick all that apply):

☐ symptoms ☐ signs ☐ blood abnormalities

☐ biopsy report ☐ other, specify:

If no, please give at least one alternative diagnosis:

☐ non specific headache
☐ migraine
☐ myofascial pain
☐ temperomandibular dysfunction
☐ cervical spondylosis
☐ fibromyalgia
☐ sinusitis
☐ orbital cellulitis
☐ shingles
☐ orbital pseudotumour
☐ metastatic disease (cancer)
☐ lymphoma
☐ Paget's disease
☐ Takayasu's arteritis
☐ large vessel vasculitis
☐ polyarteritis nodosa
☐ Granulomatosis with polyangiitis (GPA)
☐ microscopic polyangiitis
☐ Churg-Strauss syndrome
☐ cryoglobulinemic vasculitis
☐ Henoch-Schonlein purpura
☐ other vasculitis, specify:
☐ other, specify:

Steroids

Has the participant taken any steroids? ☐ Yes ☐ No

If yes, are these being given for: (tick all that apply) ☐ Suspected GCA ☐ Other condition

Please record details below:

Name of Steroid	Route†	Dose (mg)*
Current steroids:		
Previous steroid preparations:		

Immunosuppressants

Is the participant currently taking any immunosuppressants? ☐ Yes ☐ No

If yes, are these being given for: (tick all that apply) ☐ Suspected GCA ☐ Other condition

If yes, please record details below:

Name of current Immunosuppressant	Route†	Total daily dose - unit	Start date
		-	D D M M Y Y
		-	D D M M Y Y
		-	D D M M Y Y
		-	D D M M Y Y
		-	D D M M Y Y

*Total daily dose in mg †e.g. PO=Oral, IV=intravenous, IM=intramuscular; IA=intraarticular

End of visit checklist

	Yes	No
Has the participant taken any concomitant medications? If yes please complete the concomitant medications form (page 27).	☐	☐
Has the participant had any adverse events? If yes please complete the adverse event form (separate pad).	☐	☐
Has the participant had any serious adverse events? If yes please complete the adverse event form (separate pad).	☐	☐
Has the participant completed the EQ5D? If yes please store the completed EQ5D in the participant's study file.	☐	☐
Has the participant completed the study? If no please complete the discontinuation form (page 28).	☐	☐

I certify that the data contained in the visit three CRF and pages 25 to 27 are complete and accurate.
(To be signed and dated by the investigator or authorised member of the investigator's staff)

Signature _____ Date D D M M Y Y Y Y

Print name _____

Please indicate the onset and resolution dates of any of the specified conditions present at any time from baseline to the six month visit (recorded on pages 2, 9 and 17). Where possible specify the day, month and year; if day is unknown specify year and, if possible, month (please cross through any unknown days and / or months). For any conditions still present at the six month visit tick ongoing.

Condition*	Details (if required**)	Onset date	Ongoing, or:	Resolution date
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y

*Please use abbreviation from the table below (conditions as defined on pages 2, 9 and 17)

DM: Diabetes Mellitus
HT: Hypertension

Ang: Angina
MI: Myocardial infarction
CCF: Heart failure

LTFH: Low trauma fracture - of hip
LTFS: - of spine
LTFF: - of forearm
LTFO: - other, specify**

Neo: Neoplasia, specify**

** If LTFO or Neo: Please specify in the details field.

Please tick here if conditions are continued on another page (download additional pages from the TABUL website: https://weblearn.ox.ac.uk) ☐

Please indicate the onset and resolution dates of any of the specified symptoms present at any time from baseline to the six month visit (recorded on pages 3, 4, 10 and 18). Where possible specify the day, month and year; if day is unknown specify year and, if possible, month (please cross through any unknown days and / or months). For any symptoms still present at the six month visit tick ongoing.

Symptom*	If 'Oth' please give details**	Onset date	Ongoing, or:	Resolution date
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y
		D D M M Y Y Y Y	☐	D D M M Y Y Y Y

*Please use abbreviation from the table below. (Symptoms as defined on pages 3, 4, 10 and 18)

An: Anorexia Ftg: Fatigue FNS: Fever or night sweats BSP: Bilateral shoulder pain	EMS: Early morning stiffness > 1 hour BHS: Bilateral hip stiffness or pain LPH: Localized pain in the head GST: Generalised scalp tenderness	STA: Swelling over temporal artery PTA: Pain over temporal artery JC: Jaw claudication TC: Tongue claudication	RLV: Reduced or lost vision in either eye DV: Double vision AF: Amaurosis fugax Oth: Other, specify**

Please tick here if symptoms are continued on another page (download additional pages from the TABUL website: https://weblearn.ox.ac.uk) ☐

Please provide details of any medications taken throughout the duration of the study (other than steroids and immunosuppressants given for suspected GCA). Where possible specify the start and end date of each medication (day, month and year if known; if day and month is unknown specify year (please cross through any unknown days and / or months). For any medications still being taken at the six month visit, tick continuing and leave the end date blank.

Medication name	Route*	Dose	Unit*	Frequency*	Reason	Start date	Continuing or end date
						D D M M Y Y Y Y	☐ D D M M Y Y Y Y
						D D M M Y Y Y Y	☐ D D M M Y Y Y Y
						D D M M Y Y Y Y	☐ D D M M Y Y Y Y
						D D M M Y Y Y Y	☐ D D M M Y Y Y Y
						D D M M Y Y Y Y	☐ D D M M Y Y Y Y
						D D M M Y Y Y Y	☐ D D M M Y Y Y Y
						D D M M Y Y Y Y	☐ D D M M Y Y Y Y
						D D M M Y Y Y Y	☐ D D M M Y Y Y Y

*Please use codes below if applicable

Route:
PO: Oral
inh: Inhaled
SC: Subcutaneous
IV: Intravenous
IM: Intramuscular
IA: Intraarticular
TD: Transdermal
nas: Intranasal
TOP: Topical
PR: Rectal
If other, specify

Unit:
mg: milligrams
g: gram
mcg: microgram
mg/kg: milligrams/kilogram
mg/m^2: milligrams/meter squared
cap: capsule
IU: International Units
l: litre
drops: drops
patch: patch
mls: millilitres
puffs: puffs
tab: tablet
If other, specify

Frequency:
OD: Once daily
BD: Twice daily
TDS: Three times a day
QDS: Four times daily
PRN: As required
Nocte: At night
Mane: Morning
Q4H: Every 4 hours
STAT: Once only
WKY: Once weekly
If other, specify

Please tick here if medications are continued on another page (download additional pages from the TABUL website: https://weblearn.ox.ac.uk) ☐

Study discontinuation

Last date of participation in study D D M M Y Y Y Y

Discontinuation reason (tick all that apply)

- ☐ Patient withdrew consent
- ☐ Investigator discretion
- ☐ Patient lost to follow-up
- ☐ Biopsy not done
- ☐ Ultrasound scan not done
- ☐ Patient died
- ☐ Other

Details

I certify that the participant has discontinued.
(To be signed and dated by the investigator or authorised member of the investigator's staff)

Signature

Date D D M M Y Y Y Y

Print name

Appendix 8 Completion of the clinical case report form

The standard operating procedure for completion of the clinical case report form can be accessed via the following link: http://ora.ox.ac.uk/objects/uuid:0eb6d248-0fe9-47a4-b151-81bfc6dfc982 (accessed 27 May 2016).

Appendix 9 Adverse event case report form

Adverse event

Start date D D M M Y Y Y Y

End date D D M M Y Y Y Y

Reported on: ☐ V1 Baseline: ☐ V2, 2 weeks ☐ V3, 6 Months ☐ Other

Comment:

Severity	Serious?	Related to Ultrasound scan?	Ultrasound Scan interrupted as a result of this adverse event?
☐ Mild ☐ Moderate ☐ Severe	☐ Yes* ☐ No	☐ Definitely related ☐ Possibly related ☐ Not related ☐ Unable to assess	☐ No ☐ Temporarily ☐ Permanently
Expected? (e.g. related to medication)	**Outcome**	**Related to temporal arterty biopsy?**	**Temporal arterty biopsy interrupted as a result of this adverse event?**
☐ Yes ☐ No	☐ Recovered ☐ Recovered with sequelae ☐ Ongoing ☐ Died	☐ Definitely related ☐ Possibly related ☐ Not related ☐ Unable to assess	☐ No ☐ Temporarily ☐ Permanently

*Serious adverse events

☐ Death Date - D D M M Y Y Y Y

☐ Life or limb threatening event

☐ Hospitalisation required - No of days ☐☐ Comment:

☐ Hospitalisation prolonged

☐ Persistent or significant disablility / incapacity

☐ Congenital abnormality

☐ Other important medical event that may jeopardise the participant

If the adverse event is serious please contact: 01865 737221 or 01865 227326 within 24 hours (or 07905211359 out of hours); fax this form to 01865 737640 or e-mail to tabul@ndorms.ox.ac.uk and tick here to indicate this is done. ☐

Reporting person / position /

Signature Date of reporting D D M M Y Y Y Y

Appendix 10 Completion of the safety report form

The standard operating procedure for completion of the safety report form (describing any AEs or serious AEs) can be accessed via the following link: http://ora.ox.ac.uk/objects/uuid:b717083c-d287-4489-b06a-041d0000eaca (accessed 27 May 2016).

Appendix 11 Collection, processing and storage of biopsy samples

The standard operating procedure for collection, processing and storage of biopsy samples can be accessed via the following link: http://ora.ox.ac.uk/objects/uuid:b5132a1c-a1d4-4c99-8c45-7b9f43d98512 (accessed 27 May 2016).

Appendix 12 Biopsy case report form

Biopsy report

Date of biopsy D D M M Y Y Y Y Date of report D D M M Y Y Y Y

Macroscopic appearance

Which side was the biopsy taken from? ☐ Left ☐ Right

Length ☐☐ mm

Is this a bifurcated sample? ☐ Yes ☐ No

Do you have any other comments on macroscopic appearance? ☐ Yes ☐ No

If yes, specify:

Microscope

Have you used the microscope that you registered with us? ☐ Yes ☐ No

If no, please inform the study co-ordinator at their next visit.

Did you use your routine staining protocol? ☐ Yes ☐ No

If no, specify:

Microscopic appearance

Is this temporal artery? ☐ Yes ☐ No*

Which sections did you cut? (tick all that apply) ☐ Transverse ☐ Longitudinal

Were deeper levels required (because initial sections did not provide enough diagnostic information)? ☐ Yes ☐ No

*If not temporal artery, is it:

☐ fat or muscle ☐ vein ☐ nerve ☐ other, specify:

Intima

☐ Normal, or tick all that apply

☐ Arteriosclerosis present ☐ Intimal hyperplasia present

Internal elastic lamina

☐ Normal, or tick all that apply

☐ Fragmentation ☐ Reduplication

Predominant site of inflammatory cellular infiltrate

Is there an inflammatory infiltrate present in this sample? ☐ Yes ☐ No

If Yes, Predominant site of inflammation:

☐ Intima ☐ Internal elastic lamina ☐ Media ☐ Adventitia ☐ Vasa vasorum ☐ Transmural

Details

Please tick all features that are present:

☐ Normal areas ☐ Giant cells ☐ Calcification ☐ Unusual features, specify:

Thrombus and occlusion

Is the vessel completely occluded? ☐ Yes ☐ No

If yes, is it due to: (tick all that apply) ☐ Thrombus ☐ Intimal hyperplasia

If no, is there : Thrombus in at least one section? ☐ Yes ☐ No

Intimal hyperplasia in at least one section? ☐ Yes ☐ No

Is there evidence of recanalisation in at least one section? ☐ Yes ☐ No

Pathological diagnosis ☐ Normal, or tick all that apply

Compatible with a diagnosis of giant cell arteritis ☐

Compatible with a diagnosis of other vasculitis ☐

Compatible with a diagnosis of arteriosclerosis ☐

Compatible with another diagnosis, please specify: ☐

Checklist

Have all slides been sent to the TABUL Office? ☐ Yes ☐ No

If Yes, How many slides are available for the study? ☐☐ (If none, enter zero)(0).

Have all remaining blocks been sent to the TABUL Office? ☐ Yes ☐ No

If Yes, How many paraffin blocks are available for the study? ☐ (If none, enter zero)(0).

If Yes, How many frozen blocks are available for the study? ☐ (If none, enter zero)(0).

Has an anonymised copy of the original biopsy report been attached to this form? ☐ Yes ☐ No

Slides and blocks to be stored in the Nuffield Orthopaedic Centre biobank

Comments

I certify that the data contained in this biopsy report are complete and accurate. (To be signed and dated by the pathologist)

Signature Date D D M M Y Y Y Y

Print name Position ☐ Consultant ☐ Trainee

If trainee, time in post: ☐☐ yrs ☐☐ mths

Appendix 13 Completion of the biopsy report case report form

The standard operating procedure for completion of the biopsy report case report form can be accessed via the following link: http://ora.ox.ac.uk/objects/uuid:eeebc59f-9ee3-4e40-a7dd-1b3d6179f972 (accessed 27 May 2016).

Appendix 14 Statistical analysis plan

Statistical analysis plan

FINAL version

Study Title	**Temporal Artery Biopsy vs ULtrasound in diagnosis of Giant Cell Arteritis (GCA)**
Short title	**TABUL**
Funding body / Reference	**NIHR HTA Reference Number: 08/64/01**
Sponsor	**University of Oxford**

Authored by

_____ ___/___/_____

Mike Bradburn *Date*

Study Statistician

CTRU, University of Sheffield

Approved by

_____ ___/___/_____

Raashid Luqmani

Chief Investigator

Consultant Rheumatologist, Oxford Nuffield Orthopaedic Centre

Professor of Rheumatology , University of Oxford

_____ ___/___/_____

Andrew Hutchings

Co-Chief Investigator

Statistician

London School of Hygiene and Tropical Medicine

_____ ___/___/_____

Surjeet Singh

Clinical Trial Co-ordinator

Nuffield Orthopaedic Centre, University of Oxford

Table of contents

List of abbreviations used

AE	Adverse event
ANCA	Anti-neutrophil cytoplasm antibodies
BVAS	Birmingham Vasculitis Activity Score
CRF	Case report form
CRP	C-reactive protein
CTRU	University of Sheffield Clinical Trial Research Unit
DMC	Data Monitoring Committee
EQ-5D	EuroQol health utility questionnaire
ESR	Erythrocyte sedimentation rate
FAS	Full analysis set
GCA	Giant cell arteritis
GCP	Good Clinical Practice
PMR	Polymyalgia Rheumatica
PP	Per-protocol
QALY	Quality adjusted life year
SAE	Serious adverse event
ROC	Receiver-operator characteristic
TAB	Temporal artery biopsy
TMG	Trial Management Group
TSC	Trial Steering Committee
US	Ultrasound
VAS	Visual analogue scale
VDI	Vasculitis Damage Index
WBC	White blood count

1 Introduction, study design and key trial objectives

1.1 Study outline

The study will assess the performance of ultrasound (US) and temporal arterial biopsy (TAB) in the diagnosis of giant cell arteritis (GCA).

The document was compiled with reference to TABUL protocol version number: 6.0 (Effective date 22 Jan 2013)

This statistical analysis plan is written in conjunction with the International Conference on Harmonisation (ICH) topic E9 (Statistical principles for clinical trials, 1998), applicable standard operating procedures from the University of Sheffield Clinical Trials Research Unit (CTRU) and trial documents referenced in section 6. The trial will be conducted in accordance with Good Clinical Practice (GCP) in Clinical Trials (International Conference on Harmonisation, 1996).

Any deviations from the statistical analysis plan will be described and justified in the final report of the trial. The analysis should be carried out by an identified, appropriately qualified and experienced statistician, who should ensure the integrity of the data during their processing. Examples of such procedures include quality control and evaluation procedures.

1.2 Outcome measures

The objectives of the trial are given in the synopsis below:

Study title	**The Role of Ultrasound Compared to Biopsy of Temporal Arteries in the Diagnosis and Treatment of Giant Cell Arteritis (GCA).**
Short title	**T**emporal **A**rtery **B**iopsy vs **Ult**rasound in diagnosis of giant cell arteritis (TABUL)
Internal Ref No	REC: 09/H0505/132, HTA: 08/64/01
Study Design	Cohort study (Observational)
Number of Participants	**435-445** (in order to achieve **402** participants that have completed the primary end-point at Visit 2 (two weeks))
Primary Objectives	**1.** To evaluate the diagnostic accuracy (sensitivity and specificity) of ultrasound as an alternative to temporal artery biopsy for the diagnosis of GCA in patients referred for biopsy with suspected GCA. **2.** To evaluate the cost-effectiveness (incremental cost per QALY) of ultrasound instead of biopsy in the diagnosis of GCA.
Secondary Objectives	**3.** To evaluate inter-observer agreement in the assessment of ultrasound and temporal artery biopsy. **4.** To elicit expert views on the appropriateness of performing a biopsy following ultrasound using clinical vignettes. **5.** To evaluate the diagnostic accuracy (sensitivity and specificity) of the sequential diagnostic strategy from **4** as an alternative to temporal artery biopsy alone in the diagnosis of GCA. **6.** To evaluate the cost-effectiveness (incremental cost per QALY) of the diagnostic strategy from **4** instead of biopsy alone in the diagnosis of GCA.

1.3 Eligibility

The inclusion and exclusion criteria as stated in the protocol are reproduced below:

1.3.1 General considerations

We will adopt a pragmatic approach to recruitment, i.e. aim to include all patients undergoing temporal artery biopsy for suspected GCA. A lower age restriction of 18 will be applied although no age criteria is necessary for this disease type as we expect the majority of patients to be elderly. The clinicians will be using their judgment and clinical experience to determine whether or not to refer for biopsy. We will include patients with pre-existing polymyalgia rheumatica. Most patients will be treated with a standard dose of prednisolone, but some may be commenced on prednisolone plus another immunosuppressive agent. They will be included even though we suspect that the biopsy and scan results may be affected differently than when compared to steroids alone.

1.3.2 Inclusion criteria

For the cohort study

(1) A clinical suspicion of new diagnosis of GCA e.g. patients with a new onset of headache, scalp tenderness, with or without elevated CRP or ESR, jaw or tongue claudication with or without visual loss.

(2) The clinician decides that the patient requires an urgent temporal artery biopsy to determine whether or not the diagnosis is GCA.

(3) The patient agrees and provides NHS consent to undergo a temporal artery biopsy as part of standard care.

(4) Patients have been started on high dose glucocorticoids or will be started on high dose glucocorticoids.

(5) Patients must be willing to attend for an ultrasound scan of their temporal and axillary arteries.

(6) Participants must be willing to give informed written consent or willing to give permission for a nominated friend or relative to provide written informed assent if they are unable to do so because of physical disabilities e.g. sudden onset of blindness/vision loss which can be caused by GCA (this will be made clear in the ethics approval application).

(7) Must be 18 years of age or

For the training cases

(1) Patients attending hospital outpatient or in patient departments for assessment for any condition (apart from giant cell arteritis or polymyalgia rheumatica) or healthy staff volunteers.

(2) Above the age of 50 years.

(3) Willing to attend for an ultrasound scan of their temporal and axillary arteries.

(4) Willing and able to give written informed consent.

1.3.3 Exclusion criteria

For the cohort study

(1) Previous diagnosis of GCA.

(2) Use of high dose glucocorticoid (>20mg prednisolone/day) for management of current suspected GCA for more than 7 days prior to the dates of the ultrasound and biopsy.

(3) Long term (>1 month) high dose (>20mg per day at any time) steroids for conditions other than PMR, within three months prior to study entry.

(4) Inability to give informed consent (either written consent or verbal assent from a relative or carer)

(5) Inability to undergo an ultrasound scan of the temporal and axillary arteries.

(6) Patients with a known cause of headache (not due to GCA), or any condition which would preclude the need for a temporal artery biopsy.

(7) Patients who are unable to undergo an ultrasound scan and a temporal artery biopsy within 7 days of starting glucocorticoids.

For the training cases

(1) Diagnosis of suspected GCA or a previous history of diagnosed or suspected GCA.

(2) Inability to give written informed consent.

(3) Inability to undergo an ultrasound scans of the temporal and axillary arteries

1.4 Randomisation and blinding

No randomisation will be employed: all patients are scheduled to have both TAB and US.

The 2-week visit will be blind to the US findings. If however the assessor intends to withdraw steroids, they contact the TABUL team, who will provide them with the US result. The assessor may choose to continue or withdraw steroids after unblinding the US data. The CRF will capture when this occurs. This procedure ensures that the original diagnosis is based on standard care whilst also allowing additional US findings to protect the safety of the patient.

1.5 Interim analyses, data monitoring committees etc.

Three committees will be established to govern the conduct of this study:

- Trial Management Group (TMG)

This consists of the TABUL study team at the lead site in Oxford (led by the Chief Investigator – Prof Raashid Luqmani) and Andrew Hutchings (Co-Chief Investigator). A list of the TMG is given below:

Name	Function
Professor Raashid Luqmani	Chief Investigator
Mrs Shauna Masters	Research Nurse
Mr. Andrew Hutchings	Co-chief investigator / Statistician
Mrs Joanna Burchall	Research Nurse
Dr Surjeet Singh	Trial Co-ordinator
Mr Varun Manhas	Biomedical Scientist
Miss Vanshika Sharma	Biomedical Scientist
Mrs Jennifer Piper	Ultrasonographer
Mr Wulf Forrester-Barker	IT Manager

- Independent Trial Steering Committee (TSC)

The TSC are an independent group who will provide trial oversight.

Name	Function
Professor Michael Ehrenstein	Chair/ Consultant rheumatologist
Professor Bleddyn Davies	Patient Representative
Professor Karim Raza	Clinical Rheumatologist
Professor David Mant	Emeritus Professor of General Practice

Members of the TMG may attend DMC meetings as non-voting members.

- Independent Data Monitoring/Management Committee (DMC)

The DMC are an independent committee who, other than offering recommendations to the TSC (primarily around safety to patients), are not involved with the TABUL trial in way. Their membership is as follows:

Name	Function
Dr Lyn Williamson	Chair/consultant rheumatologist
Prof Jonathan Sterne	Statistician
Dr Simon Travis	Experimental medicine specialist
Kate Gilbert	Patient representative

Members of the TMG may attend the open section of DMC meetings as non-voting members.

During the study the CTRU will provide the committees with status reports detaining the data completeness, recruitment, loss to follow-up, compliance to protocol and safety outcomes. The remit of the DMC includes recommending that the trial cease on safety grounds, but other than this no interim analyses are planned.

2 Data sources, data and analysis populations

2.1 Sample Size and Power

A sample of 402 patients provides 90% power at a 5% type I error rate to test the joint hypothesis that

(i) US has greater sensitivity than TAB, based on detecting a sensitivity of 76% for TAB and 87% sensitivity for US; and
(ii) The specificity of US is no less than 83%, based on an expected specificity of 96%

The postulated sensitivity and specificity figures are based on the meta-analysis by Karassa et al (2005). This sample size will allow estimation of a one-sided rectangular confidence region for ultrasound false and true positive fractions, assuming 80% prevalence of GCA in patients having a biopsy for suspected GCA, with the sample size inflated (gamma=0.1) due to uncertainty in the proportion of cases/controls in a cohort design (Pepe, 2003).

In order to allow for losses to follow-up the original plan was to recruit 430 participants to the study. It was anticipated that most losses to follow-up will occur in recruited participants not having both ultrasound and biopsy at their appointed time within 7 days of starting steroids. Previous experience of observational studies in PMR (Hutchings et al 2007) suggest no further losses at the

week 2 assessment and 4% at the 6 month assessment. However, the primary outcome (reference standard diagnosis) can be derived in participants who do not complete the 6 month assessment. These assumptions will be checked using early monitoring of recruitment and follow-up rates, with recruitment targets modified as necessary. After monitoring the actual recruitment and withdrawals it was found that the withdrawal rate was slightly higher than expected. It was decided to change the target recruitment figure to 435-445.

2.2 Data sources

The data used in this study will come from data entered onto the following Case Report Forms (CRFs):

Visit	CRF reference	Version No/Effective date	
Training Screening	Screening – Training	1.0	6 April 2010
Training Baseline	Baseline- Training	1.0	6 April 2010
US for training	Ultrasound CRF	2.0	3 June 2010
Screening	Screening – Clinical	2.0	4 June 2010
Baseline	Baseline (Visit 1)- Clinical	2.0	20 September 2011
Baseline	Baseline (Visit 1)- EQ-5D	n/a	EuroQol 1990
< 2 weeks	Biopsy CRF	2.0	20 September 2011
< 2 weeks	Ultrasound CRF	2.0	3 June 2010
2 weeks	Two Weeks (Visit 2) – Clinical	2.0	20 September 2011
2 weeks	Two Weeks (Visit 2) – EQ-5D	n/a	EuroQol 1990
6 months	Six Months (Visit 3) – Clinical	2.0	20 September 2011
6 months	Six Months (Visit 3) – EQ-5D	n/a	EuroQol 1990
From consent to last visit	Adverse Event/Reaction reporting form	3.0	20 September 2011

This data will be stored on the Sheffield CTRU database (PROSPECT). Images from US and TAB will be stored on the Oxford database and will be made available for subsequent expert review and assessment of inter-observer agreement.

2.3 Protocol Deviations

The following are considered major deviations:

1. High dose steroids started more than 2 weeks before presentation
2. TAB not performed within 10 days of starting high dose steroids
3. US either not performed or performed after TAB

2.4 Analysis populations

Analyses will be conducted on the following groups of patients:

Name	Patients included
Training phase	
Training cases	All patients who enter into the training (pre-) phase of the study
Main study	
Primary analysis	All patients who did not deviate from the protocol, as defined in section 2.3, and for whom a clinical diagnosis has been made.
Per-protocol analysis	The subset of the primary analysis population for whom TAB was undertaken within 7 days of commencement of steroids
US/TAB agreement analysis	All patients who did not deviate from the protocol, as defined in section 2.3.
US/TAB agreement – per-protocol analysis	The subset of the US/TAB agreement analysis population for whom TAB was undertaken within 7 days of commencement of steroids.

If sufficient data are available, a separate analysis will be performed for patients in whom the TAB was more than 10 days after high dose steroids, with specific attention paid to those who were delayed for clinical reasons.

2.5 Data Management

A Data Management Plan (DMP) agreed by the CTRU and TMG will define the procedures for data entry, cleaning and validation

3 Outline of analyses

3.1 General considerations

Data will be reported and presented according to the STARD statement (Bossuyt et al 2003).

Complete details of data derivations and methods of handling multiplicity, multi-centre data and missing data are covered in section 4. No interim analyses or early stopping are planned.

All summaries will be provided on the complete case patient set unless otherwise stated

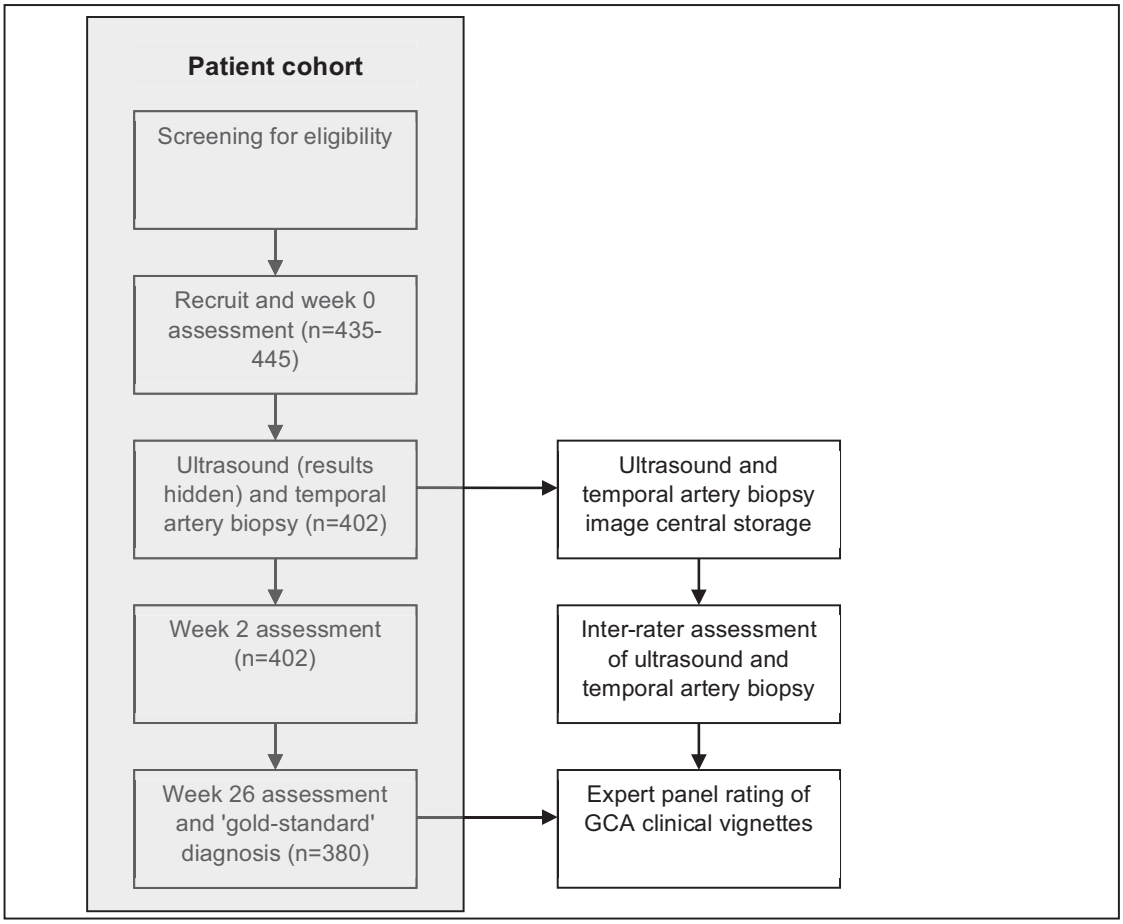
All tables will present summary statistics defined by the nature of the measurement. Summaries of continuous variables will comprise the number of observations used and either
i) mean, median, standard deviation, minimum and maximum, or
ii) median, inter-quartile range, minimum and maximum
as appropriate for the distributional form of the data. Summaries of categorical variables will comprise the number of observations used, and the number and percentage of observations in each category.
All statistical tests will be two-tailed with alpha = 0.05, and all confidence intervals will be two-sided, 95% intervals, unless one-sided tests of joint hypotheses are specifically stated.

3.2 Disposition and data completeness

The flow of patients to the various stages will be summarised by the following:

Enrolment	The number of patients screened for entry, the number of patients entered, the number of patients not entered with reasons, the diagnosis (consistent with GCA, not consistent with GCA, not available) from TAB, the diagnosis from US, the number of patients who discontinued at Visit 2 (two weeks) and Visit 3 (6 months), and the reference diagnosis (as described in the flow diagram below).

Patient recruitment and follow up with data collection



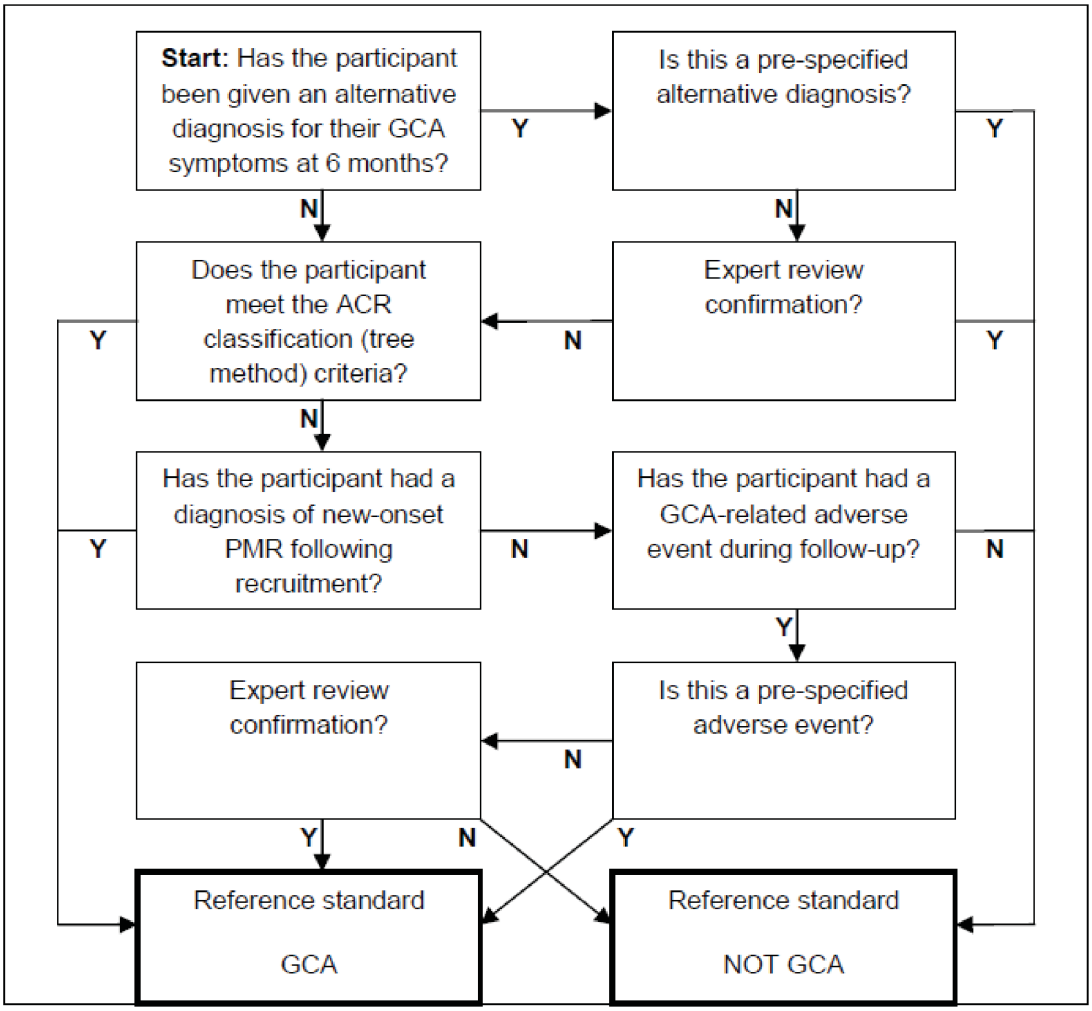


Figure 1 Algorithm for study patients

3.3 Demographics and baseline characteristics

The following summaries will be presented:

Demographics	Treatment centre, age at baseline, gender, ethnic group, current smoking status and smoking history.
Presenting and evolving medical history and conditions	The proportion of patients with each current condition at presentation, 2 weeks and 6 months.
Presenting and evolving symptoms	The proportion of patients with each symptom at presentation, 2 weeks and 6 months.
Physical and evolving examination	The proportion of patients with each abnormal feature at presentation, 2 weeks and 6 months.
Initial diagnosis and treatment	The certainty of GCA diagnosis, the proportion of patients taking steroids, and the proportion of patients taking immunosuppressants.

Should a patient transfer between centres during the study, the “centre” defined in the analysis is the place at which the patient underwent US.

3.4 Efficacy

3.4.1 Accuracy of US and TAB in relation to reference diagnosis (primary endpoint)

The principal outcome is the performance of US and TAB in relation to the reference ('gold-standard') diagnosis of GCA. The reference diagnosis will be reached as specified in the study protocol.

The following summaries will be presented

TAB	The cross-tabulation of diagnosis by TAB against reference diagnosis, together with the sensitivity, specificity, and associated 95% confidence intervals
US	The cross-tabulation of diagnosis by US against reference diagnosis, together with the sensitivity, specificity, and associated 95% confidence intervals
TAB v US	The cross tabulation of diagnosis by TAB against the diagnosis by US, overall and by final reference diagnosis, together with Kappa statistic and the McNemar test

The kappa statistic will be used to assess agreement between TAB and US, and McNemar’s test will be used to detect systematic discordance i.e. whether one method is more or less likely to diagnose (Fleiss *et al* 2003). Rectangular confidence intervals and the McNemar test statistics will be calculated using the exact Binomial methods.

The performance of US will be evaluated as defined in the protocol using a one-sided rectangular confidence region for sensitivity > TAB and specificity > 0.83 with a 5% type I error rate.

3.4.2Performance of US and TAB in relation to patient & disease characteristics

Several additional analyses are planned around the accuracy of US and TAB in relation to different characteristics of the disease and subgroups of patient.

Analyses will be carried out using logistic regression in which the outcome is the reference diagnosis. The performance of US across each characteristic will be assessed, initially one characteristic at a time, by fitting the US diagnosis, the characteristic and their interaction term. A receiver-operator characteristic (ROC) curve will be produced to assess the sensitivity and specificity across different levels of the covariate. The same procedure will be used to assess the performance of TAB.

Following on from this, further modelling will be undertaken to evaluate different strategies for the detection of GCA. The performance of the following strategies will be assessed:

1. Standard diagnosis (“how does the standard method perform?”)
1a. TAB alone
1b. TAB plus additional potential prognostic factors collected at baseline (e.g. age, history, BVAS)
2. Experimental diagnosis (“how do the standard and test methods perform together?”)
2a. TAB plus US (independently) alone
2b. TAB plus US (either/or positive) alone
2c. TAB plus US (interaction) alone
2d. TAB plus US (independently), plus additional baseline factors

2e. TAB plus US (either/or positive), plus additional baseline factors
2f. TAB plus US (interaction) , plus additional baseline factors

3. Reduced-experimental diagnosis ("is a TAB always necessary?")

3a. US alone

3b. US plus additional baseline factors

3c. US (+ vs -) plus TAB (+ vs – in US -) alone

3d. US (+ vs -) plus TAB (+ vs – in US -), plus additional baseline factors

Models will be built using logistic regression, with diagnostic ability assessed by sensitivity, specificity and the c-statistic.
Depending on the level of agreement and the number of positive cases, some of the analyses may not be possible due to collinearity. Special care will be taken to avoid overfitting and sparse cells. Models will be internally validated using bootstrap methods to assess reproducibility.

Accuracy of US in relation to scan findings

The prevalence of the following specific findings will be tabulated:

US Findings	The proportion of patients with a biopsy positive halo, stenosis, or occlusion; and the sites involved (common superficial temporal, parietal ramus, proximal frontal ramus, distal frontal ramus, axillary), by reference diagnosis

Accuracy of US and TAB in relation to GCA signs and symptoms

GCA characteristics v US	Cross-tabulation of the presence of characteristic clinical features of GCA in relation to TAB findings.
GCA characteristics v TAB	Cross-tabulation of the presence of characteristic clinical features of GCA in relation to US findings.

As stated in the protocol, the characteristics investigated will include the presence or absence of Polymyalgia Rheumatica (PMR) and visual symptoms at presentation.

Accuracy of US and TAB in relation to patient characteristics

Patient characteristics	The prevalence of GCA by age, gender, ethnicity and smoking status

3.4.3 Accuracy with respect to timing of US & TAB

The accuracy and agreement in regards to the timing of the US and TAB will be assessed. Two durations will be calculated:

Duration 1: From starting steroids to performing the US scan

Duration 2: From the US scan to having the TAB

TAB versus time	The agreement between diagnosis by TAB and reference diagnosis, by duration from starting steroids to TAB (duration 1 + duration 2).
US versus time	The agreement between diagnosis by US and reference diagnosis, by duration from starting steroids commencing to US.

Duration will be calculated in whole days and groupings will be made with attention to ensuring a reasonable number within each category. This analysis will include all entered patients, regardless of whether or not their TAB was within 8 days.

1.4.4 Inter-observer agreement

Variation in the interpretation of US and TAB will be evaluated by comparing raters' interpretations of US and TAB against expert review and agreement between raters' interpretations. The following will be reported:

Raters versus US expert review	Agreement between raters' assessment of US and expert review
Raters versus TAB expert review	Agreement between raters' assessment of TAB and expert review
US raters' agreement	Multi-rater kappa with 95% confidence interval
TAB raters' agreement	Multi-rater kappa with 95% confidence interval

1.4.5 Reference diagnosis evolution and influences

Additional analyses will investigate how the diagnosis changes across the patient's follow-up. The number and percentage of patients whose clinical diagnosis changes between two weeks and six months will be presented, and the characteristics of these patients will be described qualitatively.

The role of various diagnostic tests in reaching the diagnosis of GCA will be reported. Finally, the utility of BVAS and VDI as potential assessment tools for GCA will be investigated using ROC analysis. The BVAS score (range 0-63) is calculated as described in Mukhtyar et al (2009), details of which are provided in Appendix 1; the VDI score (range 0-64) is calculated as described in Exley et al (1997), details of which are provided in Appendix 2.

The following summaries will be presented:

Diagnosis evolution	The number and percentage of patients with a diagnosis of GCA and each alternative diagnosis at two weeks and six months, and the number and percentage of patients whose diagnosis changed between two weeks and six months.
Influences on GCA diagnosis	The proportion of GCA diagnoses which were recorded as being influenced by each of the following: symptoms, signs, blood abnormalities, TAB or other characteristics, at both 2 weeks and 6 months.
Associates with reference diagnosis	The reference diagnosis in relation to various attributes including BVAS and VDI

1.4.6 Modelling of alternative methods to diagnose GCA

The findings from 3.4.1-3.4.5 above will be brought together to input into a decision model for diagnosing GCA. The aim will be to assess whether alternative diagnostic strategies could be employed in relation to subgroups or characteristics. Specifically:

- Whether US followed by TAB is necessary in all patients (US alone may conceivably be used as a rule in/rule out for TAB)
- Which (if any) subgroups of the cohort could be diagnosed without the need for TAB and/or US

Testing strategies will be based on pre-test assessment of patients being at high, medium, or low risk of having GCA.

The accuracy of the dual US-TAB approach and alternative approaches will be reported. Models will be validated internally using cross-centre model fitting, temporal model fitting and bootstrap methods. Logistic regression will be employed, and goodness-of-fit will be tested via Hosmer-Lemeshow test (Hosmer & Lemeshow 2000).

1.4.7 The following summaries will be reported **Cost-effectiveness**

The cost-effectiveness of different models will be quantified in a separate analysis plan

1.4.8 Health-related quality of life

Health-related quality of life for the cost-effectiveness analysis will be measured at each assessment using the EuroQol EQ-5D at baseline, 2 week and 6 month visits. The EQ-5D health state will be derived from the questionnaire using UK population norms. The EQ-5D thermometer scale health state, as measured in response to the question “What is your own health state today”, will be scored between 0-100.

EQ-5D health state	The EQ-5D health state at each time point and the change from baseline at 2 weeks & six months
EQ-5D thermometer health state	The EQ-5D thermometer health state at each time point and the change from baseline at 2 weeks & six months

1.4.9 Steroid usage and side effects

Steroid usage will be recorded at baseline, 2 weeks and 6 months.

Steroid usage	The number and percentage of patients on steroids at presentation and at two week visit; the average daily dose, the average length of time spent on steroids, and the estimated cumulative exposure.
Side effects	The number and percentage of patients experiencing each side effect

3.5 Safety outcomes

The following summaries will be presented:

Conditions and symptoms by time point	The number and percentage of participants with each condition or symptom by time point
AEs	The number and percentage of participants reporting each AE following study start, by relatedness to US, TAB and overall
Serious AEs (SAEs)	The number and percentage of participants reporting an SAE following study start

Steroid side effects are described previously and will not be repeated herein.

4 Modifications to the original protocol analysis statement

Not applicable.

5 References

Bossuyt PM, Reitsma JB, Bruns DE Gatsonis PP, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, De Vet HC. Standards for Reporting of Diagnostic Accuracy. Toward complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. *British Medical Journal* 2003;326:41-4

Exley AR, Bacon PA, Luqmani RA, Kitas GD, Gordon C, Savage CO, Adu D. Development and initial validation of the Vasculitis Damage Index for the standardized clinical assessment of damage in the systemic vasculitides. Arthritis and Rheumatism 1997;40(2):371-80.

Fleiss JL, Levin B, Paik MC (eds.) . *Statistical methods for rates and proportions (3rd ed)*. New Jersey: Wiley, 2003.

Hosmer DW, Lemeshow S. *Applied Logistic Regression, Second Edition.* New York: John Wiley & Sons 2000.

Hutchings A, Hollywood J, Lamping DL, et al. Clinical outcomes, quality of life, and diagnostic uncertainty in the first year of polymyalgia rheumatica. *Arthritis & Rheumatism* 2007;57:803-809.

ICH Harmonised Tripartite Guideline Topic E6(R1): Guideline for Good Clinical Practice (1996). Available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6_R1/Step4/E6_R1__Guideline.pdf (last accessed 19June 2012)

Karassa FB, Matsagas MI, Schmidt WA, Ioannidis JP. Meta-analysis: test performance of ultrasonography for giant-cell arteritis. *Annals of Internal Medicine* 2005;142:359-369.

Mukhtyar C, Lee R, Brown D, Carruthers D, Dasgupta B, Dubey S, Flossmann O, Hall C, Hollywood J, Jayne D, Jones R, Lanyon P, Muir A, Scott D, Young L, Luqmani R A. Modification and validation of the Birmingham Vasculitis Activity Score (version 3) *Annals of Rheumatic Diseases* 2009;68:1827-1832 doi:10.1136/ard.2008.101279

Pepe MS. *The statistical evaluation of medical tests for classification and prediction.* Oxford, OUP: 2003.

6 Appendix

Appendix 1: Birmingham Vasculitis Activity Score (Adapted by permission from BMJ Publishing Group Limited. Modification and validation of the Birmingham Vasculitis Activity Score (version 3), Mukhtyar C, Lee R, Brown D, Carruthers D, Dasgupta B, Dubey S, *et al.*, 68, 2009.)

1. If all manifestations are persistent (i.e. the "persistent" box is ticked), the scores from column 2 are used. Otherwise, the scores from column 3 are used. (Items for which column 2 is marked "n/a" cannot be considered persistent)
2. The scores are applied to each ticked manifestation. If no manifestations are recorded, a score of zero is applied.
3. Within each domain, a maximum score is applied. For example, the sum of scores for "general" manifestations cannot be greater than 3 (2 if all persistent).

Manifestation	Persistent	New / Worse
1. General (Maximum score)	2	3
Myalgia	1	1
Arthralgia or arthritis	1	1
Fever ≥38° C	2	2
Weight Loss ≥2 kg	2	2

2. Cutaneous (Maximum score)	3	6
Infarct	1	2
Purpura	1	2
Ulcer	1	4
Gangrene	2	6
Other skin vasculitis	1	2

3. Mucous Membranes / eyes (Maximum score)	3	6
Mouth ulcers / granulomata	1	2
Genital ulcers	1	1
Adnexal inflammation	2	4
Significant proptosis	2	4
Scleritis / Episcleritis	1	2
Conjunctivitis / Blepharitis / Keratitis	1	1
Blurred vision	2	3
Sudden visual loss	n/a	6
Uveitis	2	6
Retinal changes (vasculitis, thrombosis / exudate / haemorrhage)	2	6

4. Ear, Nose & Throat (Maximum score)	3	6
Bloody nasal discharge / crusts / ulcers / granulomata	2	4
Paranasal sinus involvement	1	2

Subglottic stenosis	3	6
Conductive hearing loss	1	3
Sensorineural hearing loss	2	6

5. Chest (Maximum score)	3	6
Wheeze	1	2
Nodules or cavities	n/a	3
Pleural effusion / pleurisy	2	4
Infiltrate	2	4
Endobronchial involvement	2	4
Massive haemoptysis / alveolar haemorrhage	4	6
Respiratory failure	4	6

6. Cardiovascular (Maximum score)	3	6
Loss of pulses	1	4
Valvular heart disease	2	4
Pericarditis	1	3
Ischaemic cardiac pain	2	4
Cardiomyopathy	3	6
Congestive cardiac failure	3	6

7. Abdominal (Maximum score)	4	9
Peritonitis	3	9
Bloody diarrhoea	3	9
Ischaemic abdominal pain	2	6

8. Renal (Maximum score)	6	12
Hypertension	1	4
Proteinuria	2	4
Haematuria	3	6
Serum creatinine 125-249 µmol/L	2	4
Serum creatinine 250-499 µmol/L	3	6
Serum creatinine ≥500 µmol/L	4	8
>30% rise in creatinine or >25% fall in creatinine clearance	n/a	6

9. Nervous system (Maximum score)	6	9
Headache	1	1
Meningitis	1	3
Organic confusion	1	3
Seizures (not hypertensive)	3	9

Stroke	3	9
Spinal cord lesion	3	9
Cranial nerve palsy	3	6
Sensory peripheral Neuropathy	3	6
Mononeuritis Multiplex	3	9

Appendix 2: Vasculitis Damage Index (Republished with permission of John Wiley and Sons Inc, from Development and initial validation of the Vasculitis Damage Index for the standardized clinical assessment of damage in the systemic vasculitides, Exley AR, Bacon PA, Luqmani RA, Kitas GD, Gordon C, Savage CO, Adu D, 40(2), 1997; permission conveyed through Copyright Clearance Centre Inc.)

The VDI records the presence or absence of 64 specific conditions since the onset of suspected GCA. The total VDI score is defined as the total number of items scored, ranging from zero to a theoretical maximum of 64.

VASCULITIS DAMAGE INDEX (VDI)

This is for recording organ damage that has occurred in patients *since the onset of suspected GCA*
Patients often have co-morbidity before onset of suspected GCA, **which must not be scored**
Record features of active disease using the Birmingham Vasculitis Activity Score (BVAS)
A new patient should **usually have a VDI score of zero**, unless:
(a) they have had suspected GCA for more than three months **and**
(b) the damage has developed or become worse since the onset of suspected GCA

Musculoskeletal? ☐ Yes ☐ No	Pulmonary? ☐ Yes ☐ No	Gastrointestinal? ☐ Yes ☐ No
If yes:	**If yes:**	**If yes:**
☐ Significant muscle atrophy or weakness	☐ Pulmonary hypertension	☐ Gut infarction/resection
☐ Deforming/erosive arthritis	☐ Pulmonary fibrosis	☐ Mesenteric insufficiency / pancreatitis
☐ Osteoporosis/vertebral collapse	☐ Pulmonary infarction	☐ Chronic peritonitis
☐ Avascular necrosis	☐ Pleural fibrosis	☐ Oesophageal stricture/surgery
☐ Osteomyelitis	☐ Chronic asthma	Renal? ☐ Yes ☐ No
Skin/Mucous membranes? ☐ Yes ☐ No	☐ Chronic breathlessness	**If yes:**
If yes:	☐ Impaired lung function	☐ Estimated/measured GFR < 50%
☐ Alopecia	Cardiovascular? ☐ Yes ☐ No	☐ Proteinuria > 0.5g/24hr
☐ Cutaneous ulcers	**If yes:**	☐ End stage renal disease
☐ Mouth ulcers	☐ Angina/angioplasty	Neuropsychiatric? ☐ Yes ☐ No
Ocular? ☐ Yes ☐ No	☐ Myocardial infarction	**If yes:**
If yes:	☐ Subsequent myocardial infarction	☐ Cognitive impairment
☐ Cataract	☐ Cardiomyopathy	☐ Major psychosis
☐ Retinal change	☐ Valvular disease	☐ Seizures
☐ Optic atrophy	☐ Pericarditis ≥ 3 mths or pericardectomy	☐ Cerebrovascular accident
☐ Visual impairment/diplopia	☐ Diastolic BP ≥ 95 or requiring antihypertensives	☐ 2nd cerebrovascular accident
☐ Blindness in one eye	Peripheral vascular disease? ☐ Yes ☐ No	☐ Cranial nerve lesion
☐ Blindness in second eye	**If yes:**	☐ Peripheral neuropathy
☐ Orbital wall destruction	☐ Absent pulses in one limb	☐ Transverse myelitis
ENT? ☐ Yes ☐ No	☐ 2nd episode of absent pulses in one limb	Other? ☐ Yes ☐ No
If yes:	☐ Major vessel stenosis	**If yes:**
☐ Hearing loss	☐ Claudication >3 mths	☐ Gonadal failure ☐ Chemical cystitis
☐ Nasal blockage/chronic discharge/crusting	☐ Minor tissue loss	☐ Marrow failure ☐ Malignancy
☐ Nasal bridge collapse/septal perforation	☐ Major tissue loss	☐ Diabetes ☐ Other
☐ Chronic sinusitis/radiological damage	☐ Subsequent major tissue loss	Total VDI Score*
☐ Subglottic stenosis (no surgery)	☐ Complicated venous thrombosis	☐ Record the number of positive items (1 point for each).
☐ Subglottic stenosis (with surgery)		

*The VDI score can either increase or remain the same over time. Remember to carry forward any previous items of damage

Appendix 15 Diagnostic accuracy for combination of strategies for the pre-test risk groups

The definitions of strategies H0 to H4, M1 to M4 and L1 to L6 are given in *Table 42*. We considered different diagnostic strategies depending on the pre-test probability of having GCA. Patients were defined as having a high, medium or low pre-test probability of having a diagnosis of GCA. We examined the sequential strategies of performing an initial ultrasound in the high-risk group and then performing a biopsy if the scan is negative (i.e. the scan is not consistent with a diagnosis of GCA).

Patients were defined as having GCA if they met any of five possible criteria:

1. No tests were performed but the likelihood that the patient has GCA is high (defined as H0).
2. The sonographer's opinion is that the ultrasound scan is consistent with a diagnosis of GCA (H1).
3. Halo is present bilaterally (in either temporal or axillary arteries) (H2).
4. Either the sonographer's opinion is that the ultrasound is consistent with a diagnosis of GCA or there are abnormalities in the axillary arteries (regardless of the sonographer's overall opinion) (H3).
5. Halo is present bilaterally or any axillary involvement is present (H4).

In the medium-risk groups we considered the above strategies (except for the 'no test' strategy), in which ultrasound is performed first, followed by biopsy (M1 to M4 would be equivalent to H1 to H4).

In the low-risk groups, we considered the same four strategies as well as two further strategies:

1. Using a negative ultrasound result as a 'rule-out' test for GCA. If ultrasound is positive, then perform a biopsy and take the diagnosis from the biopsy result (L5).
2. Using the absence of any abnormal finding on the ultrasound as a 'rule-out' test for GCA. If there are any abnormalities, perform a biopsy and take the diagnosis from the biopsy result (L6).

Strategies M1 and M3 resulted in the same classification of participants so we deliberately omitted repeating the data because the result was identical.

TABLE 75 Sequential strategies of performing an initial ultrasound in the high-risk group and then performing a biopsy if the scan is negative

Strategy			GCA	Not GCA	
High pre-test risk	Medium pre-test risk	Low pre-test risk	Sensitivity (*N* = 257), *n* (%)	Specificity (*N* = 124), *n* (%)	Number of TABs required (*N* = 381), *n* (%)
H0	M1	L1	185 (72.0)	96 (77.4)	178 (46.7)
		L2	175 (68.1)	108 (87.1)	200 (52.5)
		L3	185 (72.0)	93 (75.0)	175 (45.9)
		L4	176 (68.5)	103 (83.1)	194 (50.9)
		L5	164 (63.8)	115 (92.7)	126 (33.1)
		L6	166 (64.6)	115 (92.7)	134 (35.2)
	M2	L1	163 (63.4)	97 (78.2)	210 (55.1)
		L2	153 (59.5)	109 (87.9)	232 (60.9)
		L3	163 (63.4)	94 (75.8)	207 (54.3)
		L4	154 (59.9)	104 (83.9)	226 (59.3)
		L5	142 (55.3)	116 (93.5)	158 (41.5)
		L6	144 (56.0)	116 (93.5)	166 (43.6)
	M4	L1	164 (63.8)	97 (78.2)	207 (54.3)
		L2	154 (59.9)	109 (87.9)	229 (60.1)
		L3	164 (63.8)	94 (75.8)	204 (53.5)
		L4	155 (60.3)	104 (83.9)	223 (58.5)
		L5	143 (55.6)	116 (93.5)	155 (40.7)
		L6	145 (56.4)	116 (93.5)	163 (42.8)
H1	M1	L1	166 (64.6)	101 (81.5)	219 (57.5)
		L2	156 (60.7)	113 (91.1)	241 (63.3)
		L3	166 (64.6)	98 (79.0)	216 (56.7)
		L4	157 (61.1)	108 (87.1)	235 (61.7)
		L5	145 (56.4)	120 (96.8)	167 (43.8)
		L6	147 (57.2)	120 (96.8)	175 (45.9)
	M2	L1	144 (56.0)	102 (82.3)	251 (65.9)
		L2	134 (52.1)	114 (91.9)	273 (71.7)
		L3	144 (56.0)	99 (79.8)	248 (65.1)
		L4	135 (52.5)	109 (87.9)	267 (70.1)
		L5	123 (47.9)	121 (97.6)	199 (52.2)
		L6	125 (48.6)	121 (97.6)	207 (54.3)
	M4	L1	145 (56.4)	102 (82.3)	248 (65.1)
		L2	135 (52.5)	114 (91.9)	270 (70.9)
		L3	145 (56.4)	99 (79.8)	245 (64.3)
		L4	136 (52.9)	109 (87.9)	264 (69.3)
		L5	124 (48.2)	121 (97.6)	196 (51.4)
		L6	126 (49.0)	121 (97.6)	204 (53.5)

TABLE 75 Sequential strategies of performing an initial ultrasound in the high-risk group and then performing a biopsy if the scan is negative (*continued*)

Strategy			GCA	Not GCA	Number of TABs required ($N = 381$), *n* (%)
High pre-test risk	Medium pre-test risk	Low pre-test risk	Sensitivity ($N = 257$), *n* (%)	Specificity ($N = 124$), *n* (%)	
H2	M1	L1	161 (62.6)	101 (81.5)	231 (60.6)
		L2	151 (58.8)	113 (91.1)	253 (66.4)
		L3	161 (62.6)	98 (79.0)	228 (59.8)
		L4	152 (59.1)	108 (87.1)	247 (64.8)
		L5	140 (54.5)	120 (96.8)	179 (47.0)
		L6	142 (55.3)	120 (96.8)	187 (49.1)
H2	M2	L1	139 (54.1)	102 (82.3)	263 (69.0)
		L2	129 (50.2)	114 (91.9)	285 (74.8)
		L3	139 (54.1)	99 (79.8)	260 (68.2)
		L4	130 (50.6)	109 (87.9)	279 (73.2)
		L5	118 (45.9)	121 (97.6)	211 (55.4)
		L6	120 (46.7)	121 (97.6)	219 (57.5)
	M4	L1	140 (54.5)	102 (82.3)	260 (68.2)
		L2	130 (50.6)	114 (91.9)	282 (74.0)
		L3	140 (54.5)	99 (79.8)	257 (67.5)
		L4	131 (51.0)	109 (87.9)	276 (72.4)
		L5	119 (46.3)	121 (97.6)	208 (54.6)
		L6	121 (47.1)	121 (97.6)	216 (56.7)
H3	M1	L1	169 (65.8)	101 (81.5)	214 (56.2)
		L2	159 (61.9)	113 (91.1)	236 (61.9)
		L3	169 (65.8)	98 (79.0)	211 (55.4)
		L4	160 (62.3)	108 (87.1)	230 (60.4)
		L5	148 (57.6)	120 (96.8)	162 (42.5)
		L6	150 (58.4)	120 (96.8)	170 (44.6)
	M2	L1	147 (57.2)	102 (82.3)	246 (64.6)
		L2	137 (53.3)	114 (91.9)	268 (70.3)
		L3	147 (57.2)	99 (79.8)	243 (63.8)
		L4	138 (53.7)	109 (87.9)	262 (68.8)
		L5	126 (49.0)	121 (97.6)	194 (50.9)
		L6	128 (49.8)	121 (97.6)	202 (53.0)
	M4	L1	148 (57.6)	102 (82.3)	243 (63.8)
		L2	138 (53.7)	114 (91.9)	265 (69.6)
		L3	148 (57.6)	99 (79.8)	240 (63.0)
		L4	139 (54.1)	109 (87.9)	259 (68.0)

continued

TABLE 75 Sequential strategies of performing an initial ultrasound in the high-risk group and then performing a biopsy if the scan is negative (*continued*)

Strategy			GCA	Not GCA	Number of TABs required (*N* = 381), *n* (%)
High pre-test risk	Medium pre-test risk	Low pre-test risk	Sensitivity (*N* = 257), *n* (%)	Specificity (*N* = 124), *n* (%)	
		L5	127 (49.4)	121 (97.6)	191 (50.1)
		L6	129 (50.2)	121 (97.6)	199 (52.2)
H4	M1	L1	165 (64.2)	101 (81.5)	225 (59.1)
		L2	155 (60.3)	113 (91.1)	247 (64.8)
		L3	165 (64.2)	98 (79.0)	222 (58.3)
		L4	156 (60.7)	108 (87.1)	241 (63.3)
		L5	144 (56.0)	120 (96.8)	173 (45.4)
		L6	146 (56.8)	120 (96.8)	181 (47.5)
	M2	L1	143 (55.6)	102 (82.3)	257 (67.5)
		L2	133 (51.8)	114 (91.9)	279 (73.2)
		L3	143 (55.6)	99 (79.8)	254 (66.7)
		L4	134 (52.1)	109 (87.9)	273 (71.7)
		L5	122 (47.5)	121 (97.6)	205 (53.8)
		L6	124 (48.2)	121 (97.6)	213 (55.9)
	M4	L1	144 (56.0)	102 (82.3)	254 (66.7)
		L2	134 (52.1)	114 (91.9)	276 (72.4)
		L3	144 (56.0)	99 (79.8)	251 (65.9)
		L4	135 (52.5)	109 (87.9)	270 (70.9)
		L5	123 (47.9)	121 (97.6)	202 (53.0)
		L6	125 (48.6)	121 (97.6)	210 (55.1)

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library