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

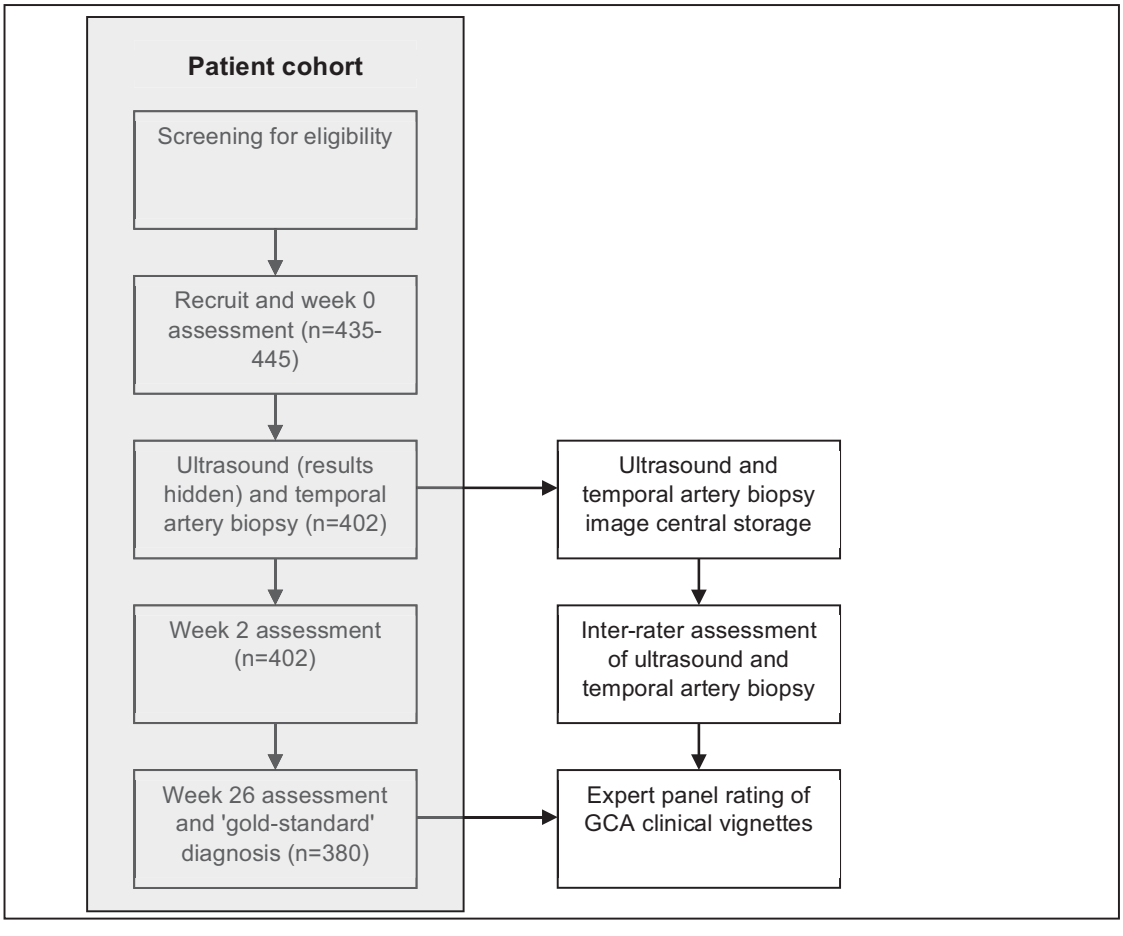
All tables will present summary statistics defined by the nature of the measurement. Summaries of continuous variables will comprise the number of observations used and either
i) mean, median, standard deviation, minimum and maximum, or
ii) median, inter-quartile range, minimum and maximum
as appropriate 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



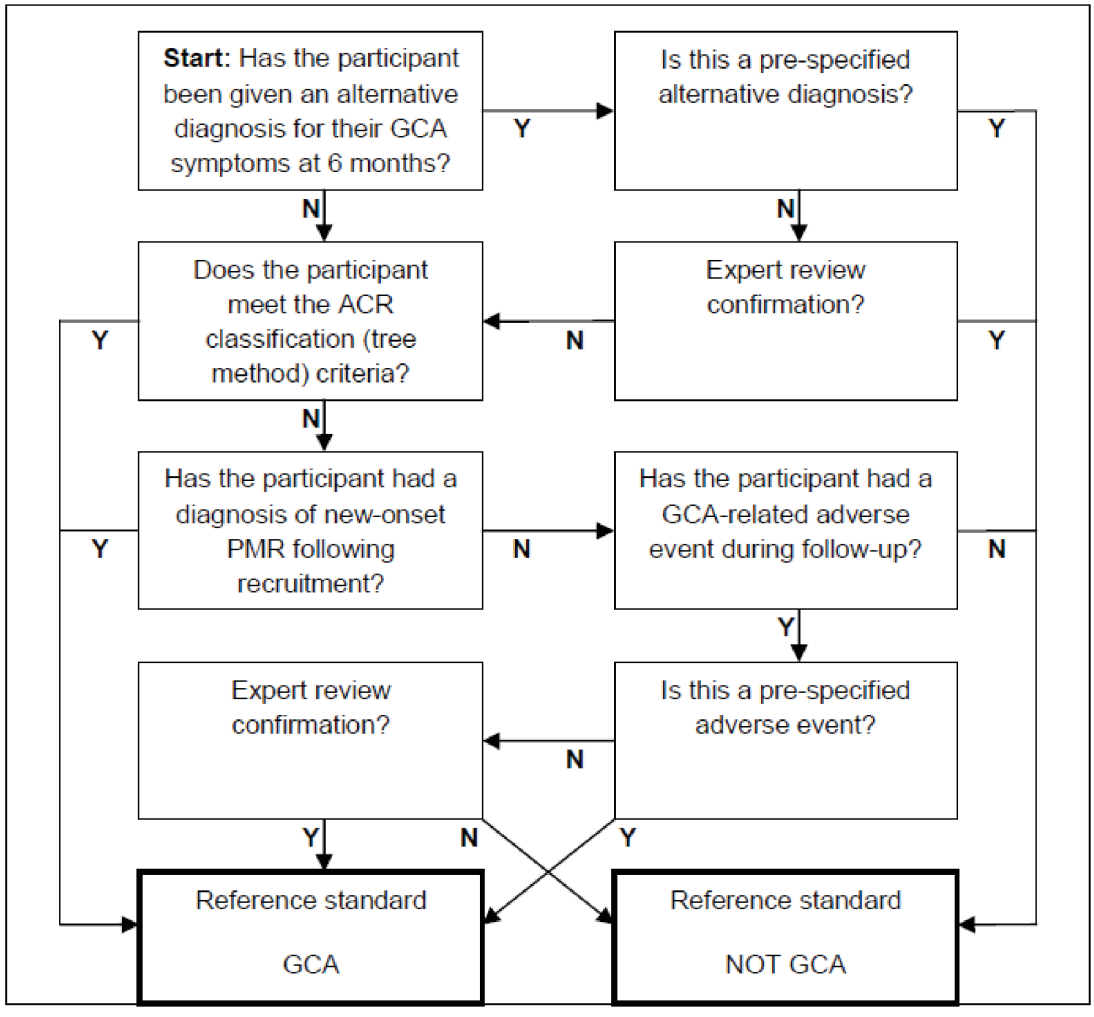


Figure 1 Algorithm for study patients

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