What can negative temporal artery biopsies tell us?

Mackie, Sarah L; Brouwer, Elisabeth

Published in:
Rheumatology

DOI:
10.1093/rheumatology/kez628

Publication date:
2020

RHEUMATOLOGY

Rheumatology 2019;0:1–3
doi:10.1093/rheumatology/kez628

Editorial

What can negative temporal artery biopsies tell us?

This editorial refers to Sensitivity of temporal artery biopsy in the diagnosis of giant cell arteritis: a systematic literature review and meta-analysis by Rubenstein *et al.* doi:10.1093/rheumatology/kez385

GCA is not always straightforward to diagnose, and this has never been more important than in this era of targeted therapies. Temporal artery biopsy (TAB) has historically been a key part of the diagnostic workup for GCA [1]. Early authors described multiple subcategories of temporal artery biopsies [2], but since publication of the 1990 ACR classification criteria [3] the primary value of TAB is usually seen as confirming the presence of disease. The results are often communicated simply as 'positive' or 'negative' [4], as to date there has been little research on whether clinically relevant pathological subtypes exist. Recent recommendations state that imaging and biopsy have 'similar diagnostic value' [4] and that it is acceptable to perform either of these tests in the diagnostic workup. The literature review underlying these recommendations focused on the diagnostic properties of imaging, rather than biopsy [5]. The imperfect sensitivity of TAB is often attributed to skip lesions [6], glucocorticoid treatment (although this particular factor may be overstated [7]) or sparing of the temporal arteries by GCA. Depending on the study methodology, TAB sensitivity has been variously estimated at 39% (based on a diagnostic accuracy study designed primarily to assess the accuracy of US rather than TAB [8]) and 87.1% (based on concordance of bilateral TAB [9]). A new meta-analysis reports the pooled sensitivity of TAB based on a systematic review of the published literature [10].

Rubenstein *et al*. [10] designed their study to estimate the proportion of TAB-positive patients within cohorts of patients who were clinically diagnosed with GCA, supported by ACR classification criteria [3] The authors took a robust, systematic approach to extract the relevant information from the literature and to identify and explore heterogeneity. The search identified all publications after 1990 (the date of publication of the ACR criteria [3]) that had at least 30 cases of GCA, including a mixture of diagnostic, epidemiological and therapeutic studies. Various pre-specified steps were taken to explore the potential sources of heterogeneity, including a systematic assessment of risk of bias. Meta-regression was used to investigate the effect of various attributes of each study on the overall estimate of sensitivity.

Notably, there was substantial clinical heterogeneity between the 32 independent cohorts identified, including substantial variation in the proportions of GCA patients that had PMR, visual manifestations, and large-vessel involvement. This heterogeneity could reflect differences in care pathways, patient selection criteria, protocols for clinical evaluation and differences in definitions of the various GCA clinical phenotypes, particularly visual manifestations. Many of the potential sources of heterogeneity were found to be difficult to assess in the published papers. For example, pathologists independently assessing the same TAB do not always agree on whether it should be classified as 'positive' or 'negative' [8], but most studies identified by Rubenstein *et al.* did not include a specific histopathological review.

It was found that the proportion of GCA with a positive TAB varied from 49.5% to 95.1%, and the overall pooled proportion was 77.3% [10]. No single study was identified as having disproportionate effect on the overall estimate. The proportion was higher in later publications, but no other clear explanation for the heterogeneity could be identified from the data extracted.

Assuming a 99% specificity of TAB (corresponding to the usual clinical assumption that a positive TAB essentially confirms the diagnosis of GCA), then a 77% sensitivity for a clinical diagnosis of GCA compares favourably with the accuracy of imaging tests for GCA (Table 1). Calculation of positive or negative predictive values requires an estimate of the prevalence of GCA in the population being studied. This was the topic of a recent systematic review [11], which gave an estimated median TAB yield of 25% (higher yield in older cohorts, lower in younger cohorts). Therefore, for calculation of negative predictive value (the probability that a negative test is a true negative), we assumed a GCA prevalence of 25%, although clearly this could vary depending on how patients with suspected GCA are identified, selected and referred for TAB. It can be seen that contrary to the oft-heard clinical assertion that 'a negative biopsy should not change management decisions because false negatives can occur', a negative TAB has a 93% negative predictive value given a 25% disease prevalence. Even if the pre-test probability were extremely high (80%), then based on these sensitivity and specificity values, a negative TAB would have a negative predictive value of 52%, and therefore would still cast doubt on the original clinical diagnosis of GCA.

Clinicians should exercise caution before applying these data on accuracy of TAB or imaging to their clinical practice because of limitations of the primary literature reveiwed. First, variation in surgical technique, specimen processing and staining protocols, and lack of standardized TAB reporting guidelines for clinical practice may produce inconsistency in classifying TAB [8, 12]. Second, with regard to the accuracy of US, if the

Table 1 Summary of recent evidence relating to sensitivity and specificity of temporal artery biopsy, temporal artery US and MRI

	Temporal artery biopsy	Temporal artery US [5]	MRI of temporal arteries [5]
Sensitivity for clinical diagnosis of GCA	77% [10]	77%	73%
Specificity for clinical diagnosis of GCA	99% (assumed)	96%	88%
Positive predictive value (assuming 25% prevalence [11])	96%	87%	67%
Negative predictive value (assuming 25% prevalence [11])	93%	93%	91%
Positive likelihood ratio	77	19	6.1
Negative likelihood ratio	0.23	0.24	0.31

Notes: For clinical context, the sensitivity/specificity results are also presented in the form of the corresponding positive and negative predictive values, assuming a prevalence of 25%, as well as the corresponding positive and negative likelihood ratios. Note that these predictive values and likelihood ratios are not necessarily directly applicable to clinical practice; in the case of TAB they depend on an assumed specificity of 99%, which may be an overestimate. Note that we did not include here 18-fluorodeoxyglucose PET studies, as few diagnostic accuracy studies have been done to date.

diagnostician and sonographer are not blinded to each others' observations, then this could introduce bias into diagnostic accuracy studies, resulting in an overestimate of the diagnostic accuracy of US for GCA. Third and most importantly with regard to the TAB data, sensitivity and specificity are paired values and have limited value when taken by themselves. Published cohorts generally focused on unequivocal cases of GCA. Cases with diagnostic uncertainty who have negative TAB are not usually included in published cohorts; this omission could introduce spectrum bias. A 'single-gate' study design of the primary studies would be preferable for calculating the more clinically useful predictive values or likelihood ratios, as this would not exclude the 'grey' (uncertain) cases. In summary, the between-study heterogeneity observed by Rubenstein *et al.* is a warning signal that hidden biases could be present in the primary literature they drew upon to produce their pooled estimate. However, genuine, between-centre variation in the diagnostic accuracy of TAB still cannot be excluded.

The bottom line here is that TAB remains a clinically valuable test that is in general well tolerated, with a very low complication rate if the temporal artery (preferably of 2–3 cm length) is taken out by an experienced surgeon. Neither TAB nor temporal artery US is a perfect test; both tests are reliant on the expertise of the operator. For patients, clinical decisions made on the basis of these tests have huge consequences. Given the imperfect reference standard, research in this area must be carefully designed to identify potential sources of variation and to minimize bias. Ultimately, rapid clinical assessment of patients with suspected GCA by a dedicated service that has the appropriate clinical, imaging and biopsy expertise is likely to provide optimal care.

Funding: S.L.M. was supported by MRC TARGET Partnership Grant (MR/N011775/1) during the period of this work. No specific funding was received from any funding bodies in the public, commercial or not-for-profit sectors to carry out the work described in this manuscript.

Disclosure statement: E.B. as an employee of the UMCG received speaker fees and consulting fees from Roche in 2017 and 2018, which were paid to the UMCG. S.L.M., as an employee of the University of Leeds, received consulting fees from Roche and Sanofi, which were paid to the University of Leeds. S.L.M. was sponsored to attend EULAR 2019 by Roche. S.L.M. is a past/present investigator on GCA clinical trials (Roche, GSK, Sanofi).

Sarah L. Mackie[1] and Elisabeth Brouwer[2]
[1]Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds and NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust, Leeds, UK and [2]Department of Rheumatology and Clinical Immunology, UMCG Groningen, Groningen, The Netherlands

Correspondence to: Sarah Mackie, University of Leeds and Leeds Biomedical Research Centre, Chapel Allerton Hospital, Harrogate Road, Leeds LS7 4SA, UK.
E-mail: s.l.mackie@leeds.ac.uk

References

1 Horton BT. An undescribed form of arteritis of the temporal vessels. Proc Staff Meet Mayo Clinic 1932;7: 700–1.

2 McDonnell PJ, Moore GW, Miller NR, Hutchins GM, Green WR. Temporal arteritis. A clinicopathologic study. Ophthalmology 1986;93:518–30.

3 Hunder GG, Bloch DA, Michel BA, Stevens MB, Arend WP *et al*. The American College of Rheumatology 1990 criteria for the classification of giant cell arteritis. Arthritis Rheum 2010;33:1122–8.

4 Hellmich B, Agueda A, Monti S *et al*. 2018 Update of the EULAR recommendations for the management of large vessel vasculitis. Ann Rheum Dis 2019; doi: 10.1136/annrheumdis-2019-215672.

5 Duftner C, Dejaco C, Sepriano A *et al*. Imaging in diagnosis, outcome prediction and monitoring of large

vessel vasculitis: a systematic literature review and meta-analysis informing the EULAR recommendations. RMD Open 2018;4:e000612.

6 Klein RG, Campbell RJ, Hunder GG, Carney JA. Skip lesions in temporal arteritis. Mayo Clin Proc 1976;51: 504–10.

7 Maleszewski JJ, Younge BR, Fritzlen JT *et al*. Clinical and pathological evolution of giant cell arteritis: a prospective study of follow-up temporal artery biopsies in 40 treated patients. Mod Pathol 2017;30:788–96.

8 Luqmani R, Lee E, Singh S *et al*. The Role of Ultrasound Compared to Biopsy of Temporal Arteries in the Diagnosis and Treatment of Giant Cell Arteritis (TABUL): a diagnostic accuracy and cost-effectiveness study. Health Technol Assess 2016;20:1–238.

9 Niederkohr RD, Levin LA. A Bayesian analysis of the true sensitivity of a temporal artery biopsy. Invest Ophthalmol Vis Sci 2007;48:675–80.

10 Rubenstein E, Maldini C, Gonzalez-Chiappe S, Chevret S, Mahr A. Sensitivity of temporal artery biopsy in the diagnosis of giant cell arteritis: a systematic literature review and meta-analysis. Rheumatology 2019;58; kez385.

11 Ing EB, Wang DN, Kirubarajan A *et al*. Systematic review of the yield of temporal artery biopsy for suspected giant cell arteritis. Neuroophthalmology 2019; 43:18–25.

12 Chakrabarty A, Mackie S, Harden C, Morgan AW. Temporal artery biopsy: audit of histological diagnosis. Rheumatology 2019;kez396.