



Health Technology Assessment

Volume 28 • Issue 46 • August 2024

ISSN 2046-4924

Real-time ultrasound elastography in the diagnosis of newly identified thyroid nodules in adults: the ElaTION RCT

*Hisham Mehanna, Jonathan J Deeks, Kristien Boelaert, Gitta Madani, Paul Sidhu,
Paul Nankivell, Neil Sharma, Rebecca Woolley, Judith Taylor, Tessa Fulton-Lieuw and
Andrew Palmer*



Real-time ultrasound elastography in the diagnosis of newly identified thyroid nodules in adults: the ElaTION RCT

Hisham Mehanna^{1*}, Jonathan J Deeks²,
Kristien Boelaert³, Gitta Madani⁴, Paul Sidhu⁵,
Paul Nankivell¹, Neil Sharma⁶, Rebecca Woolley²,
Judith Taylor⁷, Tessa Fulton-Lieuw¹ and Andrew Palmer²

¹Institute of Head and Neck Studies and Education, University of Birmingham, Birmingham, UK

²Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, UK

³Institute of Applied Health Research, University of Birmingham, Birmingham, UK

⁴Imperial College Healthcare NHS Trust, London, UK

⁵King's College Hospital NHS Foundation Trust, London, UK

⁶Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK

⁷The Thyroid Trust, London, UK

*Corresponding author

Published August 2024

DOI: 10.3310/PLEQ4874

This report should be referenced as follows:

Mehanna H, Deeks JJ, Boelaert K, Madani G, Sidhu P, Nankivell P, *et al.* Real-time ultrasound elastography in the diagnosis of newly identified thyroid nodules in adults: the ElaTION RCT. *Health Technol Assess* 2024;**28**(46). <https://doi.org/10.3310/PLEQ4874>

Health Technology Assessment

ISSN 2046-4924 (Online)

Impact factor: 3.6

A list of Journals Library editors can be found on the [NIHR Journals Library website](#)

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 3.6 and is ranked 32nd (out of 105 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

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

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal

Manuscripts 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

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

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.

This article

The research reported in this issue of the journal was funded by the HTA programme as award number 12/19/04. The contractual start date was in April 2014. The draft manuscript began editorial review in October 2022 and was accepted for publication in April 2023. 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' manuscript 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 article.

This article presents independent research funded by the National Institute for Health and Care 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, the HTA programme or the Department of Health and Social Care. 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, the HTA programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.

Copyright © 2024 Mehanna *et al.* This work was produced by Mehanna *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaptation in any medium and for any purpose provided that it is properly attributed. See: <https://creativecommons.org/licenses/by/4.0/>. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).

Abstract

Real-time ultrasound elastography in the diagnosis of newly identified thyroid nodules in adults: the ElATION RCT

Hisham Mehanna^{1*}, Jonathan J Deeks², Kristien Boelaert³,
Gitta Madani⁴, Paul Sidhu⁵, Paul Nankivell¹,
Neil Sharma⁶, Rebecca Woolley², Judith Taylor⁷,
Tessa Fulton-Lieuw¹ and Andrew Palmer²

¹Institute of Head and Neck Studies and Education, University of Birmingham, Birmingham, UK

²Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, UK

³Institute of Applied Health Research, University of Birmingham, Birmingham, UK

⁴Imperial College Healthcare NHS Trust, London, UK

⁵King's College Hospital NHS Foundation Trust, London, UK

⁶Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK

⁷The Thyroid Trust, London, UK

*Corresponding author h.mehanna@bham.ac.uk

Background: Strain and shear wave elastography which is commonly used with concurrent real-time imaging known as real-time ultrasound shear/strain wave elastography is a new diagnostic technique that has been reported to be useful in the diagnosis of nodules in several organs. There is conflicting evidence regarding its benefit over ultrasound-guided fine-needle aspiration cytology alone in thyroid nodules.

Objectives: To determine if ultrasound strain and shear wave elastography in conjunction with fine-needle aspiration cytology will reduce the number of patients who have a non-diagnostic first fine-needle aspiration cytology results as compared to conventional ultrasound-only guided fine-needle aspiration cytology.

Design: A pragmatic, unblinded, multicentre randomised controlled trial.

Setting: Eighteen centres with a radiology department across England.

Participants: Adults who had not undergone previous fine-needle aspiration cytology with single or multiple nodules undergoing investigation.

Interventions: Ultrasound shear/strain wave elastography-ultrasound guided fine-needle aspiration cytology (intervention arm) – strain or shear wave elastography-guided fine-needle aspiration cytology.

Ultrasound-only guided fine-needle aspiration cytology (control arm) – routine ultrasound-only guided fine-needle aspiration cytology (the current standard recommended by the British Thyroid Association guidelines).

Main outcome measure: The proportion of patients who have a non-diagnostic cytology (Thy 1) result following the first fine-needle aspiration cytology.

Randomisation: Patients were randomised at a 1 : 1 ratio to the interventional or control arms.

Results: A total of 982 participants (80% female) were randomised: 493 were randomised to ultrasound shear/strain wave elastography-ultrasound guided fine-needle aspiration cytology and 489 were randomised to ultrasound-only guided fine-needle aspiration cytology. There was no evidence of a

difference between ultrasound shear/strain wave elastography and ultrasound in non-diagnostic cytology (Thy 1) rate following the first fine-needle aspiration cytology (19% vs. 16% respectively; risk difference: 0.030; 95% confidence interval -0.007 to 0.066 ; $p = 0.11$), the number of fine-needle aspiration cytologies needed (odds ratio: 1.10; 95% confidence interval 0.82 to 1.49; $p = 0.53$) or in the time to reach a definitive diagnosis (hazard ratio: 0.94; 95% confidence interval 0.81 to 1.10; $p = 0.45$). There was a small, non-significant reduction in the number of thyroid operations undertaken when ultrasound shear/strain wave elastography was used (37% vs. 40% respectively; risk difference: -0.02 ; 95% confidence interval -0.06 to 0.009 ; $p = 0.15$), but no difference in the number of operations yielding benign histology – 23% versus 24% respectively, $p = 0.70$ (i.e. no increase in identification of malignant cases) – or in the number of serious adverse events (2% vs. 1%). There was no difference in anxiety and depression, pain or quality of life between the two arms.

Limitations: The study was not powered to detect differences in malignancy.

Conclusions: Ultrasound shear/strain wave elastography does not appear to have additional benefit over ultrasound-guided fine-needle aspiration cytology in the diagnosis of thyroid nodules.

Future work: The findings of the ElaTION trial suggest that further research into the use of shear wave elastography in the diagnostic setting of thyroid nodules is unlikely to be warranted unless there are improvements in the technology. The diagnostic difficulty in distinguishing between benign and malignant lesions still persists. Future studies might examine the role of genomic testing on fine-needle aspiration samples. There is growing use of targeted panels of molecular markers, particularly aimed at improving the diagnostic accuracy of indeterminate (i.e. Thy3) cytology results. The application of these tests is not uniform, and their cost effectiveness has not been assessed in large-scale trials.

Study registration: This study is registered as ISRCTN (ISRCTN18261857).

Funding: This award was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme (NIHR award ref: 12/19/04) and is published in full in *Health Technology Assessment*; Vol. 28, No. 46. See the NIHR Funding and Awards website for further award information.

Contents

List of tables	xi
List of figures	xiii
List of supplementary material	xv
List of abbreviations	xvii
Plain language summary	xix
Scientific summary	xxi
Chapter 1 Introduction	1
Trial rationale/introduction	1
<i>Thyroid nodules</i>	1
<i>Current recommended classification and investigation of thyroid nodules</i>	1
<i>Current deficiencies in the investigation and diagnosis of thyroid nodules</i>	2
<i>Technological developments – shear and strain wave elastography</i>	2
Summary of previous studies	3
Critique of previous studies	5
<i>Methodological aspects</i>	5
Subsequent research	5
<i>Rationale for the ElaTION trial</i>	5
Summary	6
Main research question	6
Chapter 2 Methods	7
Objectives	7
Trial design	7
Participants	7
<i>Inclusion criteria</i>	7
<i>Exclusion criteria</i>	7
<i>Rationale for choice of inclusion and exclusion criteria</i>	8
Outcome measures	9
<i>Primary outcome</i>	9
<i>Secondary outcomes</i>	9
<i>Outcome assessment schedule</i>	9
Sample size	9
Trial sites	11
Recruitment process	11
<i>Informed consent</i>	11
Randomisation	12
<i>Randomisation method and stratification variables</i>	12
<i>Randomisation</i>	12
Planned interventions	12
<i>Experimental arm – strain or shear wave elastography in conjunction with ultrasound-guided fine-needle aspiration cytology</i>	13
<i>Control arm – conventional ultrasound-guided fine-needle aspiration cytology</i>	13

CONTENTS

<i>Other management at the discretion of local doctors</i>	13
<i>Compatibility with other studies</i>	13
Trial assessments	14
<i>Trial schema and trial visit schedule</i>	14
<i>Follow-up assessments</i>	14
<i>Patient assessment</i>	14
<i>Complication rates</i>	14
<i>Data collection</i>	14
Statistical methods	15
<i>Primary outcome analysis</i>	15
<i>Secondary outcome analyses</i>	15
<i>Missing data and sensitivity analyses</i>	16
<i>Subgroup analyses</i>	16
<i>Defining final definitive diagnosis</i>	16
<i>Cost-effectiveness analysis</i>	17
Ethics approval, regulations and trial registration	17
<i>Ethics approval</i>	17
<i>Sponsorship</i>	17
<i>Regulations</i>	18
Monitoring and oversight	18
<i>Confidentiality of personal data</i>	18
<i>In-house data quality assurance</i>	18
<i>Independent Trial Steering Committee</i>	18
<i>Data Monitoring and Ethics Committee</i>	18
<i>Long-term storage of data</i>	18
Safety assessment and reporting	19
<i>Adverse events</i>	19
<i>Serious adverse events</i>	19
<i>Notification of deaths</i>	19
<i>Pharmacovigilance responsibilities</i>	20
<i>Notification of serious breaches of GCP and/or the protocol</i>	20
Trial withdrawals	21
Chapter 3 Results	23
Recruitment	23
Participant flow and withdrawals	23
Completeness of data	24
Baseline data	24
Adherence	28
<i>Adherence to randomised test</i>	28
<i>Adherence to final definitive diagnosis protocol definition</i>	28
Primary outcome	29
<i>Primary analysis</i>	29
<i>Sensitivity analyses</i>	29
Secondary outcomes	29
<i>Thy 1 result following any FNAC</i>	29
<i>Thyroidectomies and benign histology</i>	30
<i>Final definitive diagnosis</i>	30
<i>Time to final definitive diagnosis</i>	30
<i>Number of FNACs until final definitive diagnosis</i>	30
Patient-reported anxiety, depression, pain and quality of life outcome measures	31
<i>Hospital and Depression Scale</i>	31
<i>EQ5D-5L</i>	32
<i>Visual analogue pain scale</i>	32

Accuracy outcomes	33
<i>Accuracy of first FNAC results compared to second FNAC</i>	33
<i>Agreement between first FNAC and second FNAC</i>	33
<i>Accuracy of USE/US without FNAC and accuracy of USE/US with FNAC</i>	33
Radiologists' survey	35
Pathology quality assurance review	35
Surgery complications and serious adverse events	37
<i>Surgery complications</i>	37
<i>Serious adverse events</i>	38
Chapter 4 Discussion	39
Summary of findings	39
Comparison of demographics with other studies	39
Comparison of trial design with previous research	39
Equality, diversity and inclusion	40
Patient and public involvement	40
Adherence	40
Strengths of the trial	40
Possible weaknesses	41
Challenges of the trial	41
Recommendations for future research	41
Conclusion	43
Additional information	45
References	49
Appendix 1 Figures	51

List of tables

TABLE 1 Assessment schedule	10
TABLE 2 Summary of data recorded on CRFs	15
TABLE 3 Baseline characteristics at randomisation split by intervention group and overall	24
TABLE 4 Nodule characteristics from first FNAC assessment by intervention group and overall	25
TABLE 5 Treatment adherence to by randomised test	28
TABLE 6 Participants who obtained their final definitive diagnoses and how they obtained them	28
TABLE 7 Primary analysis of the number of Thy 1 results following first FNAC by trial arm	29
TABLE 8 Number of Thy 1 results following first FNAC by trial arm including those with staggered SE-FNACs performed	30
TABLE 9 Number of Thy 1 results following any FNAC	31
TABLE 10 Time to final definitive diagnosis	31
TABLE 11 Mean VAPS scores over follow-up by test group	32
TABLE 12 Accuracy of first FNAC compared with second FNAC in malignant nodules	33
TABLE 13 Accuracy of first FNAC compared with second FNAC in benign nodules	33
TABLE 14 Agreement between first FNAC and second FNAC	34
TABLE 15 Accuracy of US-alone compared with US-FNAC in malignant nodules	34
TABLE 16 Accuracy of US-alone compared with US-FNAC in benign nodules	34
TABLE 17 Sensitivity analyses of accuracy of US-alone compared with US-FNAC in benign nodules excluding indeterminate US and FNAC results	35
TABLE 18 Accuracy of US-USE alone compared with USE-FNAC in malignant nodules	35
TABLE 19 Accuracy of US-USE alone compared with USE-FNAC in benign nodules	35
TABLE 20 Radiologists' feedback from first FNAC	36
TABLE 21 Complication rates from thyroidectomy at 30 days and 6 months post surgery	37
TABLE 22 List of participant centres, principal investigators and ElatION radiologists	42

List of figures

FIGURE 1 Trial schema	8
FIGURE 2 Patient flow for the ITT population	23
FIGURE 3 Time to FDD Kaplan–Meier Plot (unadjusted HR included in the figure)	32
FIGURE 4 Mean HADS anxiety subscale score over time by test	51
FIGURE 5 Mean change in HADS anxiety subscale score over time by test	51

List of supplementary material

Report Supplementary Material 1 Additional tables

Supplementary material can be found on the NIHR Journals Library report page (<https://doi.org/10.3310/PLEQ4874>).

Supplementary material has been provided by the authors to support the report and any files provided at submission will have been seen by peer reviewers, but not extensively reviewed. Any supplementary material provided at a later stage in the process may not have been peer reviewed.

List of abbreviations

ACR TI-RADS	American College of Radiologists Thyroid Imaging Reporting and Data System	IMP	investigational medicinal product
AUC	area under the curve	INHANSE	Institute of Head and Neck Studies and Education
BCTU	Birmingham Clinical Trials Unit (also known as the Trial Office)	IQR	interquartile range
BTA	British Thyroid Association	ITT	intention to treat
CI	confidence interval	MD	mean difference
CRI	credible interval	MDT	multidisciplinary team
CRF	case report form	NIHR	National Institute for Health and Care Research
DMEC	Data Monitoring and Ethics Committee	OR	odds ratio
FDD	final definitive diagnosis	PIS	patient information sheet
GCP	good clinical practice	RCT	randomised controlled trial
GDPR	General Data Protection Regulation	RD	risk difference
GP	general practitioner	REC	research ethics committee
HADS	Hospital Anxiety and Depression Scale	RTE	real-time elastography
HR	hazard ratios	SAE	serious adverse event
HRA	Health Research Authority	SD	standard deviation
HTA	health technology assessment	SSI	SuperSonic shear imaging
ICH-GCP	International Conference on Harmonisation Guidelines for Good Clinical Practice	STE	strain elastography
		SWE	shear wave elastography
		TSC	trial steering committee
		USE	ultrasound shear/strain wave elastography
		US-FNAC	ultrasound-guided fine-needle aspiration cytology
		VAPS	Visual Analogue Pain Scale

Plain language summary

About half the population will have lumps in their thyroid if examined by an ultrasound scan but may not know they have one. About one in twenty people will feel a thyroid lump in their neck at some time in their life, with about one in twenty of those being malignant. Currently, the recommended way of getting a diagnosis of thyroid nodules is by using ultrasound to guide a needle to get cells from the lump, called ultrasound-guided fine-needle aspiration cytology. These cells are examined to determine the cause of the lump. If there are enough cells, Doctors can then make a diagnosis of whether the lump is benign or malignant. If not, patients will undergo another ultrasound-guided fine-needle aspiration cytology. One in five ultrasound-guided fine-needle aspiration cytologies are non-diagnostic with an overall false-positive rate of approximately 24%. This means one in five patients, with benign disease, may undergo unnecessary diagnostic operations. Thyroid surgery carries risks of complications, which could be avoided if we had better ways to diagnose which patients actually need an operation.

We conducted a randomised trial, ElaTION, to determine if a new technology called strain and shear wave elastography, commonly known as real-time elastography, would be better at helping the radiologist take a sufficient sample of cells and reduce the number of non-diagnostic results, reducing the number of fine-needle aspiration cytologies required to make a definitive diagnosis.

Nine hundred eighty-two patients were recruited between 2015 and 2018 and followed up until the end of the trial. Patients were randomised into two groups: 489 patients received the standard ultrasound-guided fine-needle aspiration cytology alone, and 493 patients received ultrasound-guided fine-needle aspiration cytology + shear wave elastography. Ultrasound shear/strain wave elastography did not reduce non-diagnostic cytology at first fine-needle aspiration cytology or improve the likelihood of determining whether the lump is benign or malignant.

The results of ElaTION do not support the use of shear wave elastography-fine-needle aspiration cytology in the diagnosis of thyroid nodules.

Scientific summary

Background

Strain elastography (STE) which is commonly used with concurrent real-time imaging known as real-time elastography (RTE) and shear wave elastography (SWE) are new diagnostic techniques that have been reported to be useful in the diagnosis of nodules in several organs. There is conflicting evidence regarding its benefit over ultrasound-guided fine-needle aspiration cytology (US-FNAC) alone in thyroid nodules.

Objectives

The primary objective was to determine if strain/shear wave ultrasound elastography (USE) in conjunction with FNAC will reduce the number of patients with thyroid nodules who have a non-diagnostic first FNAC result as compared to conventional ultrasound (US)-only guided FNAC.

Design

ElaTION was a pragmatic, multicentre randomised controlled trial.

Setting

Eighteen centres with a radiology department across England.

Participants

Adults with thyroid nodules who had not undergone previous FNAC with single or multiple nodules undergoing investigation.

Interventions

Ultrasound shear/strain wave elastography-US guided FNAC (intervention arm) – strain or shear wave elastography (USE)-guided FNAC. US-only guided FNAC (control arm) – routine US-only guided FNAC (the current standard recommended by the British Thyroid Association guidelines).

Main outcome measures

Primary outcome

The proportion of patients who have a non-diagnostic cytology (Thy 1) result following the first FNAC.

Secondary outcomes

The number of FNACs required to obtain a definitive diagnosis; time from first FNAC to a definitive diagnosis; the false-positive rate of nodules; the number of patients who have a non-diagnostic cytology result following any FNAC; the proportion of patients undergoing thyroidectomy; accuracy of US alone

(overall); accuracy of first FNAC; accuracy of USE or US without FNAC compared to accuracy of USE or US with FNAC; and patient reported anxiety, pain and quality of life [by the Hospital Anxiety and Depression Scale (HADS), Visual Analogue Pain Scale (VAPS) and EuroQol (EQ)-5D questionnaire at baseline, 3, 6 and 12 months post randomisation].

Results

A total of 982 participants were randomised: 493 were randomised to USE-US guided FNAC and 489 were randomised to US-only guided FNAC. There was no evidence of a difference between USE and US in non-diagnostic (Thy 1) rate following the first FNAC [19% vs. 16% respectively; risk difference (RD): 0.030; 95% confidence interval (CI), -0.007 to 0.066; $p = 0.11$], the number of FNACs needed [odds ratio (OR): 1.10; 95% CI, 0.82 to 1.49; $p = 0.53$] or in the time to reach a definitive diagnosis [hazard ratio (HR): 0.94; 95% CI, 0.81 to 1.10; $p = 0.45$]. There was a small, non-significant reduction in the number of thyroid operations undertaken when USE was used (37% vs. 40% respectively; RD: -0.02; 95% CI, -0.06 to 0.009; $p = 0.15$), but no difference in the number of operations yielding benign histology - 23% versus 24% respectively, $p = 0.70$ (i.e. no increase in identification of malignant cases) - or in the number of serious adverse events (2% vs. 1%). There was no difference in anxiety and depression, pain or quality of life between the two arms.

Limitations

The study was not powered to detect differences in malignancy.

Conclusions

Ultrasound shear/strain wave elastography does not appear to have additional benefit over US-FNAC in the diagnosis of thyroid nodules.

Study registration

This study is registered as ISRCTN (ISRCTN18261857).

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme (NIHR award ref: 12/19/04) and is published in full in *Health Technology Assessment*; Vol. 28, No. 46. See the NIHR Funding and Awards website for further award information.

Chapter 1 Introduction

Sections from this chapter have been reproduced from the ElATION protocol document, available on the National Institute for Health and Care Research (NIHR) Funding and Awards website.¹

Trial rationale/introduction

Thyroid nodules

Palpable thyroid nodules are common, being detected in about 5–7% of the population.² Using ultrasound (US), nodularity of the thyroid can be detected in up to 50% of the population.³ Approximately 4–7% of thyroid nodules are malignant, and therefore most national guidelines recommend the investigation of nodules larger than 5–10 mm in diameter.^{4,5} Clinically impalpable thyroid nodules, commonly incidentally identified on imaging, appear to carry a similar risk of malignancy and should therefore be investigated in the same way as palpable nodules.

Due to the increased use of imaging modalities, such as US carotid duplex, magnetic resonance imaging (MRI) for cervical spinal disease, or whole body imaging such as positron emission tomography–computed tomography (PET-CT), incidental thyroid nodules that are asymptomatic are increasingly being detected and investigated.^{6–10} This is resulting in a rapidly increasing burden of investigation of thyroid nodules.¹¹ In one average-sized hospital, over a period of 5 years, 1412 US-guided fine-needle aspirations were undertaken for the investigation of thyroid nodules – an average of 282 scans per year.¹²

Current recommended classification and investigation of thyroid nodules

Definitive investigation of thyroid nodules is by US and fine-needle aspiration cytology (FNAC) according to the 2014 British Thyroid Association (BTA) guidelines.⁴ US can detect features that are suggestive of malignancy, including an irregular or microlobulated outline, microcalcifications, hypoechogenicity, extrathyroidal extension and lesions being taller than wide. US alone can predict the risk of malignancy with accuracy varying from 22% to 89%.¹³ FNAC remains the best practice for diagnosing thyroid malignancy and the most accurately carried out under US guidance.

The 2014 BTA guidelines recommends a system for classifying the results of FNAC, with subsequent management based on this classification.⁴ The recommendations for this study are based on these guidelines, and are as follows:

- Thy 1 – non-diagnostic: BTA guidelines recommend repeat US-guided FNAC, especially if there is suspicion of malignancy, up to two times. If Thy 1 was obtained in all three instances, our recommendation was to surgically remove nodules.
- Thy 2 – benign: guidelines recommend a further benign FNAC if there are suspicious features or indeterminate features on histology or US. While discharge after a single benign FNAC result can be considered in the absence of suspicious features on US, a second US FNAC was requested for the study to allow examination of the false-negative rate of a single Thy 2 policy.
- Thy 3a denotes that a neoplasm is possible, with atypical features present but not enough to place into other categories. A repeat US FNAC is usually recommended. In the situation of repeated Thy 3a, a multidisciplinary team (MDT) discussion and possible surgery should be considered.
- Thy 3f denotes a follicular neoplastic lesion of indeterminate nature, which may be benign or malignant.
- Thy 4 is suspicious of cancer.
- Thy 5 is diagnostic of cancer.

An audit of 1412 consecutive US FNAC from a single institution demonstrated 20% of the FNAC results to be Thy 1, 70%; Thy 2, 5%; Thy 3 (3a and 3f), 5%; Thy 4, 3%; and Thy 5, 2%.¹²

The BTA recommendations are to consider offering surgery to remove nodules with Thy 3f, 4 and 5 cytology results to both obtain definitive histological diagnosis and, in some cases, provide definitive treatment.

Current deficiencies in the investigation and diagnosis of thyroid nodules

Non-diagnostic and false-positive FNAC results

While FNAC is currently the most reliable diagnostic technique, it is subject to sampling error and analysis uncertainties depending on several factors, including identifying the correct nodule and the correct part of an individual nodule to perform the FNAC. FNAC carries a non-diagnostic rate of up to 20%.^{12,14} The 2014 BTA guidelines recommend repetition of FNAC after obtaining non-diagnostic (Thy 1) samples at least once more if there is a suspicion of malignancy, preferably under US-guidance. If two or three non-diagnostic results are obtained, and there is a suspicion of malignancy, a diagnostic hemi-thyroidectomy is usually recommended. The ability to reduce Thy 1 rates and any subsequent surgery that may become necessary as a result would be highly beneficial by protecting patients from unnecessary procedures, and also reducing cost and burden to health care services.

In addition, there remains a deficiency of cytology to be able to differentiate between benign and malignant pathology in the 25% of nodules that fall into an indeterminate category. These indeterminate (Thy 3) nodules are malignant in approximately 20–35% of cases, meaning that up to 80% are benign and unless causing compressive symptoms do not need to be removed. The diagnostic performance of FNAC improves as the nodule demonstrates more features of malignancy, with sampled nodules ultimately being malignant in about 80% of Thy 4 cases and 98–99% of Thy 5 cases.³

When considering Thy 3, 4 and 5 results together, there is an overall false-positive rate of approximately 24%. This by definition implies that one in five patients with benign disease will undergo diagnostic operations that could have been avoided. All surgery exposes patients to risk, and so improving this false-positive rate and consequently reducing thyroidectomy operations done for benign nodules could have a significant impact on individual- and population-level patient care.

Diagnostic accuracy of ultrasound alone

Ultrasound alone has a variable diagnostic accuracy for predicting malignancy. The differential expression of US features has been proposed as a mechanism for differentiating between benign and malignant lesions. The BTA guidelines proposes a U-classification, ranging from U1 (normal) to U5 (malignant). The American College of Radiologists Thyroid Imaging Reporting and Data System (ACR TI-RADS) employs an assessment of similar sonographic features to classify lesions using a points-based system into five categories between TR1 (benign) and TR5 (malignant). The TI-RADS system proposes size as a criterion for proceeding to sampling with FNAC.

Head-to-head comparisons of these classification systems have been attempted. Specificity of both systems is similarly high (BTA 98%, ACR-TIRADS 95%) but both have poor sensitivity (41% and 46% respectively).¹⁵ In addition, the technique experiences a limitation in inter-rater reliability and reproducibility. There have been few large-scale prospective studies examining its accuracy as a sole diagnostic tool within a randomised multicentre setting. Ascertainment of the accuracy of US alone and also a comparison of its performance with that of the new elastography technique before widespread roll-out are required.

Technological developments – shear and strain wave elastography

Elastography is a technology that uses low-frequency vibrations to assess the elastic properties – and therefore stiffness – of tissues. Ultrasound elastography (USE) has been proposed as an adjunct to US

in the assessment of thyroid nodules. USE combines the diagnostic advantages of US assessment and US-guided FNAC with an additional assessment of the lesion's stiffness. The aim of this is to increase the accuracy of thyroid cancer diagnosis as malignant thyroid nodules are harder (therefore more stiff) than benign ones.¹⁶ The comparative amount and pattern of soft to hard areas within a nodule indicates the likelihood of malignancy. This offers the potential to aid both nodule selection when faced with multiple nodules (assisting the operator target the nodule with the highest risk of malignancy for sampling) and also may assist the radiologist undertaking the needle biopsy to target the area within a single nodule most likely to harbour malignant cells. Hence, USE may also potentially increase the yield of positive FNAC results.¹⁷ A meta-analysis of 638 patients from 11 studies showed that USE features alone had a pooled sensitivity of 92% [confidence interval (CI) 88 to 96] and specificity of 90% (CI 85 to 95) for identifying malignant thyroid nodules.¹⁸ Most studies have also shown very good correlation between radiologists.¹⁹

There are two elastography methods currently available: strain elastography (STE) [most commonly used with real-time elastography (RTE)] and shear wave elastography (SWE). STE requires application of an oscillating physical pressure by the operator on the target lesion, by exerting light manual pressure on the transducer. By comparing echo signals before and after slight compression, the strain (amount of deformation of tissue) is calculated. This is superimposed on the US image as a colour-coded map (thus providing a qualitative visual map). The harder tissue (e.g. malignant thyroid tissue) is seen as a distinct area of colour (often coded red or blue) differing from normal background tissue. A strain ratio may be calculated from the differences in strain measurements between the nodule and the normal background thyroid (providing a semi-quantitative assessment of the stiffness of the tissue).

Shear wave elastography is generated by a pulse from the US transducer and is a true quantitative measurement of the velocity of sound in tissue; shear waves travel faster in harder (and therefore presumed malignant) tissue. Measured velocities can also be displayed graphically on a colour map or numerically as shear wave indices. SWE gives similar information about tissue hardness of the lesion, and more importantly can quantify the stiffness of abnormal tissue without the need to compare to normal tissue (hence giving a true quantitative measurement as opposed to the semi-quantitative measurement from STE). Both methods of elastography provide a similar basic assessment of the underlying hardness of the tissue and may be used to assess a focal thyroid lesion.

Thus, USE assessment can be qualitative (visual assessment of a colour map), semi-quantitative (strain ratios) or quantitative (shear wave indices). No clear advantage has been identified between these approaches in a meta-analysis of 72 studies with 13,505 patients – sensitivity 84%, 83% and 78%; specificity 81%, 80% and 81%, respectively.²⁰

Summary of previous studies

Two systematic reviews and meta-analyses assessing studies comparing the diagnostic performance of USE have recently been published.

The first included studies are from January 2011 to July 2021.²⁰ To be included, studies were required to evaluate the diagnostic performance of USE to differentiate between benign and malignant nodules in a clinical setting. FNAC or histopathology was required as the gold standard comparator. Studies with fewer than 50 nodules assessed were excluded, as were non-English studies and those with insufficient diagnostic outcomes reported. Seventy-two studies with 13,505 patients and 14,015 thyroid nodules were included in the analysis. Mean age of the patients was 46 years and the mean percentage of men was 24%.

Pooled sensitivity, specificity and area under the curve (AUC) were calculated separately for studies using qualitative, semi-quantitative and quantitative assessments. There were 26, 22 and 32 studies

using these methodologies, respectively. The pooled sensitivity, specificity and AUC were 84% (95% CI 0.83 to 0.85), 81% (95% CI 0.80 to 0.82) and 0.89 (95% CI 0.87 to 0.91) respectively for qualitative USE; 83% (95% CI 0.81 to 0.84), 80% (95% CI 0.79 to 0.82) and 0.93 (95% CI 0.91 to 0.95) respectively for semi-quantitative USE; and 78% (95% CI 0.76 to 0.79), 81% (95% CI 0.80 to 0.82) and 0.87 (95% CI 0.86 to 0.88), respectively for quantitative USE.

Sub-analysis of these studies by prospective or retrospective data acquisition demonstrated a worse diagnostic performance of USE in the prospective group.

- Pooled sensitivities were 74.4% versus 89% (qualitative) and 78% versus 78.7% (quantitative) for prospective versus retrospective studies respectively.
- Pooled specificities were 82.3% versus 79.7% (qualitative) and 80.9% versus 81.8% (quantitative) for prospective versus retrospective studies respectively.
- AUC were 87% versus 92% (qualitative) and 88% versus 87% (quantitative) for prospective versus retrospective studies respectively.

Too few retrospective studies used semi-quantitative assessments and therefore this comparison could not be made. In addition, no direct comparison of the diagnostic performance of USE versus US was made.

The second meta-analysis only included randomised controlled trials (RCTs) between database inception to January 2021.²⁰ In addition to the English literature, this meta-analysis included Chinese databases as well. Inclusion criteria involved studies having pathological controls available. Eleven RCTs were identified and included; however, only eight studies analysed the diagnostic efficacy of US elastography. This included 968 cases (482 in the experimental group and 486 in the control group). The pooled sensitivity was 72.26% (95% CI 0.625 to 0.764), specificity 95.35% (95% CI 0.815 to 0.943), false-negative rate 12.5% and false-positive rate 10.3%. The AUC was 0.86, indicating that USE has high diagnostic value for benign and malignant thyroid nodules and is in line with the results from the first systematic review. Of note, this meta-analysis did not report the overall malignancy rate of subjects included in the 11 studies. A very high event rate is noted, which may suggest a selected cohort – different to the unselected cohort included in the ElaTION RCT.

The performance of SuperSonic shear imaging (SSI), which for the first time used shear SWE with RTE (prior to this only STE was available with real-time imaging) was reported in another meta-analysis. Database searches up to July 2021 identified 21 studies with 3376 patients and a total of 4296 thyroid nodules. A malignancy rate of > 40% was seen in this group. SSI showed a summary sensitivity of 74% (95% CI 67% to 79%), specificity of 82% (95% CI 77% to 87%) and area under receiver operating characteristic (AUROC) of 0.85 (95% CI 0.82 to 0.88) for the differentiation between benign and malignant thyroid nodules.

The diagnostic performance of elastography in thyroid nodules reported as indeterminate after FNAC was examined in a systematic review and Bayesian meta-analysis by Qiu *et al.*²¹ in 2019. Twenty studies with 1734 indeterminate thyroid nodules, undergoing elastography, were included. The summary estimates of sensitivity and specificity were 0.766 [95% credible interval (CrI) 0.686 to 0.835] and 0.867 (95% CrI 0.780 to 0.931), respectively. The estimate of AUC was 0.743. The authors note that quantitative shear wave indices and semi-quantitative strain ratios were more efficient than qualitative RTE colour maps, in the setting of indeterminate nodules.

No RCTs were identified since the last search dates of these systematic reviews.

No RCTs were identified directly comparing elastography and US or examining the ability of elastography in reducing the need for FNAC.

Critique of previous studies

Methodological aspects

Lack of comparison to other diagnostic techniques

No randomised studies have compared elastography in the assessment of thyroid nodules directly to US. This makes reported sensitivity and specificity values difficult to interpret. USE has been examined alone and in addition to FNAC.²² However, this study examined the diagnostic ability of defining benign versus malignant nodules and not the ability to reduce the number of benign nodules undergoing FNAC.

Poor sensitivity

Pooled sensitivities from several meta-analyses are reported between 72% and 76%. This is disappointingly low, and clinically worrying with potentially a quarter of malignant nodules being incorrectly reported as negative for cancer.

Cohort selection

Rates of malignant cases reported in the studies included in the systematic reviews are between 30% and 40%. This is significantly higher than might be expected in an unselected cohort undergoing investigation for a thyroid nodule. This likely represents highly selected cohorts in tertiary centres. Indeterminate or suspicious nodules may have already undergone community assessment before referral to thyroid cancer centres. In addition, there is wide discrepancy between cases included in the different systematic reviews. For example, some studies have only examined selected cohorts (e.g. indeterminate nodules). In others, exclusion of nodules with predominantly cystic or calcified appearance is not clearly stated.

Lack of consensus

Quantitative scoring systems have been proposed as a way of improving the diagnostic performance of elastography. There are a variety of scoring systems that have been suggested, but there is a lack of consensus between studies on how to apply these scoring systems consistently. This is further complicated by the different elastography techniques studied and the development of technology over the past few decades. Elastography can be divided into different categories according to the excitation method and way that stiffness is expressed. Two broad categories – strain (STE) and SWE exist. STE uses mechanical force which can be internal (from carotid pulsations) or external (from pressure applied by the operator). SWE depends on generation of an acoustic radiation force (ARFI) pulse. Some systems use point or two-dimensional SWE and newer systems use RTE.²³

Interobserver variability

Few studies have considered the interobserver variability when using elastography in thyroid nodule assessment. The RCTs included in the systematic review did not consider this aspect and is a potential limitation of their reproducibility.²⁴ In particular the degree of precompression applied by the operative using STE can strain tissues beyond their elastic limit causing serious spurious results.

Subsequent research

Rationale for the ElaTION trial

The ElaTION trial is important because of the following factors:

A significant health need:

Thyroid nodules affect a substantial proportion of the population and are increasingly being identified incidentally on routine imaging of the head and neck. There is a need to improve the performance of US-guided FNAC to reduce both the overall number of FNACs and importantly also the diagnostic operations needed to establish a diagnosis. This would reduce the morbidity associated with the procedures (e.g. permanent loss of voice, difficulty swallowing, permanent hypocalcaemia requiring life-long medication), and decrease the inconvenience and anxiety shown to be associated with uncertainty before diagnosis, especially on repeated tests.²⁵

Considerable potential resource and cost requirements:

Thyroid nodules are very common, affecting up to 70% of the population. They are being detected with increasing frequency due to the rising use of imaging technology. A systematic review of the literature demonstrated that in 168,876 patients, 26% of US scans and 22% of CT scans of the neck performed for non-thyroid causes showed incidental nodules in the thyroid. These data translate to approximately 1900 new patients with thyroid nodules identified over a period of a year in the one hospital in the UK, resulting in a potential cost of investigation with US-FNAC of £683,650 in one institution. Extrapolated to 400 hospitals in the UK, this could potentially cost the NHS over £272M per annum.⁶

Decreasing the number of non-diagnostic or unnecessary FNAC and consequent operations might present the opportunity for considerable cost savings to the NHS.

Outstanding issues in FNAC:

There remain some important questions regarding the use of US alone for diagnosis and the need for repetition of US-only guided FNAC. FNAC is subject to sampling and analysis uncertainties, depending on the radiologist's and cytopathologist's experiences. FNAC is reported to have a false-negative rate of about 5.2% which has, to date, not been addressed in a large prospective randomised setting.²⁵ There is therefore a lack of level 1 evidence.

Summary

The need for ElatION: a large, multicentre, RCT.

In view of conflicting results from some of the retrospective and prospective case series and the fact that most results are single institution reports, an RCT was required to provide evidence of the role of USE in the diagnosis of thyroid nodules. The rationale of the trial was to effectively reduce the need for FNAC and also the false-positive rates of FNAC. This would have the potential to reduce healthcare costs and patient distress significantly. In addition, the ElatION trial attempted to answer some of the important outstanding questions in thyroid ultrasonography – including the efficacy of US-only protocols and the need for repetition of US FNAC in the diagnosis of thyroid nodules which are Thy 2 (benign) on initial FNAC.

Main research question

In patients undergoing investigation of thyroid nodules, does strain or SWE, in conjunction with US, reduce the number of patients who have a non-diagnostic (Thy 1) FNAC result following the first FNAC assessment as compared to conventional US-only guided FNAC?

Chapter 2 Methods

Sections from this chapter have been reproduced from the ElaTION protocol document, available on the NIHR Funding and Awards website.¹

Objectives

The primary objective was to determine if shear or strain wave ultrasound elastography (USE), in conjunction with US, to guide FNAC differed in the number of patients who had a non-diagnostic (Thy 1) FNAC result following the first FNAC assessment as compared to conventional US-only guided FNAC.

The secondary objectives were as follows:

- whether USE compared to US differs in:
 - the total FNACs required to reach a definitive diagnosis
 - the time from first FNAC assessment to definitive diagnosis
 - the false-positive rate of nodules
 - the non-diagnostic rate following any FNAC
 - the number of patients having thyroidectomy
 - quality of life, anxiety and procedural pain
- whether US alone was an accurate diagnostic modality for all thyroid nodules
- whether the first FNAC result was as accurate as the second FNAC result
- whether USE and US without FNAC were as accurate as USE or US with FNAC
- to determine the value of USE in radiologist decision-making and undertaking of FNAC
- to determine complication rates of thyroidectomy
- whether SE-FNAC was cost-effective compared to current practice of US-FNAC.

Trial design

ElaTION was a pragmatic, multicentre randomised controlled diagnostic trial which compared the use of elastography in conjunction with US-guided FNAC (the intervention, SE-FNAC) with conventional US-only guided FNAC (current practice, US-FNAC).

The trial schema is shown in [Figure 1](#).

Participants

Inclusion criteria

1. Patients with single or multiple thyroid nodules, whether solid or partially cystic (i.e. mixed), undergoing investigation.
2. Patients aged 18 or over.
3. Patients able and willing to give written informed consent.

Exclusion criteria

1. Patients who had undergone previous thyroid FNAC within the last 6 months.
2. Patients with a bleeding diathesis that precluded FNAC (patients currently on warfarin and aspirin therapy were eligible).

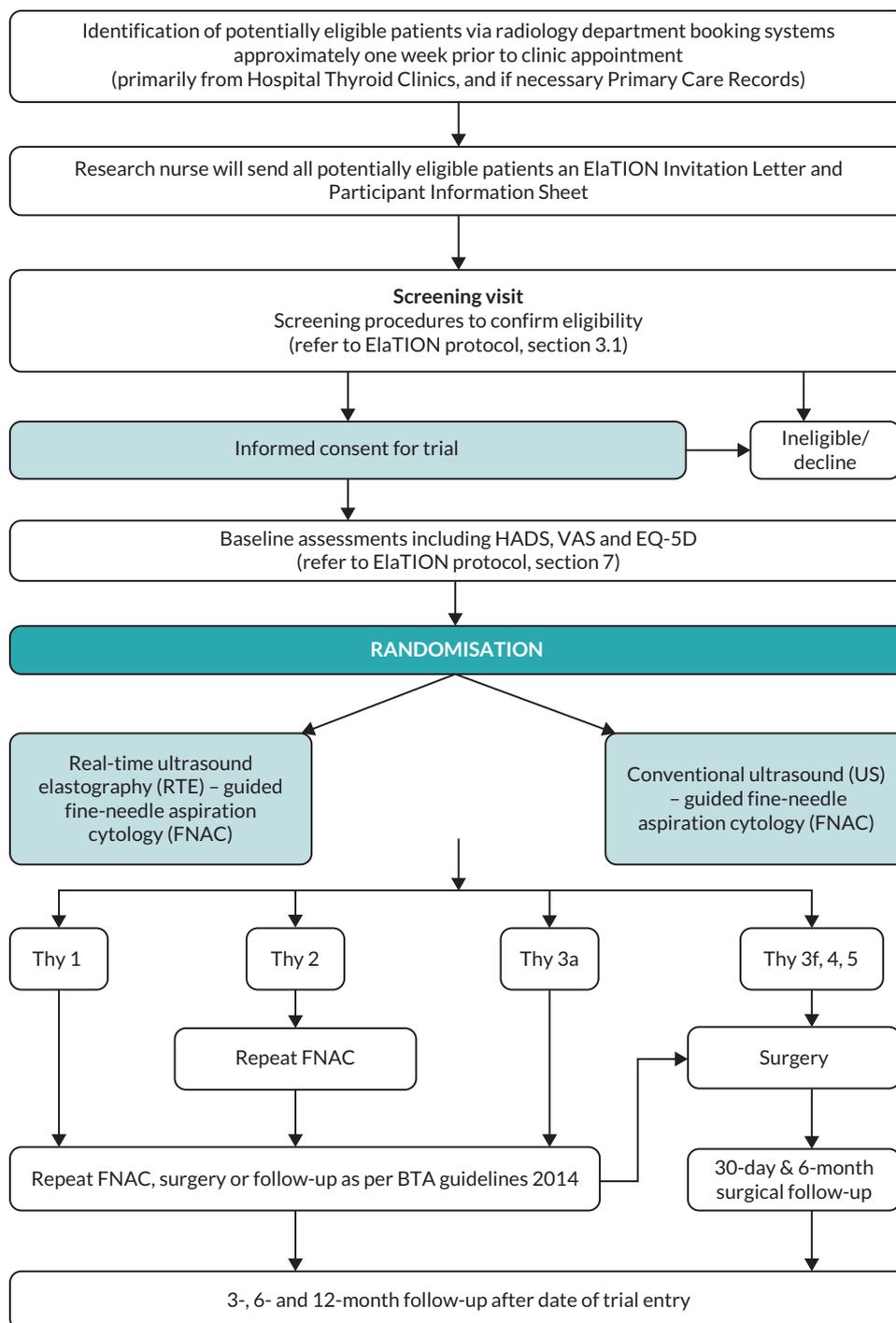


FIGURE 1 Trial schema.

3. Patients with a needle phobia.
4. Pregnant patients.
5. Patients with purely cystic nodules or with recent haemorrhage, with no solid component.
6. Thyroid nodules that appeared to have rim calcification or egg shell calcification.

Rationale for choice of inclusion and exclusion criteria

The role of USE in the diagnosis of thyroid nodules remains controversial, particularly whether this technique reduces the number of FNACs that need to be performed to determine a diagnosis of malignancy. In this trial we aimed to reflect the pragmatism of including all adult patients with thyroid nodules.

Patients with bleeding diathesis or needle phobia were excluded on the grounds that it is more risky to perform FNAC in them. FNAC is often avoided during pregnancy in particular since hormonal changes during pregnancy may alter FNAC interpretation. FNAC in nodules with rim or eggshell calcification is technically very difficult and patients with nodules with these characteristics were excluded.

The presence of a cyst or multiple cysts often precludes a USE scan being done because the cyst may not have sufficient amounts of surrounding solid thyroid tissue. Patients with these nodules were excluded from the study. However, to ensure that an accurate and representative picture of current practice and an accurate assessment of the exact usefulness of the technique in routine clinical practice were obtained, we collected anonymised data about those patients, even though they were not randomised.

Outcome measures

Primary outcome

The primary outcome measure was the proportion of patients who have a non-diagnostic (Thy 1) FNAC result following the first FNAC compared between the SE-FNAC and US-only guided FNAC.

Secondary outcomes

1. Number of FNACs needed to obtain a definitive diagnosis in each patient.
2. Time from first FNAC to obtaining a definitive diagnosis in each arm.
3. The proportion of patients with a benign histology, compared between arms.
4. The proportion of patients who had a non-diagnostic (Thy 1) cytology result following any FNAC by arm.
5. The proportion of patients who received surgery by arm.
6. Accuracy of first FNAC results and repeated FNAC results, compared between arms.
7. If USE or US without FNAC was as accurate as USE or US with FNAC, compared between arms.
8. Patients that reported anxiety, pain and quality of life [by the hospital anxiety and depression scale (HADS) questionnaire, VAPS and EQ-5D] at baseline and at 3, 6 and 12 months post randomisation.
9. Radiologist survey-completed by radiologists at the end of each procedure to identify whether radiologists found USE had contributed to their decisions, ease of use, and their prediction of malignancy of the nodule using USE or US features alone.
10. Complication rate from any thyroidectomy at 30 days and 6 months post surgery, including haematoma rate and temporary hypocalcaemia at 30 days and vocal cord palsy and permanent hypocalcaemia rate at 6 months.
11. Resource usage for consultation time and diagnostic testing procedures as well as subsequent management including consultations and surgical treatments.

Outcome assessment schedule

Participant reported assessments were made following informed consent and then at baseline (prior to FNAC), and then at 3, 6 and 12 months post randomisation (see [Table 1](#)).

Sample size

We planned to recruit a sample of 968 patients to achieve over 90% power for detecting the following difference in the primary outcome at the 5% significance level, allowing for 15% loss to follow-up. An audit of US-only guided FNACs in one institution suggested that Thy 1s made up 20% of FNAC results.¹² This is likely to be an overestimate as this was including cystic and haemorrhagic nodules (which tend to yield paucicellular specimens leading to Thy 1 cytology results), so a more conservative estimate of 10% was assumed in calculating the sample size. The hypothesis was that using USE in addition to US in

TABLE 1 Assessment schedule

Assessment schedule	Prior to trial entry	Baseline prior to first FNAC	Immediately after any FNAC	After FNAC result	After surgery	30 days post op	3 months post randomisation	3 months post op	6 months post op	6 months post randomisation	12 months post op	12 months post randomisation
Written informed consent	X											
Review inclusion/exclusion criteria	X											
EQ-5D		X			X		X			X		X
HADS		X					X			X		X
Cost collection CRF					X		X			X		X
Blood sample collection		X						X			X	
Tissue sample collection ^a			X		X							
Histology Assessment CRF ^b					X							
Visual Analogue Pain Score			X									
FNAC Assessment CRF			X									
FNAC Result CRF ^c				X								
Surgical Decision Form				X								
Surgery and surgical complications CRF ^d						X			X			

a For participants who consented to the collection of blood and tissue.

b The accompanying histopathology report was provided together with this CRF.

c The accompanying radiology report was provided together with this CRF.

d Completed using the 30-day Post-Operative Assessment CRF and 6-month Post-Operative Assessment CRF.

guiding the FNAC would reduce the number of Thy 1s to 4%. To detect this difference, with a continuity correction, required 411 patients in each arm or 822 in total. Therefore, after adjusting for a 15% drop-out rate, a total of 968 patients was required.

Trial sites

A total of 23 sites across the UK went through the set-up process and 18 of these proceeded to full approval.

Recruitment process

Patients with single or multiple thyroid nodules were identified for inclusion in the ElATION trial prior to attending a radiology session for US-FNAC of the thyroid. Nurses or researchers at participating trusts identified potentially eligible patients at their initial consultation or via review of the radiology department booking systems one week prior to their scheduled appointment.

Sites were provided with an ethics committee-approved patient invitation letter which the nurse (or other members of the research team) sent to eligible patients to invite them to participate in the study. The Patient Information Sheet (PIS) was also sent with the invitation letter. Patients were asked to attend approximately 30 minutes prior to their scheduled appointment (exact length of time was as per local preference), in order to discuss the trial and be asked to consent.

To exclude patients with largely cystic nodules with little solid component and haemorrhagic nodules, it was necessary to perform an US. Therefore, patients who otherwise met the eligibility criteria and who agreed to consent for entry into the study were consented and the radiologist performed an US to determine the presence or absence of such nodules, prior to randomisation. Those patients found to have predominantly cystic nodules or those with clearly haemorrhagic were excluded from the trial and did not receive a randomised allocation; details of these patients were recorded on the ElATION Screening Log.

Informed consent

The conduct of the trial was in accordance with the principles of good clinical practice (GCP) and applicable regulatory requirements.

The patient's written informed consent to participate in the trial was obtained prior to performing any trial-related procedure, prior to randomisation and after a full explanation of the study had been given. A PIS was provided to facilitate this process.

If the participant expressed an interest in participating in the trial, they were asked to sign and date the latest version of the Consent Form. The participant had to give explicit consent for the regulatory authorities, members of the research team, and representatives of the sponsor to be given direct access to the participant's medical records. The investigator (or delegate) then signed and dated the form. Written informed consent was obtained by a trained member of the research team with GCP training, knowledge of the trial protocol and delegated authority from the local principal investigator (PI).

Within the ElATION trial, consent was usually obtained by an ElATION research nurse on site. However, consent could also be obtained by the Consultant Radiologist or by a delegated person, for example, specialist registrar, radiographer or sonographer. Details of the informed consent discussions were recorded in the participant's medical notes. This included the date of discussion, the name of the trial, summary of the discussion, version number of the PIS given to the participant and version number of the consent form signed, and the date consent was received.

METHODS

At each visit the participant's willingness to continue in the trial was ascertained and documented in the medical notes. Throughout the trial the participant was given the opportunity to ask questions about the trial. Any new information, that may have been relevant to the participant's continued participation, was provided.

A separate consent form was signed by participants who were willing to give blood and tissue samples for the purpose of translational research.

The patient's general practitioner (GP) was notified, with the patient's consent, and a specimen 'Letter to GP' was supplied.

In order to ascertain generalisability, a log was kept of all patients who were potentially eligible and were approached to participate in the study, if they were randomised or not, and the reason for non-randomisation.

Randomisation

Randomisation method and stratification variables

Patients were randomised in a 1 : 1 ratio between USE-guided FNAC and conventional US-guided FNAC.

A minimisation procedure, using a computer-based algorithm, was used to avoid chance imbalances in important stratification variables. The stratification variables were:

1. radiologist: as US and USE scans are operator-dependent;
2. solitary nodule versus multinodular: as multiple nodules can affect the utility and accuracy of SE;
3. size of nodule (≤ 4 cm vs. > 4 cm);
4. solid versus mixed solid and cystic nodules.

A random factor was incorporated into the randomisation to reduce predictability and thus avoid selection bias. This means that each patient had a probability of either being minimised, or of receiving the opposite intervention to the one they would have received if they had been minimised.

Randomisation

Once eligibility was confirmed and after written informed consent was obtained, patients were randomised to the trial online using a bespoke trial website, or by telephone directly with the ElatION Trial Office.

Information was needed on the number of nodules and their nature (solid or mixed) to enable randomisation. If multiple nodules were found, the nodule most suspicious at the time of the US was used for randomisation.

Planned interventions

Definitive investigation of thyroid nodules is by US and FNAC according to the 2014 BTA guidelines.³ The intervention being assessed within ElatION was the use of strain or shear wave USE at the same time as the routine US examination.

Experimental arm – strain or shear wave elastography in conjunction with ultrasound-guided fine-needle aspiration cytology

Those randomised to the intervention arm underwent USE together with routine US each time they had a diagnostic evaluation and FNAC of their nodule throughout the duration of the trial.

All repeat FNACs had to be undertaken using the same US technique as the first one specified by the randomisation and ideally by the same SE-accredited radiologist.

Training and accreditation for SE

Radiologists and senior sonographers trained and accredited in USE and US-FNAC of the thyroid delivered the intervention.

As elastography was not commonly used in the UK, and many radiologists did not have experience of this technique, all participating radiologists and any senior sonographers delivering the intervention, were required to attend a training and accreditation module developed for the ElatION trial.

STEP 1: Participating radiologists had to submit an audit of the results of 20 consecutive FNACs that they have undertaken in the last 18 months and the total number of US FNACs undertaken in the previous year.

STEP 2: Participating radiologists attended a workshop on USE.

STEP 3: Following the workshop, the radiologist used USE in conjunction with normal US on 15 patients in their routine radiology lists at their hospitals to gain experience. A logbook of cases, with the outcome of the FNAC result, was required.

STEP 4: Do a 'hot case' accreditation – where the radiologist performed USE US on one patient attending a radiology list and indicated which nodule they would sample. Following the successful completion of the programme, accreditation in USE was awarded.

STEP 5: The scans of the first five USE cases, done by each radiologist, were reviewed by the Trial Central Radiology Panel.

STEP 6: Online elastography community: as part of radiologist's ongoing training, complex cases could be circulated via a centralised email list for further discussion by all the radiologists. Once the radiologists completed the accreditation process, they were added onto the list by the ElatION Trial Coordinator.

Control arm – conventional ultrasound-guided fine-needle aspiration cytology

Conventional grey scale and colour Doppler US-guided FNAC for all the FNACs was required until a definitive diagnosis was obtained.

All repeat FNACs were undertaken using the same US technique as the first one specified by the randomisation and ideally by the same USE-accredited radiologist.

Other management at the discretion of local doctors

Apart from the trial treatments allocated at randomisation, all other aspects of patient management were at the discretion of the local doctors, with no other special treatments or investigations and no additional follow-up visits.

Compatibility with other studies

A patient could be part of both ElatION and another interventional trial, provided the other trial did not affect: (1) the decision to do a US-FNAC and (2) the decision to undertake surgery based on the FNAC result.

Trial assessments

Trial schema and trial visit schedule

Follow-up assessments

Follow-up assessments were undertaken for one year from randomisation or until definitive diagnosis had been obtained if not achieved during that first year. This was sufficient time to allow for further FNACs, if required, after the first test. It also allowed sufficient time for any surgery to be undertaken and histological diagnosis to be available.

For both intervention and control arms, diagnosis and management proceeded as follows:

- Thy 1 – repeat FNAC, especially if suspicion of malignancy. If Thy 1 on three FNACs, then the recommendation was made for diagnostic surgery.
- Thy 2 – repeat FNAC within 3–6 months. If two benign (Thy 2) FNAC results were obtained, then the patient could be discharged. Consideration for discharge with a U2 scan and one Thy 2 was allowed, but not recommended for the purposes of this study.
- Thy 3a – repeat FNAC within 3–6 months or discussion with MDT. Surgery may then be advised.
- Thy 3f/4/5 – surgery was necessary.

All repeat FNAC were undertaken using the same US technique as the first one specified by the randomisation and ideally by the same SE-accredited radiologist.

All follow-up data were captured on the relevant case report form (CRF) and returned to the ElATION Trial Office.

Patient assessment

Patient-reported assessments were performed using commonly used, validated questionnaires:

1. EQ-5D questionnaire completed at: baseline (at recruitment); after surgery; and 3, 6 and 12 months after randomisation;
2. cost collection form for health resource usage, completed at baseline, 3, 6 and 12 months after first FNAC;
3. Visual Analogue Pain Score after every FNAC;
4. HADS questionnaire completed at: baseline, 3, 6 and 12 months after randomisation.

Where possible, patient questionnaires were completed when patients attended hospital appointments. If this was not feasible, the questionnaires were posted by a member of the research team at the site, to the patient for completion at home.

Complication rates

Complication rates, following thyroid operations, were recorded at 30 days and 6 months post surgery.

Data were collected on haematoma rate and temporary hypocalcaemia rate at 30 days and vocal cord palsy and permanent hypocalcaemia at 6 months post surgery only.

Data collection

The data collection comprised the forms detailed in [Table 2](#):

TABLE 2 Summary of data recorded on CRFs

Form	Summary of data recorded	Schedule for submission
FNAC assessment CRF	Type of procedure performed (USE-FNAC or US-FNAC); Date of procedure; Timepoint (baseline, second FNAC, third FNAC, fourth FNAC); If local anaesthetic was used; Number and type of needles used for biopsy; If cytopathologist was present during the FNAC procedure, and if so if the first sample was adequate for diagnostic assessment ^a ; Size of nodule; Echogenicity, Composition, Calcification type (if any); Halo; AP and Margins; US assessment; US-USE assessment (if applicable); Name of radiologist; Radiologist feedback on helpfulness of USE.	Immediately following each FNAC performed.
FNAC result CRF	Date of procedure; Date of result; Details of result (i.e. Thy 1, Thy 2, etc.); Name of reporting cytopathologist; Outcome of consultation of FNAC (repeat FNAC, surgery or discharge).	Immediately following release of cytopathology results from each FNAC performed.
Surgical decision CRF	Planned date of surgery; Number of FNACs required to obtain decision for surgery; Date of last FNAC; Result of last FNAC.	Following clinical decision to progress to thyroidectomy surgery.
Histopathology assessment CRF	Histology reference; Name of reporting histopathologist; Histopathological diagnosis (malignant or benign); Any other incidental findings.	Immediately following release of histopathology results from surgery.
30-days post operative CRF and 6-months post operative CRF	Date of surgery; Type of operation performed; Date of hospital discharge; Any subsequent readmission; Reason for readmission; Any surgical complications within the 30 days following surgery.	30 days after surgery and 6 months after surgery.
SAE form	Reason for reporting; SAE start date; SAE end date (if applicable); Details of event; Outcome of event; Local investigator review of causality to ElATION.	Within 24 hours of site being aware of any event that satisfied the SAE criteria.

^a One centre had the benefit of a cytopathologist present at the time of FNAC, who would perform a preliminary review of the aspirate before sending for further analysis, thus significantly reducing the number of Thy 1 results. By asking the question of whether the first aspirate was adequate to send for diagnosis we reasonably assumed any answers of 'No' could be used as Thy 1.

Statistical methods

The primary comparison groups were composed of those in the SE-FNAC arm compared to those in the US-FNAC arm. Patients, not biopsies, were the unit of analysis. All analyses were on the intention to treat (ITT) principle, with all patients analysed in the arms to which they were allocated irrespective of adherence to the randomised allocated diagnostic tool, and all patients were included in the analyses. For all tests, summary statistics are presented and 95% CIs were constructed where appropriate. A $p < 0.05$ was considered statistically significant, and there was no adjustment for multiple comparisons.

Primary outcome analysis

The primary outcome analysis was the proportion of patients who had a Thy 1 result following their first FNAC. A generalised linear mixed model was used to estimate the risk difference (RD) between the SE-FNAC and US-FNAC groups, with the minimisation variables: nodule multiplicity (solitary vs. multinodular); nodule composition (solid vs. mixed solid and cystic); and nodule size (≤ 4 cm vs. > 4 cm) included as fixed variables; and radiologist included as a random effect.

Secondary outcome analyses

All secondary outcomes (except the accuracy analyses) were adjusted for the minimisation variables.

METHODS

Binary secondary outcomes (e.g. number of thyroidectomies) were analysed as per the primary outcome. For the proportion of patients with benign histology, all patients randomised to the trial were included in the denominator, to avoid bias if there was an imbalance in the number of thyroidectomies between the groups.

Time from first FNAC to definitive diagnosis was analysed using a Cox-proportional hazards model. Hazard ratios (HRs) were presented as well as the median and interquartile range (IQR). Kaplan–Meier curves were also produced for visual presentation of the comparisons.

The number of FNACs needed to obtain a definitive diagnosis was analysed using an ordinal regression model to obtain an odds ratio (OR).

Quality of life outcome measures included the HADS, VAPS and EQ-5D questionnaires. These were taken at baseline, 3, 6 and 12 months post randomisation. The outcome measures were analysed using repeated measures methods using all available data. The baseline value of the measure, time and minimisation variables as per primary analysis were included in the model as fixed effects, with radiologist included as a random effect.

Accuracy outcomes were presented as tables with corresponding sensitivity and specificity provided alongside the *p*-value from McNemar's test.

The radiologist survey results, thyroidectomy complications and serious adverse events (SAEs) were presented in tables and summarised using numbers and percentages.

Missing data and sensitivity analyses

Sensitivity analyses were also performed on the primary outcome measure. These included: a per-protocol analysis, which included only those who receive the test intervention they were randomised to receive; an analysis excluding those centres where there was a cytologist present; an analysis including only those participants who had had a radiologist positive about USE; an analysis including only those radiologists who performed at least: 10 USEs; 20 USEs; 40 SEs; and 80 USEs in the trial; and finally an analysis including only those radiologists who passed their treatment quality assurance.

A sensitivity analysis was also carried out on the accuracy outcomes imputing missing participant final definitive diagnosis (FDD) with the following assumptions: if a participant has at least one FNAC result classified as 'malignant', then their FDD will be classified as malignant; and if a participant has a single Thy 2 result, or a Thy 2 result plus a Thy 1 then their FDD will be classified as benign.

Subgroup analyses

Subgroup analyses were planned on the stratification variables used for randomisation. These were: solitary nodule versus multinodular; the size of the nodule (< 4 cm vs. > 4 cm); and solid versus mixed solid and cystic nodules. Tests for interaction were performed to assess whether the intervention effect differs between the strata. The study has not been powered to detect any differences in these subgroups, so any significant results are purely hypothesis generating.

Defining final definitive diagnosis

For the accuracy outcomes, a definition of what would be considered a participant's final diagnosis was required. The following is what was used as FDD:

Benign – If participant received two Thy 2 FNAC results; or participant received a U2 FNAC assessment with a Thy 2 result; or participant had surgery but the nodule was found to be benign.

Malignant – If the participant obtained a malignant histological diagnosis.

The number of participants and how FDD was obtained is summarised and presented in tables. Assumptions were placed on those missing FDD.

Cost-effectiveness analysis

A cost-effectiveness analysis was planned. However, when the preliminary results became available, it became clear that a cost effectiveness analysis was not appropriate, given there was no difference in primary and secondary outcomes between the two arms, so would be no cost-benefit to the addition of SE. Therefore, the cost-effectiveness analysis was not undertaken.

Ethics approval, regulations and trial registration

Ethics approval

ElaTION received full ethical approval from the South Central – Berkshire Research Ethics Committee (REC) on 10 October 2014. The REC reference is 14/SC/1206.

ElaTION was brought under the Health Research Authority (HRA) approval process in August 2016. The HRA assesses governance and legal compliance, and the ElaTION Trial Office was responsible for obtaining this approval.

Local hospitals conducted internal capacity and capability checks to assess the facilities and resources needed to run the trial, in order to give host site permission. The Trial Office provided help to the local Principal Investigator in the process of obtaining trust management approval by supplying the HRA Local Documents Package. The local Principal Investigator was responsible for liaison with the Trust management with respect to locality issues.

Once hospital approval was obtained, the ElaTION Trial Office confirmed that all appropriate site approvals were in place and that the USE accreditation had been completed. When the ElaTION Trial Office, on behalf of the sponsor, verified that all applicable regulatory requirements have been met, the local PI was informed that the study was open at the hospital and potential trial participants could start to be approached. The Trial Office sent the Investigator's Site File containing all trial materials to the local Principal Investigator.

Sponsorship

Sponsorship was provided by the University of Birmingham upon signing of the Clinical Study Site Agreement with each trial site.

ElaTION was developed by the Institute of Head and Neck Studies and Education (InHANSE) team and the Birmingham Clinical Trials Unit (BCTU) and funded by the Health Technology Assessment (HTA) programme of the NIHR (12/19/04).

The University of Birmingham was the trial 'sponsor'.

There were no specific arrangements for compensation made in respect of any SAEs occurring through participation in the trial, whether from the side effects listed, or others yet unforeseen. Hospitals selected to participate in this trial provided clinical negligence insurance cover for harm caused by their employees and a copy of the relevant insurance policy or summary was to be provided to the University of Birmingham, upon request.

No liability claims were submitted against the ElaTION trial.

Regulations

The trial was conducted in compliance with the principles of the Declaration of Helsinki (1996), the principles of GCP and in accordance with all applicable regulatory requirements including but not limited to the Research Governance Framework for Health and Social Care and the Medicines for Human Use (Clinical Trials) Regulations 2004, as amended in 2006 and any subsequent amendments.

Monitoring and oversight

Confidentiality of personal data

All data were handled in accordance with the UK Data Protection Act 1998 and subsequent amendments.

Participant name was not included on any CRF used in ElaTION. The participant's initials, date of birth and trial identification number were used for identification.

Personal data and sensitive information required for the ElaTION trial were collected directly from trial participants and hospital notes. Participants were informed about the transfer of this information to the ElaTION Trial Office at the BCTU and InHANSE and asked for their consent. The data were entered onto a secure computer database, either directly via the internet using secure socket layer encryption technology or indirectly from paper by BCTU staff.

All personal information received in paper format for the trial was held securely and treated as strictly confidential according to BCTU policies. All staff involved in the ElaTION trial (clinical, academic, BCTU, InHANSE) shared the same duty of care to prevent unauthorised disclosure of personal information. No data that could be used to identify an individual will be published. Data are stored on a secure server at BCTU under the provisions of the Data Protection Act and/or applicable laws and regulations.

In-house data quality assurance

Monitoring and audit

ElaTION was centrally monitored; however, on-site monitoring occurred if triggered. Investigators and their host Trusts were required to permit trial-related monitoring and audits to take place by the ElaTION Trial Office, providing direct access to source data and documents as requested. Trusts may also have been subject to inspection by the Research and Development Manager of their own Trust and were encouraged to do everything requested by the Chief Investigator in order to prepare and contribute to any inspection or audit. Trial participants were made aware of the possibility of external audit of data they provided in the participant information sheet.

Independent Trial Steering Committee

The Trial Steering Committee (TSC) provided independent supervision for the trial, providing advice to the Chief and Co-Investigators as well as the Sponsor on all aspects of the trial and affording protection for patients by ensuring the trial was conducted in compliance with the protocol, GCP and the applicable regulatory requirements.

Data Monitoring and Ethics Committee

During the study, interim analyses of safety and outcome data were supplied, in strict confidence, to an independent Data Monitoring and Ethics Committee (DMEC) along with any other analyses that the committee requested.

Long-term storage of data

Archiving will be authorised by the BCTU on behalf of the Sponsor following submission of the end of trial report.

Principal Investigators are responsible for the secure archiving of essential trial documents (for their site) as per their NHS Trust policy. All essential documents will be archived for a minimum of 5 years after completion of the trial.

Destruction of essential documents requires authorisation from the BCTU on behalf of the Sponsor.

Safety assessment and reporting

Adverse events

The collection and reporting of data on adverse events (AE) and SAEs was done in accordance with International Conference on Harmonisation Guidelines for Good Clinical Practice (ICH GCP) and the Research Governance Framework 2005.

No investigational medicinal products (IMPs) were used as part of this trial. As all of the surgical techniques being tested in this trial were used as standard practice, there were no (serious) AEs which would be anticipated as a unique consequence of participation in the trial. Any trial-related SAEs, which required immediate reporting, were reported on a trial-specific SAE form.

Other outcomes, which may also be considered safety outcomes, but which were anticipated outcomes for this group of patients, were captured on the routine follow-up CRFs (30-day and 6-month Post-Operative CRFs), these included:

- vocal cord palsy
- temporary or permanent hypocalcaemia
- haematoma
- infection
- re-operation due to surgical complications.

Serious adverse events

Definition of serious adverse event

- Recognised side effects of the treatment or disease, or an event which is secondary to those recognised effects.
- Hospitalisations for routine treatment or monitoring of the studied indication, not associated with any deterioration in condition.
- Hospitalisations for treatment, which was elective or pre-planned, for a pre-existing condition that is unrelated to the indication under study, and did not worsen.
- Admission to a hospital or other institution for general care, not associated with any deterioration in condition.
- Treatment on an emergency, outpatient basis for an event not fulfilling any of the definitions of serious given above and not resulting in hospital admission.

For the purposes of this trial these expected SAEs did not require reporting on a SAE form. These events were recorded in the source data according to local practice and included on the routine follow-up CRFs (the 30-Day Post-Operative Assessment CRF and the 6-Month Post-Operative Assessment CRF).

Disease-related morbidity and routine treatment or monitoring of a pre-existing condition that has not worsened were not considered as SAEs and were not reported to the Trial Office.

Notification of deaths

All deaths were reported to the BCTU on the SAE Form irrespective of whether the death was related to disease progression or an unrelated event. If a participant died, any post-mortem findings were

provided to the BCTU with the SAE form. The BCTU reported all deaths to the DMEC for continuous safety review.

Pharmacovigilance responsibilities

Local principal investigator (or nominated individual in PI's absence)

- Medical judgement in assigning seriousness and causality to SAEs.
- To fax SAE forms to BCTU within 24 hours of becoming aware, and to provide further follow-up information as soon as available.
- To report SAEs to the trust if required, in line with local arrangements.
- To sign an Investigator's Agreement accepting these responsibilities.

Chief investigator (or nominated individual in CI's absence)

- To assign causality and expected nature of SAEs.
- To review all events assessed as SAEs in the opinion of the local investigator.

Birmingham Clinical Trials Unit

- To prepare annual safety reports to main REC and TSC.
- To prepare SAE safety reports for the DMEC at 12-month intervals.
- To report all fatal SAEs to the DMEC for continuous safety review.

Trial Steering Committee

- To provide independent supervision of the scientific and ethical conduct of the trial on behalf of the trial Sponsor and funding bodies.
- To review data, patient compliance, completion rates, AE (during treatment and up to the end of follow-up).
- To receive and consider any recommendations from the DMEC on protocol modifications.

Data Monitoring and Ethics Committee

- To review overall safety and morbidity data to identify safety issues which may not be apparent on an individual case basis.
- To recommend to the TSC whether the trial should continue unchanged, continue with protocol modifications, or stop.

Notification of serious breaches of GCP and/or the protocol

A 'serious breach' is a breach which is likely to affect to a significant degree:

1. the safety or physical or mental integrity of the participants of the trial; or
2. the scientific value of the trial.

The BCTU on behalf of the Sponsor notified the REC in writing of any serious breach of:

1. the conditions and principles of GCP in connection with the trial; or
2. the protocol relating to the trial, as amended from time to time, within 7 days of becoming aware of that breach.

The Sponsor was notified immediately of any case where the above definition applied during the trial conduct phase.

Trial withdrawals

Patients could withdraw at any time during the trial if they chose not to continue.

There were different types of withdrawal:

- The patient did not want to attend trial-specific follow-up visits, but agreed to be followed-up according to standard practice (i.e. agreed that follow-up data could be collected at standard clinic visits).
- The patient was not willing to be followed up for trial purposes at any further visits (i.e. agreed that any data collected prior to the withdrawal of consent can be used in the trial final analysis).

Patients were free to withdraw the trial without reason, but any reasons that were stated were reported to the ElaTION Trial Office by site staff. Patients who withdrew from trial treatment but continued with on-going follow-up and data collection were followed-up in accordance with the protocol.

Chapter 3 Results

Recruitment

ElaTION opened to recruitment on 13 February 2015, and the first participant was recruited into the trial on 27 February 2015. The trial closed to recruitment with 1008 randomised participants, the last participant randomised on 26 September 2018. An inability to assess the primary outcome was discovered for 26 participants at Ashford and St Peter's Hospitals NHS Foundation Trust before the trial was closed to recruitment, therefore these participants had to be excluded, but this meant that the trial could still continue to recruit to target.

After excluding 26 participants, 982 patients were randomised in ElaTION from 18 radiology centres across England (see [Report Supplementary Material 1: Table 1](#)).

Participant flow and withdrawals

The participant flow is presented in [Figure 2](#). There were a total of 33 participants (3%) where follow-up is incomplete. Of the 18 (2%) participants who withdrew from the trial, 13 withdrew before determining their FDD. Nine (1%) participants died, causes of death were: cancer (five participants); exacerbation of chronic obstructive pulmonary disease (COPD) (one participant); hypoxic brain injury (one participant); bowel obstruction (one participant); and an acute tracheal obstruction (one participant). Five deaths

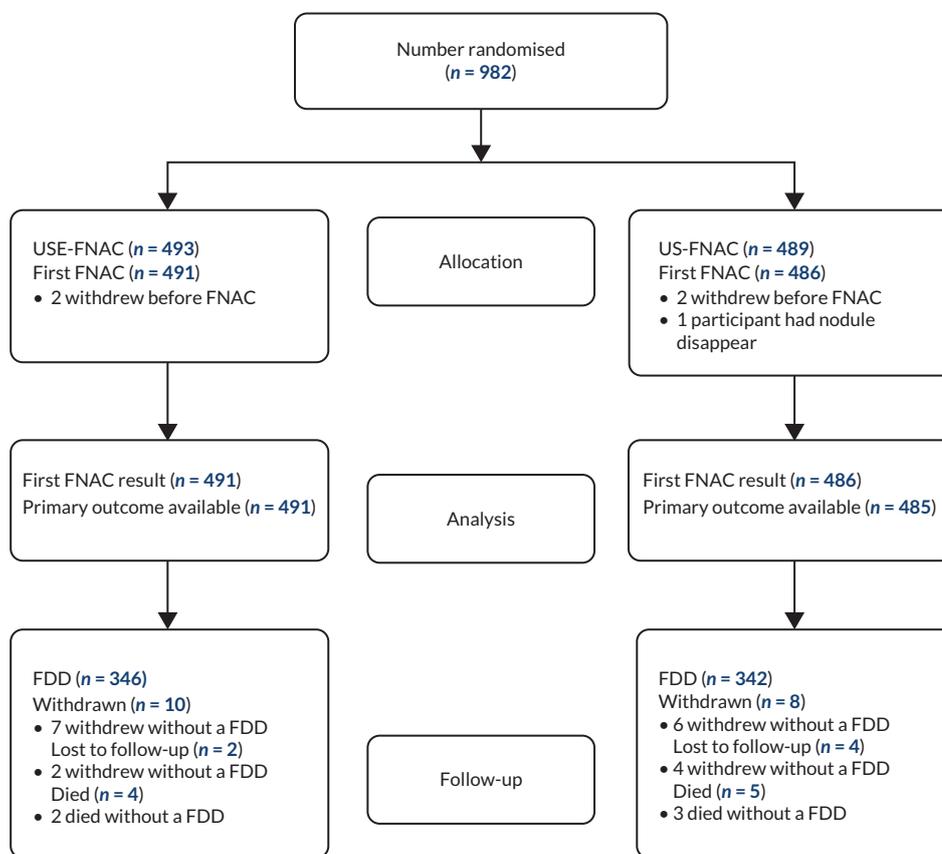


FIGURE 2 Patient flow for the ITT population.

RESULTS

were prior to receiving their FDD. Six participants (1%) were lost to follow-up, all without a FDD. There were seven (1%) participants who partially withdrew – that is, data collection was for clinical forms only (so these participants' quality of life assessment forms will be complete up until the point they partially withdrew).

Completeness of data

In general, rates of completion of first assessment forms were high. Return rates declined for the second and third assessments, though rates remained high (> 86%) for results (see [Report Supplementary Material 1: Table 2](#)) as many centres only undertook one assessment in their clinical protocols.

Post-baseline quality of life forms showed about a 30% drop in completion rate from baseline whilst return of surgery forms were all 98% or above (see [Report Supplementary Material 1: Tables 3 and 4](#)).

Baseline data

The mean age of participants in ElaTION was approximately 51 years old and 80% (789/982) were female (see [Table 3](#)). Most nodules were multinodular (71%) and less than or equal to 4 cm (79%), with an almost even split of solid versus mixed solid and cystic (53% vs. 47% respectively).

TABLE 3 Baseline characteristics at randomisation split by intervention group and overall

		SE-FNAC	US-FNAC	All
		N = 493	N = 489	N = 982
Age (years)	Mean (SD)	51.8 (15.6)	50.8 (14.8)	51.3 (15.2)
	Median			
	(IQR)	52 (39–64)	51 (40–62)	51 (39–63)
Sex	Male	92 (19%)	101 (21%)	193 (20%)
	Female	401 (81%)	388 (79%)	789 (80%)
Nodule multiplicity ^a	Solitary	140 (28%)	141 (29%)	281 (29%)
	Multinodular	353 (72%)	348 (71%)	701 (71%)
Nodule composition ^a	Solid	261 (53%)	258 (53%)	519 (53%)
	Mixed solid and cystic	232 (47%)	231 (47%)	463 (47%)
Nodule size ^a	≤ 4 cm	395 (80%)	380 (78%)	775 (79%)
	> 4 cm	98 (20%)	109 (22%)	207 (21%)
Radiologist ^a	Radiologist A	6 (1%)	6 (1%)	12 (1%)
	Radiologist B	10 (2%)	8 (2%)	18 (2%)
	Radiologist C	90 (18%)	89 (18%)	179 (18%)
	Radiologist D	19 (4%)	22 (5%)	41 (4%)
	Radiologist E	2 (0.4%)	4 (1%)	6 (1%)
	Radiologist F	26 (5%)	29 (6%)	55 (6%)
	Radiologist G	45 (9%)	40 (8%)	85 (9%)
	Radiologist H	12 (2%)	11 (2%)	23 (2%)
	Radiologist I	5 (1%)	5 (1%)	10 (1%)

TABLE 3 Baseline characteristics at randomisation split by intervention group and overall (*continued*)

	SE-FNAC N = 493	US-FNAC N = 489	All N = 982
Radiologist J	8 (2%)	6 (1%)	14 (1%)
Radiologist K	15 (3%)	16 (3%)	31 (3%)
Radiologist L	4 (1%)	5 (1%)	9 (1%)
Radiologist M	1 (0.2%)	1 (0.2%)	2 (0.2%)
Radiologist N	23 (5%)	23 (5%)	46 (5%)
Radiologist O	16 (3%)	16 (3%)	32 (3%)
Radiologist P	8 (2%)	6 (1%)	14 (1%)
Radiologist Q	1 (0.2%)	0 (0%)	1 (0.1%)
Radiologist R	1 (0.2%)	1 (0.2%)	2 (0.2%)
Radiologist S	111 (23%)	113 (23%)	224 (23%)
Radiologist T	24 (5%)	25 (5%)	49 (5%)
Radiologist U	1 (0%)	1 (0.2%)	1 (0.1%)
Radiologist V	2 (0.4%)	1 (0.2%)	3 (0.3%)
Radiologist W	8 (2%)	8 (2%)	16 (2%)
Radiologist X	12 (2%)	11 (2%)	23 (2%)
Radiologist Y	27 (5%)	26 (5%)	53 (5%)
Radiologist Z	16 (3%)	15 (3%)	31 (3%)
Radiologist AA	1 (0.2%)	1 (0.2%)	2 (0.2%)

a Minimisation variables.

Nodules from the first US assessment were mostly isoechoic (55%), 36% were hypoechoic, and 9% were hyperechoic in echogenicity (see [Table 4](#)). Nodules had a mean (SD) of 27.4 (15.6) mm with well-defined margins (89%). On the first US assessment: 290 (30%) were assessed as U2 (Benign); 532 (54%) were assessed as U3 (Indeterminate/Unequivocal); 113 (12%) were assessed as U4 (Suspicious); and 42 (4%) were assessed as U5 (Malignant).

TABLE 4 Nodule characteristics from first FNAC assessment by intervention group and overall

Nodule characteristics from first FNAC	USE-FNAC N = 491	US-FNAC N = 486	All N = 977
Local anaesthetic used			
Yes	146 (30%)	137 (28%)	283 (29%)
No	345 (70%)	348 (72%)	693 (71%)
Missing	0 (0%)	1 (0.2%)	1 (0.1%)
Number of needles used			
1	148 (30%)	147 (30%)	295 (30%)
2	257 (52%)	259 (53%)	516 (53%)

continued

RESULTS

TABLE 4 Nodule characteristics from first FNAC assessment by intervention group and overall (continued)

Nodule characteristics from first FNAC	USE-FNAC	US-FNAC	All
	N = 491	N = 486	N = 977
3	45 (9%)	51 (10%)	96 (10%)
4	28 (6%)	20 (4%)	48 (5%)
5	2 (0.4%)	1 (0.2%)	3 (0.3%)
Missing	11 (2%)	8 (2%)	19 (2%)
Type of needle used			
Spinal	23 (5%)	17 (4%)	40 (4%)
Normal (blood)	457 (93%)	455 (94%)	912 (93%)
Cyto-Foam Needle	1 (0.2%)	2 (0.4%)	3 (0.3%)
Missing	10 (2%)	12 (2%)	22 (2%)
Nodule unit of measurement			
mm	486 (99%)	481 (99%)	967 (99%)
cm	5 (1%)	5 (1%)	10 (1%)
Missing	0 (0%)	0 (0%)	0 (0%)
Nodule size (mm)			
Mean (SD)	27.1 (15.8)	27.7 (15.4)	27.4 (15.6)
Median (IQR)	25 (14–36)	25 (16–38)	25 (15–37)
Missing	0 (0%)	0 (0%)	0 (0%)
Nodule echogenicity			
Hypo	200 (41%)	155 (32%)	355 (36%)
Iso	249 (51%)	284 (58%)	533 (55%)
Hyper	41 (8%)	46 (9%)	87 (9%)
Missing	1 (0.2%)	1 (0.2%)	2 (0.2%)
Nodule composition			
Solid	233 (47%)	245 (50%)	478 (49%)
Mixed	216 (44%)	208 (43%)	424 (43%)
Spongiform	29 (6%)	26 (5%)	55 (6%)
Purely cystic nodules with no solid component	0 (0%)	1 (0.2%)	1 (0.1%)
Missing	13 (3%)	6 (1%)	19 (2%)
Calcification type			
None	383 (78%)	376 (77%)	759 (78%)
Micro	30 (6%)	39 (8%)	69 (7%)
Macro/coarse	66 (13%)	61 (13%)	127 (13%)
Rim ^a	5 (1%)	7 (1%)	12 (1%)
Missing	7 (1%)	3 (1%)	10 (1%)

TABLE 4 Nodule characteristics from first FNAC assessment by intervention group and overall (*continued*)

Nodule characteristics from first FNAC	USE-FNAC	US-FNAC	All
	N = 491	N = 486	N = 977
Halo			
Regular	165 (34%)	202 (42%)	367 (38%)
Interrupted	60 (12%)	47 (10%)	107 (11%)
Absent	265 (54%)	237 (49%)	502 (51%)
Missing	1 (0.2%)	0 (0%)	1 (0.1%)
AP > TR			
Yes	40 (8%)	42 (9%)	82 (8%)
No	450 (92%)	443 (91%)	893 (91%)
Missing	1 (0.2%)	1 (0.2%)	2 (0.2%)
Margins^b			
Well defined	432 (88%)	438 (90%)	870 (89%)
Irregular	54 (11%)	45 (9%)	99 (10%)
Missing	5 (1%)	3 (1%)	8 (1%)
Doppler blood flow			
Central	33 (7%)	36 (7%)	69 (7%)
Peripheral	167 (34%)	183 (38%)	350 (36%)
Mixed	257 (52%)	240 (49%)	497 (51%)
None	33 (7%)	27 (6%)	60 (6%)
Missing	1 (0.2%)	0 (0%)	1 (0.1%)
Metastatic lymph nodes			
Yes	11 (2%)	14 (3%)	25 (3%)
No	475 (97%)	469 (97%)	955 (97%)
Unclear	5 (1%)	3 (1%)	8 (1%)
Missing	1 (0.2%)	0 (0%)	0 (0%)
US assessment			
U1 – Normal	0 (0%)	0 (0%)	0 (0%)
U2 – Benign	147 (30%)	143 (29%)	290 (30%)
U3 – Indeterminate/Equivocal	252 (51%)	280 (58%)	532 (54%)
U4 – Suspicious	73 (15%)	40 (8%)	113 (12%)
U5 – Malignant	19 (4%)	23 (5%)	42 (4%)
Missing	0 (0%)	0 (0%)	0 (0%)

a Rim only appeared as a calcification type on version 2.0 which came into effect 2 June 2015.

b Margins had separate variables in versions 1.0, 1.1 and 1.2 for well-defined and irregular margins. Three 'no' answers for both well-defined and irregular included in missing. Two 'yes' to both well-defined and irregular answers.

Note

Numbers and percentages are presented unless otherwise specified.

Adherence

Adherence to randomised test

Adherence to the participant's randomised test was high. For the initial FNAC assessment, adherence was 99% or higher for both the USE-FNAC and US-FNAC groups (see [Table 5](#)). The adherence dropped slightly for FNAC assessments following the initial FNAC, with a larger drop observed in the USE-FNAC group compared with the US-FNAC group, but adherence was still good (85% or above) by the third FNAC.

Adherence to final definitive diagnosis protocol definition

Out of 982 participants randomised, 688 reached a FDD. There were 346 participants randomised to the USE-FNAC group who had a FDD, of which 276 had a benign FDD and 70 had a malignant diagnosis. In the US-FNAC group 342 participants had a FDD, 263 had a benign FDD and 79 had a malignant FDD. [Table 6](#) shows how these final definitive diagnoses were obtained.

TABLE 5 Treatment adherence to by randomised test

		USE-FNAC	US-FNAC
<i>Initial FNAC assessment</i>	Randomised imaging – Yes N (%)	490/491 (99.8%)	483/486 (99%)
<i>Second FNAC assessment</i>	Randomised imaging – Yes N (%)	257/270 (95%)	248/254 (98%)
<i>Third FNAC assessment</i>	Randomised imaging – Yes N (%)	39/46 (85%)	39/40 (98%)
<i>Additional FNAC assessment</i>	Randomised imaging – Yes N (%)	5/7 (71%)	7/8 (88%)
Total number of participants having at least one USE-FNAC assessment		491	11
Total number of participants having at least one US-only FNAC assessment		21	484

TABLE 6 Participants who obtained their final definitive diagnoses and how they obtained them

	USE-FNAC N = 346	US-FNAC N = 342
Benign		
2 x Thy 2 FNAC results	46 (13%)	35 (10%)
U2 US followed by a Thy 2 FNAC result	46 (13%)	45 (13%)
Histology	98 (28%)	104 (30%)
US 2 and 2 x Thy 2 results	71 (21%)	66 (19%)
2 x Thy 2 and histology result	0 (0%)	2 (1%)
US 2 US followed by a Thy 2 FNAC result and histology	11 (3%)	9 (3%)
All FDD criteria	4 (1%)	2 (1%)
Malignant		
Histology	70 (20%)	76 (22%)
US 2 US followed by a Thy 2 FNAC result and histology	0 (0%)	3 (1%)

Primary outcome

Primary analysis

The primary outcome is the proportion of participants with a Thy 1 result following the first FNAC post-randomisation. The sample size was based on a Thy 1 rate of 10% in the US-FNAC arm and a 4% Thy 1 rate in the USE-FNAC arm.

The Thy 1 rate following the first FNAC observed in the trial was higher than anticipated in both the USE-FNAC and US-FNAC groups (19% vs. 16% respectively). However, there was no evidence of a difference between the groups found (RD 0.03; 95% CI, -0.007 to 0.066; $p = 0.11$) (see [Table 7](#)).

Sensitivity analyses

There were a number of sensitivity analyses carried out to assess the effect, if any, on inferences including: adherence to the test; there being a cytologist present at the centre; radiologist opinion of elastography being positive; and radiologist passing on treatment quality assurance.

The per-protocol analysis result, which included only those who adhered to randomised test, supported the primary analysis (RD: 0.029; 95% CI, -0.007 to 0.065; $p = 0.11$). The only centres with a cytologist present were King's College Hospital and St. Peter's Hospital. However, two participants recruited from St. Peter's Hospital had their FNAC performed at Ashford Hospital, which did not have a cytologist present. Again, the results from this analysis excluding cases where a cytologist was present, whilst seeing a slight drop in Thy 1 results (17% vs. 14% for USE-FNAC and US-FNAC respectively) saw no evidence of a difference between the groups (see [Report Supplementary Material 1: Table 5](#)).

Similarly, the analyses including only those radiologists who felt elastography helped above conventional US in determining malignancy and the radiologists who passed on treatment quality assurance supposed the primary analysis (see [Report Supplementary Material 1: Table 6](#)).

We also performed an analysis to assess whether the accuracy of using elastography increased with the number of SEs. Whilst the percentage of Thy 1 rates decreased in the USE-FNAC group (see [Table 8](#)), they also decreased in the US-FNAC group. The differences between USE and US remained broadly the same as per the primary analysis, despite the volume of USE scans done, and the direction of effect was in favour of US-FNAC.

Secondary outcomes

Thy 1 result following any FNAC

When considering the number of Thy 1 results following any FNAC performed, a borderline significant difference was observed in favour of US-FNAC when compared with USE-FNAC (RD: 0.041; 95% CI

TABLE 7 Primary analysis of the number of Thy 1 results following first FNAC by trial arm

	USE-FNAC	US-FNAC	Adjusted ^a RD (95% CI)	RD (95% CI)	<i>p</i> -value ^b
Primary outcome	N = 491	N = 485			
N (%)	91 (19%)	78 (16%)	0.030 (-0.007 to 0.066)	0.025 (-0.023 to 0.072)	0.11

a Adjusting for minimisation variables.

b *p*-value taken from adjusted model.

Note

A negative difference favours the USE-FNAC arm.

TABLE 8 Number of Thy 1 results following first FNAC by trial arm including those with staggered SE-FNACs performed

	USE-FNAC	US-FNAC	Adjusted ^a RD (95% CI)	RD (95% CI)	p-value ^b
Primary outcome					
	N = 444	N = 442			
SEs performed > 10	N (%) 76 (17%)	N (%) 69 (16%)	0.021 (-0.015 to 0.058)	0.015 (-0.034 to 0.064)	0.25
	N = 344	N = 345			
SEs performed > 20	N (%) 60 (17%)	N (%) 47 (14%)	0.039 (0.003 to 0.075)	0.038 (-0.016 to 0.092)	0.03
	N = 244	N = 242			
SEs performed > 40	N (%) 35 (14%)	N (%) 31 (13%)	0.004 (-0.030 to 0.038)	0.015 (-0.046 to 0.076)	0.82
	N = 199	N = 202			
SEs performed > 80	N (%) 21 (11%)	N (%) 17 (8%)	0.022 (-0.005 to 0.048)	0.021 (-0.036 to 0.079)	0.12

a Adjusting for minimisation variables.

b p-value taken from adjusted model.

Note

A negative difference favours the USE-FNAC arm.

0.002 to 0.081; $p = 0.04$) (see [Table 9](#)). However, the number of follow-up FNACs would be influenced by the result of the first FNAC result, so this result is to be treated with caution as it is subject to bias. Indeed, the Thy 1 rate appears to increase with successive FNACs, and that occurs for both USE and US guided. This suggests that as non-Thy 1 cases are excluded, more Thy 1 cases remain in the pool, and therefore represent a high proportion of the pool. It also suggests that a large proportion of the Thy 1 cases remain Thy 1 despite successive FNAs.

Thyroidectomies and benign histology

Patients in the USE arm received fewer thyroidectomies than those in the US-FNAC group (37% vs. 40% respectively), but this was not statistically significant (see [Report Supplementary Material 1: Table 7](#)). Although not powered sufficiently, there was no evidence of a difference in the rate of benign histology between the groups in the trial (RD: -0.007; 95% CI -0.04 to 0.03; $p = 0.70$).

Final definitive diagnosis

There were 688/982 (70%) participants who had a FDD. Details of how these were reached can be found in [Report Supplementary Material 1: Table 8](#). For the 294 without a FDD, 102 (35%) had a Thy 2 result following first FNAC and of these, 76 did not receive a follow-up FNAC (see [Report Supplementary Material 1: Table 9](#)).

The percentage of participants who reached FDD was the same in each group: 346/493 (70%) in the USE-FNAC group and 342/489 (70%) in the US-FNAC group.

Time to final definitive diagnosis

No difference was observed in the trial between the groups when examining time to FDD (HR: 0.94; 95% CI 0.81 to 1.10; $p = 0.45$) (see [Table 10](#), [Figure 3](#)).

Number of FNACs until final definitive diagnosis

The median number of FNAC required to reach FDD was higher in the USE-FNAC group (median: 2.0; IQR 1.0–2.0) than the US-FNAC group (median: 1.0; IQR 1.0–2.0). The odds of having more FNACs in the USE-FNAC was 1.10 when compared with the US-FNAC group holding all other variables constant, however, this was not statistically significant (see [Report Supplementary Material 1: Table 10](#)).

TABLE 9 Number of Thy 1 results following any FNAC

	USE-FNAC	US-FNAC	Adjusted ^a RD (95% CI)	RD (95% CI)	p-value ^b
	N = 493	N = 489			
	N = 491	N = 485			
Thy 1 result following: first FNAC	91 (19%)	78 (16%)			
	N = 25	N = 239			
Second FNAC	57 (22%)	45 (19%)			
	N = 44	N = 37			
Third FNAC	15 (34%)	12 (32%)			
	N = 7	N = 7			
Fourth FNAC	0 (0%)	0 (0%)			
	N = 492	N = 485			
Any FNAC	122 (25%)	100 (21%)	0.041 (0.002 to 0.081)	0.042 (-0.011 to 0.094)	0.04

a Adjusting for minimisation variables.

b p-value taken from adjusted model.

Note

Numbers and percentages are presented unless stated otherwise. A negative difference favours the USE-FNAC arm.

TABLE 10 Time to final definitive diagnosis

	USE-FNAC N = 493	US-FNAC N = 489	Adjusted ^a HR (95% CI)	HR (95% CI)	p-value ^b
N (%)	346 (70%)	342 (70%)	0.94 (0.81 to 1.10)	1.07 (0.92 to 1.24)	0.45
Median time in months to FDD (IQR)	3.3 (1.5–6.4)	3.4 (1.5–6.2)			

a Adjusting for minimisation variables.

b p-value taken from adjusted model.

Note

A ratio < 1 favours the USE-FNAC arm.

Patient-reported anxiety, depression, pain and quality of life outcome measures

Hospital and Depression Scale

Patients having USE were generally slightly more anxious and reported more depression at baseline (see [Report Supplementary Material 1, Table 11](#)). There appeared to be some divergence between the arms in both anxiety and depression at 6 months in favour of USE-FNAC (see [Appendix 1, Figures 4 and 5](#)), and there was evidence of an interaction between test and assessment on the depression scale (see [Report Supplementary Material 1: Table 12](#)).

However, there was no evidence of an overall difference between USE-FNAC and US-FNAC in either the anxiety subscale [mean difference (MD): -0.18; 95% CI -0.43 to 0.07; $p = 0.16$] or the depression subscale (MD: -0.006; 95% CI -0.27 to 0.26; $p = 0.97$).

RESULTS

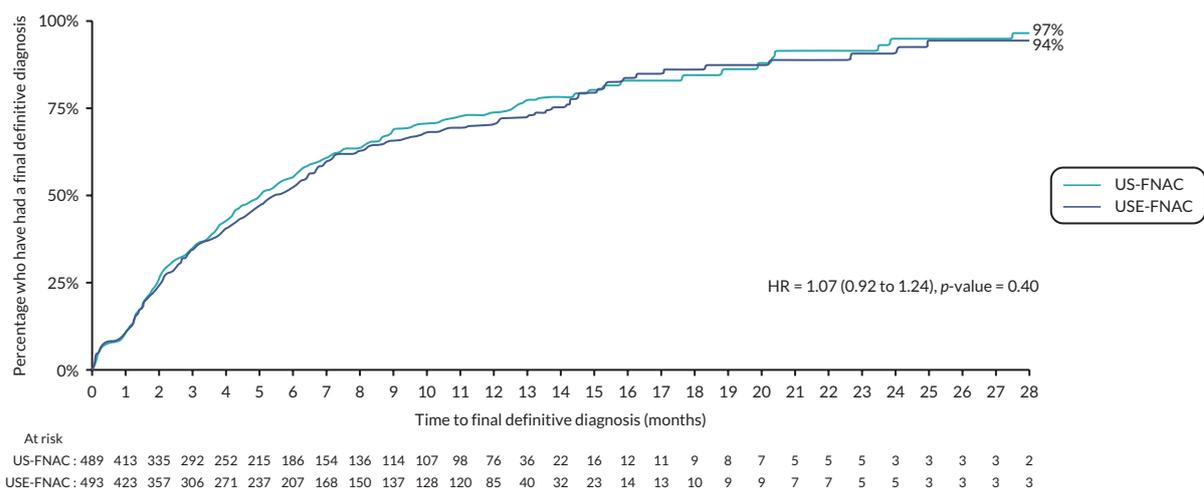


FIGURE 3 Time to FDD Kaplan–Meier Plot (unadjusted HR included in the figure).

EQ5D-5L

EQ5D-5L was included as a quality-of-life outcome measure predominantly for the health economic analysis. The statistical analyses showed no differences between the groups at any assessment point (see [Report Supplementary Material 1: Table 13](#)), nor when looking at overall assessments (see [Report Supplementary Material 1: Table 14](#)).

Visual Analogue Pain Scale

If greater accuracy was obtained by using USE-FNAC compared with US-FNAC, then this might have resulted in participants experiencing less pain from the procedure.

Slightly less pain was observed following the first FNAC in the USE-FNAC group (see [Table 11](#)), however, not significantly so. The direction of the pain score favours US-FNAC for the third and extra FNAC; however, numbers are smaller and the differences still remain insignificant.

TABLE 11 Mean VAPS scores over follow-up by test group

	USE-FNAC	US-FNAC	MD ^a (95% CI)	p-value
	N = 493	N = 484		
	N = 477	N = 475		
Immediately after the first FNAC	22.7 (22.1)	25.2 (25.2)	-2.6 (-5.6 to 0.5)	0.10
	N = 225	N = 199		
Immediately after the second FNAC	26.7 (24.7)	29.4 (26.0)	-2.7 (-7.5 to 2.2)	0.28
	N = 37	N = 25		
Immediately after the third FNAC	29.9 (22.3)	28.5 (27.4)	1.4 (-11.2 to 14.1)	0.82
	N = 9	N = 8		
Immediately after extra FNAC	24.6 (22.5)	20.6 (23.4)	3.9 (-19.8 to 27.7)	0.73

a Using two-sample t-test.

Note

Estimates are presented as means (SD) unless otherwise specified.

Accuracy outcomes

Accuracy of first FNAC results compared to second FNAC

For those with a FDD of malignant, there were only 13 participants with a first and second FNAC that was not a Thy 1 result (and therefore could be included in the analysis). Due to this, of those included in the analysis, all 13 have a malignant diagnosis on their second FNAC (see [Table 12](#)). The sensitivity was 0.54 following the first FNAC, compared with 1.00 for the second FNAC. Sensitivity analyses were performed where assumptions were placed on both Thy 1 results and those without a FDD. These can be found in [Report Supplementary Material 1: Table 15](#).

For those with a benign FDD, there is some evidence that the specificity is better on the first FNAC (specificity = 0.89) compared with the second FNAC (specificity = 0.84), $p = 0.03$ (see [Table 13](#)). The sensitivity analyses for benign nodules are presented in [Report Supplementary Material 1: Table 16](#).

Agreement between first FNAC and second FNAC

Agreement between the first FNAC and second FNAC was examined in those who had at least two FNACs, and in the first instance looking at only those who had a diagnostic result. The result of the first FNAC changed on the second FNAC in 72/330 (21.8%) cases. Importantly in 13% of the cases, it changed from benign to malignant, and in 8.8% it changed from malignant to benign (see [Table 14](#)). Secondary analyses of agreement between the first and second FNACs were performed (see [Report Supplementary Material 1: Table 17](#)).

Accuracy of USE/US without FNAC and accuracy of USE/US with FNAC

Accuracy of US-only compared with US-FNAC

The sensitivity of US-alone was better than US-FNAC (see [Table 15](#)). In malignant modules, US-alone would have missed seven cases (9% of malignant cases). On the other hand, US-FNAC would have missed 10 cases (13% malignant cases). Though this result was not statistically significant, the sensitivity analyses still remained in favour of US-alone in malignant nodules (see [Report Supplementary Material 1: Table 18](#)). However, again this was not statistically significant.

Conversely, when diagnosing benign nodules, US-FNAC had statistically higher specificity than US-alone (0.67 vs. 0.48 respectively, $p < 0.0001$) (see [Table 16](#)). Of the 61 that were malignant on US-alone, but benign on final diagnosis, diagnosis was reached in 43 by repeat FNACs, and in 18 after surgery.

TABLE 12 Accuracy of first FNAC compared with second FNAC in malignant nodules

	Second FNAC, malignant	Second FNAC, benign	Total	First FNAC sensitivity	Second FNAC sensitivity
First FNAC malignant	7	0	7	0.54 (95% CI 0.25 to 0.81)	1.00
First FNAC benign	6	0	6		
Total	13	0	13		

TABLE 13 Accuracy of first FNAC compared with second FNAC in benign nodules

	Second FNAC, malignant	Second FNAC, benign	Total	First FNAC specificity	First FNAC specificity
First FNAC malignant	17	10	27	0.89 (95% CI 0.85 to 0.93)	0.84 (95% CI 0.80 to 0.89)
First FNAC benign	22	198	220		
Total	39	208	247	$p = 0.03$	

TABLE 14 Agreement between first FNAC and second FNAC

	Second FNAC malignant	Second FNAC benign	Total
First FNAC malignant	60	29	89
First FNAC benign	43	198	241
Total	103	227	330

TABLE 15 Accuracy of US-alone compared with US-FNAC in malignant nodules

	US-FNAC malignant	US-FNAC benign	Total	US-alone sensitivity	US-FNAC sensitivity
US-alone malignant	63	7	70	0.91 (96% CI 0.85 to 0.97)	0.87 (95% CI 0.80 to 0.95)
US-alone benign	4	3	7		
Total	67	10	77	$p = 0.37^a$	

a p -value from McNemar's test.

TABLE 16 Accuracy of US-alone compared with US-FNAC in benign nodules

	US-FNAC malignant	US-FNAC benign	Total	US-alone specificity	US-FNAC specificity
US-alone malignant	61	61	122	0.48 (95% CI 0.40 to 0.52)	0.67 (95% CI 0.61 to 0.73)
US-alone benign	16	97	113		
Total	77	158	235	$p < 0.0001$	

Of the 16 that were benign on US-alone but malignant on final diagnosis, 15 were reached after surgery and one on repeat FNAC. The sensitivity analyses where assumptions were placed on Thy 1 results and missing FDD also supported the primary analysis (see [Report Supplementary Material 1: Table 19](#)).

However, if we exclude US U3 (indeterminate on US) cases, then the specificity of US-alone rises considerably, and indeed may be better than US-FNAC, though with no statistically significant difference, when diagnosing benign nodules.

In benign nodules, if we also exclude Thy 3 (Thy 3a, Thy 3f and Thy 3), as well as U3 indeterminate results, then the specificity of US-alone is very high (92%) and only slightly less than US-FNAC (0.99). This is borderline statistically significant at $p < 0.05$ level, but not significant at the more stringent 0.001 level when accounting for multiple analyses (see [Table 17](#)).

Accuracy of US-USE compared with USE-FNAC

In contrast to when comparing US-alone with US-FNAC, the sensitivity of USE-FNAC was better than US-USE-alone (0.92 vs. 0.75, $p = 0.008$) (see [Table 18](#)). Though this benefit reduces if Thy 1 results are assumed to be benign, it is consistent with the results where assumptions are placed on missing FDD (see [Report Supplementary Material 1: Table 19](#)). Similarly, when comparing US-alone to US-FNAC in benign nodules, USE-FNAC had better specificity than US-USE alone (see [Table 19](#), [Report Supplementary Material 1: Table 20](#)). Again, this was supported by the sensitivity analyses (see [Report Supplementary Material 1: Table 21](#)).

TABLE 17 Sensitivity analyses of accuracy of US-alone compared with US-FNAC in benign nodules excluding indeterminate US and FNAC results

	US-FNAC malignant	US-FNAC benign	Total	US-alone sensitivity	US-FNAC sensitivity
Excluding U3 results					
US-alone malignant	5	8	13	0.90	0.83
US-alone benign	16	97	113		
Total	21	105	126		
Excluding U3 and Thy 3a, Thy 3 and Thy 3f					
US-alone malignant	1	8	9	0.92	0.99
US-alone benign	0	97	97		
Total	1	105	106		

TABLE 18 Accuracy of US-USE alone compared with USE-FNAC in malignant nodules

	USE-FNAC malignant	USE-FNAC benign	Total	US-USE sensitivity	USE-FNAC sensitivity
US-SE-alone malignant	45	3	48	0.75 (95% CI 0.66 to 0.86)	0.92 (95% CI 0.86 to 0.99)
US-SE-alone benign	14	2	16		
Total	59	5	64		

TABLE 19 Accuracy of US-USE alone compared with USE-FNAC in benign nodules

	USE-FNAC malignant	USE-FNAC benign	Total	US-USE specificity	USE-FNAC specificity
US-SE-alone malignant	34	65	99	0.60 (95% CI 0.54 to 0.66)	0.73 (95% CI 0.68 to 0.79)
US-SE-alone benign	32	115	147		
Total	66	180	246		

Radiologists' survey

Radiologists felt able to biopsy 95% of nodules on the first FNAC (see [Table 20](#)). For 407/491 (83%) FNACs, radiologists did not feel elastography helped in identifying the nodule required for FNAC, and similarly, 76% felt it did not help identify the specific part within the nodule for FNAC. Only 36% found USE helpful over and above traditional US in determining malignancy. Survey results from any FNAC, whilst subject to bias, were in line with results seen from the first FNAC (see [Report Supplementary Material 1: Table 22](#)).

Pathology quality assurance review

Of the 486 cases identified for pathology review, 94% (458) cases were received. Of those 458 cases, 368 have been reviewed, constituting 85% of the original sample identified. Of those reviewed, there were 163 histopathology cases following surgery and 205 FNAs.

TABLE 20 Radiologists' feedback from first FNAC

		USE arm
		N = 491 (%)
Was the nodule suitable to be biopsied in your view?	Yes	466 (95)
	No	23 (5)
	Missing	2 (0.4)
If no, please select why:		
	<i>Pure cystic lesion</i>	1 (4)
	<i>No normal tissue surrounding it</i>	8 (35)
	<i>Pathology separate to thyroid</i>	0 (0)
	<i>Benign nodule</i>	3 (13)
	<i>Other</i>	11 (48)
	<i>Missing</i>	0 (0)
Did the elastography technique help you in identifying the nodule required for FNAC?	Yes	69 (14)
	No	407 (83)
	N/A	6 (1)
	Missing	9 (2)
Did the elastography technique help you identify the specific part within the nodule for FNAC?	Yes	111 (23)
	No	373 (76)
	N/A	2 (0.4)
	Missing	5 (1)
Did you find the RET to be helpful over and above the conventional US in determining malignancy?	Yes	178 (36)
	No	311 (63)
	Missing	2 (0.4)

Of the 205 FNA cases, eight showed some disagreement between the original hospital result and quality assurance review, but only two were significant: one changing from a benign Thy 2 to a Thy 4 (potentially malignant), and the other changing from a potentially malignant to atypical. The rest were minor changes that do not affect whether the cases are benign or malignant.

Of the 163 surgical histopathology cases: only one was changed from malignant to benign.

In that case the original diagnosis of malignancy was because there was an incidental 3 mm papillary thyroid carcinoma within the sample, which would not have been picked up on US. The nodules that were picked up on US were all benign. So whilst the diagnosis by the local team is NOT incorrect, this case should be considered as a benign case for the purposes of the study.

These results show that the histological and cytological reporting of samples within ElATION were very accurate and show less than 1% change in results on review.

Surgery complications and serious adverse events

Surgery complications

Out of 982 participants in the trial, 379 received surgery. Complications were reported at 30 days and 6 months post surgery. There were two deaths incorrectly reported on the surgery complication forms: one participant died due to cancer; the other due to a bowel obstruction. Overall, the surgical complications were very low (see [Table 21](#)). Further details on the complications can be found in [Report Supplementary Material 1: Table 23](#).

TABLE 21 Complication rates from thyroidectomy at 30 days and 6 months post surgery

		Surgery N = 379 (%)
30-day and/or 6-month post operative form		
Bleeding requiring return to theatre	Yes	2 (1)
	No	315 (83)
	Missing	62 (16)
Haematoma not requiring evacuation	Yes	1 (0.3)
	No	316 (83)
	Missing	62 (16)
Wound infection	Yes	8 (2)
	No	309 (82)
	Missing	62 (16)
Hypocalcaemia requiring replacement	Yes	15 (4)
	No	317 (84)
	Missing	47 (12)
Vocal cord palsy ^a	Yes	2 (1)
	No	295 (78)
	Missing	82 (22)
Keloid scarring ^a	Yes	3 (1)
	No	294 (78)
	Missing	82 (22)
Death	Yes	2 ^b (1)
	No	330 (87)
	Missing	47 (12)

a Only reported on 6-month post operative form.

b Incorrectly reported as surgical complications. One participant died due to cancer; and one participant died due to a bowel obstruction.

Serious adverse events

As anticipated, the number of SAEs was very low. Thirteen out of the 982 participants (1%) experienced an SAE (see [Report Supplementary Material 1: Table 24](#)). There were marginally more SAEs in the USE-FNAC group than the US-FNAC group (10 in 9 participants vs. 4 in 4 participants respectively), though not significantly so ($p = 0.26$). Details on the SAEs by group can be found in [Report Supplementary Material 1: Table 25](#).

Chapter 4 Discussion

Summary of findings

Elastography does not appear to add any benefit to the use of US alone when assessing thyroid nodules. In particular, USE does not reduce the Thy 1 (non-diagnostic rate) of US fine-needle aspiration. In addition, it does not result in a reduction in the number of FNACs needed to reach a definitive diagnosis, nor in the time to reach that diagnosis. There was a reduction in the number of thyroidectomies performed in the USE arm compared to the US-only arm (37% vs. 40%), but this was not statistically significant. In addition, there was no reduction in the number of unnecessary diagnostic operations, that is, operations resulting in a benign diagnosis. Furthermore, there was no reduction in the number of complications or SAEs seen in either arm.

There may be a slight improvement in terms of depression and anxiety with SE, but the overall levels were the same across both arms over time, and there was no difference in pain or overall quality of life.

About one-third of radiologists found USE useful overall, and only 14% found USE useful in identifying the nodule to biopsy. This is supported by the fact that USE seemed to assess more malignant nodules as benign than US-alone.

Comparison of demographics with other studies

There have been two large recent meta-analyses. The meta-analysis by Zhou *et al.*²⁶ did not report on the demographics of the patients included. However, in the meta-analysis by Cantisani *et al.*²⁰ which included both RCTs and cohorts, the proportion of male patients was 24%, compared to 20% in this study. The mean age of the patients in that meta-analysis was 46 years, and in this study it was 51 years. No other baseline characteristics were reported.

The Cantisani *et al.* meta-analysis reported a high malignancy rate of 33% which suggests bias in patient selection by the 72 studies included in the meta-analysis. The rates of malignancy were not expressly reported in the Zhou *et al.* study, but Figure 9 shows a very high event rate especially in the experimental arm.

Comparison of trial design with previous research

There are 11 randomised control trials on USE versus US in thyroid nodules reported in the literature. This study is by far the largest study, recruiting five times more patients than the largest RCT previously reported. Indeed, this trial is larger than the total number of patients recruited to all 11 trials reported in the meta-analysis by Zhou *et al.* in 2021 put together.

This study found no benefit to the use of USE over the use of US alone when used in conjunction with FNA. This is in contrast to many studies published before. However, most previous studies have reported on sensitivity and specificity in a cohort without comparing it to a control arm with US-only. Therefore most of the published studies could not report on the additional benefit of using USE over using US alone.

The Zhou *et al.*²⁶ meta-analysis, however, concentrated on RCT with experimental and control arms. That meta-analysis did report a benefit to using USE with an OR of 14.67 compared to control in establishing

a histological diagnosis (defined as FNA or Histology). However, Zhou *et al.* themselves note that the 11 studies, included, were all small in size, ranging from 52 to 265 subjects. The meta-analysis did not report the overall malignancy rate of subjects included in the 11 studies, but Figure 9 shows a very high event rate – especially in the experimental arm – which suggests strong selection bias. This study of ours was a pragmatic study of all comers, and therefore less likely to suffer from selection bias, and likely to be more representative of the actual population in routine clinical practice.

The results of trials previously reported could also reflect that they may have been performed by early adopters and USE enthusiasts, who have a lot of experience with SE. The radiologists in this study reflected the general population of hospital radiologists with a wide range of experience of USE from long experiences to almost none before the trial. This could have affected our results; however, we performed sensitivity analyses by throughput – and whilst the rates of Thy 1 reduce with increased throughput, there was still no difference between USE and US alone.

Equality, diversity and inclusion

The majority of patients involved in the trial were women since thyroid nodules are more common in women. There was no data collection on ethnicity or socioeconomic status.²⁷⁻²⁹

Patient and public involvement

Patients were involved in the design of the study from the very beginning and in oversight of the study throughout its lifetime. The results of the trial have been shared with the patient and public advisers.

Adherence

Participants showed very high adherence to the test as per randomisation this was especially apparent in the first two assessments with over 95% adherence. High adherence was found in the third FNAC as well but less so for USE (85%) than US (98%). Furthermore, FNACs showed a drop in adherence to 70% for USE and 85% for US scans.

Strengths of the trial

This study is the largest randomised trial to report on USE in thyroid nodules. It is almost five times bigger than the next largest RCT published in the literature, and indeed this study contained more than the total combined number of patients recruited for previous RCTs.²⁶ Importantly, it was sufficiently powered to identify differences in non-diagnostic rates (Thy 1).

In addition, this study had a pragmatic real-world design that allowed all types of centres to participate reflecting normal NHS practice and allowed both types of USE (strain and shear wave). Notably the Thy1 rates observed in the USE (19%) and US-only (16%) arms were higher than the assumed 4% and 10%. These findings reflect this study's pragmatic multicentre UK trial with real-world data representative of patients seen in routine UK practice.

Since USE was a new diagnostic technique and only a few centres had experience in it, this study implemented a strong training and certification programme. Radiologists were required to submit an

audit of their last 20 FNACs and the number of US FNACs that they had undertaken in the last year to ensure that they were experienced in US FNAC of the thyroid. They then had to attend a workshop and following that had to provide a log book of USE use on 15 patients prospectively, before being able to participate in the trial. They then performed one 'hot case' accreditation and once approved started participating in the study. They then provided their first five cases for assessment.

Finally, sensitivity analyses of the results by surrogate measures were undertaken for experience (throughput) and quality of radiologists undertaking SE. In addition, there was a quality assurance programme for pathology results.

The final strength of the study was that radiologists were also asked for their opinions as to whether they found USE useful or not, therefore the study included qualitative insights from the radiologists as well as quantitative efficacy measures.

Possible weaknesses

The main weakness of the study was that it was not powered to identify differences in the numbers of benign and malignant nodules diagnosed between the two arms. However, the rates identified were very similar between the two arms, and indeed were better for US alone. Accuracy rates were also very similar. Therefore, it is highly unlikely that there would be an undetected statistically significant difference between the two arms in terms of identifying malignancy.

A total of 688 of the 982 cases (70%) were recruited by only seven of the 23 centres (see [Table 22](#)). These seven centres recruited 50 or more cases each. Centres that recruited large numbers of patients may have better results. However, a sensitivity analysis by throughput to assess this was undertaken, and this did not change the results.

Challenges of the trial

One of the main challenges of the trial, which also resulted in a weakness, is that radiology is a research-naïve specialty in the UK and therefore obtaining research support was difficult and delayed set-up and recruitment. It also affected the ability to obtain scans and pathology slides for quality assurance in a timely fashion, and therefore the study could not undertake the assessment of quality prospectively in real time. The pathology review also experienced several other unexpected delays. One such delay was due to changes of staff at the original reviewing centre requiring the appointment of new consultant pathologists at a different Trust in order to perform the review; a further delay to the commencement of the review was due to the introduction of General Data Protection Regulation (GDPR) in 2018 which prohibited the transport of pseudoanonymised patient material outside of the Trials Unit based on the robustness of the consent forms used for the trial, prompting a significant revision of the review process. A final delay was caused by the COVID-19 pandemic in 2020, which resulted in a short-term pause of the collection of pathology material, and further long-term delays due to hospital staffing changes and impacts on local pathology departments. Recruitment was further delayed because of difficulties in obtaining funding for USE in some centres.

Recommendations for future research

The findings of the ElATION trial suggest that further research into the use of SWE in the diagnostic setting of thyroid nodules is unlikely to be warranted unless there is a change in technology. The clinical problems that the trial set out to address of non-diagnostic needle aspirations from thyroid nodules, and the difficulty in diagnosing benign from malignant lesions, still persist. Future studies might examine

TABLE 22 List of participant centres, principal investigators and ElATION radiologists

Hospital	NHS trust	Principal investigator	Radiologist/s ^a	Number recruited
Lister Hospital	East and North Herts NHS Trust	Mr George Mochloulis	Dr Kanchana Rajaguru	224
Charing Cross Hospital Hammersmith Hospital St Mary's Hospital	Imperial College Healthcare NHS Trust	Dr Gitta Madani	Dr Gitta Madani; Dr Kunwar Bhatia	210
Kings College Hospital	King's College Hospital NHS Foundation Trust	Prof Paul Sidhu	Prof Paul Sidhu; Dr Annamaria Deganello	88
Leicester Royal Infirmary	University Hospitals of Leicester NHS Trust	Dr Ram Vaidhyanath	Dr Ram Vaidhyanath; Ms Amy Barnes	60
Queen Alexandra Hospital	Portsmouth Hospitals NHS Trust	Dr Jasper Bekker	Dr Jasper Bekker; Dr Janine Domjan; Dr Daren Gibson; Dr Chris Bowles	58
Basildon University Hospital	Basildon and Thurrock University Hospitals NHS Foundation Trust	Dr Thaj Rehman	Dr Thaj Rehman	55
University Hospital of North Tees	North Tees and Hartlepool Hospitals NHS Foundation Trust	Dr Arun Batra	Dr Arun Batra	53
North Manchester General Hospital	Pennine Acute Hospitals NHS Trust	Mr Sharan Jayaram	Dr Niranjana Desai	49
Northwick Park Hospital St Mark's Hospital	London North West University Healthcare NHS Trust	Dr Ravi Lingam	Dr Ravi Lingam	41
Ipswich Hospital	The Ipswich Hospitals NHS Trust	Dr James Hathorn	Dr James Hathorn	32
Queen Elizabeth Hospital Birmingham	University Hospitals of Birmingham NHS Foundation Trust	Prof Hisham Mehanna	Dr Steve Colley	24
Southend University Hospital	Southend University Hospital NHS Foundation Trust	Dr Mohammad Aslam	Dr Mohammad Aslam	23
Royal Berkshire Hospital	Royal Berkshire NHS Foundation Trust	Dr Farhan Ahmad	Dr Farhan Ahmad	18
Luton and Dunstable University Hospital	Luton and Dunstable University Hospitals NHS Foundation Trust	Dr Thayahlan Iyngkaran	Dr Thayahlan Iyngkaran	16
Norfolk and Norwich University Hospital	Norfolk and Norwich University Hospitals NHS Foundation Trust	Dr Davina Pawaroo	Dr Davina Pawaroo	14
James Paget Hospital	James Paget University Hospitals NHS Foundation Trust	Mr Carl Philpott	Dr Nabil Mahmood	10
Ashford Hospital St Peter's Hospital	Ashford and St Peter's Hospitals NHS Foundation Trust	Dr Vineet Prakash	Dr Vineet Prakash; Dr Oliver Wignall; Ms Gurjit Rai-Tidbury	4
Castle Hill Hospital	Hull and East Yorkshire Hospitals NHS Trust	Mr James England	Ms Jean Bainbridge	3

^a Includes radiographers and sonographers.

the role of genomic testing on FNA samples. Internationally there is growing use of targeted panels of molecular markers, particularly aimed at improving the diagnostic accuracy of indeterminate (i.e. Thy 3) cytology results. Improved certainty of the benign nature of a thyroid nodule would allow patients to avoid the risks associated with such unnecessary surgery. The application of these tests is not uniform, and their cost effectiveness has not been assessed in large-scale trials. Furthermore, improvements in scoring systems to raise the inter-rater reliability between radiologists may improve radiological accuracy in the setting of thyroid nodule grading.

Conclusion

In conclusion, USE does not appear to add benefit to US in the assessment of thyroid nodules and the diagnosis of thyroid malignancy in a multicentre setting in the NHS.

Additional information

Contributions of authors

Hisham Mehanna (<https://orcid.org/0000-0002-5544-6224>) (Chair of Head and Neck Surgery; Director, Institute of Global Innovation; Director, Institute of Head and Neck Studies and Education, University of Birmingham), Chief Investigator.

Jonathan J Deeks (<https://orcid.org/0000-0002-8850-1971>) (Professor of Biostatistics, Institute of Applied Health Research, University of Birmingham), Trial Design and Statistics.

Kristien Boeleart (<https://orcid.org/0000-0002-6435-177X>) (Professor of Endocrinology, University of Birmingham), Endocrinology.

Gitta Madani (<https://orcid.org/0000-0001-8109-419X>) (Consultant Radiologist, Imperial College Healthcare NHS Trust), Radiology.

Paul Sidhu (<https://orcid.org/0000-0003-1928-4077>) (Professor of Imagine Sciences, King's College London; Consultant Radiologist, King's College Hospital NHS Foundation Trust), Radiology.

Paul Nankivell (<https://orcid.org/0000-0003-4664-6117>) (Senior Clinical Lecturer, Institute of Head and Neck Studies and Education, Institute of Cancer and Genomic Sciences, University of Birmingham; Consultant ENT/Head and Neck Surgeon, University Hospital Birmingham), Surgery.

Neil Sharma (<https://orcid.org/0000-0001-6776-0977>) (Clinical Scientist, University of Birmingham; Consultant Head and Neck Surgeon, University Hospital Birmingham, Translational Research.

Rebecca Woolley (<https://orcid.org/0000-0001-5119-1431>) (Senior Medical Statistician, Birmingham Clinical Trials Unit, University of Birmingham), Statistics.

Judith Taylor (<https://orcid.org/0000-0001-8024-271X>) (Co-director and Secretary, International Thyroid Cancer Alliance; Chair, The Thyroid Trust, UK), Patient and Public Involvement.

Tessa Fulton-Lieuw (<https://orcid.org/0000-0002-9857-1834>) (Team Leader, Institute of Head and Neck Studies and Education, University of Birmingham), Trial Management.

Andrew Palmer (<https://orcid.org/0000-0003-0837-4624>) (Senior Trial Manager, Birmingham Clinical Trials Unit, University of Birmingham), Trial Management.

Acknowledgements

See [Table 22](#) for a list of participating centres and radiologists.

Patient and public representatives

Judith Taylor (Thyroid Cancer Alliance)

Oversight Committees

Trial Steering Committee

Chair: Professor James McCaul (Queen Elizabeth Hospital Glasgow)

Mr Joel Smith (Royal Devon and Exeter NHS Foundation Trust)

Dr Sriram Vaidhyanathan (Leeds Teaching Hospitals NHS Trust)

Dr Liz Hensor (University of Leeds)

Mrs Fiona Riley (British Thyroid Foundation)

Data Monitoring Committee

Chair: Dr Wai Lup Wong (Mount Vernon Cancer Centre)

Dr Laura Moss (Velindre Hospital)

Dr Gillian Lancaster (Lancaster University)

Contributions of authors

Lister Hospital, Stevenage – G. Mochloulis, K. Rajaguru, P. Baker, C. Cruz; *Imperial Hospitals, London* – G. Madani, K. Bhatia, T. Stoycheva, L. Honeyfield, M. Mazzola; *King's College Hospital, London* – P. Sidhu, A. Deganello, D. Quinlan, A. Rahman, A. Chhabra; *Queen Alexandra Hospital, Portsmouth* – J. Bekker, J. Domjan, D. Gibson, C. Bowles, D. Barnes, B. Holding; *Leicester Royal Infirmary* – R. Vaidhyanath, A. Barnes; *Basildon Hospital* – T. Rehman, K. Goodsell; *University Hospital of North Tees* – A. Batra, L. Poole; *North Manchester General Hospital* – S. Jayaram, N. Desai, J. Rothwell, J. Allsop; *Northwick Park Hospital/St Mark's Hospital, London* – R. Lingam, T. Adedoyin, R. Baldwin; *Ashford and St Peter's Hospital, Chertsey* – V. Prakash, O. Wignall, G. Rai-Tidbury, V. Frost, C. Gray; *Ipswich Hospital* – J. Hathorn, P. Ridley, A. Williams, J. Wells; *Queen Elizabeth Hospital Birmingham* – H. Mehanna, S. Colley, J. Jones, A. Pease; *Southend Hospital* – M. Aslam, E. Mphansi; *Luton and Dunstable Hospital* – T. Iyngkaran, A. Rafiq; *Royal Berkshire Hospital* – F. Ahmad, S. MacGill, R. Carson; *Norfolk and Norwich University Hospital* – D. Pawaroo, D. Archer, A. Dann; *James Paget Hospital, Great Yarmouth* – C. Philpott, N. Mahmood, J. Woods; *Castle Hill Hospital, Hull* – J. England, J. Bainbridge, C. Abernethy.

Equality, diversity, and inclusion

Patients were identified for ElaTION from a number of NHS Trusts across England. Gender and age were recorded at randomisation but no other factors such as ethnicity, religion or political/sexual orientation were collected. ElaTION was designed in 2014 with binary gender classifications of 'Male or Female'. This was representative of the time and would be treated with more sensitivity if the trial was set up today. ElaTION was not set up to use non-English versions of licensed patient completed documents (EQ5D-3L and HADS). This could have caused a potential barrier to entry for groups of potential participants, particularly at sites in larger metropolitan areas. A consideration for future research would be to liaise with potential sites identified in trial set-up to determine the most commonly-spoken non-English languages within the patient population, and to seek relevant translated licences prior to trial set-up. ElaTION was designed with, and managed with regular input from Patient and Public Involvement Representative Judith Taylor. Judith Taylor, Co-Director of the Thyroid Cancer Alliance and Chair of the Thyroid Trust, was integral in helping us design patient-facing documents such as the PIS, invitation letters and posters, and reviewing the draft final report.

Patient data statement

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety and plan NHS services. Patient data are kept safe and secure, to protect everyone's privacy, and we implement safeguards to make sure that it is stored and used responsibly.

Data-sharing statement

All data requests should be submitted to the corresponding author for consideration. Access to anonymised data may be granted following review.

Ethics statement

ElaTION received full ethical approval from the South Central – Berkshire Research Ethics Committee on 10 October 2014. The REC reference is 14/SC/1206.

Information governance statement

The University of Birmingham is committed to handling all personal information in line with the UK Data Protection Act (2018) and the General Data Protection Regulation (EU GDPR) 2016/679. Under the Data Protection legislation, The University of Birmingham is the Data Controller, and you can find out more about how we handle personal data, including how to exercise your individual rights and the contact details for our Data Protection Officer here (<https://www.birmingham.ac.uk/privacy>).

Disclosure of interests of authors

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at <https://doi.org/10.3310/PLEQ4874>.

Primary conflicts of interest: Hisham Mehanna reports membership of HTA Commissioning Trials Committee 2013–7; honoraria from AstraZeneca; travel support from MSD, and speaker's bureau from MSD and Merck. Hisham Mehanna also reports Advisory Board membership for several organisations, and DMC membership for the MACRO study. Hisham Mehanna is also Chairman of the Head and Neck Cancer InterGroup, and reports financial interests from Warwickshire Head Neck Clinic Ltd. Paul Sidhu reports book royalties from Thieme, Hodder, Elsevier and Springer for eight books; consulting fees from Bracco and Samsung; payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events for Bracco, Philips, Samsung and Siemens; Editor-in-Chief, *Ultrasound in Medicine and Biology*; and receipt of machines from Philips, and Samsung Ultrasound for his institution. Paul Nankivell reports participation on a Data Safety Monitoring Board or Advisory Board for the SAVER and TALGiTs trials. Judith Taylor reports grants or contracts from Elli Lilly, and Bayer, paid to Thyroid Cancer Alliance; consulting fees from Workgroup of European Cancer Advocacy Networks, Bayer, and Eisai made to Thyroid Cancer Alliance, and fees from Merck made to The Thyroid Trust; review of materials for Patients in Publications training course for WECAN, review of graphical aids for Bayer, lunchtime lectures for Eisai with fees donated to Thyroid Cancer Alliance, member of patient advisory board for Merck with fees donated to The Thyroid Trust; personal payment for attending the following meetings: American Society for Clinical Oncology, European Society for Medical Oncology,

ADDITIONAL INFORMATION

reimbursement for travel to ERN-EURACAN meetings, reimbursement for attendance at European Thyroid Association meetings, attendance fee and travel reimbursements to attend from Tests to Targeted Treatments board meetings, and Cancer Drug Development Forum patient advocacy travel and accommodation grant to attend conference; Judith Taylor is an unpaid Chair of The Thyroid Trust, unpaid de facto treasurer of Thyroid Cancer Alliance, and has received payment to attend ERN-EURACAN annual meeting as patient advocacy representative.

References

1. Mehanna M. *ElaTION Trial Protocol. Version 5.0*. URL: <https://njl-admin.nihr.ac.uk/document/download/2011879> (accessed April 2023).
2. Deandrea M, Mormile A, Veglio M, Motta M, Pellerito R, Gallone G, *et al*. Fine-needle aspiration biopsy of the thyroid: comparison between thyroid palpation and ultrasonography. *Endocr Pract* 2002;**8**(4):282–6. <https://doi.org/10.4158/EP.8.4.282>
3. Hegedüs L. Clinical practice. The thyroid nodule. *N Engl J Med* 2004;**351**(18):1764–71. <https://doi.org/10.1056/NEJMcp031436>
4. Perros P, Boelaert K, Colley S, Evans C, Evans RM, Gerrard G, *et al*. Guidelines on the management of thyroid cancer. *Clin J Soc Endocrinol* 2014;**81**(1):1–122. <https://doi.org/10.1111/cen.12515>
5. Cooper D, Doherty G, Haugen B, Kloos R, Lee S, Mandel S, *et al*. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2009;**19**(11):1167–214. <https://doi.org/10.1089/thy.2009.0110>
6. Rad MP, Zakavi SR, Layegh P, Khooei A, Bahadori A. Incidental thyroid abnormalities on carotid color Doppler ultrasound: frequency and clinical significance. *J Med Ultrasound* 2015;**23**(1):25–8. <https://doi.org/10.1016/j.jmu.2014.04.005>
7. Ezzat S, Sarti DA, Cain DR, Braunstein GD. Thyroid incidentalomas: prevalence by palpation and ultrasonography. *Arch Intern Med* 1994;**154**(6):1838–40. <https://doi.org/10.1001/archinte.154.16.1838>
8. Ahmed S, Horton KM, Jeffrey RB Jr., Sheth S, Fishman EK. Incidental thyroid nodules on chest CT: review of the literature and management suggestions. *Am J Roentgenol* 2012;**195**(5):1066–71. <https://doi.org/10.2214/AJR.10.4506>
9. Yousem DM, Huang T, Loevner LA, Langlotz CP. Clinical and economic impact of incidental thyroid lesions found with CT and MR. *AJNR Am J Neuroradiol* 1997;**18**(8):1423–28. URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/pmc8338147/> (accessed April 2023).
10. Soelberg KK, Bonnema SJ, Brix TH, Hegedüs L. Risk of malignancy in thyroid incidentalomas detected by 18F-fluorodeoxyglucose positron emission tomography: a systematic review (2012). *Thyroid Off J Am Thyroid Assoc* 2015;**22**(9):918–25. <https://doi.org/10.1089/thy.2012.0005>
11. Colley S, Olliff J, Shimal A, Page A. Rapid responses: radiological and cytological burden. *BMJ* 2009;**338**:b733.
12. Al Maqbali T, Tedla M, Dekker A, Weickert MO, Mehanna H. Malignancy Risk Analysis in patients with inadequate fine-needle aspiration cytology (FNAC) of the thyroid. *PLOS ONE* 2012;**7**(11):e49078. <https://doi.org/10.1371/journal.pone.0049078>
13. Bastin S, Bolland MJ, Croxson MS. Role of ultrasound in the assessment of nodular thyroid disease. *J Med Imaging Radiat Oncol* 2009;**53**(2):177–87. <https://doi.org/10.1111/j.1754-9485.2009.02060.x>
14. Cesur M, Corapcioglu D, Bulut S, Gursoy A, Yilmaz AE, Erdogan N, Kamel N. Comparison of palpation-guided fine-needle aspiration biopsy to ultrasound-guided fine-needle aspiration biopsy in the evaluation of thyroid nodules. *Thyroid* 2006;**16**(6):555–61. <https://doi.org/10.1089/thy.2006.16.555>
15. Watkins L, O'Neill G, Young D, McArthur C. Comparison of British Thyroid Association, American College of Radiology TIRADS and Artificial Intelligence TIRADS with histological

- correlation: diagnostic performance for predicting thyroid malignancy and unnecessary fine-needle aspiration rate. *Br J Radiol* 2021;**94**(1123):20201444. <https://doi.org/10.1259/bjr.20201444>
16. Lyschchik A, Higashi T, Asato R, Tanaka S, Ito J, Mai JJ, *et al.* Thyroid gland tumor diagnosis at US elastography. *Radiology* 2005;**237**(1):202–11. <https://doi.org/10.1148/radiol.2363041248>
 17. Rago T, Scutari M, Santini F, Loiacono V, Piaggi P, Di Coscio G, *et al.* Real-time elastosonography: useful tool for refining the presurgical diagnosis in thyroid nodules with indeterminate or nondiagnostic cytology. *J Clin Endocrinol Metab* 2010;**95**(12):5274–80. <https://doi.org/10.1210/jc.2010-0901>
 18. Bojunga J, Herrman E, Meyer G, Weber S, Zeuzem S, Freidrich-Rust M. Real-time elastography for the differentiation of benign and malignant thyroid nodules: a meta-analysis. *Thyroid* 2010;**20**(10):1145–50. <https://doi.org/10.1089/thy.2010.0079>
 19. Kim JK, Baek JH, Lee JH, Kim JL, Ha EJ, Kim W, *et al.* Ultrasound elastography for thyroid nodules: a reliable study? *Ultrasound Med Biol* 2012;**38**(9):1508–13. <https://doi.org/10.1016/j.ultrasmedbio.2012.05.017>
 20. Cantisani V, De Silvestri A, Scotti V, Fresilli D, Tarsitano M, Polti G, *et al.* US-elastography with different techniques for thyroid nodule characterization: systematic review and meta-analysis. *Front Oncol* 2022;**12**:845549. <https://doi.org/10.3389/fonc.2022.845549>
 21. Qiu Y, Xing Z, Liu J, Peng Y, Zhu J, Su A. Diagnostic reliability of elastography in thyroid nodules reported as indeterminate at prior fine-needle aspiration cytology (FNAC): a systematic review and Bayesian meta-analysis. *Eur Radiol* 2020;**30**(12):6624–34. <https://doi.org/10.1007/s00330-020-07023-0>
 22. Zhu L, Chen Y, Ai H, Gong W, Zhou B, Xu Y, *et al.* Combining real-time elastography with fine-needle aspiration biopsy to identify malignant thyroid nodules. *J Int Med Res* 2020;**48**(12):97602. <https://doi.org/10.1177/0300060520976027>
 23. Davies G, Koenen M. Acoustic radiation force impulse elastography in distinguishing hepatic haemangiomas from metastases: preliminary observations. *Br J Radiol* 2011;**84**(1006):939–43. <https://doi.org/10.1259/bjr/97637841>
 24. Swan KZ, Nielsen VE, Bo MB, Bonnema SJ. Is the reproducibility of shear wave elastography of thyroid nodules high enough for clinical use? A methodological study. *Clin Endocrinol (Oxf)* 2017;**86**(4):606–13. <https://doi.org/10.1111/cen.13295>
 25. Lang EV, Berbaum KS, Lutgendorf SK. Large-core breast biopsy: abnormal salivary cortisol profiles associated with uncertainty of diagnosis. *Radiology* 2009;**250**(3):631–7. <https://doi.org/10.1148/radiol.2503081087>
 26. Zhou Y, Chen H, Qiang J, Wang D. Systematic review and meta-analysis of ultrasonic elastography in the diagnosis of benign and malignant thyroid nodules. *Gland Surg* 2021;**10**(9):2734–44. <https://doi.org/10.21037/gs-21-492>
 27. Alexander EK, Doherty GM, Barletta JA. Management of thyroid nodules. *Lancet Diabetes Endocrinol* 2022;**10**(7):540–8. [https://doi.org/10.1016/S2213-8587\(22\)00139-5](https://doi.org/10.1016/S2213-8587(22)00139-5)
 28. Chehade JM, Silverberg AB, Kim J, Case C, Mooradian AD. Role of repeated fine-needle aspiration of thyroid nodules with benign cytologic features. *Endocr Pract* 2001;**7**(4):237–43. <https://doi.org/10.4158/EP.7.4.237>
 29. Kobaly K, Kim CS, Mandel SJ. Contemporary management of thyroid nodules. *Annu Rev Med* 2022;**73**:517–28. <https://doi.org/10.1146/annurev-med-042220-015032>

Appendix 1 Figures

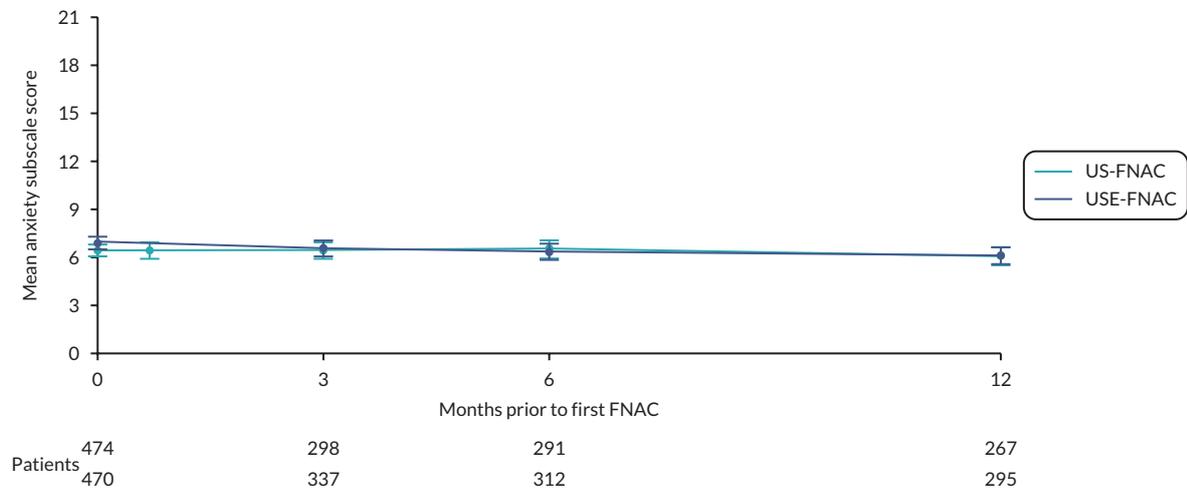


FIGURE 4 Mean HADS anxiety subscale score over time by test.

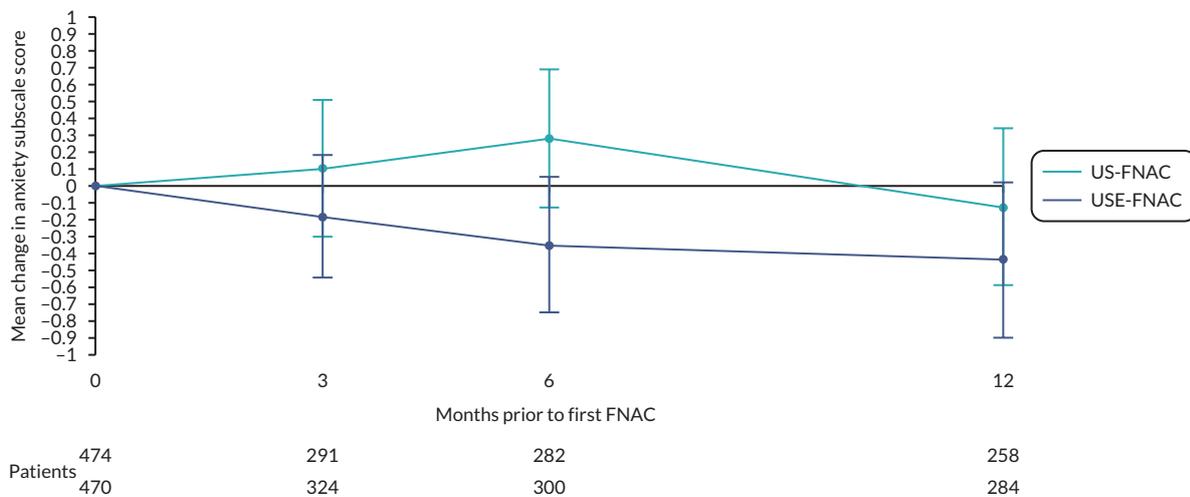


FIGURE 5 Mean change in HADS anxiety subscale score over time by test.

EME
HSDR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

*This report presents independent research funded by the National Institute for Health and Care Research (NIHR).
The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the
Department of Health and Social Care*

Published by the NIHR Journals Library