This is a repository copy of *Interobserver variation in the diagnosis of fibroepithelial lesions of the breast: a multicentre audit by digital pathology*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/127677/

Version: Accepted Version

Title: *Interobserver variation in the diagnosis of fibroepithelial lesions of the breast: A multicentre audit by digital pathology.*

Authors
Dessauvagie, BF [1,2,3,4]
Lee, AHS[5]
Meehan, K[4]
Nijhawan, A[1]
Tan, PH[6]
Thomas, J[7]
Tie, B[3]
Treanor, D[1,2,8]
Umar, S[1]
Hanby, AM[1,2]
Millican-Slater, R[1]

Affiliations
1. Department of Cellular Pathology, St James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, UK.
2. Leeds institute of Cancer and Pathology, Leeds University, Cancer Genetics Building, St James's University Hospital, Leeds, UK.
3. Department of Anatomical Pathology, PathWest Laboratory Medicine, Fiona Stanley Hospital, Murdoch, WA, Australia.
4. UWA Medical School and School of Biomedical Science, University of Western Australia, Perth, Australia.
5. Department of Histopathology, Nottingham University Hospitals NHS Trust, Nottingham, UK.
6. Division of Pathology, Singapore General Hospital, Singapore.
7. Department of Pathology, Western General Hospital, Edinburgh, UK.
8. Linköping University, Linköping, Sweden.

Corresponding Author
Dessauvagie, BF
Department of Anatomical Pathology, PathWest Laboratory Medicine, Fiona Stanley Hospital, Murdoch, WA, Australia.
(w) +61 8 6152 8484
(m) +61 402 344 578
(e) ben.dessauvagie@health.wa.edu.au

ABSTRACT

Aim: Fibroepithelial lesions (FELs) of the breast span a morphological continuum including lesions where distinction between cellular fibroadenoma (FA) and benign phyllodes tumour (PT) is difficult. The distinction is clinically important with FAs managed conservatively while equivocal lesions and PTs are managed with surgery. We sought to audit core biopsy diagnoses of equivocal FELs by digital pathology and to investigate whether digital point counting is useful in clarifying FEL diagnoses.

Method. Scanned slide images from cores and subsequent excisions of 69 equivocal FELs were examined in a multicentre audit by eight pathologists to determine the agreement and accuracy of core needle biopsy diagnoses (CNB) and by digital point counting of stromal cellularity and expansion to determine if classification could be improved.

Results. Interobserver variation was high on CNB with a unanimous diagnosis from all pathologists in only 8 cases of FA, diagnoses of both FA and PT on the same CNB in 15 and a 'weak' mean kappa agreement between pathologists (k=0.36). 'Moderate' agreement was observed on CNBs amongst breast specialists (k=0.44) and on excision samples (k=0.49). Up to 23% of lesions confidently diagnosed as FA on CNB were PT on excision and up to 30% of lesions confidently diagnosed as PT on CNB were FA on excision. Digital point counting did not aid in the classification of FELs.

Conclusion. Accurate and reproducible diagnosis of equivocal FELs is difficult, particularly on CNB, resulting in poor interobserver agreement and suboptimal accuracy. Given the diagnostic difficulty, and surgical implications, equivocal FELs should be reported in consultation with experienced breast pathologists as a small number of benign FAs can be selected out from equivocal lesions.

Keywords: Breast pathology, fibroepithelial lesions, fibroadenoma, phyllodes tumour, quality assurance, digital pathology

INTRODUCTION

Fibroepithelial lesions (FELs) of the breast are biphasic neoplasms of epithelium and stroma which exist on a morphologic continuum from clearly benign fibroadenomas (FA) to clearly malignant phyllodes tumours (PT). At the extremes of this spectrum, diagnoses are straightforward. However in the middle are equivocal FELs where the distinction between cellular FA, and benign PT is difficult (Figure 1).

Distinction of cellular FA and benign PT in this grey zone is subjective requiring assessment of features including stromal cellularity, stromal overgrowth, stromal cell atypia, stromal mitotic count, leaf-like architecture and tumour circumscription, some of which are not clearly defined. There is little to be gained from immunohistochemistry or molecular studies as these equivocal lesions exhibit overlapping staining profiles and genetic alterations.[1] The difficulty is greater in core needle biopsies (CNB) where sampling of these lesions, known to show tumour heterogeneity, is limited. Studies consistently show poor interobserver agreement in the classification of these lesions in surgical specimens[2 3] and discrepancies between CNB and final excision diagnoses.[4-6]

The distinction between FA and benign PT on CNB is clinically important. In the United Kingdom a CNB diagnosis of FA is given a 'B2' benign category, and managed conservatively often with observation alone.[7] By contrast, unless overtly malignant, a CNB diagnosis of PT is designated 'B3' (lesion of uncertain malignant potential) and complete excision of PT regardless of grade is recommended. Sometimes a definite diagnosis of a cellular FEL is not possible on CNB. Such equivocal FELs are also categorised as 'B3' and surgical excision is recommended. Clearly, correctly identifying 'B2' FAs amongst FELs is important for minimising unnecessary and potentially deforming breast surgery.

An audit of breast screening pathology practice at Leeds Teaching Hospitals National Health System Trust (LTHT) between 2012-2013 identified a rate of CNB diagnosed 'B3' lesions (which also includes radial scars and atypical intraduct epithelial proliferations among others) in excess of UK Royal College of Pathologists (RCPath) guidelines (12% versus <9%).[7] More specifically an audit of 'B3' FELs found that 62% were FA on excision which is higher than what is reported in most large studies.[4 5 8-11]

The primary aim of this work was to expand on this initial audit of 'B3' FELs at LTHT by conducting a multicentre review of the CNB and excisions of these lesions with two main goals: 1) to assess the degree of interobserver variability in diagnosing these lesions across a large group of pathologists; specifically looking at whether variability is reduced amongst specialist breast pathologists and 2), to determine whether it was possible to reduce the number of diagnostic excisions with a final diagnosis of fibroadenoma. A secondary aim was to examine whether assessment of stromal cellularity and stromal expansion in FELs could be quantified by digital point counting of cells resulting in a more objective way of assessing these features and thus aiding in their classification.

METHOD

Case selection and digitisation

Slides of 69 FELs categorised as 'B3' on CNBs and their subsequent excisions were retrieved from archives of the Leeds Teaching Hospital NHS Trust from 2009 to 2011. Because of the practicalities of dealing with multiple geographic locations, all cases were assessed on a digital pathology platform. One representative H&E slide from the original CNB and the subsequent matched excision were anonymised, given a project code, and then scanned at ×400 magnification (effective resolution = 0.25 microns per pixel) on an Aperio AT2 slide scanner (Aperio, Leica Biosystems, Vista, CA, USA). A pragmatic approach was used to choose the slides and none were screened in any way for a specific diagnostic feature. Images were hosted on the 'Virtual Pathology at the University of Leeds' website (http://www.virtualpathology.leeds.ac.uk/, last accessed 23/11/2017) with private domain login details made available to participants. Slides were viewed on standard resolution monitors either online using Webscope (Aperio, Leica Biosystems, Vista, CA, USA) or downloaded and viewed locally with Imagescope (Aperio, Leica Biosystems, Vista, CA, USA).

Audit participants and instructions

In the audit, cases were viewed by eight pathologists from four tertiary pathology institutions in three countries. Pathologists ranged in experience from recently qualified pathologists, with less than 10 years specialist experience (four participants), to specialised breast pathologists, the latter defined as having more than 10 years specialist experience with a dominant practice in breast pathology (four participants). Each observer was provided with a set of guidelines for the semi-quantitative scoring of a series of morphological features (Supplementary table 1), guidelines for the use of diagnostic categories for CNBs (Supplementary table 2), and suggested diagnostic terms for FELs on excision specimens (Supplementary table 3). In brief, expected diagnostic categories for CNB were; 1) 'FA' (i.e. B2 code), 2) 'FEL' (i.e. B3 code; incorporating cellular FELs favouring FA, favouring PT and those truly indeterminate), and 3) 'PT' (i.e. B3 if considered benign or uncertain or B5 if considered malignant.). More specific diagnoses were expected for excisions, although the category of 'benign FEL', again to capture cellular FELs indeterminate between FA and PT was accepted. Diagnoses of mammary hamartoma and fibroadenoma were grouped together. Evaluations were first performed on CNBs blinded to the excision specimen. Once CNB evaluation was completed and results returned to the collating pathologist (BD), the excision samples were provided to participants for evaluation. The scans were coded such that CNBs and excisions could not be matched by participants. Clinical details (age, clinical features, imaging findings) were not available to the pathologist as these data, although available when reporting in clinical practice, should not affect CNB coding as B2 or B3 as this should be based on pathological features alone.[7]

Statistical Analysis

Kappa statistics for interobserver agreement were determined between all possible pairings between pathologists using Analyse-it software for Microsoft Excel (Microsoft, Redmond, USA). Accuracy of CNB for each pathologist was reflected by the proportion of cases correctly and incorrectly classified on CNB with respect to their excision biopsy diagnosis. A weighted 'diagnosis score (DS)' was calculated for the excisions by the summation of a numerical value for each diagnosis offered by the eight pathologists (FA = 1; FEL = 2; PT = 3) divided by 8. FA or probable FA was defined as DS<1.5. The DS was a pragmatic solution (to allow statistical analysis) given the poor agreement we

observed even on the excision specimens. Spearman correlation was calculated between each characteristic and the DS using SPSS statistical analysis software (SPSS; Chicago, USA.

Digital point counting
All CNB and excision slides were viewed and scored by one pathologist (BD) within ImageScope (Aperio) on a high resolution screen (Figure 2). The lesional area(s) for point counting were outlined within the software using the pen tool. Stereological point counting was performed using a digital graticule of approximately 900 points superimposed on the outlined area(s) using RandomSpot software (University of Leeds, Leeds, UK).[12] Individual points were viewed one-by-one at high power magnification (x40) on screen within ImageScope and assessed as to whether they overlayed stromal cell nuclei, stromal connective tissue, duct epithelium, duct lumen or were non-informative. In so doing RandomSpot generated an unbiased sample of the distribution of point classes within the outlined area whilst minimising the number of observations required.[12] Percent stromal cellularity and the percent epithelial to stromal ratio (as a surrogate for stromal expansion) were calculated as these were parameters amenable to evaluation by point counting. Spearman or Pearson correlation was calculated between these measurements on core and 1) their subjective assessment on core, 2) their biomorphometric measurement on matched excision, and 3) the DS on excision, all using SPSS statistical analysis software (SPSS; Chicago, USA). The necessary number of measurement points was established in a pilot study of 10 cases (5 cores and 5 excisions; not shown) where approximately 850 points was the least number required in any of the 10 samples to reach a stable running mean of either percent stromal cellularity or percent epithelial to stromal ratio +/- 10% error.

Digital pathology survey
At the completion of their analyses, each pathologist was surveyed regarding their experience, comfort and confidence of using digital pathology both in this audit and in their general pathology practice and was given an opportunity to feedback specific positive and/or negative experiences (Supplementary table 4).

RESULTS

Pathologist agreement in core needle biopsies
Unanimous agreement was obtained in only 8/69 CNBs (11%) and for each of these cases the diagnosis was FA (Figure 3). In 15 (21%) cases both FA and PT were offered as diagnoses by different pathologists on the same core. The range of kappa scores for agreement of diagnosis between pathologists for CNBs was 0.15 (very weak) to 0.55 (moderate) with a mean kappa of 0.36 (weak). For specialist pathologists, the kappa values were 0.33 (weak) to 0.51 (moderate) with a mean of 0.44 (moderate) and 0.22 (weak) to 0.51 (moderate) with a mean of 0.35 (weak) for generalist pathologists (Table 1).

Pathologist agreement in excision biopsies
Unanimous agreement was obtained in 21/69 (30%) excision biopsies; 19 FAs and 2 PTs (Figure 4). In 37 cases (54%) both FA and PT were offered as diagnoses by separate pathologists. The range of kappa scores for agreement of diagnosis between pathologists was 0.32 (weak) to 0.74 (strong) with a mean of 0.49 (moderate) for all

pathologists, 0.44 (moderate) to 0.74 (strong) with a mean of 0.54 (moderate) for specialists, and 0.23 (weak) to 0.63 (strong) with an average of 0.44 (moderate) for generalists (Table 2).

Accuracy of core needle biopsies
There were only four cases (9%) where unanimous agreement was observed among all eight pathologists in both the core biopsy and the excision biopsy. In three cases (8%) there was unanimous FA diagnosis on CNB but at least one pathologist diagnosed PT on the matched excision. The proportion of cases confidently classified as an FA on the CNB but as PT on the matched excision by each pathologist ranged from 5 to 23%. The proportion of cases definitively diagnosed as PT on the CNB but as FA on the matched excision by each pathologist ranged from 0 to 50%. There was little improvement in the range of these values when specialists only or generalists only were considered (Table 3).

Correlation of core needle biopsy morphology to diagnosis on excision biopsy
There were significant, moderate correlations between the excision diagnosis score (DS) and the perceived presence of stromal hypercellularity (0.40), cleft like spaces (0.38), infiltrative margins (0.33) and stromal cell atypia (0.52) on CNB (p<0.01 for each). Each of these parameters (except cellular atypia) was considered present in the CNB by at least one pathologist in cases unanimously diagnosed as FA on excision. No significant association was observed for periductal hypercellularity, stromal expansion or mitoses. Necrosis and malignant elements were observed too infrequently for meaningful statistical analysis.

Utility of digital point counting
A significant moderate correlation (0.57) was observed between stromal cellularity assessment on CNBs assessed by pathologists and by point counting. A significant moderate correlation of stromal cellularity as determined by point counting was observed between the core and matched excision (0.57) (p<0.001). There was no significant difference in the percent core stromal cellularity between cases with an excision diagnosis score (DS) of <1.5 (i.e. unanimous or near unanimous diagnosis of FA on excision) or DS≥1.5 (Figure 5).
A significant moderate correlation (0.44) was observed between pathologists' assessment of stromal expansion and epithelial:stromal ratio by point counting on CNBs. By point counting there was a moderate significant correlation of epithelial:stromal ratio between the core and matched excision (0.51) (p<0.001). There was no significant difference in CNB epithelial:stromal ratio between cases with an excision of DS<1.5 and DS≥1.5 (Figure 6).

Participant experience of digital pathology
Seven of the eight participants reported feeling comfortable utilising the Aperio digital platform for the audit and that they were satisfied with the quality of the digital images provided. Fifty percent reported use of digital pathology in routine diagnostic work while all reported use in some aspect of their clinical practice. Six pathologists overall felt comfortable using digital pathology for diagnosis in routine practice, while one was neutral and one felt uncomfortable.
It was reported that visual scanning of slides at low power was slow and resolution was worse compared to conventional microscopy, exacerbated by poor viewing screen

quality. Mitoses were harder and more time consuming to identify, and where found were difficult to quantify by microscopic fields.

DISCUSSION

The accurate classification of FELs on CNB carries significant clinical implications as patients diagnosed with benign FA are typically managed without surgery. On the contrary, equivocal FELs and PTs necessitate further sampling depending on local preferences; the former, either mammotome or open diagnostic biopsy for classification, while the latter is typically managed with wide local excision due to risk of long term local recurrence and occasional long term malignant transformation. Yet the distinction between these lesions is far from perfect as the subjective nature of distinguishing criteria has led to poor reproducibility, a problem exacerbated in limited CNB material.[2-6]

The audit reiterates these results. Poor interobserver agreement in FELs on CNB is demonstrated by only 8/69 (11%) cores where unanimous diagnoses from all pathologists occurred (all FAs), 15 (21%) cases where both FA and PT were simultaneously offered as diagnoses by different pathologists and 'weak' mean kappa statistics between diagnoses (0.36). Similarly, in their recent study, Bandyopadhyay *et al*.[4] demonstrated only 'fair' (0.20) core biopsy diagnostic agreement of FELs between five pathologists within their institution which showed minimal improvement to 0.27 with departmental education. The interobserver agreement on cores when assessed specifically by specialist breast pathologists was higher (mean kappa of 0.44 versus 0.35), yet in a study of surgical specimens by Lawton et al, only a poor agreement between specialist breast pathologists was observed, with unanimous agreement in only 2/15 cases.[3] The agreement in our study was higher on the excision specimens (mean kappa of 0.48), an agreement identical to that observed by Cserni et al, in their study of six pathologists classifying 30 equivocal FEL excisions.

Accuracy of CNB as a diagnostic test for FELs was low. Unanimous agreement across core and excision for all eight pathologists occurred in only four cases, and in three cases where a unanimous diagnosis of FA was achieved on the core, at least one pathologist diagnosed PT on the excision. Inaccuracy is also reflected in the proportion of cases thought to be FA on CNB but ultimately diagnosed as PT on excision, and vice versa. Regarding the former, our proportion ranging from 5-21% is similar to previous reports of 3.5-25%[5 8-10 13 14] and regarding the latter, our range of 0-50% is higher than the 0-8% reported by others.[5 10 13]

These results again reflect difficulties in applying subjective criteria on a small amount of tissue, but also in applying them on potentially non-representative sampling of a tumour type that is often heterogeneous.[15] And while it was reassuring that moderate objective biomorphometric correlations between core and excision were observed, these might have been stronger if not for the established fact that FAs can have areas resembling PT and vice versa.[15]

Our audit suggests the most helpful criteria in the classification of FELs in CNBs include a subjective appreciation of infiltrative margins, hypercellularity and stromal cell atypia, all of which have been considered predictive of a subsequent PT diagnosis in various

prior studies, with increasing cellularity often identified as one of, if not the strongest predictor of PT on subsequent excision.[5 6 11 16 17]

Our attempt to provide more objective measures demonstrated a significant moderate correlation between stromal hypercellularity and stromal expansion assessed by pathologists and digital point counting, yet failed to stratify FEL diagnoses. Digital analysis of different morphological characteristics of FELs has also been investigated by McKenna *et al.*[18] Using automated image analysis algorithms, the authors demonstrated values for stromal cellularity, stromal:epithelial ratio, stromal overgrowth, and mean nuclear diameter were significantly different between classes of FELs. Yet, like our study, a critical examination of their data shows significant overlap in values observed between FEL classes for these parameters, so do not offer a solution to the problem of equivocal FELs.[18]

The primary implication of our study is that the substratification of equivocal FELs on a core is extremely difficult and the 'B3' category for FELs necessitating surgical excision for diagnosis is frequently inevitable. However, within this group there were a small number of cases in which all reviewers provided a diagnosis of FA on CNB suggesting that these cases could have safely avoided a diagnostic excision. This should be interpreted with some caution, however, as in a study scenario one may be more confident at making a definitive diagnosis of FA than in clinical practice. This represents a potential limitation in our study and may also affect agreement statistics. Potential identification of such cases requires a collegiate workplace, with specialist input, which encourages corroboration to reach diagnosis. The study also reminds us that core sampling of a heterogeneous tumour can lead to a non-representative diagnosis. In turn this highlights the imperative function of the multi-disciplinary team which might identify unusual clinical and radiological features, such as tumour size or rapid growth that might flag such cases. Nevertheless it is important to remember that most fibroepithelial lesions are FAs and most of these are accurately diagnosed on core biopsy. This study has looked at the difficult boundary between FA and PT.

Our study also highlights the strengths of digital pathology. Scanned slides were quickly and easily accessed from pathologists across the globe, and in so doing, the inclusion of pathologists from different geographical locations countered against diagnostic drift within a department or local region. This is clearly beneficial in an audit setting, but has potential benefits in routine diagnostic pathology, consult cases, remote and telepathology and for quality assurance programs. But there are limitations too. Scanning across slides can be slow, resolution was worse than conventional microscopy, mitoses can be harder to find, and microscopic quantification, often based on 'per high power field', was difficult. All participants in this study are experienced in the utilisation of digital slides and despite these minor difficulties, most reported that the digital viewing platform and image quality were acceptable for interpretation of FELs. Thus, while the use of a digital platform to conduct this audit is a potential limitation of our study, it is unlikely to make a significant contribution of the low diagnostic agreement. Scanning slides at high power, use of medical grade high resolution screens for viewing and computers of sufficient processing power can mitigate some of the deficiencies of digital pathology. The conversion of digital fields to microscopic fields affects multiple organ systems and needs pragmatic resolution if digital pathology is going to be more widely adopted for routine diagnostic purposes.

CONCLUSION

The classification of equivocal FELs along the margin of cellular FA and low grade PT is difficult, particularly on CNB, resulting in poor interobserver agreement and suboptimal accuracy of diagnosis. We demonstrated stromal hypercellularity on core as a morphologic parameter associated with PT on excision, but in both subjective and objective assessment showed considerable overlap with benign FAs. Given the diagnostic difficulty, potential surgical implications and improved agreement in these lesions amongst experts, equivocal FELs should be reported in conjunction with experienced breast pathologists.

KEY MESSAGES

- Classification of equivocal FELs at the boundary between cellular FA and benign/borderline PT is difficult, particularly on CNB, resulting in suboptimal interobserver agreement and suboptimal diagnostic accuracy
- Digital point counting of stromal cellularity and epithelial:stromal ratio does not aid in the classification of these lesions
- Difficult FELs should be reported in collaboration with experienced breast pathologists, where identification of benign FAs from equivocal lesions could reduce unnecessary surgery.

REFERENCES

1. Hanby AM, Millican-Slater R, Dessauvagie BF. Fibroepithelial neoplasms of the breast. Diagn Histopathol 2017;**23**(4):149-58
2. Cserni G, Orosz Z, Kulka J, et al. Divergences in diagnosing nodular breast lesions of noncarcinomatous nature. Pathology oncology research : POR 2006;**12**(4):216-21 doi: PAOR.2006.12.4.0216[published Online First: Epub Date]|.
3. Lawton TJ, Acs G, Argani P, et al. Interobserver variability by pathologists in the distinction between cellular fibroadenomas and phyllodes tumors. International journal of surgical pathology 2014;**22**(8):695-8 doi: 10.1177/1066896914548763[published Online First: Epub Date]|.
4. Bandyopadhyay S, Barak S, Hayek K, et al. Can problematic fibroepithelial lesions be accurately classified on core needle biopsies? Human pathology 2016;**47**(1):38-44 doi: 10.1016/j.humpath.2015.09.001[published Online First: Epub Date]|.
5. Lee AH, Hodi Z, Ellis IO, et al. Histological features useful in the distinction of phyllodes tumour and fibroadenoma on needle core biopsy of the breast. Histopathology 2007;**51**(3):336-44 doi: 10.1111/j.1365-2559.2007.02786.x[published Online First: Epub Date]|.
6. Tsang AK, Chan SK, Lam CC, et al. Phyllodes tumours of the breast - differentiating features in core needle biopsy. Histopathology 2011;**59**(4):600-8 doi: 10.1111/j.1365-2559.2011.03939.x[published Online First: Epub Date]|.
7. Lee AH, Carder P, Deb R, et al. Guidelines for non-operative diagnostic procedures and reporting in breast cancer screening. London, UK: The Royal College of Pathologists, 2016.
8. Abdulcadir D, Nori J, Meattini I, et al. Phyllodes tumours of the breast diagnosed as B3 category on image-guided 14-gauge core biopsy: analysis of 51 cases from a

single institution and review of the literature. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology 2014;**40**(7):859-64 doi: 10.1016/j.ejso.2014.02.222[published Online First: Epub Date]|.

9. Dillon MF, Quinn CM, McDermott EW, et al. Needle core biopsy in the diagnosis of phyllodes neoplasm. Surgery 2006;**140**(5):779-84 doi: 10.1016/j.surg.2006.03.022[published Online First: Epub Date]|.

10. Komenaka IK, El-Tamer M, Pile-Spellman E, et al. Core needle biopsy as a diagnostic tool to differentiate phyllodes tumor from fibroadenoma. Archives of surgery 2003;**138**(9):987-90 doi: 10.1001/archsurg.138.9.987[published Online First: Epub Date]|.

11. Yasir S, Gamez R, Jenkins S, et al. Significant histologic features differentiating cellular fibroadenoma from phyllodes tumor on core needle biopsy specimens. American journal of clinical pathology 2014;**142**(3):362-9 doi: 10.1309/AJCPZUZ96RESGPUP[published Online First: Epub Date]|.

12. Wright AI, Grabsch HI, Treanor DE. RandomSpot: A web-based tool for systematic random sampling of virtual slides. Journal of pathology informatics 2015;**6**:8 doi: 10.4103/2153-3539.151906[published Online First: Epub Date]|.

13. Bode MK, Rissanen T, Apaja-Sarkkinen M. Ultrasonography and core needle biopsy in the differential diagnosis of fibroadenoma and tumor phyllodes. Acta radiologica 2007;**48**(7):708-13 doi: 10.1080/02841850701367911[published Online First: Epub Date]|.

14. Foxcroft LM, Evans EB, Porter AJ. Difficulties in the pre-operative diagnosis of phyllodes tumours of the breast: a study of 84 cases. Breast 2007;**16**(1):27-37 doi: 10.1016/j.breast.2006.05.004[published Online First: Epub Date]|.

15. Chng TW, Lee JY, Lee CS, et al. Validation of the Singapore nomogram for outcome prediction in breast phyllodes tumours: an Australian cohort. Journal of clinical pathology 2016;**69**(12):1124-26 doi: 10.1136/jclinpath-2016-203951[published Online First: Epub Date]|.

16. Jacobs TW, Chen YY, Guinee DG, Jr., et al. Fibroepithelial lesions with cellular stroma on breast core needle biopsy: are there predictors of outcome on surgical excision? American journal of clinical pathology 2005;**124**(3):342-54 doi: 10.1309/5N2C-4N5X-CB8X-W8JL[published Online First: Epub Date]|.

17. Jara-Lazaro AR, Akhilesh M, Thike AA, et al. Predictors of phyllodes tumours on core biopsy specimens of fibroepithelial neoplasms. Histopathology 2010;**57**(2):220-32 doi: 10.1111/j.1365-2559.2010.03607.x[published Online First: Epub Date]|.

18. McKenna AM, Pintilie M, Youngson B, et al. Quantification of the morphologic features of fibroepithelial tumors of the breast. Archives of pathology & laboratory medicine 2007;**131**(10):1568-73 doi: 10.1043/1543-2165(2007)131[1568:QOTMFO]2.0.CO;2[published Online First: Epub Date]|.

FOOT NOTES

conceived and oversaw project, participated in audit, statistical analysis, critically reviewed manuscript.

Licence for Publication: The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in JCP and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence (http://group.bmj.com/products/journals/instructions-for-authors/licence-forms).

Competing Interests: PT has patents titled 'Exome Sequencing of Breast Fibroadenomas Reveals Highly Recurrent MED12 Exon 2 Somatic Mutations (Ref: TEA-P001WO)', 'Genomic Progression of Breast Fibroepithelial Tumors (Ref: TEA-P002WO)', and 'Multigene qPCR Assay Classifying Breast Fibroepithelial Lesions'. DT is an advisory board member of Sectra Medical and Leica and has research contracts with Leica, FFEI Life Science and Roche. DT receives no personal remuneration for any of these activities.

Ethics approval: This is a diagnostic audit, a mandated requirement for ongoing quality assurance in pathology diagnostic laboratories in the UK. As such ethics approval was not required. Audit approval was obtained from The Clinical Director of Pathology, St James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, UK.

TABLES

Table 1. Kappa statistics for the core biopsy diagnoses between the eight pathologists (A to H). Dark grey shading represents correlations between specialist breast pathologists.

	A	B	C	D	E	F	G	H
A								
B	0.29							
C	0.32	0.51						
D	0.20	0.41	0.45					
E	0.44	0.25	0.34	0.17				
F	0.22	0.15	0.28	0.21	0.28			
G	0.51	0.46	0.52	0.55	0.40	0.25		
H	0.41	0.48	0.47	0.33	0.37	0.16	0.55	

Table 2. Kappa statistics for the excision biopsy diagnoses between the eight pathologists (A to H). Dark grey shading represents the correlations between specialist breast pathologists.

	A	B	C	D	E	F	G	H
A								
B	0.48							
C	0.55	0.69						
D	0.42	0.44	0.74					
E	0.56	0.31	0.18	0.15				
F	0.32	0.46	0.58	0.62	0.23			
G	0.63	0.63	0.72	0.64	0.41	0.51		
H	0.49	0.47	0.45	0.47	0.50	0.39	0.58	

Table 3. Accuracy of a core biopsy diagnoses. Dark grey shading represents values of specialist breast pathologists.

%	A	B	C	D	E	F	G	H
FA on core biopsy/PT on excision	7	8	7	8	23	14	5	23
PT on core biopsy/FA on excision	22	50	33	0	0	0	0	0

FIGURE LEGENDS

Figure 1: Unanimously diagnosed 'B3' fibroeptihelial lesion on core needle biopsy. A. At screening magnification, diffusely increased stromal cellularity and core fragmentation are observed (H&E). B-D. At high power, increased stromal cellularity (B), equivocal periductal stromal condensation (C), zones of stromal expansion (D) and an occasional 'leaf-like' formation are evident. E. At screening magnification, the corresponding excision, unanimously diagnosed as a phyllodes tumour (most participants classifying it as borderline), shows areas of stromal hypercellularity, stromal expansion, predominant leaf like growth and a poorly defined edge which irregularly extends into adjacent adipose tissue.

Figure 2. Digital point counting. A. The lesional area for analysis was outlined with the pen tool in Aperio software. B. The region of interest, exported from Aperio, was uploaded to RandomSpot software which generated the specified number of points (+/- a set tolerance amount) for analysis. C and D. These points were viewed within Aperio. E-I. Points were classified as stromal cell nucleus (top left), stroma (top right), epithelium (bottom left), duct lumen (bottom right) and non-informative (centre inset).

Figure 3. Core biopsy diagnoses submitted for each case. The cases are represented along the horizontal axis. The vertical columns represent the diagnoses submitted, with a different colour representing each diagnosis (white=fibroadenoma, grey=equivocal FEL lesion, Black=phyllodes tumour). The height of each colour represents the number of pathologists who submitted that diagnosis.

Figure 4. Excision biopsy diagnoses submitted for each case. The cases are represented along the horizontal axis. The vertical columns represent the diagnoses submitted, with a different colour representing each diagnosis (white=fibroadenoma, grey=equivocal FEL, Black=phyllodes tumour). The height of each colour represents the number of pathologists who submitted that diagnosis.

Figure 5. Box and whisker plot showing no significant difference in cellularity biomorphometrics on core biopsy between FEL cases with an excision diagnosis score (DS) of <1.5 (i.e. unanimous or near unanimous diagnosis of FA on excision) or DS≥1.5.

Figure 6. Box and whisker plot showing no significant difference in stromal expansion (epithelial:stromal ratio) biomorphometrics on core biopsy between FEL cases with an excision diagnosis score (DS) of <1.5 (i.e. unanimous or near unanimous diagnosis of FA on excision) or DS≥1.5.