Reproducibility in the diagnosis of needle core biopsies of non-palpable breast lesions: an international study using virtual slides

Histopathology 2010, **56**, 720–726. DOI: 10.1111/j.1365-2559.2010.03548.x

Reproducibility in the diagnosis of needle core biopsies of non-palpable breast lesions: an international study using virtual slides published on the world-wide web

Francesco Alfredo Zito, Paolo Verderio,[1] Giovanni Simone, Vito Angione,[2] Paola Apicella,[3] Simonetta Bianchi,[4] Antonio Felix Conde,[5] Omar Hameed,[6] Julio Ibarra,[7] Antony Leong,[8] Natale Pennelli,[9] Ezio Pezzica,[10] Vania Vezzosi,[4] Vincenzo Ventrella,[11] Sara Pizzamiglio,[1] Angelo Paradiso[12] & Ian Ellis[13]

Department of Pathology, National Cancer Institute 'Giovanni Paolo II', Bari, [1]Unit of Medical Statistics and Biometry, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, [2]Department of Pathology, Azienda Ospedaliera-Universitaria, S. Maria della Misericordia, Udine, [3]Department of Pathology, Pistoia Hospital, Pistoia and [4]Department of Human Pathology and Oncology, Azienda Ospedaliero-Universitaria Careggi, Florence, Italy, [5]Department of Pathology, University Hospital Perpetuo Socorro, Badajoz, Spain, [6]Department of Pathology, University of Alabama School of Medicine, Birmingham, AL and [7]Department of Pathology, Memorial Care Breast Center, at Orange Coast, Fountain Valley, CA, USA, [8]Department of Pathology, University of Newcastle, Newcastle, Australia, [9]Department of Pathology, University of Padua, Padua, [10]Department of Pathology, Treviglio Hospital, Treviglio, [11]Women's Department and [12]Clinical Experimental Oncology Laboratory, National Cancer Institute Giovanni Paolo II, Bari, Italy, and [13]Department of Histopathology, University of Nottingham, Nottingham, UK

Date of submission 29 December 2008
Accepted for publication 26 August 2009

Zito F A, Verderio P, Simone G, Angione V, Apicella P, Bianchi S, Conde A F, Hameed O, Ibarra J, Leong A, Pennelli N, Pezzica E, Vezzosi V, Ventrella V, Pizzamiglio S, Paradiso A & Ellis I

(2010) *Histopathology* **56**, 720–726

Reproducibility in the diagnosis of needle core biopsies of non-palpable breast lesions: an international study using virtual slides published on the world-wide web

Aims: To conduct an internet-based study using virtual slides (VS) of sterotactic core biopsy specimens of non-palpable breast lesions in order to evaluate inter-observer reproducibility between pathologists.

Methods and results: A total of 18 breast lesions, determined to be histologically complex by two pathologists, were selected. Digitized VSs were then created using QuickTime Virtual Reality technology (Apple, Cupertino, CA, USA) and posted on the world-wide web. In all, 10 pathologists completed the evaluations of 18 VSs using the five diagnostic categories (B1–B5) from the *European guidelines for quality assurance in breast cancer screening and diagnosis*. Their results were compared with those of every other participating pathologist, and were then individually compared with the results of a highly experienced breast pathologist (referee). Of the 18 cases, 10 (56%) were classified by the referee as borderline (B3 and B4). Comparisons with reference values showed a less than satisfactory level of reproducibility (median $\kappa_w = 0.60$). As regards interobserver reproducibility, results showed that, in general, the level of agreement was not satisfactory (median $\kappa_w = 0.53$).

Conclusions: Overall, the findings are comparable to those quality control studies using circulating slides when analysis is done on borderline cases.

Keywords: atypical ductal hyperplasia, core breast biopsy, quality control, telepathology, virtual slide

Abbreviations: ADH, atypical ductal hyperplasia; DCIS, ductal carcinoma *in situ*; QA, quality assurance; VR, virtual reality; VS, virtual slide

Address for correspondence: Dr F A Zito, Istituto Tumori 'Giovanni Paolo II', via Hahnemann 10, 70126 Bari, Italy. e-mail: fazito@fastwebnet.it

Introduction

The use of new computer-based interactive technologies in medicine, such as virtual reality (VR), is increasing. VR applications play an important role in medical training because it is possible to learn in an environment of absolute safety. This is obviously important in the field of surgery and for other invasive medical procedures. It is also important for training in pathology because it makes access to a large variety of digitized cytological and histological (virtual slide) cases possible. Furthermore, virtual slides (VSs) enable one to overcome several difficulties related to pathology practice. For example, one problem associated with the use of traditional glass slides is that their distribution among pathologists during quality assurance (QA) programmes is often inefficient, requiring much time to complete conventional quality control programmes. Furthermore, there is the risk of significant differences between samples cut from multiple tissue sections from the same paraffin block. There have been very few quality control studies that have verified the diagnostic reproducibility of VS.[1–14] The purpose of this study was to assess whether pathologists are able to make an accurate and reproducible diagnosis on-line using VSs of selected core breast lesion biopsy specimens with particular regard to borderline lesions. These lesions are difficult to diagnose and published interobserver reproducibility studies have shown a low level of agreement among pathologists.[15–19] In 1991, Rosai[18] highlighted this problematic aspect, but as yet there has been no significant improvement in the diagnosis of these lesions. These data suggest that new and more feasible models of QA are necessary in order to improve agreement in the diagnosis of borderline lesions. To this end, the authors carried out an on-line interobserver study of VSs generated from stereotactic breast core biopsies specimens with the goal of determining the interobserver agreement between general pathologists and an expert in the field of breast pathology.

Materials and methods

VSs were generated from 18 cases of large core needle breast biopsy specimens of complex non-palpable lesions originally evaluated at the Department of Pathology, National Cancer Institute (Bari, Italy). The chosen cases resulted from routine referrals to the department, and were chosen for this study as a result of their particular complexity, which required the opinion and expertise of more than one pathologist. Large areas of the haematoxylin and eosin-stained slides were digitized to high power (20×) using a digital camera (Olympus DP20, resolution: 1600×1200 pixel) mounted on a microscope (Olympus BX 41) and connected to a computer in order to generate VSs using QuickTime VR (Apple, Cupertino, CA, USA) technology.[20] A web-page was created (http://www.oncologico.bari.it/istopatologia/index.html) to allow access to the server that contained the VSs. The participants were recruited via e-mail using a mailing list of Italian pathologists (ITAPAT; http://www.siapec.it) and by personal invitation via e-mail. At the time of registration into the study, the pathologists provided data regarding the institution to which they belonged, and the number of cases of breast pathology per year. For each VS, a reference diagnosis was provided by an experienced breast pathologist (I.E.), from the Department of Histopathology, University of Nottingham, UK. The participating pathologists were then required to classify the VSs according to the five categories of the *European guidelines for quality assurance in breast cancer screening and diagnosis*[21]: B1, normal tissue/unsatisfactory; B2, benign lesion; B3, uncertain malignant potential; B4, suspicious of malignancy; B5, malignant. In addition, an open box was provided for a written diagnostic description of each case for the purpose of verifying the appropriateness of the diagnostic category. Interobserver reproducibility was assessed, as well as the reproducibility between each pathologist, and those of the referee, by computing the weighed κ statistic (κ_w).[22] The κ_w statistic is the most widely accepted measure of agreement when, as in our case, the data in question arise from an ordinal scale. Its values usually lie between 0 (absence of agreement) and 1 (absolute agreement). A negative value may be obtained in situations where the actual agreement is less than a chance one. In addition, we investigated the contribution of each diagnostic category of observed reproducibility by determining the κ category-specific statistics (κ_{cs}) and their weighted average (Cohen's κ statistic, κ_c). Finally, we evaluated the interchangeability between any two different diagnostic categories using the κ interchangeability statistics (κ_i) computed as suggested by Holman.[23] As previously reported,[24] κ_w values were considered satisfactory if equal to 0.80, whereas values of κ_{cs} and κ_c were interpreted in a qualitative manner on the basis of the Landis and Koch[25] classification criteria. Positive κ_i values indicate a level of interchangeability between the two categories under consideration which is greater than expected by merely a chance effect. Conversely, negative values have to be considered satisfactory, as they indicate a good discriminatory capability between the two

considered categories. Furthermore, in order to validate the QuickTime VR technology, the reproducibility between the reference VS diagnoses and the diagnoses based on traditional histological slides made by two experienced breast pathologists (F.A.Z. and G.S.) was evaluated. The pathologists were given 6 months to complete the diagnoses.

Results

A satisfactory level of reproducibility ($\kappa_w = 0.83$) between the diagnoses made in a blind study of traditional glass slides was carried out by two pathologists from the same institution, and the diagnoses made from the corresponding VSs by the referee.

Table 1. Distribution of the diagnoses made from the referee and from the pathologists

		Pathologists' diagnoses			
Virtualslide (VS)	Referee's diagnoses	B1–2	B3	B4	B5
VS-1	B5: Clear cell ductal carcinoma *in situ*	4	4	1	1
VS-2	B2: Sclerosing adenosis	10			
VS-3	B3: Adenosis with epithelial atypia	8	2		
VS-4	B3: Flat epithelial atypia with at least atypical ductal hyperplasia	1	6	1	2
VS-5	B4: Flat epithelial atypia with at least atypical ductal hyperplasia and features suspicious of ductal carcinoma *in situ*	1	3		6
VS-6	B5: Invasive carcinoma with classical lobular features				10
VS-7	B5: Invasive carcinoma with tubular features				10
VS-8	B3: Unusual case with glandular structures involving peripheral nerve-like structures. Could be invasive carcinoma, benign hamartoma or adenosis involving nerve	3	3		4
VS-9	B5: Invasive carcinoma with tubulolobular features. Background flat epithelial atypia and probable low-grade micropapillary ductal carcinoma *in situ*		3	2	5
VS-10	B5: Low-grade ductal carcinoma *in situ* with background flat epithelial atypia	1			9
VS-11	B4: Flat epithelial atypia with features of at least atypical ductal hyperplasia and areas suspicious of mucin hypersecretory micropapillary ductal carcinoma *in situ*	1	5		4
VS-12	B3: Adenosis with epithelial atypia	3	6		1
VS-13	B3: Adenosis with epithelial atypia. May be part of a fibroepithelial lesion or hamartoma	2	5	2	1
VS-14	B3: Adenosis with epithelial atypia. Probable lobular neoplasia. Would look at E-cadherin and cytokeratin expression	3	4	1	2
VS-15	B2: Focal lactational change	8		1	1
VS-16	B2: Sclerosing adenosis	7	1		2
VS-17	B3: Papillary lesion		5	1	4
VS-18	B3: Radial scar	4	5		1

A total of 36 pathologists from various parts of the world originally enrolled in the study; however, only 10 provided diagnoses for all of the 18 diagnostic cases under consideration (Table 1).

Reproducibility analysis was performed by adopting a classification criterion based on four categories obtained by jointly considering B1 and B2 categories. Lesions were classified by the referee as follows: three cases B1/B2; eight cases B3; two cases B4; five cases B5.

The open box, which was inserted into the web page, demonstrated a good level of appropriateness between the descriptive diagnoses and the diagnostic categories. In those diagnostic categories with a high level of agreement, there was also a high level of agreement in the descriptive diagnosis furnished in the open box (e.g. in VS 2, 100% of pathologists indicated the B2 category, and in the open box all of the pathologists made a diagnosis of adenosis). Five of the pathologists did not identify any lesions in the B4 category.

In four cases (VS-1, VS-3, VS-8 and VS-11) the prevalent diagnostic category chosen by the pathologists was different from that of the referee (Table 1). Comparisons with reference diagnoses (Table 2) showed a less than satisfactory level of reproducibility with a median κ_w value of 0.60 (range 0.19–0.82) that did not correlate with the numbers of cases per year of each department (Table 3). For only one pathologist was a κ_w value >0.80 observed.

Table 2. Reproducibility versus referee and interobserver reproducibility – κ_w values

Pathologist's code	Reproducibility versus referee	Interobserver reproducibility		
		Minimum	Median	Maximum
P01	0.42	−0.18	0.35	0.46
P02	0.67	0.28	0.51	0.79
P03	0.60	0.12	0.63	0.82
P04	0.60	0.18	0.51	0.72
P05	0.61	0.07	0.55	0.79
P06	0.19	−0.18	0.07	0.36
P07	0.56	0.07	0.53	1.00
P08	0.82	0.01	0.63	0.82
P09	0.62	0.19	0.63	0.74
P10	0.56	0.07	0.53	1.00
Overall	0.60	−0.18	0.53	1.00

Table 3. Number of breast samples per year compared with the value of reproducibility versus referee (κ_w)

Pathologist's code	Breast stereotactic biopsies (*n*)	Surgical breast specimens (*n*)	κ_w
P01	330	270	0.42
P02	120	100	0.67
P03	3000	800	0.60
P04	150	80	0.60
P05	1000	3500	0.61
P06	405	240	0.19
P07	200	380	0.56
P08	80	130	0.82
P09	300	400	0.62
P10	3000	800	0.56

Regarding interobserver reproducibility, results showed that, in general, the level of agreement was not satisfactory, with a median κ_w value of 0.53 (range −0.18 to 1.00). Table 2 reports the distribution of κ_w values observed for each pathologist compared with all the others.

As shown in Table 4, the most significant contribution to overall agreement ($\kappa_c = 0.30$) was provided by the two extreme diagnostic categories showing fair/moderate agreement (B1/B2, $\kappa_{cs} = 0.40$; B5, $\kappa_{cs} = 0.41$). The least significant contribution was found in the intermediate categories B3 ($\kappa_{cs} = 0.15$) and B4 ($\kappa_{cs} = -0.001$), showing slight agreement and

Table 4. κ_{cs} and κ_i values

Category	B1 + B2	B3	B4	B5
B1 + B2	**0.40***	−0.07†	−0.01	−0.34
B3	−0.08	**0.15**	*0.03*	−0.13
B4	−0.09	*0.21*	−0.001	−0.13
B5	−0.26	−0.09	−0.01	**0.41**

*κ_{cs} values (in bold) measure the contribution of each category to the overall agreement.

†κ_i values (in italic) measure the discriminatory capacity between any two different categories.

Table 5. Time taken for each pathologist to complete the diagnoses

Pathologist's code	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	Mean
Time taken (days)	51	82	78	141	43	42	4	98	103	75	71.7

disagreement, respectively. Table 4 shows that positive κ_i values were observed only for the B3 versus B4 pair ($\kappa_i = 0.03$) and for the B4 versus B3 pair ($\kappa_i = 0.21$), indicating that the discriminatory diagnostic capacity related to categories B3 and B4 was not satisfactory.

The mean time needed to complete the diagnoses was 71.7 days (range 4–141 days) (Table 5). The way in which the web page was set allowed us to know whether the site was accessed one or more times by each participant, but not the precise number of times.

Discussion

The detection of minimally invasive carcinoma, ductal carcinoma *in situ* (DCIS) and borderline lesions of the breast has increased dramatically since the introduction of mammographic screening. Notwithstanding the fact that the histopathological diagnostic accuracy of stereotactic core biopsy specimens is high for malignant and benign lesions, it remains low for borderline lesions. Studies carried out on circulating slides of core biopsy specimens, surgical specimens, and even on pre-sampled, still digital images have shown a low level of agreement among pathologists for borderline lesions.[15–19,26]

The results of this study on VSs seem to reproduce the data of such studies on conventional circulating slides of core biopsy specimens.[15,19] Agreement among pathologists is high in the two extreme categories of benign and malignant lesions, respectively (B1/B2 and B5), and low in B3 and B4 lesions (Table 4).

The low discriminatory capacity to distinguish B3 from B4 categories (positive value of κ_i), and the absence of any diagnoses in the B4 category by five pathologists shows a use of the two categories which is not clear-cut. The B4 category, which includes technical problems, does not allow for a definitive diagnosis of a probable carcinoma, nor cases of intraductal proliferation with extensive and severe atypia.[27] This latter condition was reported by the referee due to a high degree of suspicion: 'at least atypical ductal hyperplasia (ADH), suspicious of low grade DCIS' (VS-5 and VS-11). It is clear that this pragmatic approach is subjective, resulting in a case being categorized as either B3 or B4.

In our study, the prevalent diagnostic category chosen by the pathologists was in concordance with the referee in 14 cases. The other four cases were discordant regarding the prevalent diagnoses, probably for the following reasons: cases VS-1 (clear cell DCIS) and VS-8 (indefinite neoplasia involving nerve) were unusual and particular cases; in case VS-11 (flat epithelial atypia with features of at least ADH and areas suspicious of mucin hypersecretory micropapillary intraductal carcinoma) none of the pathologists chose B4, choosing instead either B3 or B5, probably because of the contemporary presence of multiple complex lesions; and in case VS-3 (adenosis with atypia) there were no clear standardized diagnostic criteria to evaluate lesions with benign architectural features, but rather with cytological atypia. In this case the diagnostic category chosen fluctuated between B2 and B3.

The absence of correlation between the number of cases of breast pathology per year and κ_w values is a confounding variable, which can probably be explained by each of the participants' technological abilities. However, when the overall agreement on VSs is considered, comparing these results with the data reported in the literature on interobserver diagnostic reproducibility on circulating slides, the use of VSs would seem to determine a low level of agreement among pathologists with respect to conventional studies. As a matter of fact, the two studies reported in the literature[15,19] on interobserver reproducibility on breast core needle biopsy specimens have shown good agreement among pathologists when all cases are considered. However, when analysis is done exclusively on borderline cases, the degree of agreement falls dramatically. Verkooijen,[19] in a multicentre study, showed that the concordance between the general and expert pathologists was only 24% for borderline cases. Collins,[15] in another multicentre study, showed lower levels of agreement for ADH (63%). These two studies have a low percentage of borderline cases (4–5%), which explains the high level of agreement between observers. The only single study conducted on-line in which the diagnoses on VSs were compared with glass slides has shown a concordance of 35.3% in the B3 category, even though it had only one case included in the B3 category, and no cases of B4.[7] Conversely, the

high percentage (56%) of B3 and B4 cases in our study, most certainly influenced the results.

There was a high drop-out rate in this study, and it was noted that the great majority of the original 36 participants accessed the web page only once without making any diagnoses. This may have been due to curiosity regarding the technology itself.

In conclusion, it would seem that our results are similar to those QA studies in which circulating glass slides were used. This could be a good starting point for other QA rounds to improve the skill of pathologists in the diagnosis of borderline lesions. Actually, VS technology has some advantages over circulating traditional glass slides: less time required to complete the studies; easy distribution of cases without geographical limitations; and easy repetition of the study once transformed into a web format. Furthermore, on-line reproducibility studies on VSs would not lose their efficacy at the end of the study because they can remain active in on-line document repositories that are available on the internet in order to improve the skills and knowledge of interested pathologists (http://oncologico.bari.it/istopatologia/index.html). Similar studies could be useful permanent educational on-line resources for the improvement of pathologists' performance.

Acknowledgements

This study is supported by research funding from the Italian Minister of Health (PF-PIO no. 5). The authors would like to thank the technical staff of 'Bioengineering and Medical Informatics Consortium (CBIM) – Pavia' and Francesco Pacoda of 'Istituto Tumori Giovanni Paolo II-Bari' for their technological support. They also thank Michael Kolk for his help in the preparation and editing of this manuscript.

References

1. Alli PM, Ollayos CW, Thompson LD *et al.* Telecytology: intraobserver and interobserver reproducibility in the diagnosis of cervical-vaginal smears. *Hum. Pathol.* 2001; **32;** 1318–1322.
2. Cross SS, Burton JL, Dubé AK *et al.* Offline telepathology diagnosis of colorectal polyps: a study of interobserver agreement and comparison with glass slide diagnoses. *J. Clin. Pathol.* 2002; **55;** 305–308.
3. Odze RD, Goldblum J, Noffsinger A *et al.* Interobserver variability in the diagnosis of ulcerative colitis-associated dysplasia by telepathology. *Mod. Pathol.* 2002; **15;** 379–386.
4. Marchevsky AM, Nelson V, Martin SE *et al.* Telecytology of fine-needle aspiration biopsies of the pancreas: a study of well-differentiated adenocarcinoma and chronic pancreatitis with atypical epithelial repair changes. *Diagn. Cytopathol.* 2003; **28;** 147–152.
5. Lee ES, Kim IS, Choi JS *et al.* Accuracy and reproducibility of telecytology diagnosis of cervical smears. A tool for quality assurance programs. *Am. J. Clin. Pathol.* 2003; **119;** 356–360.
6. Molnar B, Berczi L, Diczhazy C *et al.* Digital slide and virtual microscopy based routine and telepathology evaluation of routine gastrointestinal biopsy specimens. *J. Clin. Pathol.* 2003; **56;** 433–438.
7. Costello SSP, Johnston DJ, Dervan PA *et al.* Development and evaluation of the virtual pathology slide: a new tool in telepathology. *J. Med. Internet Res.* 2003; **5;** e11.
8. Burthem J, Brereton M, Ardern J *et al.* The use of digital 'virtual slides' in the quality assessment of haematological morphology: results of a pilot exercise involving UK NEQAS(H) participants. *Br. J. Haematol.* 2005; **130;** 293–296.
9. Helin H, Lundin M, Lundin J *et al.* Web-based virtual microscopy in teaching and standardizing Gleason grading. *Hum. Pathol.* 2005; **36;** 381–386.
10. Odze RD, Tomaszewski JE, Furth EE *et al.* Variability in the diagnosis of dysplasia in ulcerative colitis by dynamic telepathology. *Oncol. Rep.* 2006; **16;** 1123–1129.
11. Furness P. A randomized controlled trial of the diagnostic accuracy of internet-based telepathology compared with conventional microscopy. *Histopathology* 2007; **50;** 266–273.
12. Ayatollahi H, Khoei A, Mohammadian N *et al.* Telemedicine in diagnostic pleural cytology: a feasibility study between universities in Iran and the USA. *J. Telemed. Telecare* 2007; **13;** 363–368.
13. Massone C, Peter Soyer H, Lozzi GP *et al.* Feasibility and diagnostic agreement in teledermatopathology using a virtual slide system. *Hum. Pathol.* 2007; **38;** 546–554.
14. Słodkowska J, Chyczewski L, Wojciechowski M. Virtual slides: application in pulmonary pathology consultations. *Folia Histochem. Cytobiol.* 2008; **46;** 121–124.
15. Collins LC, Connoly JL, Page DL *et al.* Diagnostic agreement in the evaluation of image-guided breast core needle biopsies. *Am. J. Surg. Pathol.* 2004; **28;** 126–131.
16. Elston W, Sloane JP, Amendoeira I *et al.* Causes of inconsistency in diagnosing and classifying intraductal proliferations of the breast. *Eur. J. Cancer* 2000; **36;** 1769–1772.
17. Palli D, Galli M, Bianchi S *et al.* Reproducibility of histological diagnosis of breast lesions: results of a panel in Italy. *Eur. J. Cancer* 1996; **4;** 603–607.
18. Rosai J. Borderline epithelial lesions of the breast. *Am. J. Surg. Pathol.* 1991; **15;** 209–221.
19. Verkooijen HM, Schipper ME, Buskens E *et al.* Interobserver variability between general and expert pathologists during the histopathological assessment of large-core needle and open biopsies of non-palpable breast lesions. *Eur. J. Cancer* 2003; **39;** 2187–2191.
20. Zito FA, Marzullo F, D'Errico D *et al.* Quicktime virtual reality technology in light microscopy to support medical education in pathology. *Mod. Pathol.* 2004; **17;** 728–731.
21. EC Working Group on Breast Screening Pathology. Quality assurance guidelines for pathology. In Perry N, Broeders M, de Wolf C, Tornberg S, Holland R, von Karsa L eds. *European guidelines for quality assurance in breast cancer screening and diagnosis*, 4th edn. Luxemburg: European Union, 2006; 219–312.

22. Fleiss JL. *Statistical methods for rates and proportions,* 2nd edn. New York: Wiley and Sons, 1981.
23. Holman CDJ. Analysis of interobserver variation on a programmable calculator. *Am. J. Epidemiol.* 1984; **120;** 154–161.
24. Italian Network for Quality Assessment of Tumor Biomarkers (INQAT) Group. Interobserver reproducibility of immunohistochemical HER-2/neu assessment in human breast cancer: an update from INQAT round III. *Int. J. Biol. Markers* 2005; **20;** 189–194.
25. Landis R, Koch G. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33;** 117–127.
26. Amendoeira I, Apostolikas N, Bellocq JP *et al.* Consistency achieved by 23 European pathologists from 12 countries in diagnosing breast disease and reporting prognostic features of carcinomas. *Virchows Arch.* 1999; **434;** 3–10.
27. Ellis IO, Humphreys S, Michell M *et al.* Guidelines for breast needle core biopsy handling and reporting in breast screening assessment. *J. Clin. Pathol.* 2004; **57;** 897–902.

