PERGAMON

European Journal of Cancer 36 (2000) 1769–1772

European Journal of Cancer

www.ejconline.com

Causes of inconsistency in diagnosing and classifying intraductal proliferations of the breast

European Commission Working Group on Breast Screening Pathology: C.W. Elston [a,*], J.P. Sloane [b¤], I. Amendoeira [c], N. Apostolikas [d], J.P. Bellocq [e], S. Bianchi [f], W. Boecker [g], G. Bussolati [h], D. Coleman [i], C.E. Connolly [j], P. Dervan [k], M. Drijkoningen [l], V. Eusebi [m], D. Faverly [n], R. Holland [o], J. Jacquemier [p], M. Lacerda [q], J. Martinez-Penuela [r], C. de Miguel [s], S. Moss [i], C. Munt [i], J.L. Peterse [t], F. Rank [u], A. Reiner [v], M. Sylvan [w], C.A. Wells [x], B. Zafrani [y]

[a] *Department of Pathology, City Hospital, Nottingham NG5 1PB, UK*
[b] *Department of Pathology, University of Liverpool, Duncan Building, Daulby Street, Liverpool L69 3GA, UK*
[c] *Universidade do Porto, Portugal*
[d] *Saint Savvas Hospital, Athens, Greece*
[e] *Hopital de Hautepierre, Strasbourg, France*
[f] *Azienda Ospedaliera, Florence, Italy*
[g] *Universitat Munster, Germany*
[h] *Instituto di Anatomia e Istologia Patologica, Turin, Italy*
[i] *Cancer Screening Evaluation Unit, Sutton, UK*
[j] *University College Hospital, Galway, Ireland*
[k] *Mater Hospital, Dublin, Ireland*
[l] *University Hospital, Leuven, Belgium*
[m] *Universita di Bologna, Italy*
[n] *CMP Laboratory, Brussels, Belgium*
[o] *Academisch Ziekenhuis, Nijmegen, The Netherlands*
[p] *Institut Paoli Calmettes, Marseille, France*
[q] *Centro Regional De Oncologia De Coimbra, Portugal*
[r] *Hospital de Navarra, Pamplona, Spain*
[s] *Hospital Virgen del Camino, Spain*
[t] *The Netherlands Cancer Institute, Amsterdam, The Netherlands*
[u] *Copenhagen University Hospital, Denmark*
[v] *Donauspital, Vienna, Austria*
[w] *Huddinge University Hospital, Stockholm, Sweden*
[x] *St Bartholomew's Hospital, London, UK*
[y] *Institut Curie, Paris, France*

Received 16 November 1999; received in revised form 2 May 2000; accepted 19 June 2000

Abstract

It is now widely recognised that classifying ductal carcinoma *in situ* (DCIS) of the breast and diagnosing atypical ductal hyperplasia are associated with significant interobserver variation. Two possible reasons for this inconsistency are differences in the interpretation of specified histological features and field selection where morphology is heterogeneous. In order to investigate the relative contribution of these two factors to inconsistent interpretation of intraductal proliferations, histological sections of 32 lesions were sent to 23 European pathologists followed 3 years later by images of small parts of these sections. Kappa statistics for diagnosing hyperplasia of usual type, atypical ductal hyperplasia and ductal carcinoma *in situ* were 0.54, 0.35 and 0.78 for sections and 0.47, 0.29 and 0.78 for images, respectively, showing that most of the inconsistency is due to differences in morphological

* Corresponding author. Tel.: +44-115-691-169, ext. 46871; fax: +44-115-627-768.
E-mail address: chris.elston@nottingham.ac.uk (C.W. Elston).

PII: S0959-8049(00)00181-7

interpretation. Improvements can thus be expected only if diagnostic criteria or methodology are changed. In contrast, kappa for classifying DCIS by growth pattern was very low at 0.23 for sections and better at 0.47 for images, reflecting the widely recognised variation in the growth pattern of DCIS. Higher kappa statistics were obtained when any mention of an individual growth pattern was included in that category, thus allowing multiple categories per case; but kappa was still higher for images than sections. Classifying DCIS by nuclear grade gave kappa values of 0.36 for sections and 0.49 for images, indicating that intralesional heterogeneity has hitherto been underestimated as a cause of inconsistency in classifying DCIS by this method. More rigorous assessment of the proportions of the different nuclear grades present could lead to an improvement in consistency.

Keywords: Breast; Ductal carcinoma *in situ*; Atypical hyperplasia; Non-atypical hyperplasia

1. Introduction

The detection of ductal carcinoma *in situ* (DCIS) of the breast and associated epithelial hyperplasia has increased dramatically since the introduction of mammographic screening. DCIS is a heterogeneous group of proliferations varying in cytological and architectural detail, some of which are thought to have a major influence on clinical outcome [1,2]. The ductal hyperplasias are associated with an increased risk of developing subsequent breast cancer, which is greater for atypical ductal hyperplasia (ADH) than non-atypical hyperplasia (hyperplasia of usual type — HUT) [3]. It is therefore of considerable importance that pathologists are able to make their diagnoses accurately, consistently and reproducibly. However, the classification of DCIS and the epithelial hyperplasias are associated with significant interobserver inconsistency [4–7], which makes prognostication unreliable. The two major reasons for pathological inconsistency are interpretation of the morphological features and selection of appropriate fields for examination. In the former there is disagreement among a group of observers about the diagnosis or classification of selected pathological changes. The latter occurs when diagnoses are based on different parts of a heterogeneous lesion. Many intraduct epithelial proliferations exhibit significant intralesional heterogeneity and the present study was designed to investigate the relative roles of the interpretative and field selection aspects on the lack of consistency in the diagnosis of these lesions, using digitised images selected from a group of previously circulated histological sections.

2. Patients and methods

In the previous study [7] cases were submitted to the co-ordinating centre at the University of Liverpool, UK, and were selected according to certain predetermined diagnostic categories. No selection based on histological appearances was made within the groups, the cases being chosen in strict chronological sequence following a specified accession date. Care was taken to exclude cases with suboptimal section quality and to ensure that the 23 participants received virtually identical sections without significant difference in the histological appearances. One haematoxylin and eosin stained section from each case was sent to 23 members of the European Commission Working Group on Breast Screening Pathology (ECWGBSP) who reported them using a standard proforma and following agreed published guidelines [8]. The ECWGBSP comprises pathologists with a specialist interest in breast pathology from nearly all the member states of the European Union. The slides were not marked in any way and no specific areas were selected for examination. There was no precirculation of a training set of slides nor was any advice given concerning the interpretation of the guidelines.

The criteria for diagnosing epithelial hyperplasia of usual type (HUT) and ADH were those of Page and Rogers [9]. DCIS was classified by nuclear grade following the criteria of the EC guidelines [8].

The present study was based on digitised images from 32 cases of intraductal proliferation from the original study. The histological slides from these cases had been circulated 3 years previously and it was considered highly unlikely that the opinion of the participants then would influence their interpretation of the appearances. One or two digitised images of colour photomicrographs taken at medium or high magnification of one small area were prepared from each of the 32 sections and printed on high quality photographic paper. The areas were selected by the pathologist at the co-ordinating centre and were deemed to be the most representative of the lesions in the sections. The participants were asked to complete a similar proforma to that used in the original study and the responses were analysed using kappa statistics [10] as detailed previously [7].

3. Results

The kappa values for diagnosing the major categories of intraductal epithelial proliferation are shown in Table 1. The values for all categories were similar for each type of preparation and almost identical to those reported in previous studies of DCIS and ADH [4,6,7]. Kappa values for HUT have not been reported

Table 1
Comparison of consistency of making overall diagnoses expressed as kappa statistics for ductal carcinoma *in situ* (DCIS), epithelial hyperplasia of usual type (HUT), atypical ductal hyperplasia (ADH)

	DCIS	HUT	ADH	Overall
Sections: κ (number)[a]	0.78 (463)	0.54 (74)	0.35 (72)	0.60 (609)
Digitised images: κ (number)[a]	0.78 (514)	0.47 (100)	0.29 (92)	0.56 (706)

[a] The numbers in brackets are the total numbers of readings in each category.

previously to our knowledge and are intermediate — between those for ADH and DCIS.

The consistency of classifying DCIS by growth pattern was analysed in two ways. The first was to use a single category per case by assigning to a 'mixed' group those cases where more than one pattern was reported. This gave kappa statistics of 0.23 for histological sections and 0.47 for images, demonstrating an improvement with the latter. In the second, any mention of an individual growth pattern was included in that category, thus allowing multiple categories per case. Notably higher kappa statistics were generally obtained when the analysis was conducted in this way, with values of 0.49 or more except for comedo (Table 2). Consistency was better for images than for histological sections, with the exception of the comedo pattern.

Finally, the kappa values for classifying DCIS according to nuclear grade were also better for the images than histological slides, although the difference was less marked than with architectural pattern (Table 3).

4. Discussion

A number of important conclusions may be drawn from these data. The poor consistency in the diagnosis of ADH must raise serious doubts concerning the robustness of the current diagnostic criteria. This has, of course, been reported previously [4,6,11], although Schnitt and colleagues [12] (whose study included some of the participants in the Rosai publication) found better agreement by using a training set and marking specific areas on a slide. Our use of digitised images fulfilled a similar function to marking specific fields and yet this did not improve the kappa value. There were, however, a larger number of pathologists in our group and a training set was not used in order to make our findings more representative of everyday practice. As discussed in our previous study [7] there is no obvious solution to the problem. External Quality Assessment schemes such as this one are, in essence, somewhat artificial exercises and this must be borne in mind. In routine diagnostic practice, further material would be examined, including deeper levels and additional tissue blocks where appropriate. In our collective experience, this usually resolves the problem, and often reveals more extensive changes that permit a definite diagnosis of DCIS rather than ADH. Nevertheless, it must be recognised that some cases will remain unresolved and it is clear that more robust, internationally agreed criteria for making these diagnoses are required.

In the previous study [7], the participants achieved an acceptably high kappa value for the overall classification of DCIS (0.78) and it is, perhaps, not surprising that an identical figure was obtained using digitised images. Both results indicate very good agreement and seem to show that the general diagnostic criteria for DCIS are sufficiently robust. However, the subclassification of DCIS remains problematical. This study has shown that field selection causes difficulties in classifying DCIS by growth pattern (higher kappa value for digitised images compared with that for histological sections) and also for nuclear grade. The former is less surprising than the latter since it is recognised that growth patterns can vary considerably within lesions, whereas nuclear grade is considered to be relatively constant in the majority of cases. Even with the digitised images, the overall kappa values were only in the average range (0.47 for growth pattern, 0.49 for nuclear grade) and the figures for intermediate nuclear grade with both methods of observation were poor. The data for nuclear grade were close to those obtained recently by Sneige and colleagues [13] in a similar study that investigated the reproducibility of the 'Lagios' nuclear

Table 2
Comparison of kappa values for architectural patterns in ductal carcinoma *in situ* (DCIS) when multiple patterns per case are analysed

	Cribriform	Micropapillary	Solid	Comedo
Sections: κ (number)[a]	0.49 (210)	0.56 (95)	0.50 (120)	0.38 (184)
Digitised images: κ (number)[a]	0.60 (265)	0.85 (82)	0.63 (184)	0.32 (254)

[a] The numbers in brackets are the total numbers of readings in each category.

Table 3
Comparison of kappa values for the classification of ductal carcinoma *in situ* (DCIS) by nuclear grade

	Nuclear grade			
	High	Intermediate	Low	Overall
Sections: κ (number)[a]	0.41 (174)	0.19 (156)	0.51 (105)	0.36 (435)
Digitised: κ (number)[a]	0.57 (211)	0.33 (184)	0.61 (88)	0.49 (483)

[a] The numbers in brackets are the total numbers of readings in each category.

grading system for DCIS. There were fewer pathologists (only six), a prior training set was circulated, diagnostic criteria agreed at a consensus meeting and specific marked areas on glass slides (essentially equivalent to our digitised images) were assessed. Despite this attention to detail the participants only achieved an overall kappa value of 0.46.

In contrast, our study has shown that individual growth patterns can be recognised with moderate to substantial consistency except, surprisingly, comedo necrosis. There appear to be two main reasons for the latter: firstly the difficulty in distinguishing comedo from non-comedo necrosis, which suggests that more defined criteria are required; secondly, the distinction between necrosis and intraluminal secretion. The higher kappa values for growth pattern recognition in the digitised images compared with the slides suggests that minority growth patterns tend to be overlooked or ignored by observers in the latter. Furthermore, the kappa value for selecting single categories of growth patterns on slides (0.23) was identical to that previously reported by the UK National Co-ordinating Group for Breast Screening Pathology [6] and shows poor consistency even when a mixed category is used. However, it should also be pointed out that the relevance of these data to clinical patient management will only become apparent when the results of ongoing randomised clinical trials are analysed. Present evidence indicates that cytological features are of greater importance in predicting clinical outcome than architecture [2].

In general, the data from the analysis of the digitised images are encouraging. They show that this medium can generate images of sufficiently high quality for teaching, consultation, consistency analysis and external quality assessment. Field selection can, however, lead to artificially high levels of consistency and overestimation of performance unless an adequate number of representative fields is selected. The study has suggested that some improvement in the subclassification of DCIS can be achieved, although there may be a plateau in the level of agreement among observers in the moderate/good range. Indeed, it may be unrealistic to expect a higher level of agreement until more clearly defined diagnostic criteria are agreed or alternative methods are found. For example, there needs to be more rigorous assessment of the proportions of the different elements present, stricter adherence to classification guidelines and a clearer definition of comedo necrosis. In contrast, it is difficult to see how histological criteria for diagnosing ADH can be improved using existing methodology. The solution may lie in the application of molecular pathology, as much is being learned about the molecular changes associated with the early development of human breast cancer [14,15].

References

1. Silverstein MJ, Poller DN, Waisman JR, *et al*. Prognostic classification of breast ductal carcinoma *in situ*. *Lancet* 1995, **345**, 1154–1157.
2. Badve S, A'Hern R, Ward AM, *et al*. A long-term comparative study of the ability of five classifications of ductal carcinoma *in situ* of breast to predict local recurrence after surgical excision. *Hum Pathol* 1992, **29**, 915–923.
3. Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. *N Engl J Med* 1985, **312**, 146–151.
4. Rosai J. Borderline epithelial lesions of the breast. *Am J Surg Pathol* 1991, **15**, 209–221.
5. Douglas-Jones AG, Gupta SK, Attanoos RL, *et al*. A critical appraisal of six modern classifications of ductal carcinoma *in situ* of the breast (DCIS): correlation with grade of associated invasive carcinoma. *Histopathology* 1996, **29**, 397–409.
6. Sloane JP, Ellman R, Anderson TJ, *et al*. Consistency of histopathological reporting of breast lesions detected by screening: findings of the UK National EQA Scheme. *Eur J Cancer* 1994, **30A**, 1414–1419.
7. European Commission Working Group on Breast Screening Pathology. Consistency achieved by 23 European pathologists in categorizing ductal carcinoma *in situ* of the breast using five classifications. *Hum Pathol* 1998, **29**, 1056–1062.
8. European Commission. European Guidelines for Quality Assurance in Mammography Screening, 2nd edn. Office for Official Publications of the European Communities, 1996, II-C-15–II-C-16.
9. Page DL, Rogers LW. Combined histologic and cytologic criteria for the diagnosis of mammary atypical hyperplasia. *Hum Pathol* 1992, **23**, 1095–1097.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977, **33**, 159–174.
11. Wells WA, Carney PA, Eliassen MS, *et al*. Statewide study of diagnostic agreement in breast pathology. *J Natl Cancer Inst* 1998, **90**, 142–145.
12. Schnitt SJ, Connolly JL, Tavassoli FA, *et al*. Interobserver reproducibility in the diagnosis of ductal proliferative breast lesions using standardised criteria. *Am J Surg Pathol* 1992, **16**, 1133–1143.
13. Sneige N, Lagios MD, Schwarting R, *et al*. Interobserver reproducibility of the Lagios nuclear grading system for ductal carcinoma *in situ*. *Hum Pathol* 1999, **30**, 257–262.
14. Lakhani SR, Collins N, Stratton MR, Sloane JP. Atypical ductal hyperplasia of the breast: clonal proliferation with loss of heterozygosity on chromosomes 16q and 17p. *J Clin Pathol* 1995, **48**, 611–615.
15. Lakhani SR. The transition from hyperplasia to invasive carcinoma of the breast. *J Pathol* 1999, **187**, 272–278.