Virchows Arch (1999) 434:3–10

ORIGINAL ARTICLE

European Commission Working Group on Breast Screening Pathology: J.P. Sloane (Chairman)
I. Amendoeira · N. Apostolikas · J.P. Bellocq
S. Bianchi · W. Boecker · G. Bussolati · D. Coleman
C.E. Connolly · V. Eusebi · C. De Miguel · P. Dervan
R. Drijkoningen · C.W. Elston · D. Faverly · A. Gad
J. Jacquemier · M. Lacerda · J. Martinez-Penuela
C. Munt · J.L. Peterse · F. Rank · M. Sylvan
V. Tsakraklides · B. Zafrani

Consistency achieved by 23 European pathologists from 12 countries in diagnosing breast disease and reporting prognostic features of carcinomas

Abstract A detailed analysis of the consistency with which pathologists from 12 different European countries diagnose and classify breast disease was undertaken as part of the quality assurance programme of the European Breast Screening Pilot Network funded by the Europe against Cancer Programme. Altogether 107 cases were examined by 23 pathologists in 4 rounds. Kappa (κ) statistics for major diagnostic categories were: benign (not otherwise specified) 0.74, atypical ductal hyperplasia (ADH) 0.27, ductal carcinoma in situ (DCIS) 0.87 and invasive carcinoma 0.94. ADH was the majority diagnosis in only 2 cases but was diagnosed by at least 2 participants in another 14, in 9 of which the majority diagnosis was benign (explaining the relatively low κ for this category), DCIS in 4 (all low nuclear

J.P. Sloane (✉)[1]
University of Liverpool, United Kingdom

I. Amendoeira
Instituto De Patologia & Immunologia Molecular,
Da Universidade Do Porto, Portugal

N. Apostolikas
Saint Savvas Hospital, Athens, Greece

J.P. Bellocq
Hopital de Hautepierre, Strasbourg, France

S. Bianchi
Instituto di Anatomia e Istologia Patologica, Firenze, Italy

W. Boecker
Gerhard-Domagk Institut fur Pathologie, Munster, Germany

G. Bussolati
Instituto di Anatomia e Istologia Patologica, Torino, Italy

D. Coleman · C. Munt
Cancer Screening Evaluation Unit, Sutton, Surrey,
United Kingdom

C.E. Connolly
Clinical Sciences Institute, University College Hospital, Galway,
Ireland

V. Eusebi
Universita di Bologna, Italy

C. De Miguel
Hospital Virgen del Camino, Pamplona, Spain

P. Dervan
Mater Hospital, Dublin, Ireland

R. Drijkoningen
UZ St. Rafael, Leuven, Belgium

C.W. Elston
City Hospital, Nottingham, United Kingdom

D. Faverly
CMP Laboratory, Bruxelles, Belgium

A. Gad · M. Sylvan
Huddinge University Hospital, Stockholm, Sweden

J. Jacquemier
Institut Paoli Calmettes, Marseille, France

M. Lacerda
Centro Regional De Oncologia De Coimbra, Portugal

J. Martinez-Penuela
Hospital de Navarra, Pamplona, Spain

J.L. Peterse
The Netherlands Cancer Institute, Amsterdam, The Netherlands

F. Rank
Rigshospitalet, Copenhagen, Denmark

V. Tsakraklides
Hygeia Hospital, Athens, Greece

B. Zafrani
Institut Curie, Paris, France

Mailing address:
[1] Department of Pathology, University of Liverpool,
Duncan Building, Daulby Street, Liverpool, L69 3GA,
United Kingdom

grade) and invasive carcinoma (a solitary 1-mm focus) in 1. The histological features of these cases were extremely variable; although one feature that nearly all shared was the presence of cells with small, uniform, hyperchromatic nuclei and a high nucleo-cytoplasmic ratio. The majority diagnosis was DCIS in 33 cases; κ for classifying by nuclear grade was 0.38 using three categories and 0.46 when only two (high and other) were used. When ADH was included with low nuclear grade DCIS there was only a slight improvement in κ. Size measurement of DCIS was less consistent than that of invasive carcinoma.The majority diagnosis was invasive carcinoma in 57 cases, the size of the majority being 100% in 49. The remainder were either special subtypes (adenoid cystic, tubular, colloid, secretory, ductal/medullary) or possible microinvasive carcinomas. Subtyping was most consistent for mucinous (κ, 0.92) and least consistent for medullary carcinomas (κ, 0.56). Consistency of grading using the Nottingham method was moderate (κ=0.53) and consistency of diagnosing vascular invasion, fair (κ=0.38). There was no tendency for consistency to improve from one round to the next, suggesting that further improvements are unlikely without changes in guidelines or methodology.

Key words Breast · Invasive carcinoma · Ductal carcinoma in situ · Atypical hyperplasia · Diagnostic consistency · Kappa statistics

Introduction

In order to encourage breast cancer screening in the European Union, the European Commission set up a Pilot Breast Screening Network, funded under the Europe against Cancer Programme. Rigorous quality assurance (QA) arrangements covering all professional disciplines were put in place to ensure that the highest possible standards were reached. The European Commission Working Group on Breast Screening Pathology (ECWGBSP) was formed to deal with the pathological aspects of QA. The group produced guidelines on reporting breast specimens which were derived from those already published in the UK [4, 9] and set up a slide exchange type of external quality assessment (EQA) scheme involving, amongst others, the pathologists working in the centres funded under the pilot programme. As part of its activities the Working Group has undertaken detailed studies of the consistency with which its members diagnose breast diseases and report prognostic features.

The present study had three aims. The first was to determine whether an adequate level of consistency could be achieved by 23 pathologists from 12 European countries in diagnosing major categories of breast disease and reporting those histological features of prognostic significance that are used for determining how individual patients should be managed. This was important if results from different European screening centres were to be compared. The second was to compare the findings with those obtained several years previously in a similar study by a similar number of pathologists from the UK. This would enable us to determine whether there were any major international differences in pathological reporting among the different countries of the European Union. The third was to determine whether the considerable improvements in the guidelines used for the previous UK study had resulted in any improvement in diagnostic consistency.

Materials and methods

Cases were submitted to the co-ordinating centre (University of Liverpool) by the 23 members of the Working Group and were selected according to certain predetermined diagnoses made at the referring centres: benign (not otherwise specified), atypical ductal hyperplasia (ADH), ductal carcinoma in situ (DCIS) of high, intermediate and low nuclear grades and invasive carcinoma of different types and grades. No selection based on histological appearance was made within these groups, the cases being chosen in strict chronological sequence following a specified accession date. Blocks were not used if sections of adequate quality could not be prepared from them or if the lesions they contained were of inadequate size to prepare 23 virtually identical sections without significant change in the lesion's size or histological characteristics. One H&E-stained section from each case was then sent to each member of the Working Group, who reported them using a pro forma and following the guidelines published by the Working Group [4]. These guidelines were derived from those produced for the UK National Health Service Breast Screening Programme and are identical to those now used in the UK [9]. The pro-forma was analysed electronically by the Cancer Screening Evaluation Unit, Sutton, Surrey, UK. The slides were not marked in any way and no specific areas were selected. A learning set of slides was not circulated at the beginning and there was no detailed dis cussion about how the guidelines should be followed before the study began. The cases were, however, discussed in detail after each circulation at Working Group meetings. Altogether 107 sets of H&E-stained sections were examined in four circulations.

The criteria for diagnosing ADH in the guidelines were those of Page and Rogers [11]. The classification of DCIS was based entirely on nuclear grade, as defined by Holland et al. [6], but cell polarisation was not taken into account. Invasive carcinomas were graded by the Nottingham method [3]. In situ and invasive carcinomas were measured on the circulated slides. The maximum diameter was recorded and was defined as the greatest distance between two points on the periphery of the lesion. No guidance was given on the method of performing the measurements, which was thus subject to some variation e.g. using the Vernier scale on the microscope stage, using a ruler to measure the lesion directly, with or without marking the slide.

The agreement between participants on the categorization of cases using each classification was measured by calculating κ statistics, which take into account the level of agreement expected purely by chance, and which also require no knowledge of the true diagnosis. For 2-way classifications there is only a single value of κ but for consistency of presentation this is given in all columns of the tables in this paper. For the 3-way classifications, an overall κ value was calculated from the κ values for individual categories, weighted by the proportion of reports in each category. Values of κ range from 0 for chance agreement only to +1 for perfect agreement, with a negative value implying systematic disagreement. Landis and Koch [8] suggest the following interpretation of different ranges of κ: 0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, 0.81–1.00 almost perfect. One disadvantage of κ statistics is their dependence on the prevalence of cases in each category; in particular, this will influence comparisons between different circulations.

Results

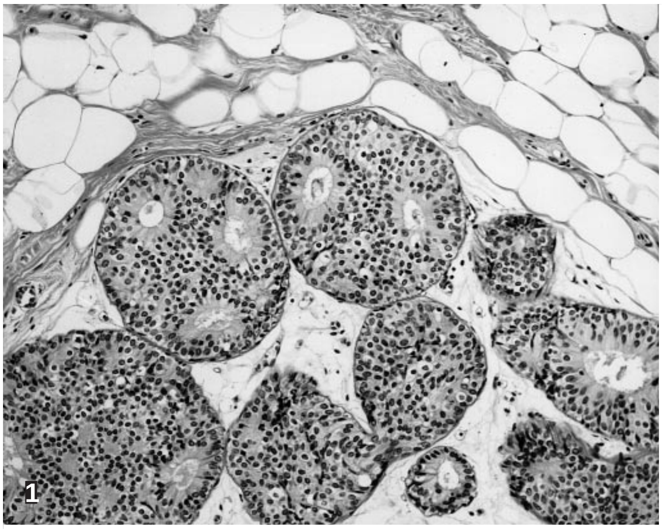
Overall Diagnoses

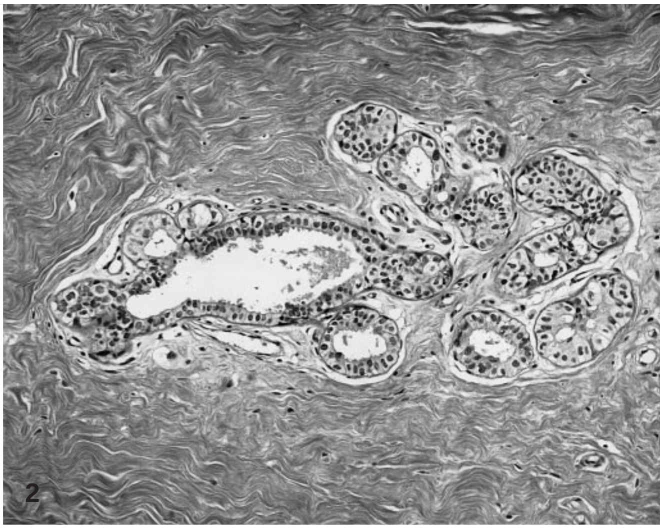
The κ statistics for the overall diagnoses are summarised in Table1

The majority diagnosis was *benign* in 15 cases, and the size of the majority varied between 40% and 100% of the participants (mean 78%). The κ statistic for all four circulations was 0.74. The relatively low level of consistency associated with this diagnosis was largely explained by cases originally selected as ADH but diagnosed as benign by the majority of participants.

ADH was the majority diagnosis in only 2 cases, where the majorities were 65% and 50% of the total numbers of readings. The κ statistic for all four circulations was 0.27. ADH was, however, diagnosed by at least 2 participants in another 14 cases, in which the majority diagnosis was benign in 9, DCIS in 4 and invasive carcinoma in 1 (where the invasive component was restricted to a solitary 1-mm focus). The percentage of atypical hyperplasia diagnoses ranged from 15% to 32% (mean 23%) in the benign cases and from 9% to 42% (mean 18%) in the DCIS cases, and was 10% in the case of invasive carcinoma. All 4 cases of DCIS were of low nuclear grade. In 6 of the 9 benign cases with diagnoses of ADH there were also diagnoses of DCIS (range 5–25%, mean 7%). The converse was also true, with 2 of the 4 DCIS cases with diagnoses of ADH also having benign diagnoses (5% in each case). Thus, of the 15 cases in the present study in which the majority diagnosis was benign, at least 2 of the 23 diagnoses were of ADH in 9 cases.

The histological features of the cases where at least 2 diagnoses of ADH were made were extremely variable. All those where the majority diagnosis was DCIS exhibited some degree of dilatation of the involved structures and a micropapillary or cribriform growth pattern (Fig. 1). Those cases where the majority diagnosis was benign were even more variable; at one extreme there was no expansion of the structures involved or any significant intraluminal proliferation (as in Fig. 2), whereas at the other there was significant intraluminal proliferation with cribriform or micropapillary growth patterns (Fig. 3). One feature that nearly all cases had in com-





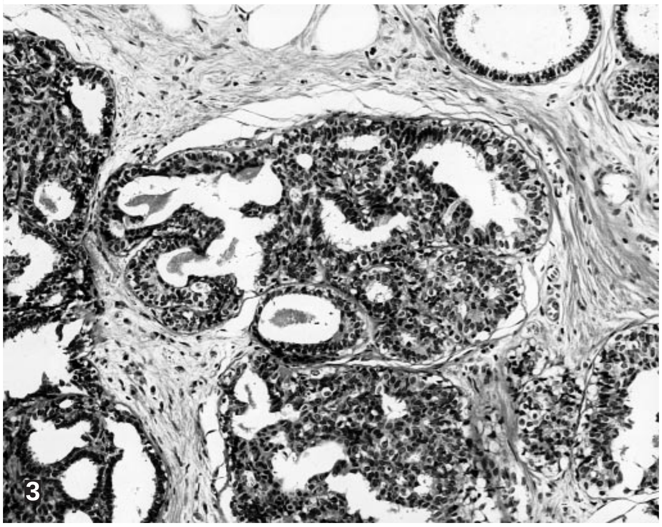


Fig. 1 In this case, 90% of participants made a diagnosis of DCIS and 10%, one of ADH. There is marked distension of the acini, which are filled with cells with small, uniform, hyperchromatic nuclei arranged in a well-developed cribriform growth pattern with pronounced cell polarisation around the secondary lumina. H&E

Fig. 2 In this case, 10% of the pathologists recorded a diagnosis of benign disease, 65% one of ADH and 25% one of DCIS. Numerous terminal ductlobular units were lined with cells with small, uniform, hyperchromatic nuclei and generally low (but somewhat variable) nucleo-cytoplasmic ratio. There is no significant intraluminal proliferation, however, and consequently no lobular distension. H&E

Fig. 3 In this case, 75% of diagnoses were of benign disease, 20% of ADH and 5% of DCIS. The cells are essentially similar to those seen in Fig. 2, but there is significant intraluminal proliferation with cribriform and micropapillary growth patterns, sometimes associated with delicate fibrovascular stroma. Compared with Fig. 1, the distribution of nuclei is uneven and there may be more than one cell type. H&E

Table 1 Consistency of making overall diagnoses expressed as κ statistics (*AH* Atypical ductal hyperplasia, *In situ/micro* in situ or microinvasive carcinoma, *Invasive* invasive carcinoma)

Round	Diagnosis				
	Benign	ADH	In situ/ micro	Invasive	Overall
1	0.66	0.17	0.83	0.96	0.79
2	0.73	0.29	0.91	0.95	0.87
3	0.83	0.33	0.87	0.97	0.86
4	0.50	0.29	0.84	0.91	0.80
All 4	0.74	0.27	0.87	0.94	0.84

Table 2 Consistency of classifying DCIS into three nuclear grades expressed as κ stastistics

Round	Nuclear grade			
	High	Intermediate	Low	Overall
1	0.47	0.16	0.44	0.35
2	0.40	0.10	0.53	0.33
3	0.38	0.13	0.70	0.41
4	0.44	0.17	0.24	0.28
All 4	0.43	0.17	0.49	0.35
All 4 (with AH included with low nuclear grade)	0.44	0.18	0.55	0.38

Table 3 Consistency of classifying DCIS into two nuclear grades, expressed as κ stastistics.

Nuclear grade			
Round	High	Other	Overall
1	0.48	0.48	0.48
2	0.47	0.47	0.47
3	0.40	0.40	0.40
4	0.42	0.42	0.42
All 4	0.46	0.46	0.46

mon, however, was the presence in the dominant cells of small, uniform, hyperchromatic nuclei and a high nucleo-cytoplasmic ratio; that is, they bore some resemblance to those of low nuclear grade DCIS.

DCIS was the majority diagnosis in 33 cases, the size of the majority ranging from 53% to 100% (mean 95%). The overall κ statistic was 0.87.

The majority diagnosis was *invasive carcinoma* in 57 cases, the size of the majority varying from 80% to 100% (mean 99%). The overall κ statistic was 0.94. A majority of 100% was encountered in 49 cases. Of the remainder, 5 were special subtypes (adenoid cystic, tubular, colloid, secretory, ductal/medullary) and 2 were very small carcinomas on which opinion was divided between invasive and microinvasive.

Table 1 shows that there was no tendency for consistency to improve from one round to the next.

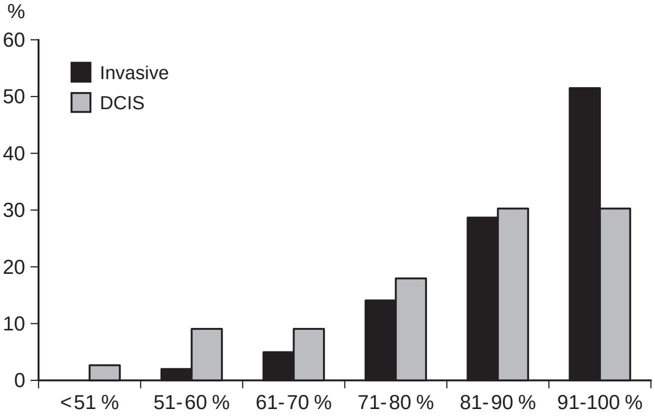


Fig. 4 Percentage of size measurements ±3 mm of the median in 57 cases of invasive carcinoma and 33 cases of DCIS

In situ carcinoma

All cases were selected as DCIS and all were classified as such. Two aspects were studied: (1) consistency of classification using the system adopted by the Working Group based on nuclear grade [4] and (2) consistency of measuring the maximal diameter.

Classification

The results are summarised in Tables 2 and 3. Three points are worthy of note: (1) intermediate-grade DCIS was diagnosed very much less consistently than high- or low-grade DCIS; (2) when atypical hyperplasia was included with low nuclear grade DCIS there was only a slight improvement in the κ statistic; and (3) a higher overall level of consistency was obtained when a 2-way (high vs other) rather than a 3-way system (high, intermediate and low) was used. There was no tendency for consistency to improve from one round to the next.

Size measurement

The consistency with which the maximal diameter of DCIS was measured is summarised in Fig. 4, where the results are expressed as the proportion of cases falling into groups defined by the percentage of measurements within 3 mm of the median. Thus, in only 60% of cases were at least 80% of measurements made by all 23 participants within 3-mm of the median. Discussions of individual cases following the slide circulations revealed that the main reasons for differing size measurements were poor circumscription and accompanying ADH, with which the DCIS merged, creating uncertainty about the precise boundaries of the latter process. There was no clear relationship between consistency of measurement and nuclear grade of the lesion. The mean percentage measurements within 3 mm of the median were

Table 4 Consistency of subtyping invasive carcinomas, expressed as κ statistics

Round	NST	Lobular	Medullary	Tubular	Mucinous	Other	Overall
1	0.57	0.74	0.45	0.73	0.97	–	0.62
2	0.46	0.71	0.78	0.56	0.93	0.16	0.56
3	0.58	0.80	0.48	0.45	0.83	0.26	0.58
4	0.37	0.82	0.31	–	–	0.43	0.48
All 4	0.51	0.76	0.56	0.61	0.92	0.25	0.58

Table 5 Consistency of grading invasive carcinomas expressed as κ statistics

Round	Grade 1	Grade 2	Grade 3	Overall
1	0.60	0.34	0.72	0.55
2	0.67	0.43	0.68	0.59
3	0.45	0.23	0.66	0.43
4	0.29	0.39	0.81	0.51
All 4	0.56	0.35	0.70	0.53

Table 6 Consistency of diagnosing vascular invasion expressed as κ statistics

Round	Present	Not seen	Overall
1	0.36	0.36	0.36
2	0.29	0.29	0.29
3	0.51	0.51	0.51
4	0.33	0.33	0.33
All 4	0.38	0.38	0.38

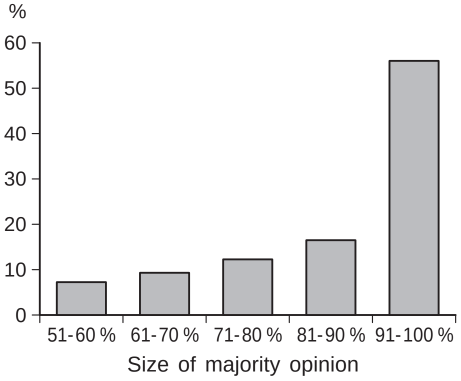


Fig. 5 Percentage agreement on the presence of vascular invasion

82% (range 55–100%) for high, 82% (range 60–95%) for intermediate and 70% (range 25–100%) for low nuclear grade. The lower mean value for low nuclear grade cases was due to their relatively small number of 6 and one very low value of 25%. Figure 4 shows that DCIS was measured with less consistency than invasive carcinoma.

Invasive Carcinoma

Several features of prognostic significance were evaluated.

Subtype

The majority diagnosis was ductal/no special type (NST) carcinoma in 27 cases, lobular in 10, medullary or atypical medullary in 6, tubular in 5 and mucinous in 5. The size of the majority for ductal/NST carcinoma ranged from 45% to 95% (mean 78%), that for lobular carcinoma from 65% to 100% (mean 86%), that for medullary/atypical medullary carcinoma from 55% to 95% (mean 78%), that for tubular carcinoma from 45% to 91% (mean 68%), and that for mucinous carcinoma from 65% to 100% (mean 90%). The κ statistics for the four rounds are summarised in Table 4. There was no improvement in overall κ from one round to the next.

Grade

The majority grade was 1 in 23 cases, 2 in 19 cases and 3 in 15 cases. The size of the majority varied from 50% to 100% (mean 79%) for grade 1, from 52% to 94% (mean 71%) for grade 2, and from 60% to 100% (mean 87%) for grade 3 tumours. The κ statistics for the four rounds are summarised in Table 5. There was no improvement in overall κ from one round to the next.

Vascular invasion

The consistency with which vascular invasion was identified is expressed as κ statistics in Table 6. Kappa for all four circulations was 0.38. Figure 5 shows that this low value is explained by a high level of disagreement in a relatively small number of cancers. Over 90% of participants agreed on the presence or absence of vascular invasion in more than half the cases.

Size

The size of the carcinoma was measured on the histological sections. In those cases where the whole tumour was not included in the section the maximal diameter of the part of it that was present was measured. The consistency with which the invasive tumour sizes were measured is summarised in Fig. 4, where it is compared with that observed for DCIS. The results are expressed as the proportion of cases falling into groups defined by the percentage of measurements that were within 3 mm of the median. At least 80% of measurements were within 3 mm of the median in 79% of cases.

Discussion

This investigation was similar to that previously reported by the UK National Co-ordinating Group for Breast Screening Pathology in terms of the study design, the types of analyses performed and the number of pathologists involved [17]. There were, however, several major differences. First, the present study involved an international group of pathologists from 12 European countries. Second, new histological features were assessed, including the subtype of invasive carcinoma, vascular invasion and a new classification of DCIS. Third, the group were following new guidelines with extensively rewritten sections on ADH, DCIS and grading and measuring invasive carcinomas.

The κ statistics for the major diagnostic categories were virtually identical to those achieved by the 22 coordinators in the UK study (benign 0. 76, atypical hyperplasia 0.25, in situ/microinvasive carcinoma 0.81, invasive carcinoma 0.94). This and the ability of the group to agree on a set of guidelines for use in the EC-supported breast screening programmes indicate that there are no significant differences among the countries of the European Union in the way breast disease is diagnosed and classified. A second, and less encouraging point is that there appear to have been no major improvements in consistency since the UK study was undertaken, although poor performance in diagnosing major categories is almost entirely due to persistent difficulties in diagnosing ADH and related intraductal proliferations, which formed a significantly greater proportion of cases than would be encountered in everyday practice.

The diagnostic criteria for atypical ductal hyperplasia adopted in the guidelines were those of Page and Rogers [11], who regard ADH as a positive diagnosis rather than one of exclusion. The disorder is defined as partial involvement of a basement-bound space by cells indistinguishable from those of low nuclear grade DCIS. Usually the second, nonatypical cell population consists of columnar, polarised cells located immediately above the basement membrane. Where there is doubt about whether the diagnosis should be atypical hyperplasia or DCIS, the more benign diagnosis is appropriate. To qualify as ADH as opposed to florid nonatypical hyperplasia the characteristic cells should comprise an entire nontapering bar crossing a space or a group of at least six or seven cells. Defined in this way, ADH is almost invariably a tiny lesion not exceeding 3 mm in its maximal dimension. The small size of the proliferation made it difficult to find suitable lesions from which the required number of sections could be cut, but none of the cases analysed exhibited major histological differences among the circulated slides.

We calculated the consistency which would have been achieved if ADH had been included in the same diagnostic category as low nuclear grade DCIS. The rationale for doing this was based on several considerations. First, at least some of the cells in ADH are identical to those of low nuclear grade DCIS following the criteria used [11]. Second, recent studies have shown that ADH shares molecular genetic abnormalities in common with DCIS and have provided evidence that it is a clonal (neoplastic) disorder [7]. Finally, it has been suggested that ADH and DCIS should be incorporated into a single classification of ductal intraepithelial neoplasia [19]. A modest improvement in the diagnostic consistency of low nuclear grade DCIS was achieved when ADH was included and a slight overall improvement in the consistency with which DCIS was classified (see Table 2). No further improvement could be achieved because atypical hyperplasia diagnoses were more likely to be encountered in cases where the majority diagnosis was benign. The data in the present study thus provide no obvious solution to this problem. Some improvement in consistency might be achieved if greater emphasis were placed on cytological features as virtually all cases in which at least 10% of diagnoses were ADH were characterised by cells with small, uniform, hyperchromatic nuclei and a high nucleo-cytoplasmic ratio. It is clear, however, that the present diagnostic criteria are not sufficiently robust to enable an acceptable level of consistency to be achieved among a reasonable number of pathologists. This view is supported by a recent study assessing diagnostic agreement among community-based general pathologists in the USA, where a κ stastistic of 0.22 was obtained for atypical hyperplasia [20].

The lack of agreement in diagnosing ADH contrasts with the high level of consistency achieved by the group in other areas, particularly in diagnosing DCIS, where the overall κ statistic was 0.87, somewhat greater than that obtained in the UK study. This is particularly gratifying given the frequency with which this disorder is detected in mammographic screening. Not surprisingly, however, classifying DCIS was not associated with the same degree of reproducibility. The overall κ statistics for the 3-way and 2-way systems based on nuclear grade were 0.38 and 0.46, respectively. These values are not particularly high and suggest that further refinements of histological classification are necessary. Nevertheless, they represent a significant improvement over the 0.23 obtained by the UK Working Group with the old system based entirely on growth pattern.

The classification was investigated using two and using three categories, as it is simpler to use a 2-way system and evaluate its clinical significance, given the relatively small number of clinical events that occur after excision of DCIS. Another reason was that the middle category of a 3-way system could be associated with a lower level of diagnostic consistency than the other two because extremes are easier to recognise. This contention was supported by the present study. Reducing the number of categories does not automatically improve the κ statistics, which take into account the number of categories used. The 2- and 3-way versions of the system were applied prospectively, which accounts for the slightly different κ statistic for high nuclear grade in each version. Retrospective examination of the 3-way version indicates that the overall κ might have been as high or even higher if the division had been between low nuclear grade and the remainder. The relatively few cases of low nuclear grade DCIS would make such a classification very unevenly balanced, however, even though it might (arguably) have greater clinical and biological significance. Our findings on the relative consistency achieved using different classifications of DCIS are reported elsewhere [18].

Greater consistency was achieved in measuring invasive carcinomas than DCIS, mainly because the latter sometimes merged with intaductal proliferation more in keeping with ADH and this gave rise to uncertainty about the extent of the lesion. In these circumstances, some participants included all the intraductal proliferation in the measurement, whereas others attempted to define the boundary between the two processes. Another problem was poor circumscription. Size variation from one slide to another was not a major factor. Overall, at least 80% of measurements were within 3 mm of the median in 45 of the 57 invasive carcinomas and in 20 of the 33 cases of DCIS. The consistency of measuring the size of invasive carcinomas was similar to that obtained in the previous UK study, but that associated with DCIS, was significantly greater in the present investigation. It is important to bear in mind the difficulties in measuring the size of DCIS as this feature is related to the risk of recurrence after local excision and has been incorporated into a recently reported prognostic index [16].

We are not aware of any previous study that might have evaluated the ability of a large group of pathologists to subtype a range of invasive carcinomas. Mucinous carcinomas were identified with a very high level of consistency (κ 0.92), followed by lobular carcinomas (κ 0.76). A lower level of reproducibility was found with tubular carcinomas (κ 0.61); this is somewhat surprising given their distinctive appearance, but is explained by the uncertainty with which they are sometimes distinguished from grade 1 ductal/NST carcinomas. This is not, however, a particularly important problem from the prognostic point of view. The relatively low κ value associated with ductal/NST carcinomas reflects the fact that they enter into the differential diagnosis with all the other types. These findings vindicate the reporting of these variants of invasive carcinoma which have been shown to have been shown to have prognostic significance [12].

There have been several previous reports documenting inconsistency in diagnosing medullary carcinoma, the subtype associated with the lowest level of consistency in the present study(κ 0.56) . This is perhaps surprising, given the tumour's striking appearance and the attention that has been given to its diagnostic criteria. In the study of Rigaud et al. [14], 9 pathologists examined 16 cases originally diagnosed as medullary carcinoma, using the criteria of Ridolfi et al. [13]. Both inter- and intraobserver agreement were relatively low, with κ values of less than 0.5. The only histological criterion on which there was more than 50% agreement was the presence or absence of in situ carcinoma. The authors concluded that although the criteria of Ridolfi et al. are clear and detailed, they are not easy to apply, and that medullary, atypical medullary and invasive ductal (NST) carcinomas form a continuous spectrum rather than discrete entities. This at least partly explains why it is not generally agreed that medullary carcinoma is associated with an excellent prognosis [12, 13].

Vascular invasion has been shown to be a powerful predictor of lymph node status and to be related to the probability of developing recurrent breast carcinoma in both node positive and negative patients [2, 10, 15]. There have been previous studies studies of the consistency with which vascular invasion is recognised, but we are unaware of one involving as many observers as the present one. In that of Gilchrist et al. [5], several slides from each of 35 node-negative modified radical mastectomy specimens were examined by three pathologists. All three concurred on the presence or absence of intralymphatic tumour in only 12 (34%) of the 35 cases, which led the authors to conclude that the identification of intralymphatic disease is not a reliably reproducible prognostic finding on which to base a recommendation for systemic chemotherapy. In the later study by Orbo et al. [10], however, two pathologists achieved a κ statistic of 0.6 in identifying lymphatic invasion in 95 invasive carcinomas. The small numbers of pathologists involved in these two studies is probably a major reason for their disparate findings. The data in the present study are more in keeping with those of Gilchrist et al., although they represent an improvement as we obtained complete agreement on vascular invasion in 22 (39%) of cases (data not shown) with a larger number of observers.

Discussions of circulated slides revealed two major reasons for the lack of agreement: (1) differences in interpretation and (2) sampling problems. The former resulted from the difficulties of distinguishing clumps of tumour cells in small vessels from those in artefactual spaces produced by retraction. Immunohistological staining can help in cases where the morphological appearances are equivocal. In the present study the evaluation was made on one H&E-stained slide only. The latter were almost invariably encountered when unequivocal vascular invasion was limited to one or two vessels (generally but not invariably of small calibre), which were consequently present in some sections but not others. This problem can be overcome to some extent by taking several blocks of tumour, particularly from the periphery where vascular invasion is easier to recognise. The level of consistency we encountered in the present study is thus likely to be an underestimate of what can be achieved in everyday practice.

The level of consistency we observed in grading invasive carcinomas was somewhat higher than that reported from the previous UK study (overall κ 0.53 vs 0.46). The guidelines had been improved, however, and were significantly more detailed for the present investigation. Table 5 shows that grade 1 and 3 tumours were reported more consistently than grade 2 tumours, demonstrating the greater ease of recognising the extremes. Discussion of individual cases revealed, unsurprisingly, that most problems were encountered with tumours on the borderlines between grades 1 and 2 and grades 2 and 3 (overall scores

5/6 and 7/8). Dalton et al. [1] sent one slide from each of 10 invasive breast carcinomas to 25 pathologists working in six centres. The slides were selected from 30 cases of invasive carcinomas of no special type to represent a spectrum of differentiation. Section quality was adequate in all cases. The slides were accompanied by a written description of the Nottingham grading system. There was unanimity in 3 cases and more than 87% agreement in another 6. The median weighted κ was 0.7. This is higher than that obtained in the present study, but fewer cases were studied and fewer centres were involved. It is presently difficult to see how the guidelines used in the present study can be improved much further, and the degree of consistency we have achieved is probably similar to or greater than that achievable in everyday practice. It would seem advisable to place greater weight on grades 1 and 3 than grade 2 in planning patient management and to use grade in combination with other prognostic features.

References

1. Dalton LW, Page DL, Dupont WD (1994) Histologic grading of breast carcinoma. Cancer 73:2765–2770
2. Davis BW, Gelber R, Goldhirsh A, Hartmann WH, Holloway L, Russell I, Rudenstam CM (1985) Prognostic significance of peritumoral vessel invasion in clinical trials of adjuvant therapy for breast cancer with axillary lymph node metastasis. Hum Pathol 16:1212–1218
3. Elston CW, Ellis IO (1991) Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long term follow up. Histopathology 19:403–410
4. European Commission. (1996) European guidelines for quality assurance in mammography screening, 2nd edn. Office for Official Publications of the European Communities, Luxembourg, pp II-C-15–II-C-16
5. Gilchrist KW, Gould VE, Hirschl S, Imbriglia JE, Patchefsky AS, Penner DW, Pickren J, Schwartz IS, Wheeler JE, Barnes JM, Mansour EG (1982) Interobserver variation in the identification of breast carcinoma in intramammary lymphatics. Hum Pathol 13:170–172
6. Holland R, Peterse JL, Millis RR, Eusebi V, Faverly D, van de Vijver MJ, Zafrani B (1994) Ductal carcinoma in situ: a proposal for a new classification. Semin Diagn Pathol 11:167–180
7. Lakhani SR, Collins N, Stratton MR, Sloane JP (1995) Atypical ductal hyperplasia of the breast: a clonal proliferation with loss of heterozygosity on chromosomes 16q and 17p. J Clin Pathol 48:611–615
8. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
9. National Co-ordinating Group for Breast Screening Pathology (1997) Pathology reporting in breast cancer screening, 2nd edn. NHSBSP publications, Sheffield
10. Orbo A, Stalsberg H, Kunde D (1990) Topographic criteria in the diagnosis of tumor emboli in intramammary lymphatics. Cancer 66:972–977
11. Page DL, Rogers LW (1992) Combined histologic and cytologic criteria for the diagnosis of mammary atypical ductal hyperplasia. Hum Pathol 23:1095–1097
12. Pereira H, Pinder SE, Sibbering DM, Galea MH, Elston CW, Blamey RW, Robertson JFR, Ellis IO (1995) Pathological prognostic factors in breast cancer. Should you be a typer or grader? A comparative study of two histological prognostic features in operable breast carcinoma. Histopathology 27:219–226
13. Ridolfi RL, Rosen PP, Port A., Kinne D, Mike V (1977) Medullary carcinoma of the breast: a clinicopathologic study with 10-year follow-up. Cancer 40:1365–1385
14. Rigaud C, Theobald S, Noel P, Badreddine J, Barlier C, Delobelle A, Gentile A, Jacquemier J, Maisongrosse V, Peffault de Latour M, Trojani M, Zafrani B (1993) Medullary carcinoma of the breast: A multicenter study of its diagnostic consistency. Arch Pathol Lab Med 117:1005–1007
15. Roses DF, Bell DA, Flotte TJ, Taylor R, Ratech H, Dubin N (1982) Pathologic predictors of recurrence in stage 1 (T1N0M0) breast cancer. Am J Clin Pathol 78:817–820
16. Silverstein MJ, Lagios MD, Craig PH, Waisman JR, Lewinsky BS, Colburn WJ, Poller DN (1996) A prognostic index for ductal carcinoma in situ of the breast. Cancer 77:2267–2274
17. Sloane JP, Ellman R, Anderson TJ, Brown CL, Coyne J, Dallimore NS, Davies JD, Eakins D, Ellis IO, Elston CW, Humphreys S, Lawrence D, Lowe J, McGee JO'D, Millis RR, Nottingham J, Ryley N, Scott DJ, Sloan JM, Theaker J, Trott PA, Wells CA, Zakhour H (1994) Consistency of histopathological reporting of breast lesions detected by screening: findings of the UK National EQA Scheme. Eur J Cancer [A]30:1414–1419
18. Sloane JP, Amendoeira I, Apostolikas N, Bellocq J-P, Bianchi S, Boecker W, Bussolati G, Coleman D, Connolly CE, Eusebi V, De Miguel C, Dervan P, Drijkoningen R, Elston CW, Faverly D, Gad A, Jacquemier J, Lacerda M, Martinez-Penuela J, Munt C, Peterse JL, Rank F, Sylvan M, Tsakraklides V, Zafrani B (1998) Consistency achieved by 23 European pathologists in categorising ductal carcinoma in situ of the breast using 5 classifications. Hum Pathol (in press)
19. Tavassoli FA (1997) Mammary intraepithelial neoplasia: a translational classification system for the intraductal epithelial proliferations. Breast J 3:48–58
20. Wells WA, Carney PA, Eliassen MS, Tosteson AN, Greenberg ER (1998) Statewide study of diagnostic agreement in breast pathology. J Natl Cancer Inst 90:142–145