

0959-8049(95)00609-5

Original Paper

Reproducibility of Histological Diagnosis of Breast Lesions: Results of a Panel in Italy

D. Palli, M. Galli, S. Bianchi, G. Bussolati, S. Di Palma, V. Eusebi, M. Gambacorta and M. Rosselli Del Turco

¹Epidemiology Unit and ⁸Breast Unit, Centro per lo Studio e la Prevenzione Oncologica, Viale A. Volta 171, Florence; ²Sezione Distaccata IST, CSPO, Florence; ³Istituto di Anatomia e Istologia Patologica, Universita' di Firenze, Florence; ⁴Istituto di Anatomia Patologica, Universita' di Torino, Turin; ⁵Divisione di Anatomia Patologica e Citologia, Istituto Nazionale Tumori, Milan; ⁶Istituto di Anatomia e Istologia Patologica, Universita' di Bologna, Bologna; and ⁷Servizio di Anatomia Patologica, Ospedale Ca' Granda Niguarda, Milan, Italy

Sixteen pathologists independently examined and classified a series of 81 breast lesions selected from the files of several Italian Pathology Departments in the context of a national task force on breast cancer (FONCAM). A four category classification system was used for analysis; according to the majority diagnosis (MD), simply defined as the most frequently reported in the panel, the series included 37 benign lesions without atypia (45.7%), nine atypical hyperplasias (11.1%), 18 in situ (22.2%) and 17 invasive carcinomas (21.0%). Concordance, estimated for all possible pair-wise comparisons between pathologists, was good (mean kappa value: 0.59). A comparison between the diagnoses of each pathologist and the panel majority diagnosis was also made. Overall, a global kappa value of 0.72 was found (range 0.57–0.85), with category-specific values being excellent for invasive carcinoma (0.89) and benign lesions without atypia (0.77), relatively good for in situ carcinoma (0.69) but poor for atypical hyperplasia (0.38). These results confirm that quality assurance procedures are particularly indicated for large screening programmes for breast cancer, and suggest that for atypical lesions strict diagnostic criteria should be adopted. Copyright © 1996 Elsevier Science Ltd

Key words: breast lesions, atypical hyperplasia, breast cancer, reproducibility, histology, diagnostic accuracy

Eur J Cancer, Vol. 32A, No. 4, pp. 603-607, 1996

INTRODUCTION

Breast cancer is the most common tumour and the leading cause of cancer death in the female population worldwide [1, 2]. A reproducible histological definition of different breast lesions, ranging between benign disease and invasive cancer, is necessary in order to treat individual patients correctly. Different histological classifications are commonly used for breast disease by groups of pathologists. Furthermore, diagnostic interobserver variability within the same classification is found [3].

In situ carcinomas are increasing because of more widespread early diagnosis and screening procedures [4]. In contrast, atypical hyperplasia (AH) is a rare condition (about 2– 4% of all benign lesions) but because of the high risk of subsequent invasive breast cancer (BC) observed in several studies, it is considered an indication for a closer follow-up of patients [5–7]. The need of a common standardised terminology is one of the major issues in the field of breast pathology.

In recent years, the value of assessing the reproducibilty of diagnoses in pathology has become widely accepted [3, 8]. A procedure often used by pathologists in order to test their diagnostic reproducibility is to circulate a set of slides. Several reports have been published in the current literature, regarding quality assurance practices in diagnostic surgical pathology [9–20]. Traditionally, concordance between observers has been assessed by use of the kappa statistic [21], which includes a correction for the amount of agreement expected by chance alone. Recently, new approaches have also suggested the possibility of taking into consideration the individual characteristics of a specific observer [22].

The aim of this study was to evaluate the reproducibility of histological diagnoses in a series of 81 selected breast lesions among 16 pathologists in Italy.

D. Palli et al.

MATERIALS AND METHODS

Selection of histological material

Histological slides of breast lesions were selected in 16 local Pathology Departments in Italy. The selection was performed with the aim of including a representative sample of all the major categories of breast pathology. The participating members were from the University or Hospital Pathology Departments located in Arezzo (three pathologists), Bari, Bologna, Firenze, Genova, Leno, Mantova, Milano (three), Torino (three) and Trieste. Only some of the participants are primarily involved in diagnostic breast pathology. A set of 82 haematoxylin-eosin stained slides (usually one for each case) were randomly numbered and circulated among participants after an initial meeting in which general criteria were discussed. One specimen (no. 10) was damaged during the circulation and was not considered. Pathologists were asked to reach a diagnosis for each case and to report it in a form for the concordance analysis. No information about age, the original diagnosis or the diagnosis of the other pathologists was provided to the participants.

Classification

Pathologists reported their diagnoses for each case on a study form in which five diagnostic categories were available. The benign breast disease (BBD) classification originally proposed by Dupont and Page [23] and supported at the consensus meeting of the College of the American Pathologists [24] was modified, combining the three main categories of BBD into two: non-proliferative or proliferative lesions without atypia (A) and atypical hyperplasia (B). Two additional categories for *in situ* carcinoma were utilised [ductal non comedo and lobular (C) and ductal comedo (D)]. A fifth category was used for invasive carcinoma (E). Description of each specimen was allowed but only the final diagnostic decision (coded 1 to 5) was considered as follows:

- A. Non-proliferative or proliferative without atypia benign breast disease;
- B. Ductal and/or lobular atypical hyperplasia (AH);
- C. Ductal non comedo and/or lobular in situ carcinoma;
- D. Ductal comedo in situ carcinoma;
- E. Invasive carcinoma.

The original forms were first checked for completeness and consistency and then computerised for the analysis. No correction was allowed for errors which might have occurred when pathologists reported their diagnoses (i.e. using wrong codes) once the original forms were obtained.

Analysis

Analysis was carried out using four categories, after combining all *in situ* carcinomas (1 = A; 2 = B; 3 = C+D; 4 = E).

The histological diagnoses of each pathologist were first compared to those of all the other 15. Overall, 120 such pairwise comparisons were performed $(n^*(n-1)/2)$, where n=16 readers). No attempt to reach a final consensus among panel members was made. The diagnoses of each pathologist were therefore compared to the MD, simply defined as the most frequent diagnosis expressed for each slide, and a category-specific analysis was also performed. As measure of agreement, either between the 16 pathologists and between each pathologist and the MD, the kappa statistic was used [21], which includes a correction for the amount of agreement which would be expected by chance alone. Values of kappa

near the maximum of 1 indicate perfect agreement. Traditionally, values higher than 0.40 are considered acceptable and higher than 0.75 excellent. Standard errors (S.E.) of the overall kappa values were calculated according to Fleiss and associates [25].

The specimens were circulated only once; therefore no data were available on intra-observer reproducibility.

RESULTS

The distribution of the diagnoses for each of the 81 breast lesions and the resulting MD are shown in Table 1: 37 benign (45.7%), 9 AH (11.1%), 18 in situ carcinomas (22.2%) and 17 invasive carcinomas (21.0%). All 16 pathologists reached the same diagnostic conclusion in 22 cases (27.2%): 11 were classified as invasive carcinoma, one as in situ carcinoma and

Table 1. Distribution of individual histological diagnoses (16 pathologists) and resulting majority diagnosis of 81 breast lesions (FONCAM, Italy)

Case	Benign*	АН	In situ	Invasive	Majority diagnosis‡
1	3	8	4	1	2
2	0	9	7	0	2
3	0	0	0	16	4
4	0	0	0	16	4
5	0	1	13	2	3
6	12	1	3	0	1
7	16	0	0	0	1
8	0	0	0	16	4
9	0	0	0	16	4
10†		_			
11	4	0	0	12	4
12	15	0	0	1	1
13	0	3	13	0	3
14	6	10	0	0	2
15	15	1	0	0	1
16	11	5	0	0	1
17	12	4	0	0	1
18	16	0	0	0	1
19	1	4	10	1	3
20	15	0	1	0	1
21	0	0	0	16	4
22	0	0	1	15	4
23	1	1	14	0	3
24	14	1	1	0	1
25	14	1	1	0	1
26	10	6	0	0	1
27	10	5	1	0	1
28	15	1	0	0	1
29	1	1	8	6	3
30	16	0	0	0	1
31	12	3	1	0	1
32	0	0	0	16	4
33	1	0	1	14	4
34	8	7	1	0	1
35	9	6	1	0	1
36	15	0	1	0	1
37	14	1	1	0	1
38	5	6	5	0	2
39	16	0	0	0	1
40	1	2	11	2	3
41	0	0	0	16	4

Table 1 continued overleaf

Table 1. Continued

Case	Benign*	АН	In situ	Invasive	Majority diagnosis‡
42	0	0	0	16	4
43	16	0	0	0	1
44	1	3	12	0	3
45	16	0	0	0	1
46	13	3	0	0	1
47	9	4	0	3	1
48	1	2	13	0	3
49	4	7	5	0	2
50	0	2	13	1	3
51	0	0	14	2	3
52	0	0	0	16	4
53	16	0	0	0	1
54	0	0	16	0	3
55	2	6	8	0	3
56	3	3	10	0	3
57	0	0	1	15	4
58	8	7	1	0	1
59	0	0	0	16	4
60	3	10	3	0	2
61	16	0	0	0	1
62	15	1	0	0	1
63	16	0	0	0	1
64	13	3	0	0	1
65	0	5	11	0	3
66	10	5	1	0	1
67	0	2	14	0	3
68	0	0	9	7	3
69	0	0	0	16	4
70	3	1	12	0	3
71	3	0	0	13	4
72	15	0	1	0	1
73	15	0	1	0	1
74	16	0	0	0	1
75	1	1	6	8	4
76	4	10	2	0	2
77	0	4	12	0	3
78	5	10	1	0	2
79	1	7	6	2	2
80	15	1	0	0	1
81	11	5	0	0	1
82	14	1	1	0	1

*Benign lesion without atypia. †Damaged during the panel. ‡1, benign lesion without atypia; 2, atypical hyperplasia; 3, in situ carcinoma; 4, invasive carcinoma. AH, atypical hyperplasia.

10 as benign disease without atypia, but in no case was AH diagnosed by all 16 participants. In the remaining 59 cases, at least one panellist was not in agreement with the others: all the four available diagnostic categories were used for 6 cases, three categories in 26 cases and two categories in the remaining 27 cases.

In the latter group, the same diagnostic conclusion was reached for 11 cases by 15 of the 16 pathologists, while only one was not in agreement (in 4 cases 15 panellists reported a diagnosis of "benign without atypia" versus 1 diagnosis of "AH"; in 4 cases 15 benign without atypia versus 1 in situ carcinoma; in 1 case 15 benign without atypia versus 1 invasive carcinoma; in 2 cases 15 invasive versus 1 in situ carcinoma).

Overall, only 9 cases were diagnosed as AH by the majority:

in 2 such cases, all the four available diagnostic categories were reported, in 5 cases three categories and in 2 cases two categories.

Kappa values for all possible pair-wise comparisons between the 16 pathologists are shown in Table 2. The mean and median kappa values are also reported for each pathologist. In these 120 pair-wise comparisons, kappa values ranged between a minimum of 0.37 and a maximum of 0.91. Mean overall kappa values for each pathologist ranged between 0.48 and 0.66. The overall mean kappa value was 0.59.

In Table 3 the cumulative distribution of the diagnoses of all the participants versus the MD is reported. Overall, 1296 (16×81) diagnostic comparisons were performed: 1042 (80.4%) were in perfect agreement. Category specific crude agreement was 84% (499/592) for the benign lesion without atypia category, 53% (77/144) for AH, 74% (213/288) for in situ carcinoma and 93% (253/272) for invasive carcinoma. The global kappa value was 0.72. On nine occasions, a case was classified by an individual reader as benign lesion without atypia while the resulting MD was invasive carcinoma (cases no. 11, 33, 71 and 75 in Table 1 with four, one, three and one diagnoses, respectively). A careful review of the 2 cases with multiple benign diagnoses after the data analysis showed morphological patterns consistent with the diagnosis of tubular carcinoma (case 11) and ductal invasive carcinoma (case 71). In the latter, it was also noted that two sections of different areas of the surgical specimen were available for review on the same slide and only normal (non-malignant) tissue was present in one section. In contrast, four diagnoses of invasive carcinoma were followed by a MD of benign lesion without atypia (cases 12 and 47, with one and three diagnoses, respectively).

In Table 4, category-specific and global kappa values estimated for individual comparisons between each pathologist and the MD are reported; mean values for category specific and overall kappa statistics are also shown. Global kappa values for individual pathologists varied between 0.57 and 0.85. The category-specific kappa value was lowest for atypical hyperplasia (range 0.14–0.70; mean value 0.38) and highest for invasive carcinoma (range 0.78–1.00; mean value 0.89). Benign lesions without atypia and *in situ* carcinoma showed intermediate values (range 0.62–0.95 and mean value 0.77; range 0.40–0.89 and mean value 0.69, respectively).

DISCUSSION

Reproducibility of breast histological diagnoses was shown to be relatively good in this large panel. In particular, an excellent agreement was found for the diagnosis of invasive carcinoma and of benign lesions without atypia. However, in the two intermediate categories (AH and in situ carcinoma), serious difficulties emerged resulting in a very low kappa value for AH, while the diagnosis of in situ carcinoma was shown to be more consistent. In none of the 9 cases with a MD of AH was the same diagnostic conclusion simultaneously reached by all participating pathologists, in agreement with other previous reports [3]. These findings suggest that no clear definition of AH is currently available and pathologists give a different meaning to this diagnosis. Because of the low agreement obtained for this specific category, it has been recently suggested, but not widely accepted, that terminology such as "mammary intra-epithelial neoplasia" combining hyperplasia, AH and in situ carcinomas is adopted [3].

Not all diagnostic errors affect patient care to the same

D. Palli et al.

Table 2. Concordance for all possible pair-wise comparisons between 16 pathologists; median, mean and overall mean kappa values for each pathologist (FONCAM, Italy)

	3	5	10	12	14	15	16	17	18	20	21	23	26	27	28	30	
5	0.56																
10	0.54	0.71															
12	0.48	0.68	0.64														
14	0.56	0.70	0.60	0.58													
15	0.47	0.62	0.59	0.48	0.50												
16	0.56	0.61	0.67	0.69	0.56	0.53											
17	0.63	0.71	0.67	0.73	0.61	0.48	0.68										
18	0.53	0.57	0.53	0.46	0.56	0.38	0.55	0.57									
20	0.61	0.71	0.68	0.67	0.68	0.58	0.64	0.70	0.52								
21	0.49	0.67	0.64	0.57	0.57	0.64	0.58	0.59	0.41	0.65							
23	0.57	0.62	0.48	0.53	0.60	0.54	0.57	0.66	0.50	0.59	0.46						
26	0.39	0.53	0.51	0.48	0.49	0.46	0.49	0.46	0.43	0.53	0.52	0.37					
27	0.56	0.69	0.61	0.62	0.71	0.55	0.62	0.71	0.44	0.79	0.61	0.65	0.52				
28	0.50	0.73	0.64	0.61	0.62	0.59	0.60	0.61	0.49	0.76	0.59	0.57	0.58	0.68			
30	0.60	0.76	0.66	0.72	0.63	0.51	0.67	0.91	0.49	0.68	0.60	0.70	0.42	0.73	0.63		
Median	0.56	0.67	0.64	0.58	0.60	0.53	0.60	0.66	0.50	0.67	0.59	0.57	0.49	0.62	0.61	0.66	Overall
Mean	0.54	0.66	0.61	0.60	0.60	0.53	0.60	0.65	0.49	0.65	0.57	0.56	0.48	0.63	0.61	0.65	0.59

Table 3. Cumulative distribution of histological diagnoses of 81 breast lesions in the comparison between majority diagnosis and all the 16 pathologists (FONCAM, Italy)

	Majority diagnosis								
All pathologists	Benign*	AH	In situ	Invasive carcinoma	Total				
Benign*	499	31	14	9	553				
AH	72	77	40	1	190				
In situ	17	33	213	9	272				
Invasive carcinon	na 4	3	21	253	281				
Total	592	144	288	272	1296				

^{*}Benign lesion without atypia. AH, atypical hyperplasia.

degree: minor disagreements may have minimal or no effect at all on patient care, while major discrepancies could seriously affect the patients. Either under-diagnosis or over-diagnosis may occur. In this study, four specimens were classified as benign lesions without atypia by a total of nine pathologists (four in 1 case, three in another) while the resulting MD was invasive carcinoma. The opposite was found in 2 different cases: a total of four pathologists (three in 1 case and one in the other) classified these specimens as invasive carcinoma while the panel MD was benign lesions without atypia. These major disagreements might have seriously affected the therapeutical decisions in individual patients.

All the slides were haematoxylin-eosin stained but were retrieved from the archives of 16 Pathology Departments in Italy and the technical quality of the different specimens might not have been optimal in every case. Other factors could have influenced the results. When the pathologists sampled the slides for this study, they might have included 'problem' cases more frequently than the routine daily diagnostic session of a general surgical pathologist. The artificial reading conditions might have also contributed to the disagreement, in particular because the series tended to be large and the time available for the classification was relatively short.

Table 4. Category-specific and global kappa values estimated for individual comparisons between each pathologist and the majority diagnosis of 81 breast lesions. Mean values for specific and overall kappa (FONCAM, Italy)

	Cate				
Pathologist	Benign*	АН	In situ carcinoma	Invasive carcinoma	Global kappa (S.E.)
3	0.62	0.37	0.63	0.89	0.63 (0.06)
5	0.95	0.65	0.72	0.93	0.84 (0.07)
10	0.75	0.25	0.82	0.92	0.75 (0.07)
12	0.85	0.29	0.69	0.92	0.76 (0.07)
14	0.72	0.41	0.84	0.89	0.72 (0.06)
15	0.73	0.37	0.52	0.78	0.64 (0.07)
16	0.85	0.14	0.64	0.89	0.71 (0.07)
17	0.80	0.25	0.77	0.96	0.75 (0.07)
18	0.67	0.27	0.45	0.82	0.58 (0.07)
20	0.87	0.70	0.89	0.88	0.85 (0.07)
21	0.80	0.54	0.67	0.85	0.74 (0.07)
23	0.64	0.27	0.64	0.82	0.62 (0.06)
26	0.63	0.24	0.40	0.88	0.57 (0.07)
27	0.75	0.44	0.86	0.88	0.75 (0.07)
28	0.87	0.47	0.72	1.00	0.80 (0.07)
30	0.80	0.31	0.73	0.92	0.73 (0.07)
Mean	0.77	0.38	0.69	0.89	0.72
(S.E.)	(0.10)	(0.15)	(0.14)	(0.05)	(0.08)

^{*}Benign lesion without atypia. AH, atypical hyperplasia.

In conclusion, we consider quality assurance in pathology mandatory, particularly in the context of a screening programme, when a higher proportion of *in situ* and more differentiated invasive cancers is expected. Our results suggest that agreement on the morphological characteristics of atypical lesions should be improved among pathologists in order to guarantee a homogeneous diagnostic classification and treatment to patients with a borderline breast lesion.

- Parkin DM, Pisani P, Ferlay J. Estimates of worldwide incidence of eighteen major cancers in 1985. Int J Cancer 1993, 54, 594– 606.
- Pisani P, Parkin DM, Ferlay J. Estimates of the worldwide mortality from eighteen major cancers in 1985. Implications for prevention and projections of future burden. *Int J Cancer* 1993, 55, 891-903.
- Rosai J. Borderline epithelial lesions of the breast. Am J Surg Pathol 1991, 15, 209-221.
- Donegan WL. Recent trends in the management of breast cancer.
 Carcinoma in situ of the breast. Can J Surg 1992, 35, 361–365.
- Tavassoli FA, Norris HJ. A comparison of the results of longterm follow-up for atypical intraductal hyperplasia and intraductal hyperplasia of the breast. *Cancer* 1990, 65, 518-529.
- Palli D, Rosselli del Turco M, Simoncini R, Bianchi S. Benign breast disease and breast cancer: a case-control study in a cohort in Italy. *Int J Cancer* 1991, 47, 703-706.
- Dupont WD, Parl FF, Hartmann WH, et al. Breast cancer risk associated with proliferative breast disease and atypical hyperplasia. Cancer 1993, 71, 1258-1265.
- 8. Palli D, Bianchi S, Linell F, et al. Histopathologic classification of breast cancer in Sweden and Italy: a comparison between two pathologists. *Tumori* 1992, 78, 247–249.
- Macartney JC, Henson DE, Codlinch BW. Quality assurance in anatomic pathology. Am J Clin Pathol 1981, 75(suppl), 467–475.
- Rickert RR. Quality assurance in anatomic pathology. Clin Lab Med 1986, 6, 697–706.
- Schned AR, Mogielicki RP, Stauffer ME. A comprehensive quality assessment program on the autopsy service. Am J Clin Pathol 1986, 86, 133-138.
- 12. Penner DW. Quality assurance in diagnostic surgical pathology. *Arch Pathol Lab Med* 1987, 111, 513.
- College of American Pathologists. Surgical Pathology/Cytopathology Quality Assurance Manual. Skokie, III, College of American Pathologists, 1988.
- Bonfiglio TA. Quality assurance in cytopathology: recommendations and ongoing quality assurance activities of the American Society of Clinical Pathologists. Acta Cytol 1989, 33, 431–433.
- Lee FD. External quality assessment of histopathology. J Clin Pathol 1989, 42, 1009-1011.
- Travers H. Quality assurance in anatomic pathology. Lab Med 1989, 20, 85-92.
- Cowan DF. Quality assurance in anatomic pathology. Arch Pathol Lab Med 1990, 114, 129-134.
- Schnitt SJ, Connolly JL, Tavassoli FA, et al. Interobserver reproducibility in the diagnosis of ductal proliferative breast lesions using standardized criteria. Am J Surg Pathol 1992, 16, 1133–1143.

- Bianchi S, Palli D, Galli M, et al. Reproducibility of histological diagnoses and diagnostic accuracy of non palpable breast lesions. Path Res Pract 1994, 190, 69-76.
- 20. U.K. National Coordinating Group for Breast Screening Pathology. Consistency of histopathological reporting of breast lesions detected by screening: findings of the U.K. National External Quality Assessment (EQA) Scheme. Eur J Cancer 1994, 30A, 1414-1419.
- Fleiss JL. Statistical Method for Rates and Proportions, 2nd edn. New York, John Wiley, 1981.
- 22. Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. Am J Epidemiol 1993, 137, 1251-1258.
- Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. N Engl J Med 1985, 312, 146– 151.
- Hutter RVP. Consensus meeting: is "fibrocystic disease" of the breast precancerous? Arch Pathol Lab Med 1986, 110, 171–173.
- Fleiss JL, Cohen J, Everitt BS. Large sample standard errors for Kappa and weighted kappa. Psychol Bull 1968, 72, 323–327.

Acknowledgement—This study was supported by the National Research Council-CNR Applied Project ACRO (93.02 332.PF39).

APPENDIX

Participating pathologists

Alberto Bellomi, Servizio di Anatomia Patologica, Ospedale C. Poma, Mantova

Simonetta Bianchi, Istituto di Anatomia Patologica, Universita' di Firenze, Firenze

Gianni Bussolati, Istituto di Anatomia Patologica, Universita' di Torino, Torino

Enrico Cardesi, Ospedale Martini Nuovo, USL III, Torino

Vincenzo Eusebi and Donatella Santini, Istituto di Anatomia e Istologia Patologica, Universita' di Bologna, Bologna

Silvana Di Palma, Divisione di Anatomia Patologica e Citologia, Istituto Nazionale Tumori, Milano

Giovanni Falconieri, Istituto di Anatomia Patologica, Ospedale Maggiore, Trieste

Elio F. Gagliano, Servizio di Senologia, USL 43, Leno

Marcello Gambacorta, Servizio di Anatomia Patologica, Ospedale Ca' Granda Niguarda, Milano

Bruno Ghiringhello, Servizio di Anatomia Patologica, Torino Carmela Giardina, Istituto di Anatomia Patologica, Bari

Guido Nicolo', Istituto Nazionale per la Ricerca sul Cancro, Genova Enrico Turolla, Presidio Sanitario, USL 70, Legnano, Milano

Galliano Tinacci, Servizio di Anatomia Patologica, USL 20/A, Arezzo Enzo Zini, Servizio di Anatomia Patologica, USL 20/A, Arezzo