Pathol. Oncol. Res. (2009) 15:725–732
DOI 10.1007/s12253-009-9176-7

Diagnostic Concordance in Reporting Breast Needle Core Biopsies using the B Classification—A Panel in Italy

Simonetta Bianchi · Saverio Caini · Maria Grazia Cattani · Vania Vezzosi · Mauro Biancalani · Domenico Palli

Received: 22 January 2009 / Accepted: 28 April 2009 / Published online: 17 May 2009

Abstract The widespread implementation of mammography screening has resulted in an increased frequency of needle core biopsies (NCB). The aim of this study was that of evaluating the diagnostic reproducibility on breast NCB, according to the B-classification, among several pathologists from different Italian regions. Fifty single slides of NCBs performed for non palpable breast lesions were selected to evaluate the diagnostic reproducibility, according to the B classification, among 31 pathologists from different Italian areas, involved in the pathologic diagnosis of screen-detected breast lesions. According to the study majority diagnosis (MD), 21 cases were classified as B2 (benign lesion), 23 B3 (lesion of uncertain malignant potential) and 6 B5 (malignant lesion). Overall, individual kappa coefficients in comparison to MD were good (mean 0.61, range 0.31–0.88). The level of inter-observer agreement, however, appeared lower in differentiating the two intermediate categories B2 and B3, thus potentially leading to over-treatment (false-positives: 26%) or under-treatment (false-negatives: 17%) of individual patients. Specific sub-types of B3 need an improvement of the diagnostic definition. A multidisciplinary approach and consultation with expert colleagues are recommended.

Keywords Breast · Non palpable lesions · Needle core biopsy · Reporting · Diagnostic concordance

Abbreviations

ADH	atypical ductal hyperplasia
AEPDT	atypical epithelial proliferation of ductal type
BI-RADS	breast imaging reporting and data system
CCCa	columnar cell change with atypia
CCHa	columnar cell hyperplasia with atypia
COBRA	core biopsy after radiological localisation
DCIS	ductal carcinoma in situ
DIOS	diagnosis optimisation study
LCIS	lobular carcinoma in situ
LIN	lobular intraepithelial neoplasia
MD	majority diagnosis
NCB	needle core biopsy
PL	papillary lesion
PPV	positive predictive value
PT	phylloid tumor
ROC	receiver operating characteristic
RS	radial scar

V. Vezzosi
Department of Human Pathology and Oncology,
University of Florence,
Florence, Italy

S. Caini · D. Palli
Molecular and Nutritional Epidemiology Unit,
Cancer Research and Prevention Institute-ISPO,
Florence, Italy

M. G. Cattani
Pathologic Anatomy Unit, Bellaria Hospital,
Bologna, Italy

M. Biancalani
Pathologic Anatomy Unit, Empoli Hospital,
Florence, Italy

S. Bianchi (✉)
Department of Human Pathology and Oncology,
University of Florence,
AOU Careggi, Viale Morgagni 85,
50134 Florence, Italy
e-mail: simonetta.bianchi@unifi.it

Introduction

The widespread implementation of mammography screening programmes in the last 10 years, and the introduction of the percutaneous needle core biopsy (NCB) procedure in the preoperative assessment of breast lesions, has resulted in an increased frequency of NCBs performed for non palpable breast abnormalities, especially microcalcifications [1].

NCB has been introduced with the double purpose of maximising, in comparison with fine needle aspiration cytology, the number of accurate and definitive preoperative diagnoses, especially in non palpable breast lesions, and avoiding a considerable number of open breast biopsies.

Currently, English [2] and European [3] guidelines recommend to apply a pathology categorisation scheme that includes five (B1-B5) reporting categories of NCB (i.e. B1: normal tissue, B2: benign lesion, B3: lesion of uncertain malignant potential, B4: suspicious of malignancy, B5: malignant) the so-called B-classification.

Although most NCB specimens can be classified as normal (B1), benign (B2), suspicious or malignant (B4 and B5 respectively), a small proportion of lesions cannot fit in these categories and are reported in the borderline category of B3 [4].

From a clinical point of view, B1-B2 categories usually do not necessitate surgical excision unless there is discordance with the radiological findings (mammographic or ultrasound pattern); it is essential in fact the comparison between radiological and histological findings in order to ensure that the NCB is representative of the screen-detected lesion. On the other hand, the B4 and B5 categories implicate therapeutic surgical treatment of the breast lesion having high positive predictive values (PPV): 74.2% (range from 62.5% to 90.6%) and more than 99% respectively. The PPV for B3 has been reported, in a recent study, to be 19.1% (range from 13.3% to 30%) [1].

Several studies, rather unsuccessfully, have tried to identify B3 subgroups with particularly high or low risk of malignancy, facilitating further management. Additional research, looking at molecular markers or predictive models, could be useful in differentiating between benign and malignant outcomes in patients with B3 lesions [5]. At present, diagnostic surgical excision remains the method of choice for managing B3 cases as more advanced lesions often co-exist [6].

A number of prior studies have addressed the issue of interobserver reproducibility in the diagnosis of open breast biopsy specimens and the results were relatively good. Because a smaller amount of tissue is obtained, the diagnostic assessment of NCB might be more difficult and more inconsistencies in the diagnosis of pathologists might occur.

Even though NCB procedure has been introduced in the clinical practice about 10–15 years ago, there are only a few published studies concerning the diagnostic agreement on NCB pathological evaluation, and only one of these, to date, used B-classification to evaluate inter-observer reproducibility.

The aim of this study was that of evaluating the diagnostic reproducibility on breast percutaneous NCB, according to the B-classification, among several pathologists from different Italian regions, involved in the pathologic diagnosis of screen-detected breast lesions.

Materials and Methods

Selection of Histological Material

Fifty NCB performed for non palpable breast lesions were selected for the study: 25 cases from the files of the Department of Human Pathology and Oncology of the Careggi University Hospital in Florence, and 25 cases from the files of the Pathology Department of the Ospedale Maggiore in Bologna. Case selection included both common (B2 and B5 categories) and borderline cases (B3 category) likely to elicit differences in interpretation. Guidance methods for selected NCB consisted of stereotactic mammography and all NCBs were performed with a vacuum-assisted biopsy device equipped with a 11-gauge needle.

With regard to the mammographic pattern, the selected data set contained 42 cases of microcalcification (6 R2, 29 R3, 5 R4 and 2 R5), 6 cases of opacity (4 R2 and 2 R5), 1 case of parenchymal distorsion (R4) and 1 case of asymmetrical density (R4) classified according to BI-RADS (Breast Imaging Reporting And Data System) mammographic five-point classification system (i.e. R1: normal/benign, R2: a lesion having benign characteristics, R3: an abnormality of indeterminate significance, R4: features suspicious of malignancy, R5: malignant features) [7].

All available routinely prepared hematoxylin and eosin-stained slides obtained from paraffin-embedded tissue blocks from selected cases were retrieved and reviewed by one of us (MB), and for each case the most representative slide was selected to build the data set of fifty cases/slides. Cases were anonymized and randomly re-labelled from 1 to 50.

Histological Review/Classification

Breast pathologists from 3 Italian regions (Emilia Romagna, Piedmont and Tuscany) with an active mammographic screening programme were invited to participate in the study: 31 agreed, 8 from Emilia Romagna, 9 from Piedmont and 14 from Tuscany.

Pathologists participating in the study were asked to classify the 50 slides according to the so-called B-

classification [2, 3] and also to provide, for each B category, the type of lesion according to the European guidelines [3]. Concerning the diagnosis of columnar cell lesions, without or with atypia, pathologists were asked to follow the diagnostic criteria described by Schnitt et al. [8]. The unique slide set remained for 1 month in the Pathology Department of each regional capital city (Bologna, Turin and Florence), and participants from each region were asked to go to that specific Pathology Department in order to perform the histopathological assessment of the slides.

An electronic spreadsheet (Microsoft Excel file) was used to collect diagnoses according to the B-classification (B1-B5). In case of a B5 diagnosis, each reader was invited to specify if it was an "in situ" or invasive cancer (B5a and B5b respectively); in case of a B2 diagnosis, it was subsequently requested to specify the type of lesion, i.e. fibroadenoma, fibrocystic change, sclerosing adenosis, columnar cell change, columnar cell hyperplasia, ductal adenoma or inflammation; finally, in case of a B3 diagnosis each pathologist was asked to specify one of the following sub-types: columnar cell lesion (change or hyperplasia) with atypia (CCCa/CCHa), atypical epithelial proliferation of ductal type (AEPDT), lobular intraepithelial neoplasia (LIN), phylloid tumor (PT), papillary lesion (PL), radial scar (RS).

No information about age of the cases selected for this panel, nor the diagnoses made by the other pathologists were provided to study participants, who had access to the original mammographic pattern leading to referral.

Statistical Analysis

Following a majority criterion, each slide was labelled with a majority diagnosis (MD) corresponding to the B category (B1-B5) most frequently reported. For all cases, except one (n. 41), the resulting MD represented a diagnostic agreement between 50% to 100% of participants. Quality control showed that, overall, only 3 diagnoses (out of 1,550) were missing, and for these readings the specific MD was considered.

Based on the *a priori* consideration that only patients with a diagnosis equal to or more severe than B3 usually undergo surgery, statistical analyses were carried out using only two major diagnostic categories: slides with B1 and B2 as MD were classified as "negative", and slides diagnosed as B3, B4 or B5 as "positive".

A ROC plot was produced with values of sensitivity and specificity computed for each individual observer. For each pathologist, sensitivity and specificity values correspond to the proportions of correctly diagnosed slides among all the slides respectively labelled as "positive" and "negative" according to MD. In the ROC plot, high values for sensitivity and specificity tend to correspond to points towards the top and the left side respectively, while a random diagnostic process would produce a cloud of points distributed along the diagonal line from the left bottom to the top right corner. A similar ROC plot was produced also for data grouped on a regional criterion.

The histological diagnoses of each reader were compared to those of all the other 30 (results not shown). As a measure of agreement, we used the standard Cohen's kappa coefficient [9]. Values for kappa statistics can range from 0 to 1, and a rough guideline for interpreting the degree of agreement based on kappa values [10] is as follows: 0–0.20 = very low, 0.21–0.40 = low, 0.41–0.60 = moderate, 0.61–0.80 = good, 0.81–0.99 = excellent, 1.00 = perfect. It should be pointed out that high values of inter-individual kappa statistics do not imply anything about the individual performance: two subjects could have a perfect agreement and a very bad performance if they assigned the same *wrong* diagnosis to all slides.

The fifty diagnoses of each reader were then compared, with the same methodology, to the majority diagnoses (MD). The individual agreement with the MD is influenced by both the sensitivity and the specificity of each reader but also by the diagnostic attitude most prevalent in the study group, and can be interpreted as a measure of individual overall performance.

In addition, for the sub-group of slides with B3 as MD, a majority diagnostic criterion was also used for the assignment of a specific histological subtype. Thus, we calculated the percentage of readings on which this subtype majority diagnosis was based, and the average number of readings per slide for each histological subtype; the distribution of these results were reported in a specific table.

Results

The distribution of the diagnoses and of the follow-up outcomes for each of the 50 slides and the resulting MD are shown in Table 1. According to the B-classification, the MDs of the 50 slides are distributed as follows: 21 (42%) B2, 23 (46%) B3, and 6 (12%) B5. In 16 cases major discrepancies were observed as more than two consecutive B categories were reported (Figs. 1 and 2). On the other hand, all 31 pathologists reached the same diagnosis only in four cases (8%): n.6, n.18, n.31 and n.46.

For the slide number 12, it could be hypothesized that an error of interpretation of B-classification occurred, which led many readers to attach an erroneous B5 code, instead of B3, to a LCIS.

Figure 3 shows a ROC plot with performance values of each participating pathologist: sensitivity values range between 0.72 and 0.97, while specificity values range between 0.43 and 1.00. Overall, 75% of the pathologists have a sensitivity higher than 0.80; specificity is not always

Table 1 Cumulative distribution of individual diagnoses into 5 main diagnostic categories (by 31 readers), majority diagnosis for each of the 50 study cases, and study outcome

Slide	B1	B2	B3	B4	B5	Majority diagnosis	Outcome[a]
1	0	**21**	10	0	0	B2	NED
2	0	**26**	5	0	0	B2	NED
3	0	10	**18**	3	0	B3	NED
4	0	2	**26**	1	2	B3	ADH
5	0	7	**24**	0	0	B3	NED
6	0	**31**	0	0	0	B2	NED
7	0	**16**	15	0	0	B2	CCCa/CCHa
8	0	5	**26**	0	0	B3	Intraductal papilloma
9	0	0	**30**	0	1	B3	NED
10	0	7	**21**	2	1	B3	DCIS
11	0	7	**24**	0	0	B3	DCIS
12	0	0	**17**	0	14	B3	LCIS
13	0	14	**17**	0	0	B3	CCCa/CCHa
14	0	0	**26**	1	4	B3	ADH
15	0	**22**	9	0	0	B2	NED
16	0	**19**	10	1	1	B2	LCIS + ADH
17	0	**22**	9	0	0	B2	NED
18	0	0	0	0	**31**	B5	Invasive carcinoma
19	0	5	**20**	2	4	B3	Benign breast disease
20	0	**23**	8	0	0	B2	NED
21	0	7	**20**	1	3	B3	NED
22	0	1	10	3	**17**	B5	DCIS
23	0	**19**	12	0	0	B2	NED
24	0	0	0	2	**29**	B5	DCIS
25	0	15	**16**	0	0	B3	DCIS
26	3	**27**	1	0	0	B2	NED
27	0	**30**	1	0	0	B2	NED
28	0	6	**17**	0	8	B3	RS
29	0	2	**29**	0	0	B3	NED
30	0	**20**	10	0	1	B2	NED
31	0	0	**31**	0	0	B3	Benign breast disease
32	0	**23**	8	0	0	B2	NED
33	0	**17**	14	0	0	B2	NED
34	0	0	**25**	4	2	B3	DCIS
35	0	**27**	4	0	0	B2	CCCa/CCHa
36	0	**18**	13	0	0	B2	NED
37	0	**18**	13	0	0	B2	NED
38	0	**16**	15	0	0	B2	Benign breast disease
39	0	2	**28**	0	1	B3	Benign breast disease
40	3	**28**	0	0	0	B2	NED
41	5	11	**15**	0	0	B3	Microinvasive carcinoma + DCIS
42	0	5	**19**	3	4	B3	Benign breast disease
43	0	0	0	1	**30**	B5	Invasive carcinoma
44	0	6	**25**	0	0	B3	LCIS
45	0	0	0	2	**29**	B5	Invasive carcinoma
46	0	**31**	0	0	0	B2	NED
47	0	1	**30**	0	0	B3	Malignant phylloides tumour
48	0	4	**25**	0	2	B3	Invasive carcinoma
49	0	**23**	6	0	2	B2	NED
50	0	0	4	0	**27**	B5	DCIS

[a] diagnosis at surgical excision or clinical follow-up with no evidence of disease (NED)

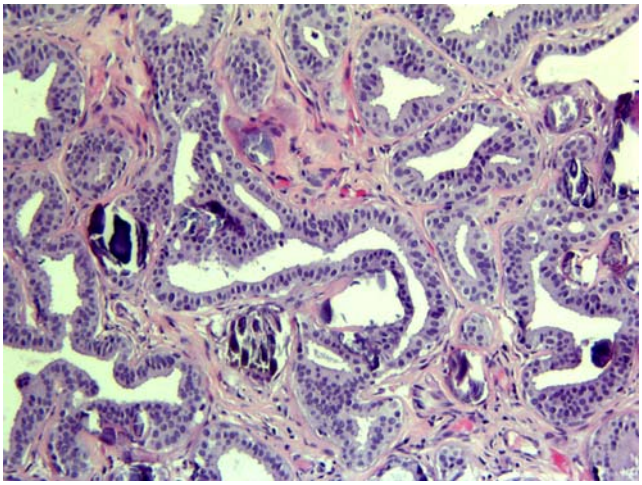

Fig. 1 Case n° 3-Columnar cell change with atypia: this case was classified as B3 by 18 pathologists, as B2 by 10 pathologists and B4 by 3 pathologists (H&E)

as good, with 4 pathologists showing a specificity less than 0.60 (sensitivity and specificity values for each pathologist were not shown in detail).

In Fig. 4 performance values according to regional groups are reported, again as a ROC plot. The group of pathologists from Emilia Romagna shows the best values of both sensitivity and specificity; pathologists from Piedmont and Tuscany have very similar performance values.

The values of inter-individual diagnostic agreement, as measured by Cohen's Kappa, only occasionally exceeded the threshold of 0.80 (data not shown), and only when the two pathologists to be compared belonged to the same geographic group. For two pathologists, a complete agreement ($K=1$) was observed with regard to primary diagnosis, while some small differences existed with regard to B3 subtype diagnosis.

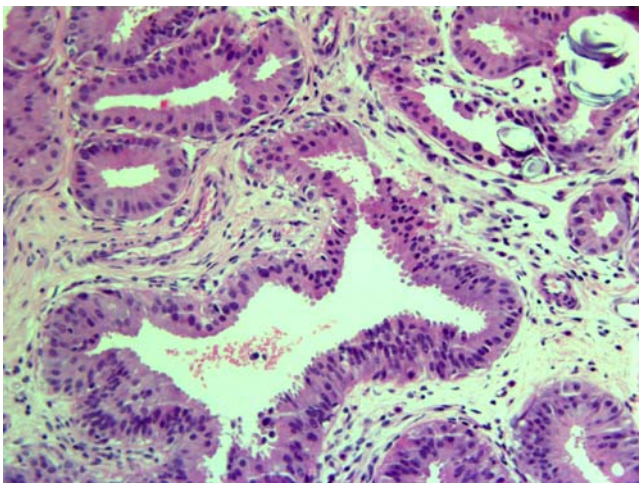

Fig. 2 Case n° 49-Cysts with epithelial lining showing apocrine metaplasia : this case was classified as B2 by 23 pathologist, as B3 by 6 pathologists and as B5 by 2 pathologists (H&E)

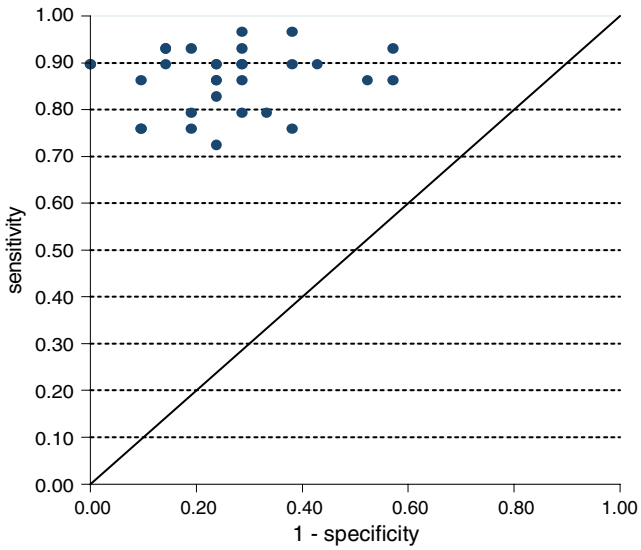

Fig. 3 ROC plot for performance values (sensitivity and 1-specificity) for the 31 individual readers

In Table 2, individual kappa coefficients in comparison to MD are shown: the degree of agreement is excellent for 2 (6%) pathologists (these two pathologists are the same that have a perfect interindividual agreement), good for 16 (52%) pathologists, moderate for 9 (29%) pathologists and low for 4 (13%) pathologists.

Table 3 shows the overall ability of this group of readers to correctly diagnose a slide as "negative" or "positive" when the MD is B2, B3 or B5 respectively. As expected, for slides with B2 or B3 as MD the concordance among different readers tends to be much lower than in the cases when MD is B5, with a false positive proportion of 26% for

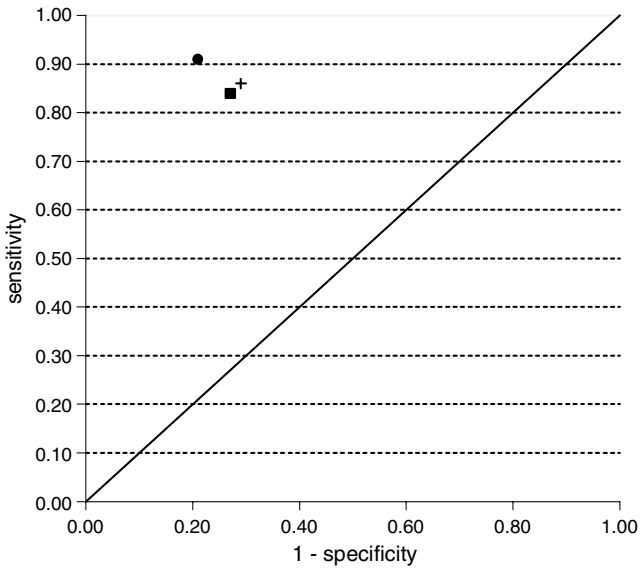

Fig. 4 ROC plot for performance values (sensitivity and 1-specificity) for the 3 regional groups in which individual readers were considered. (● Emilia Romagna, + Piedmont, ■ Tuscany)

Table 2 Values of the kappa statistic for concordance between each reader's diagnosis and Majority Diagnosis (values below 0.4 and above 0.8 are reported in bold)

Reader ID	Kappa	Reader ID	Kappa	Reader ID	Kappa
R1	0.76	**R12**	**0.35**	**R23**	0.59
R2	0.79	**R13**	0.48	**R24**	0.51
R3	0.75	**R14**	**0.38**	**R25**	0.67
R4	0.67	**R15**	0.53	**R26**	0.62
R5	0.79	**R16**	0.63	**R27**	0.66
R6	0.75	**R17**	0.62	**R28**	**0.31**
R7	0.49	**R18**	0.64	**R29**	**0.38**
R8	0.70	**R19**	0.61	**R30**	0.63
R9	**0.88**	**R20**	0.59	**R31**	0.58
R10	**0.88**	**R21**	0.64	**Mean value**	**0.61**
R11	0.46	**R22**	0.56		

slides with B2 as MD (range 0–48 %) and a false negative proportion of 17% when the MD is B3 (range 0–48 %).

Finally, Table 4 shows the diagnostic agreement in relation to the histological subtype of 23 cases with B3 as primary MD. When the most frequently diagnosed subtype was PT or PL, the agreement was nearly complete (100% and 98.2% respectively). On the contrary, the percentage of agreement based on which the attribution of a majority diagnosis was made was much lower for the other subtypes, with a minimal value of 52% for AEPDT. In the column to the right, it can be seen that the diagnosis of B3 is more frequent when a histological pattern such as PT, PL or AEPDT can be identified (more than 28 readers assign the diagnosis B3 in these cases), while it presents some difficulties when the most assigned subtype is CCCa/CCHa (on average, only 21 readers assign B3 as principal diagnosis in this case).

Table 3 Cumulative percent distribution of "negative" and "positive" individual diagnoses reported by all the 31 readers for the 3 subsets of slides with different Majority Diagnosis, and percent range for individual slides. Percentages of "false negative" and "false positive" diagnoses are reported in bold

Majority diagnosis	N° of slides	"negative" diagnoses % (range)	"positive" diagnoses % (range)
B2	21	74 (52–100)	**26 (0–48)**
B3	23	**17 (0–48)**	83 (52–100)
B5	6	**1 (0–3)**	99 (97–100)

Discussion

While reproducibility studies on breast pathology have been the subject of several published reports in the last decades, there are only few published studies dealing with diagnostic reproducibility in reporting NCB, and, at our knowledge, only one of these used the B-classification to evaluate inter-observer concordance [11], but with only two pathologists involved. In the German mammography screening the Diagnosis Optimisation Study (DIOS) has been planned to evaluate the reliability and validity of the histopathological evaluation of core biopsy specimens using the B-classification [12].

The present study suggests that the level of inter-observer agreement among several pathologists in the diagnosis of breast NCB is dependent on the B category, and that serious difficulties may emerge in differentiating the two intermediate categories (B2 and B3). Similar results have been reported by Collins et al. [13]: in their study, five categories similar to B-classification were used, i.e. benign, atypical ductal hyperplasia (ADH), lobular neoplasia, ductal carcinoma in situ (DCIS) and invasive carcinoma, and the lowest levels of diagnostic agreement between local and central pathologists were observed in cases diagnosed by central pathologists as ADH and lobular neoplasia, with percentage of diagnostic agreement of 63% and 53% respectively, denoting a proneness to interobserver variability in diagnosis particularly when standardized criteria are not employed.

An international study concerning reproducibility in the diagnosis of 18 cases of non palpable breast complex lesion NCB, using virtual slides on the world-wide web, among 10 pathologists with a reference diagnosis provided by an experienced breast pathologist, reported a median K value of 0.60 for the agreement with the reference diagnosis, and a median K value of 0.53 for the inter-observer reproducibility (Zito 2008, personal communication).

It should be emphasized that cases selected for the present study do not really represent a consecutive series of breast NCBs, due to an extremely high prevalence (46% according to MD) of B3 diagnoses. In fact, in the breast NCB daily practice the reported prevalence of this diagnosis is around 10%, and was 9.2% in our previous series [14]. The choice of a series of slides in which B2 and B3 types were over-represented has been dictated by a clinical relevance consideration; a diagnostic error is more likely to lead to inadequate clinical outcomes (i.e., the advice to perform an unnecessary surgical excision on one hand, and the failure to excise a "borderline" breast lesion on the other hand) when the slides belong to the intermediate categories, B2 or B3, rather than to the other extreme types (B1 or B4-B5). It is nonetheless likely that with a random series of unselected cases, as observed in daily histopathological NCB diagnostic

Table 4 Percent agreement between individual diagnoses and Majority Diagnosis for specific histological subtypes in a sub-set of 23 cases with B3 as primary MD; readers with primary diagnosis other than B3 were not considered (exact agreement reported in bold in the diagonal) §

		Individual diagnosis (B3 subtypes)						N° of slides	N° of readings	Mean N of readings per slide
		CCC/CCH with A	AEPDT	LIN	PT	PL	RS			
Majority diagnosis (B3 subtypes)	**CCC/CCH with A**	**71.3%**	22.2%	1.8%	–	0.6%	0.6%	8	167	20.9
	AEPDT	24.6%	**52.1%**	6.3%	–	9.2%	4.9%	5	142	28.4
	LIN	10.9%	6.3%	**80.5%**	–	–	–	5	128	25.6
	PT	–	–	–	**100.0%**	–	–	1	30	30.0
	PL	–	–	–	–	**98.2%**	1.8%	2	57	28.5
	RS	6.3%	16.7%	–	–	–	**75.0%**	2	48	24.0
								23	**572**	**24.9**

§ The % values of each line do not always add up to 100 because in a few cases the reader did not report any specific sub-type but only the main diagnostic category (B3)

practice, the level of overall interobserver reproducibility could be higher than that reported in the present study.

The COBRA study [15] assessed the interobserver variability between general and expert pathologists in a large series of large-core needle and open biopsies of non palpable breast lesions. In this study, pathological lesions were classified into five categories that do not exactly correspond to B-classification, even if a borderline category is reported that can be considered similar to B3 category of the B-classification. The Authors observed low levels of interobserver agreement in the category of borderline lesions not only in open breast biopsies but also in large-core needle biopsies, where only 24% of cases with an expert diagnosis of "borderline" were diagnosed similarly by the general pathologists. In addition, the Authors underline that, while in large-core needle biopsies a diagnosis of "borderline" lesion is always followed by a surgical excision, this is not the case for open biopsy specimens, although the increased risk of breast cancer associated with this diagnosis.

Previous studies [16, 17] have shown that a diagnostic surgical excision is indicated when a breast NCB is classified as B3, because in 23%–50% of the cases the excision biopsy will yield malignancy; in our series of B3 cases the rate of malignancy after a B3 diagnosis on NCB is 35% [14]. It is well known, in fact, that B3 category mainly consists of lesions which either are known to show heterogeneity or to have an increased risk (albeit low) of associated malignancy [2, 3].

Misclassification of NCB of the breast by the pathologist may result in two main errors. Firstly, benign NCBs (i.e. B1-B2) may be misclassified as biopsies that implicate further surgical excision of the breast lesion (false positive rate). Secondly, lesions that implicate further surgical excision of the breast lesion (i.e. B3-B5) may be misclassified as benign biopsies, which results in a lack of further surgical diagnostic evaluation (false-negative rate) [12]. In general, inconsistencies in diagnosing NCB as B3 or as B2 will lead to overtreatment or undertreatment of individual patients; therefore, consistency in reporting NCB diagnosis becomes particularly important because of these clinical implications.

Actually in our series, cases classified in the two central categories B2 and B3 showed a proportion of readings in the "false positive" or in the "false negative" categories that approximately represented one fourth or one sixth of the total number of diagnoses, respectively.

A high level of inter-observer agreement was observed in our study for the specific B5 category. While there were a few minor disagreements (i.e., pathologists assigning B4 or even B3 to slides with B5 as MD), only one single reading reported a B2 diagnosis in a case with B5 as MD. Also the opposite situation occurred, with a very few cases of B5 diagnoses in presence of a B2 majority diagnosis. Such major discrepancies might have serious impact on the therapeutic decisions made for individual patients in the sense of failing to diagnose invasive cancer or diagnosing a benign lesion as malignant and performing unnecessary surgical excision. The frequency of these major discrepancies was quite low and seemed to be independent on the type of specimen (NCB versus surgical excision) as reported by Verkooijen et al. [15], who found a similar incidence of major discrepancies in large-core needle biopsy specimens and open breast biopsies.

Finally, our study shows a low level of inter-observer agreement on NCB concerning specific subtypes of lesions within the B3 category, i.e. AEPDT and CCCa/CCHa. At present, this does not have a clinical relevance as all cases

diagnosed as B3 are referred for an open breast surgical excision; however, this confirms the low levels of inter-observer agreement when dealing with proliferative breast lesions (without /with atypia) even in NCB, as previous studies have widely reported for breast surgical specimens. A recent study on reproducibility of pathological diagnosis of columnar cell lesions performed on digitised images [18] reports a moderate to good agreement with a range for kappa values from 0.44 to 0.71. One of the categories with the lowest numbers of complete agreement for individual images was CCCa, and the Authors concluded that for this category more efforts are needed to improve diagnostic consistency .

In conclusion, our study, although based on a series of selected NCB cases (with a clear over-representation of "borderline" cases), suggests the necessity to further investigate the reproducibility and diagnostic accuracy in this particular setting of preoperative assessment of breast pathology, with the specific aim to avoid under or over-diagnosis in NCB reporting, with all their clinical implications. In addition, our results suggest that the agreement on the morphological diagnostic criteria of some B3 subtype lesions, particularly those concerning proliferative atypical breast lesions (AEPDT and CCCa/CCHa), should be improved among pathologists in order to guarantee a more homogeneous diagnostic classification and treatment of patients and a more appropriate comparison among different histological NCB series.

From a practical point of view, multiple step sections and multidisciplinary meetings with radiologists and surgeons to assess the radio-pathological correlation between NCB and mammographic findings appear to be mandatory. A consultation with colleagues with extensive expertise in breast pathology should be recommended when some degree of uncertainty emerges about a specific NCB diagnosis (particularly when the two central B2 and B3 categories are involved) in order to avoid undertreatment or overtreatment of the patients.

Conflict of interest statement The authors declare that they have no conflict of interest

References

1. El-Sayed ME, Rakka EA, Reed J et al (2008) Audit of performance of needle core biopsy diagnoses of screen detected breast lesions. Eur J Cancer 44:2580–2586
2. NHSBSP Breast Screening Programme (2001) Guidelines for Non-Operative Diagnostic Procedures and Reporting in Breast Cancer Screening. Sheffield: NHSBSP Pub. No.50
3. European Commission Working Group on Breast Screening Pathology (2006) 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. Volume 6, 4th edn. Office for Official Publications of the European Communities, Luxemburg, pp 219–312
4. Pinder SE, Reis-Filho JS (2007) Non-operative breast pathology. J Clin Pathol 60:1297–1299
5. Verkooijen HM (2008) Needle core biopsy for screen detected breast lesions: time to raise the bar. Eur J. Cancer 44:2540–2541
6. Kumaroswamy V, Liston J, Shaaban AM (2008) Vacuum assisted stereotactic guided mammotome biopsies in the management of screen detected microcalcifications: experience of a large breast screening centre. J Clin Pathol 61:766–769
7. Lazarus E, Mainiero MB, Schepps B et al (2006) BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value. Radiology 239:385–391
8. Schnitt SJ, Vincent-Salomon A (2003) Columnar cell lesions of the breast. Adv Anat Pathol 10:113–124
9. Cohen J (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20:37–46
10. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:59–174
11. Bonk U, Gohla G, Sauer U et al (2005) Grosse Ubereinstimmung Eergebnisse des ersten Mammographie-Screening-Projekts in Bremen aus histopathologischer Sicht. Deutsche Arzteblatt online (http://www.aerzteblatt.de/aufsaetze/0501)
12. Kluttig A, Trocchi P, Heinig A et al (2007) Reliability and validity of needle biopsy evaluation of breast-abnormalities using the B-categorization—design and objectives of the Diagnosis Optimisation Study (DIOS). BMC Cancer 7:100–108
13. Collins LC, Connolly JL, Page DL et al (2007) Diagnostic agreement in the evaluation of image-guided breast core needle biopsies. Am J Surg Pathol 28:126–131
14. Houssami N, Ciatto S, Bilous M et al (2007) Borderline breast core needle histology: predictive values for malignancy in lesions of uncertain malignant potential (B3). Br J Cancer 96:1253–1257
15. Verkooijen HM, Peterse JL, Schipper MEI et al (2003) 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 39:2187–2191
16. Lee AD, Denley HE, Pinder SE et al (2003) Excision biopsy findings of patients with breast needle core biopsies reported as suspicious of malignancy (B4) or lesion of uncertain malignant potential (B3). Histopathology 42:331–336
17. Dillon MF, McDermott EW, Hill AD et al (2007) Predictive value of breast lesions of "Uncertain Malignant Potential" and "Suspicious for malignancy" determined by needle core biopsy. Ann Surg Oncol 14:704–711
18. Tan PH, Ho BCS, Selvarajan S et al (2005) Pathological diagnosis of columnar cell lesions of the breast: are there issues of reproducibility? J Clin Pathol 58:705–709

Pathologists participating into the study

Armando Andreozzi, Paola Apicella, Simonetta Bianchi, Angelo Giovanni Bonadio, Rodolfo Brizio, Isabella Castellano, Maria Grazia Cattani, Paolo De Giuli, Raffaella Fabbri, Guido Ficarra, Maria Pia Foschini, Giorgio Gardini, Roberto Incensati, Luigia Macrì, Tiziana Megha, Fiorella Nuzzo, Alessandra Papucci, Francesca Pietribiasi, Valentina Rapicano, Nicoletta Ravarino, Monica Ricci, Antonio Rizzo, Adriana Salvadori, Anna Sapino, Roberta Sarnelli, Umberto Simi, Mara Truglia, Vania Vezzosi, Laura Viberti, Carla Zavattero, Vanna Zucchi.