FREQUENCY OF DIAGNOSTIC CONCORDANCE IN CORE BREAST BIOPSIES OF CATEGORY B3 AND B4 AT A TERTIARY CARE CENTRE

Ghazala Sadaf, Muhammad Tahir Khadim, Rabia Ahmed, Muhammad Asif, Safina Ahmed, Shahid Jamal, Sammeen Salim

Armed Forces Institute of Pathology/National University of Medical Sciences (NUMS) Rawalpindi Pakistan

ABSTRACT

Objective: To evaluate interobserver variability in core breast biopsies of category B3 and B4 at a tertiary care centre

Study Design: Retrospective, observational, cross-sectional study.

Place and Duration of Study: Histopathology Department, Armed Forces Institute of Pathology (AFIP), Rawalpindi, from Mar 2017 to Jun 2018.

Methodology: A total of thirty cases of histologically confirmed lesions of category B3 and B4 on core breast biopsies from March 2017 to June 2018 were retrieved from archive of Histopathology department, AFIP Rawalpindi. Patients' age, histologic diagnosis and reporting B category were noted. The selected cases were anonymized and distributed among 3 pathologists for independent review. The participating pathologists were kept unaware of the findings given by fellow participants and previously agreed reference diagnoses. Study cases were reviewed by the participating pathologists. The results were analyzed and overall concordance rate, discordance rate, over interpretation rate and under interpretation rate were calculated.

Results: A total of 30 (n=30) patients were enrolled of which 22 were assigned category B3 (73%) and 8 were assigned category B4 (27%) on initial microscopic evaluation (original opinion). The ages of the study patients ranged from 25 to 85 years. The average concordance rate of morphologic findings of study pathologists compared to the reference diagnosis was 63% (19/30). Among these the average concordance rate compared to reference diagnosis among category B3 lesions was 73% (16/22) and among category B4 lesions was 37.5% (3/8). The average rate of disagreement of morphologic findings of study pathologists compared with the reference diagnosis was 37% (11/30). Among these the average rate of disagreement compared to reference diagnosis among category B3 lesions 27% (6/22) and among category B4 lesions was 62.5% (5/8). The overall over interpretation rate was 44% and overall under interpretation rate was 56%. The overall rate of unanimous agreement of independent diagnoses among the three panel consultants was 50% (15/30). Among these 15 cases 12 were assigned category B3 and 3 were assigned category B4. The overall rate of disagreement of independent diagnoses among the three panel consultants was 50% (15/30).

Conclusion: In our study the opinion was established by interpretation of a single slide. Average concordance rate among the participating pathologists' findings and the previously agreed upon final diagnoses was 63% and disagreement was 37%, the rate of agreement being higher in the category B3 and lower in the category B4 lesions. However effect of these results on the therapy and prognosis of patients still needs to be evaluated.

Keywords: Core breast biopsy, Category B3 and B4, Interobserver variability.

INTRODUCTION

Breast cancer is the commonest detected cancer worldwide and the main cause of cancer related mortality among women[1-3]. As distant metastases are considered as the chief reason of death, prompt diagnosis of breast cancer followed by appropriate therapy can effectively decrease the cancer related mortality[4,5]. Generally for the diagnosis histopathological evaluation of tissue biopsy is regarded as the most reliable tool which forms the basis of subsequent treatment of the patients with breast lesions[6,7]. Recently there has been a rise in the number of core biopsies being performed as well as frequency of non invasive breast lesions being picked up in the sub-clinical and pre-symptomatic stage owing to

Correspondence: Dr Ghazala Sadaf, Histopathology Department, Armed Forces Institute of Pathology, Rawalpindi Pakistan
Received: 01 Feb 2019; revised received: 20 Oct 2019; accepted: 31 Oct 2019

introduction of screening programs using mammography and other new imaging techniques for early diagnosis of breast cancer[8-10]. These lesions are high risk type and include intraductal proliferative, non invasive and precursor breast lesions having uncertain behavior[8,11,12]. These are the lesions which we included in our study and which have been categorized as B3 and B4 according to B classification. The category B3 i.e lesion of uncertain malignant potential includes atypical intraductal epithelial proliferations (AIDEP) comprising atypical ductal hyperplasia (ADH), flat epithelial atypia (FEA), apocrine atypia and atypia that does not conform to one of these patterns; lobular neoplasia comprising atypical lobular hyperplasia (ALH) and lobular carcinoma in situ (LCIS); Phyllodes tumour; papillary lesions; radial scar; mucocoele-like lesions (mucin in the stroma); some rare lesions such as adenomyoepithelioma, microglandular adenosis, spindle cell lesions such as fibromatosis and vascular lesions such as hemangioma. The category B4 i.e suspicious includes core biopsies containing probable carcinoma but cannot be diagnosed definitively because of sampling, processing or artefactual limitations, scanty non-high grade DCIS and other high grade lesions short of being diagnosed as LCIS or DCIS[13].

As far as interobserver variability is concerned, there has been an increasing trend of seeking second opinions due to increased focus on diagnostic discrepancies in medical writings, publications and mass media[14]. Therefore various medical institutes have formulated and applied guidelines and protocols for acquiring second opinions as this may increase diagnostic precision and enhance standard of patient management[15]. In practice of pathology breast lesions are generally considered to be among the most challenging and problematic as far as interpretation and diagnosis is concerned[16]. And among breast lesions, the most difficult ones are those that are borderline and fall between atypical ductal hyperplasia (ADH) and limited extent low-grade ductal carcinoma in situ (DCIS); the lesions which lie in a single pathologic sequence and still are segregated technically owing to very minute histologic variations and their degree[17,18]. Owing to their borderline nature these breast lesions has always been problematic and disparity in opinions is inevitable in their assessment and diagnosis[19]. These variations among opinions of different pathologists on these "grey zone" or uncertain and suspicious lesions can be attributed to various aspects like the biopsy needle size, degree of atypia, associated atypia, pathologist's proficiency and experience, and organizational setup (eg, tertiary care center vs local hospital)[20]. Although these variations can be well explained but these result in consequent open biopsy of the lesion for final diagnosis, create apprehension among the patients and their families because of false positive reports and result in erroneous judgements in patient management[12,21].

The evidence from local population is limited and there is hardly any study on evaluation of inter observer variability of opinions on breast biopsies. Therefore present study was designed to study the inter observer variability in histopathological evaluation of breast biopsies of these non invasive lesions which are difficult to diagnose with certainty so that we are able to determine the proportion of patients whose diagnosis is changed on review resulting in change in subsequent management plan and outcome.

METHODOLOGY

This retrospective, observational, cross-sectional study was carried out at department of Histopathology, Armed Forces Institute of Pathology, Rawalpindi, from March 2017 to June 2018 after taking approval from the Institutional Review Board. All the reports of the core breast biopsies from March 2017 to June 2018 of female patients of all ages were reviewed and a total of 30 cases were identified to be reported as category B3 and B4. Among these 22 cases had been originally reported as B3 and 8 had been originally reported as B4. These 30 cases of core breast biopsies which were diagnosed on routine histopathology as category B3 and B4 lesions of breast were included in the study. The initial diagnosis

of each case which had been given after consensus on intradepartmental consultation was considered as the reference diagnosis. Other B categories of breast lesions such as B1, B2 and B5 i.e definitely benign or malignant were excluded. The original case slides were retrieved from archive of Histopathology department, AFIP Rawalpindi. Patients' age, histologic diagnosis and reporting B category were noted. A panel of three pathologists who had expertise in interpretation of breast pathology i.e were FCPS qualified, were practicing as consultant histopathologists for at least two years in tertiary care centre and had high volume exposure (≥10 breast biopsy specimens weekly) was selected. The selected cases were anonymized and distributed among 3 panel pathologists for independent interpretation and review. The participating pathologists were kept unaware of the findings given by fellow participants and previously agreed reference diagnoses however they were informed about the age of patients. The results were analyzed and overall concordance rate, discordance rate, over interpretation rate and under interpretation rate were calculated.

RESULTS

A total of 30 patients were enrolled of which 22 were assigned category B3 (73%) and 8 were assigned category B4 (27%) on initial microscopic evaluation (agreed upon final diagnosis). The age range of the study patients was 25 to 85 years. The average rate of agreement of morphologic evaluations of study pathologists compared to the reference diagnosis was 63% (19/30). Among these the average concordance rate compared to reference diagnosis among category B3 lesions was 73% (16/22) and among category B4 lesions was 37.5% (3/8). The average frequency of disagreement of morphologic evaluations of study pathologists compared with the reference diagnosis was 37% (11/30). Among these the average rate of disagreement compared to reference diagnosis among category B3 lesions was 27% (6/22) and among category B4 lesions was 63.5% (5/8). The overall under interpretation rate was 45.5% (5/11) and among these underinterpreted cases all the cases were originally reported as B4 which were later reported as B5. The overall over interpretation rate was 54.5% (6/11) and among these over interpreted 5 cases were originally reported as B3 which were later reported as B2 whereas 1 case was originally reported as B4 which was later diagnosed as B3. The overall rate of unanimous agreement of independent diagnoses among the three panel consultants was 50% (15/30). Among these 15 cases 12 were of category B3 and 3 were of category B4. The overall rate of disagreement of independent diagnoses among the three panel consultants was 50% (15/30). These results are shown in tables-I & II.

Table-I: Frequency of concordance and discordance rates.

Categories	Concordance Rate, n (%)	Discordance Rate, n (%)
Category B3 (n=22)	16 (73%)	6 (27%)
Category B4 (n=8)	3 (37.5%)	5 (63.5%)
Total (n=30)	19 (63%)	11 (37%)

Table-II: Over interpretation and under interpretation rates.

Categories	Over-interpretation Rate, n (%)	Under-interpretation Rate, n (%)
Category B3 (n=5)	5 (83%)	-
Category B4 (n=6)	1 (17%)	5 (100%)
Total (n=11)	6 (54.5%)	5 (45.5%)

DISCUSSION

This study showed that the average agreement rate of morphologic evaluations of study pathologists compared to the reference diagnosis was 63% (19/30). Among these the average concordance rate compared to reference diagnosis among category B3 lesions was 73% (16/22) and among category B4 lesions was 37.5% (3/8). Study conducted by Elmore *et al* showed that among a total of 6900 different evaluations the pathologists participating in the study gave the same opinion as the previously agreed final diagnosis in 75.3% of the cases. The average rate of agreement was highest in the categories of invasive carcinoma followed by benign high being 96% and 87% respectively. The rate of agreement of participating pathologists to the previous final

diagnosis was lower in the category of atypia being only 48% whereas rate of agreement for DCIS was high being 84%[6]. According to Perez *et al* moderate agreement was observed between the original histopathological diagnosis and the second opinion i.e percentage concordance being 83%. After the review, the diagnosis of malignancy was confirmed in 140/163 cases (86%) and the diagnosis of benign lesions was confirmed in 34/46 cases (74%)[12]. Elmore *et al* reported that the overall rate of diagnostic agreement as compared to reference diagnosis was 75.3%. The agreement rate within diagnostic categories was 47.8% for cases of atypia, 84.1% for DCIS, 87.1% for benign without atypia and 96.1% for invasive carcinoma[14]. In a study conducted by Allison *et al* the diagnostic agreement in cases of atypical ductal hyperplasia which is a category B3 lesion with the consensus diagnosis of ADH occurred for 48% of case interpretations. The remainder of interpretations were quantified as follows: 25% benign, 10% fibroepithelial atypia/lobular neoplasia, 17% DCIS and <10% invasive carcinoma[17]. Elmore *et al* (Ann Intern Med 2016) in their study reported that the overall rate of agreement with reference diagnosis was 92.3%. Verification of invasive breast cancer and benign without atypia diagnoses was highly probable; values were 97.7% and 97.1% respectively. Verification was less probable for atypiabeing 37.8% and ductal carcinoma in situ (DCIS) being 69.6%[18].

In this study overall rate of disagreement of diagnostic interpretations of participating pathologists compared with the reference diagnosis was 37% (11/30). Among these the average rate of discrepancy compared to reference diagnosis among category B3 lesions was 27% (6/22) and among category B4 lesions was 63.5% (5/8). The overall over interpretation rate was 45.5% (5/11) and among these over interpreted cases all the cases were originally reported as B4 which were later reported as B5. The overall under interpretation rate was 54.5% (6/11) and among these under interpreted 5 cases were originally reported as B3 which were later reported as B2 whereas 1 case was originally reported as B4 which was later diagnosed as B3. According to Elmore *et al* the average rate of disagreement among the morphologic evaluations of the study pathologists and previous final diagnosis was 24.7%. Generally the tendency of over interpreting and under interpreting the histopathologic findings was not related to the pathologic type of lesion or a certain pathologist but it showed wide variations among the pathologists and type of lesions. Although the rate of over-interpretation of the cases of DCIS as invasive was as low as 3%, the rate of over interpretation of atypia was recorded up to 17% and over interpretation rate of benign without atypia was seen in 13% cases. The rate of under-interpretation of invasive carcinoma was 4%, of DCIS was 13% and that of atypia was 35%[6]. Perez *et al* reported that overall discordance rate of diagnosis between the original reference diagnosis and review was low being 17%. The highest disagreement was observed in cases of ductal carcinoma in situ with microinvasion (6/6 cases; 100%). Important discordance was observed in cases of atypical ductal hyperplasia (16/30 cases; 53%) and ductal carcinoma in situ (25/75 cases; 33%). Regarding the ductal carcinoma in situ, good agreement was observed between the original diagnosis and the review (29/39 cases, percent agreement = 74%)[12]. According to Elmore *et al* the overall al disagreement rate of second opinion as compared to the original reference diagnosis 24.7%. The highest misclassification rate within diagnostic categories after single interpretation was for cases of atypia (52.2%), followed by DCIS (15.9%), benign without atypia (12.9%), and invasive carcinoma (3.9%). The overall over interpretation rate after second opinion as compared to the original reference diagnosis was 9.9%. The highest over interptretation rate within diagnostic categories after single interpretation was for cases of atypia (17.4%), followed by benign without atypia (12.9%) and DCIS (2.6%). The overall under interpretation rate after second opinion as compared to the original reference diagnosis was 14.8%. The highest under interptretation rate within diagnostic categories after single interpretation was for cases of atypia

(34.7%), followed by DCIS (13.3%) and invasive (3.9%)[14]. In a study conducted by Allison *et al* the diagnostic disagreement in cases of atypical ductal hyperplasia which is a category B3 lesion with the consensus diagnosis of ADH occurred for 52% of case interpretations. Of these of 25%were under interpreted as benign, 10% were assigned the same category with a different diagnosis like fibroepithelial atypia/lobular neoplasia and rest were over interpreted as DCIS (17%) and invasive carcinoma (<10%)[17]. According to Elmore *et al* the overall rate of disagreement between the review and reference diagnosis was 7.7%. The overall over interpretation rate being 4.6% and under interpretation rate being 3.2%. Over interpretation rates were 26% for benign non-proliferative, 18% for proliferative without atypia, 17% for atypia, and 3% for DCIS. Under interpretation rates were 8% for proliferative without atypia, 35% for atypia, 13% for DCIS, and 4% for invasive breast cancer[18].

According to our study the overall rate of unanimous agreement of independent diagnoses among the three panel consultants was 50% (15/30). Among these 15 cases 12 were assigned category B3 and 3 were assigned category B4. The overall rate of disagreement of independent diagnoses among the three panel consultants was 50% (15/30). Elmore *et al* (JAMA 2015) reported that the 3 study pathologists agreed to each other completely on the diagnosis in 75% (180/240) of the cases after the initial individual assessment[6]. According to Elmore *et al* the average between pathologist pair wise agreement rate for single interpretations of the same case was 70.4%[14].

No study conducted on interobserver variability in histopathological dignosis of breast biopsies in local population was found in literature however the results of this study were consistent with the majority of published studies. In our study only category B3 and B4 lesions were included which are atypical and suspicoius lesions and these are the lesions which are most likely to be misinterpreted. This is consistent with the published studies including Elmore *et al*, Perez *et al*, Elmore *et al*, Ellison *et al* and Elmore *et al*[6,12,14,17,18].

CONCLUSION

In this study of pathologists the overall discordance between the individual pathologists' assessment and the expert consensus-derived reference diagnoses was 37% which is significant and can induce clinical errors in treatment decisions. This diagnostic variability is related to multiple factors but consensus conferences, standardized electronic reporting formats and comments on suboptimal specimen quality can be used to reduce diagnostic variability. Particularly second opinions can statistically significantly improve diagnostic agreement for pathologists' interpretations of breast biopsy specimens; however, variability in diagnosis will not be completely eliminated, especially for breast specimens with atypia. Therefore this finding may be useful in formulating management plans for women who have an indeterminate biopsy result and more research is required to evaluate the effect of these results on the management of patients.

ACKNOWLEDGMENT

We would like to acknowledge all the studied patients for their cooperation. We wish to pay special gratitude to our families and friends for their continuous support and motivation.

CONFLICT OF INTEREST

This study has no conflict of interest to be declared by any author.

REFERENCES

1. Morse JM, Pooler C, Vann-Ward T, Maddox LJ, Olausson JM, Roche-Dean M, et al. Awaiting diagnosis of breast cancer: strategies of enduring for preserving self. Oncol Nurs Forum 2014; 41(4): 350-59.
2. Melzer C, Ohe J, Hass R. Breast carcinoma: From initial tumor cell detachment to settlement at secositesettlement at secondary sites. Biomed Res Int 2017; 2017: 8534371.
3. Tao YK, Shen D, Sheikine Y, Ahsen OO, Wang HH, Schmolze DB, et al. Assessment of breast pathologies using nonlinear microscopy. Proc Natl Acad Sci USA 2014; 111(43): 15304-09.
4. Campuzano S, Pedrero M, Pingarrón JM. Non-invasive breast cancer diagnosis through electrochemical biosensing at different molecular levels. Sensors (Basel) 2017; 17(9): s17091993.
5. Wang L. Earlydiagnosis of breast cancer. Sensors (Basel) 2017; 17(7): e1572.
6. Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson ANA, et al. Diagnostic concordance among pathologists

interpreting breast biopsy specimens. J Am Med Assoc 2015; 313(11): 1122–32.
7. Allison KH, Reisch LM, Carney PA, Weaver DL, Schnitt SJ, O'Malley FP, et al. Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel. Histopathology 2014; 65(2): 240-51.
8. Jackson SL, Frederick PD, Pepe MS, Nelson HD, Weaver DL, Allison KH, et al. Diagnostic reproducibility: What happens when the same pathologist interprets the same breast biopsy specimen at two points in time? Ann Surg Oncol 2017; 24(5): 1234-41.
9. Bianchi S, Caini S, Cattani MG, Vezzosi V, Biancalani M, Palli D. Diagnostic concordance in reporting breast needle core biopsies using the B classification-A panel in Italy. Pathol Oncol Res 2009; 15(4): 725-32.
10. Gomes DS, Porto SS, Balabram D, Gobbi H. Inter-observer variability between general pathologists and a specialist in breast pathology in the diagnosis of lobular neoplasia, columnar cell lesions, atypical ductal hyperplasia and ductal carcinoma in situ of the breast. Diagn Pathol 2014; 9(1): 121.
11. Murray M. Pathologic high risk lesions, diagnosis and management. Clin Obstet Gynecol 2016; 59(4): 727-32.
1. 12.Perez AA, Balabram D, SallesMde A, Gobbi H. Consultation in breast surgical pathology: interobserver diagnostic variability of atypical intraductal proliferative lesions. Rev Bras Ginecol Obstet 2013; 35(4): 164-70.
12. Lee AHS, Anderson N, Carder P, Cooke J, Deb R, Ellis IO, et al. Guidelines for non-operative diagnostic procedures and reporting in breast cancer screening. 2016. The Royal College of Pathologists. 2016; 1-74. Available from: https://www.rcpath.org/uploads/assets/4b16f19c-f7bd-456c-b212f557f8040f66/G150-Non-op-reporting-breast-cancer-screening-Feb17.pdf
13. Elmore JG, Tosteson AN, Pepe MS, Longton GM, Nelson HD, Geller B, et al. Evaluation of 12 strategies for obtaining second opinions to improve interpretation of breast histopathology: simulation study. BMJ 2016; 353: i3069.
14. Geller BM, Nelson HD, Carney PA, Weaver DL, Onega T, et al. Second opinion in breast pathology: policy, practice and perception. J Clin Pathol 2014; 67(11): 955-60.
15. Quintyn-Ranty ML, Gordien K, Caveriviere P, Mery É, Jamme-Lallemand M, Wuithier P. Improving practice in breast pathology: 34-months experience of the regional SENOPATH network and webinars as a tool for diagnosis of difficult lesions of the breast. Bull Cancer 2015; 102(10): 823-33.
16. Allison KH, Rendi MH, Peacock S, Morgan T, Elmore JG, Weaver DL. Histologic features associated with diagnostic agreement in atypical ductal hyperplasia of the breast: Illustrative cases from the B-Path study. Histopathol 2016; 69(6): 1028-46.
17. Elmore JG, Nelson HD, Pepe MS, Longton GM, Tosteson AN, Geller B. Variability in pathologists' interpretations of individual breast biopsy slides: A population perspective. Ann Intern Med 2016; 164(10): 649-55.
18. Semwal S, Joshi D, Khare A, Goel G, Kapoor N. Validation and modification of the Masood scoring index for the diagnosis of atypical breast lesions. Acta Cytol 2017; 61(2): 111-16.
19. Darvishian F, Singh B, Simsir A, Ye W, Cangiarella JF. Atypia on breast core needle biopsies: reproducibility and significance. Ann Clin Lab Sci 2009; 39(3): 270-76.
20. Hunt RJ, Steel JR, Porter GJ, Holgate CS, Watkins RM. Lesions of uncertain malignant potential (B3) on core biopsy in the NHS Breast Screening Programme: is the screening round relevant. Ann R Coll Surg Engl 2012; 94(2): 108-11.