Special ISSUE • 2016 • vol.1 • 9 - 16

BREAST ULTRASONOGRAPHY: WHO MAKES THE BEST DIAGNOSIS?

Serra, N.[1*]; Gagliardi, G.[1]; Faggian, A.[1]; Grassi, R.[2]

[1] Radiology Department, Second University of Naples
[2] Radiology and Radiotherapy Department, University of Sassari

* nicola.serra@unina2.it

Abstract

To evaluate the most efficient between breast and general radiologists and any significant or clinically relevant differences in breast ultrasound diagnosis. Four hundred consecutive patients attending for breast ultrasound were included. Each patient was examined by a breast radiologist and subsequently by general radiologist. Both operators noted their findings and wrote a concluding report without conferring. Reports were compared, with histological and biological analysis . 400 patients with 384 females and 16 males, with age range 16-86 y.o. and mean age about 48 y.o., are examined. The breast radiologist results most accurate in comparison the general radiologist in breast ultrasound (p-value = 0.0625) with a significance level of 10%.
The breast radiologist is more accuracy in comparison to general radiologist in the evaluation of breast pathology in the US diagnoses.

Key words: ultrasonography, breast cancer, radiologist

Introduction

Breast ultrasonography is a quick, cheap, noninvasive and safe diagnostic method; it is the first choice in underforty women and a complementary examination for overforty women also to evaluate palpable abnormalities and characterize masses detected at mammography and magnetic resonance imaging (MRI) [1]. It was primarily used to distinguish cystic from solid masses. Solid masses with spiculation, taller-than-wide orientation, angular margins, microcalcifications, and posterior acoustic shadowing are to be consider malignant or suspicious; instead, if a mass presents two or three lobulations, ellipsoid shape, a thin capsule and a homogeneously echogenic echotexture, could be consider benign [2]. But often breast lesions form an extremely heterogeneous group consisting of benign lesions mimicking carcinoma [3,4] with a poor interobserver agreement in US BI-RADS classification [5]; therefore, sonography is also currently used to guide for interventional breast procedures, such as US-guided fine needle citology or biopsies [6]. Fine-needle aspiration plays an important role also in characterization, in tumor grading and in immunocytochemical identifying specific tumor markers, even if biopsies allow precise and accurate diagnosis avoiding repeated examinations causing the patient anxiety and unnecessary surgical procedures [7].

However, the main limitation of breast ultrasonography are operator dependent nature and its low specificity, leading to a high rate of false positive results [8]. Several methods are developed to improve the performance of US and cost/effectiveness ratio, such as higher resolution transducers, Doppler imaging, harmonic imaging, spatial and frequency elastosonography, contrast US, 3D US and automated breast ultrasonography. Although these advanced technology, interobserver agreement between radiologists in breast ultrasonography is poor and only a fellowship training in breast imaging could improve the accuracy of image interpretation [1, 9]. Purpose of our study is to analyse the accuracy of general radiologists in breast disease diagnosis and to assess the quality of their work. In particular we wanted to know whether there was agreement between the findings of the breast radiologists and those of general radiologists.

Methods

This research is a prospective study on accuracy between general radiologist and breast radiologist.

This study included 400 patients in the period between 01 December 2010 and 01 December 2012. The patients sample was composed from 96 % females and 4 % males with ages into range 16-86 and with mean ages about 48 and standard deviation about 15. Inclusion criteria were: breast and ovarian familiarity, BRCA 1 and/or 2 mutation, history of breast surgery, history of ≥2 years hormone replacement therapy and recurrent cyclical mastodynia. Exclusion criteria were: proliferative benign breast diseases, breast implants. To start by our sample data (400 patients), we have considered four random subgroups of 100 patients. Each subgroup was examined by a breast radiologist with 25 years of experience of breast diseases (that have read at least 480 screening mammograms or breast exams (ultrasonography, MRI) per year or 960 exams every two years, according to Mammography Quality Standards Act) [10] and subsequently by general radiologist, with 25 years of experience who did not work exclusively in breast diagnostic field. Each breast radiologist and general radiologist had examined one subgroup only. In the diagnostic phase, there was no communication between the general and breast radiologist. For each group both the breast radiologist and general radiologist separately has recorded the number of images taken, the number of images saved, their confidence in their diagnostic findings and conclusions. The examinations in the study were limited to the breast.

US was performed using a Logiq S6 scanner (GE Healthcare, Waukesha, WI, USA) with a multifrequency matrix-array linear transducer (7-14 MHz). Colour Doppler US was also performed to study intralesional vascularity.

For each exam, radiologists defined the age intervals (≤40, 40-70, ≥ 70 y.o.), the number of the lesions detected (1, 2 or 3), the lesion size (<5 or ≥5mm), breast pattern (fibroglandular or no fibroglandular), the localization (UOQ, LIQ, LOQ, UIQ, RA) and placed lesions in two categories: not suspicious (negative exams, cysts or fibroadenoma) or suspicious lesions (cancer). The effective diagnosis was performed with 2 years follow-up (152 patients), cytological diagnosis (66 patients) after US-guided fine needle aspiration, histopatological diagnosis (32 patients) after US-guided biopsies or surgical excision. US-guided biopsies were performed with 14-gauge needles. Two to seven samples were obtained (mean two to three). Histological examination was performed by an experienced pathologist.

The effective diagnosis was performed with 2 years follow-up (US examination every six months) for lesions included in BI-RADS 1 and 2 category,

US-guided citology or biopsies for lesions included in BI-RADS 3, 4 and 5 categories.

Statistical analysis

The statistical analysis were performed by Matlab statistic toolbox ver. 2008 for Windows at 32 bit. The chi-square test [11] were performed to define if the subgroups were statistically equal, instead the McNemar's exact test [12] was performed to evaluate the diagnostic accuracy between breast radiologists and general radiologists in breast ultrasound examination. Finally Sensitivity and specificity with confidence intervals [13], were computed on the all results obtained by breast radiologists and general radiologists.

Results

Our goal was to evaluate the most efficiency between breast and general radiologist in the evaluation of breast pathology with US diagnosis [14]. In table 1 we have described the characteristics and the diagnostic responses to the breast ultrasound for each subgroup. The subgroup 1 had a major patients number with ages into range 40-70 (56%), one lesion (40%), lesion size 5 mm (32%), breast pattern NO FG (52%), 9% of multicentric benign pathologies and 3 % of the multifocal cancers. The subgroup 2 had a major numbers of patients with ages into range 40-70 (52%), one lesion (46%), lesion size 5 mm (32%), breast pattern NO FG (60%), 6% of multicentric benign pathologies and 4 % of multifocal cancers. The subgroup 3 had a major numbers of patients with ages into range 40 (44%), one lesion (44%), lesion size 5 mm (36%), breast pattern FG (52%), 5% of multicentric benign pathologies and 3 % of the multifocal cancers. Finally the subgroup 4 had a major patients number with ages into range both 40 and 40-70 (44%), one lesion (48%), lesion size 5 mm (40%), breast pattern NO FG (52%), 5% of multicentric benign pathologies and 3 % of the multifocal cancers. In table 2 we had reported the diagnosis responses for each breast and general radiologists. We observe by Table 2 that both the breast radiologist 1 and 4 were perfect in all diagnoses; the breast radiologist 2, had not identified one cyst evaluating the patient negative; the breast radiologist 3 had not identified one fibroadenoma, evaluating the patient negative. The general radiologist 1, had not identified one fibroadenoma evaluating the patient negative; the general radiologist 2, had not identified two fibroadenomas, evaluating one negative case and one cist, while one cancer was evaluated like cyst; the general radiologist 3 had not identified two fibroadenomas and one cyst, evaluating the three patients negative to the diagnosis; finally the general radiologist 4, had not identified one cyst evaluating the patient negative. In table 3 and in figure 1, we have showed the results on total group of 400 patients, for the breast and general radiologists. Particularly in table 3 we have expressed in parentheses the total number of errors, considering both false positive and false negative.

We observe by Table 3, that the breast radiologists had a total error in breast diagnosis of 0.5% (they hadn't correctly identified one fibroadenoma and one cyst), 46.50% of negative responses with 0.5% of error (two false negative cases: one fibroadenoma and one cyst), 13.75% of fibroadenoma responses with 0.25% of error (one fibroadenoma was not correctly identified), 8.00% of cancer responses with 0.00% of error and finally 31.75% of cyst responses with 0.25% of error (one cyst was not correctly identified). Indeed the general radiologists had a total error in breast diagnosis of 2.00% (they hadn't correctly identified five fibroadenomas, one cancer and two cysts), 47.50% of negative responses with 1.50% of error (six false negative cases: five fibroadenomas and one cyst), 12.75% of fibroadenoma responses with 1.25% of error (five fibroadenomas were not correctly identified), 7.75% of cancer responses with 0.25% of error (one cancer was not correctly identified) and finally 32.00% of cyst responses with 1.00% of error (two false positive cases: one Fibroadenoma and one cancer; and two cysts were not correctly identified). In Figure 1 we had shown graphically the results of Table 3. To start Table 2, we had compared the proportion of correct diagnoses between breast and general radiologists. One tail McNemar's exact test was performed with a significance level 10%. The result was that the breast radiologists had a proportion of correct diagnoses (53.5 %) statistically major in comparison to the proportion of correct diagnoses of general radiologists (52.5 %), p-value = 0.0625 (Table 3).

In conclusion, we had compared the accuracy between breast radiologist and general radiologist in terms of Sensitivity (Se) and specificity (Sp)[14] and computed the confidence intervals with score method with continuity correction. The results on total sample were for the breast radiologists, Se = 0.99 with confidence interval (0.973, 0.998) and Sp = 0.989 with confidence interval (0.971, 0.997), while for the general radiologists, Se = 0.972 with confidence interval (0.949, 0.986) and Sp = 1 with confidence interval (0.988, 1.00), i.e. with a significance level of 5%, there were no significant

differences in the evaluations both true-positive and true-negative between breast radiologists and general radiologists (Table 4).

Discussion

By this study we have observed that the breast radiologists identifies all cases of breast cancers (32/32), 55/56 fibroadenomas, 127/128 cysts and 186/184 negative cases, with global error of 0,5% (2/400). Instead, the general radiologist reported 190/184 negative cases, 31/32 cancers, 128/128 cases of breast cysts (but with four errors) and 51/56 fibroadenomas, with global error of 2% (8/400). Two false negative cases reported by breast radiologists were one cyst and one fibroadenoma, both with a diameter <5mm.

Instead, 6 false negative cases were evaluated by general radiologists and were four fibroadenomas (with diameters < 5mm) and two cysts (one with diameters < 5mm and one with diameter 5mm); one fibroadenoma was evaluated like a cyst (with diameter < 5mm) and one cancer was evaluated like a cyst (with diameter <5mm).

We had that the subgroups were not statistically different about patients number responses (p-value = 0.69, with ^{2} test. We don't had significant differences among the breast radiologists diagnoses p-value = 0.368(^{2} test) and among the general radiologists diagnoses, p-value = 0.836 (^{2} test) in comparison to effective diagnosis (Table 2). In each group the patients underwent the standard breast examination by a breast radiologist and subsequent in the same day, by a general radiologist. These tests and procedures, imply that on our results there could be a very low probability of statistical bias. By Table 2, we had compared the proportion of correct diagnoses between breast radiologists (53.5 %) and general radiologists (52.5 %), with one tail McNemar's exact test. The results were, with a significance level 10%, that the breast radiologists had a proportion of correct breast diagnoses statistically major in comparison to the proportion of correct diagnoses of general radiologists, p-value = 0.0625 (Table 3). Finally, we had compared the performances both breast and general radiologists in terms of Sensitivity (Se) and specificity (Sp) and computed the confidence intervals with score method with continuity correction. On total group the breast radiologists had Se = 0.99 with confidence interval (0.973, 0.998) and Sp = 0.989 with confidence interval (0.971, 0.997), instead the general radiologists had Se = 0.972 with confidence interval (0.949, 0.986) and Sp = 1 with confidence interval (0.988, 1.00), i.e. with a significance level of 5%, on total sample there were not significant differences in the evaluations both true-positive and true-negative between breast radiologists and general radiologists (Table 4). In comparison with cytological, histological and/or 2 years follow-up diagnosis performed on all patients, breast radiologists are more accurate in comparison to general radiologists in the evaluation of breast pathologies with US methodology with a probability of 90 %. Our study complies with other studies proving that a higher volume of procedures among individual phisicians are associated with better outcomes across a variety of conditions [15,16]; despite United States requires interpretation of only 960 breast exams over a 2 years period, European Commission guidelines recommend a minimum volume of 5000 mammograms per year for interpreting radiologist [17-19]. Thèberge in 2014 affirms that only if a interpretive volume is consistently less than this requirement, radiologist accuracy may be compromised and that accuracy improves with increases in volume of up to approximately 3000 breast exams interpreted annually [20,21].

Some studies did not find evidence that great volume or experience at interpreting breast exams is associated with better performance [21] probably because these radiologists may not receive feedback on the outcome of all women. Disagreement is probably related also to the absence of learning curves in radiology [22], because it has been demonstrated that radiologists with more years of experience in interpreting mammograms have low false-positive rates, without significatively effects on sensitivity [23, 24]. The irreplaceable breast training demostrates also that automated breast volume scanning cannot be preferred to a manual US examination [25,26]. Residency breast imaging training, dedicated post-residency courses, continuing medical education, annual interpretive volume could improve the radiologist learning curve and accuracy of breast exams with an earlier stage breast cancer detection with low rate of unnecessary biopsies that result in both patient anxiety and increased medical costs. [1,27]. Even if this study was monocentric and analyzed only one techique , the main advantages were a big sample and a prospective study; furthermore, we examined breast US accuracy within individual radiologists, comparing with cytological, histological and/or 2 years follow-up diagnosis; we compared the results obtained by breast and general radiologists in four different subgroups of 100 patients to minimize a possible statistic bias and to obtain statistically significant

results. and we used only two categories for breast lesions, avoiding any specific breast classification, such as BI-RADS classification.

In conclusion, a multicentric prospective study in future could confirm the more accuracy of breast radiologist in comparison to general radiologist in the evaluation of breast pathology in the US diagnoses.

References

1. Di Grezia G, Somma F, Serra N, et al., Reducing Costs of Breast Examination: Ultrasound Performance and Inter-Observer Variability of Expert Radiologists Versus Residents. Cancer Invest 2016;20:1-6.
2. Stavros AT, Thickman D, Rapp CL, et al., Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. Radiology 1995;196(1):123–134
3. Gatta G, Iaselli F, Parlato V, et al., Differential diagnosis between fibroadenoma, giant fibroadenoma and phyllodes tumour: sonographic features and core needle biopsy Radiologia Medica 2011;116:905-918
4. Cho SH, Park SH, Mimickers of breast malignancy on breast sonography J Ultrasound Med 2013;32(11):2029-36
5. ELverici E, Zengin B, Barca AN et al Interobserver and Intraobserver Agreement of Sonographic BIRADS Lexicon in the Assessment of Breast Masses Iran J Radiol, 2013; 10(3):122–127
6. Gatta G, Di Grezia G, Ancona A, et al., Underestimation of atypical lobular hyperplasia and lobular carcinoma in situ at stereotaxic 11-gauge vacuum-assisted breast biopsy Euro J Inflam 2013;11(3):825-835
7. R. Plantade, Interventional radiology: The corner-stone of breast management Diagnostic and Interventional Imaging 2013; 94 (6); 575–591
8. Leconte I, Fellah L US and dense breasts: where do we stand? J Radiol 2008;89:1169-79
9. Elmore JG et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy Radiology. 2009;253(3):641-51
10. Hendrick RE Quality assurance in mammography. Accreditation, legislation, and compliance with quality assurance standards Radiol Clin North Am 1992;30:243-55
11. Sokal R.R. Rohlf F.J. Biometry. The principles and practice of statistics in biological research, Third edition, Freeman, New York, 1995;724-74.
12. McNemar, Q. Note of the sampling error of the difference between correlated proportions or percentages. Psychometrika; 1947(12)153-157
15. Newcombe R.G. Improved Confidence Intervals for the difference between binomial proportions based on paired data. Statist. Med 1998; 17: 2635-2650
16. Griner PF, Mayewski RJ, Mushlin AI, Greenland P Selection and interpretation of diagnostic tests and procedures. Annals Int Med 1981(94) 555-600
17. Haim, Is volume related to outcome in healthcare? A systematc review and methodologic critiuqe of the literature. Ann Inter Med 2002 17;137 (6):511-20
18. Kan, Olivotto – Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program, Radiology 2000; 215(2):563-7
19. CI Lee JG Elmonre Increasing value by increasing volume: call for changes in US breast screening practices J Natl Cancer Inst 2014 doi:10.1093/jnci/dju028
20. Smith Bindman Comparison of screening mammography in the United States and the united Kingdom Jama 2003 22;290(16):2129-37
21. European breast cancer network – European guidelines for quality assurance in breast cancer screening and diagnosis. Dec 2, 2013
22. Thèberge Radiologist interpretive volume and breast cancer screening accuracy in a canadian organized screening program JNCI 2014; 106 (3); djt461
23. Barlow WE, Accuracy of screening mammography interpretation by characteristics of radiologists JNCI 2004;15;96(24):1840-50
24. Taylor PM. A review of research into the development of radiologic expertise: implications for computer-based training. Acad Radiol 2007;14:1252–1263
25. Sickles EA, Wolverton DE, Dee KE, Performance parameters for screening and diagnostic mammography: specialist and general radiologists. Radiology 2002;224:861–869
26. Gatta G, Di Grezia G, Lieto R, Ancona A, Laieta MT, Brunese L, Grassi R, Rotondo A No evidence for an association between trauma and breast carcinoma: A retrospective cohort study Global Adv J Med Medical Sci 2012;7:172-180
27. Ferraro GA, Romano T, De Francesco F, et al., Management of prepubertal gynecomastia in two monozygotic twins with Peutz-Jeghers syndrome: from aromatase inhibitors to subcutaneous mastectomy. Aesth Plast Surg Aesth Plast Surg 2013;37(5):1012-22
28. Di Grezia G, Romano T, De Francesco F, et al., Breast ultrasound in the management of gynecomastia in Peutz-Jeghers Syndrome:a case report. J Med Case Report 2014;(18) 8:440
29. Di Grezia G, Prisco V, Iannaccone T, et al., Personality disorder and temperamental traits in patients with breast disease: preliminary results Minerva Psichiatrica 2016 Sept; 57(3):85-92

Table 1. *Subgroups description (*in parenthesis the patients number with multifocal cancer)

Parameters	Subgroup 1 %	Subgroup 2 %	Subgroup 3 %	Subgroup 4 %
Ages intervals				
≤40	40.00	36.00	44.00	44.00
40-70	56.00	52.00	40.00	44.00
≥70	4.00	12.00	16.00	12.00
Nr. Lesion				
1	40.00	46.00	44.00	48.00
2	12.00	8.00	8.00	8.00
3	0.00	2.00	0.00	0.00
Lesion size				
≥ 5 *mm*	32.00	32.00	36.00	40.00
< 5 *mm*	20.00	24.00	16.00	16.00
Breast pattern				
FG	48.00	40.00	52.00	48.00
NO FG	52.00	60.00	48.00	52.00
Localization				
UOQ	5.00 (1.00)	16.00	19.00	11.00 (1.00)
LIQ	7.00 (1.00)	4.00 (2.00)	1.00 (1.00)	7.00 (1.00)
LOQ	8.00	8.00 (1.00)	9.00	10.00 (1.00)
UIQ	10.00	12.00	10.00	8.00
RA	10.00 (1.00)	6.00 (1.00)	5.00 (2.00)	12.00
LOQ- LIQ	0.00	2.00	0.00	1.00
LOQ-UOQ	2.00	1.00	0.00	1.00
LOQ-UIQ	0.00	0.00	2.00	0.00
LOQ-RA	2.00	1.00	0.00	0.00
LIQ-UIQ	2.00	0.00	0.00	1.00
LIQ-UOQ	1.00	0.00	0.00	1.00
LIQ-RA	0.00	1.00	2.00	0.00
UOQ-UIQ	1.00	0.00	1.00	0.00
UOQ-RA	1.00	0.00	0.00	0.00
UIQ-RA	0.00	1.00	0.00	1.00
US Response				
Cancer	4.00	12.00	8.00	8.00
Fibroadenoma	12.00	16.00	12.00	16.00
Cyst	36.00	28.00	32.00	32.00
Negative	48.00	44.00	48.00	44.00

Table 2. US response for each subgroup of Breast Radiologist and General Radiologist, and effective diagnosis.

US response	% Total error	% Negative	% Cyst	% Fibroadenoma	% Cancer
Breast Radiologist 1	0.00	48.00 (0.00)	36.00 (0.00)	12.00 (0.00)	4.00 (0.00)
Breast Radiologist 2	1.00	45.00 (+1.00)	27.00 (-1.00)	16.00 (0.00)	12.00 (0.00)
Breast Radiologist 3	1.00	49.00 (+1.00)	32.00 (0.00)	11.00 (-1.00)	8.00 (0.00)
Breast Radiologist 4	0.00	44.00 (0.00)	32.00 (0.00)	16.00 (0.00)	8.00 (0.00)
General Radiologist 1	1.00	49.00 (+1.00)	36.00 (0.00)	11.00 (-1.00)	4.00 (0.00)
General Radiologist 2	3.00	45.00 (+1.00)	30.00 (+2.00)	14.00 (-2.00)	11.00 (-1.00)
General Radiologist 3	3.00	51.00 (+3.00)	31.00 (-1.00)	10.00 (-2.00)	8.00 (0.00)
General Radiologist 4	1.00	45.00 (+1.00)	31.00 (-1.00)	16.00 (0.00)	8.00 (0.00)
Effective diagnosis 1		48.00	36.00	12.00	4.00
Effective diagnosis 2		44.00	28.00	16.00	12.00
Effective diagnosis 3		48.00	32.00	12.00	8.00
Effective diagnosis 4		44.00	32.00	16.00	8.00

Table 3. US response for breast radiologists, general radiologists and effective diagnosis for total group

US response	% Total error	NOT SUSPICIOUS			SUSPICIOUS
		% Negative	% Cyst	% Fibroadenoma	% Cancer
Breast Radiologists	0.50	46.50 (0.50)	31.75 (0.25)	13.75 (0.25)	8.00 (0.00)
General Radiologists	2.00	47.50 (1.50)	32.00 (1.00)	12.75 (1.25)	7.75 (0.25)
Effective diagnosis		46.00	32.00	14.00	8.00

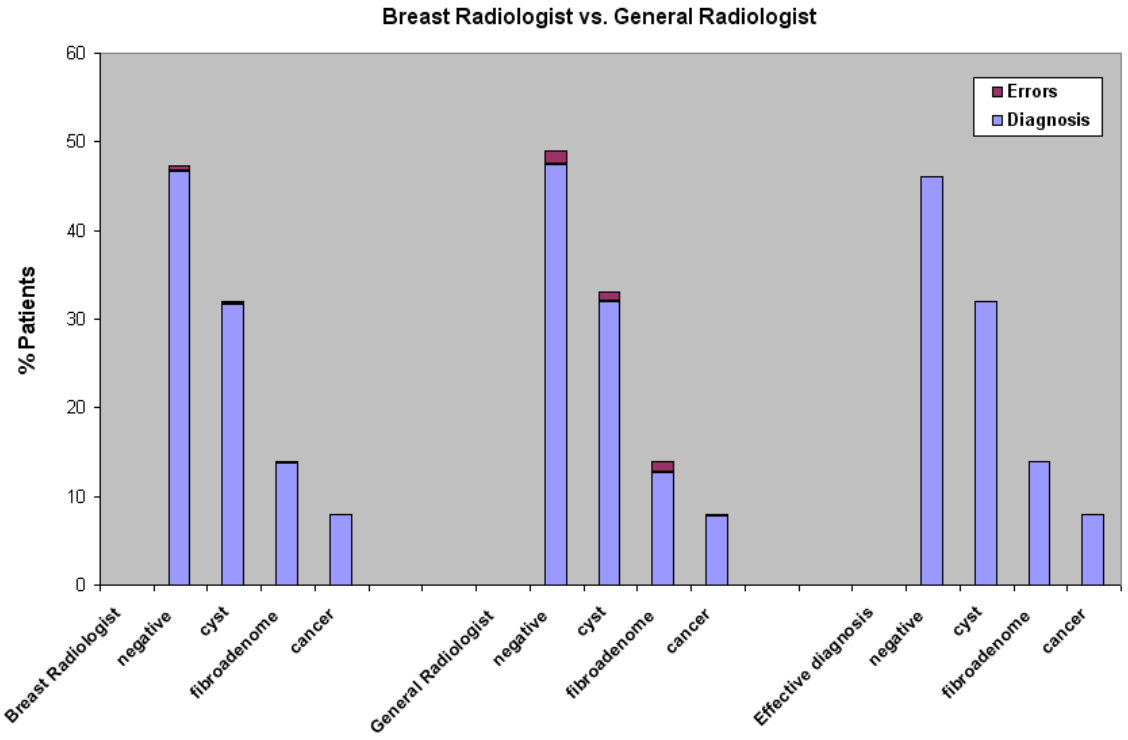

Figure 1. US response for breast radiologists, general radiologists and effective diagnosis, for total group

Table 3. One tail McNemar's exact test in the evaluation the accuracy between breast and general radiologists.

Test : McNemar's exact test	**significant test**
H_0 : general radiologist $\geq$ breast radiologist	False
H_1 : general radiologist < breast radiologist	True
p-value one-tail (< α)	0.0625
α (significance level)	0.1
Proportion of correct diagnosis	
Breast Radiologist	53.5 %
General Radiologist	52.5 %

Table 4. Sensitivity and Specificity with confidence intervals, for breast and general radiologists in the breast pathologies diagnoses with US ultrasound .

Breast pathologies diagnoses	**Sensitivity *(Se)***	**Confidence interval at 95% for *Se***	**Specificity *(Sp)***	**Confidence interval at 95% for *Sp***
Breast Radiologists	0. 990	(0.973, 0.998)	0.989	(0.971, 0.997)
General radiologists	0. 972	(0.949, 0.986)	1.000	(0.988, 1.000)