Title

Diagnostic Reproducibility: What Happens When the Same Pathologist Interprets the Same Breast Biopsy Specimen at Two Points in Time?

Journal

Annals of Surgical Oncology, 24(5)

ISSN

1068-9265

Authors

Jackson, Sara L
Frederick, Paul D
Pepe, Margaret S
et al.

Publication Date

2017-05-01

DOI

10.1245/s10434-016-5695-0

HHS Public Access

Author manuscript

Ann Surg Oncol. Author manuscript; available in PMC 2017 August 01.

Published in final edited form as:

Ann Surg Oncol. 2017 May ; 24(5): 1234–1241. doi:10.1245/s10434-016-5695-0.

Diagnostic Reproducibility: What Happens When the Same Pathologist Interprets the Same Breast Biopsy Specimen at Two Points in Time?

Sara L. Jackson, MD, MPH[1], **Paul D. Frederick, MPH, MBA**[1], **Margaret S. Pepe, PhD**[5], **Heidi D. Nelson, MD, MPH**[2], **Donald L. Weaver, MD**[9], **Kimberly H. Allison, MD**[4], **Patricia A. Carney, PhD**[3], **Berta M. Geller, EdD**[6], **Anna N. A. Tosteson, ScD**[7,8], **Tracy Onega, PhD**[7], and **Joann G. Elmore, MD, MPH**[1]

[1]Department of Medicine, University of Washington School of Medicine

[2]Providence Cancer Center, Providence Health and Services Oregon, and Departments of Medical Informatics and Clinical Epidemiology and Medicine, Oregon Health & Science University

[3]Department of Family Medicine, Oregon Health & Science University

[4]Department of Pathology, Stanford University School of Medicine

[5]Program in Biostatistics and Biomathematics, Fred Hutchinson Cancer Research Center

[6]Department of Family Medicine, University of Vermont

[7]Department of Community and Family Medicine, The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Norris Cotton Cancer Center

[8]Department of Medicine, Geisel School of Medicine at Dartmouth

[9]Department of Pathology and University of Vermont Cancer Center, University of Vermont, Burlington, VT

Abstract

Background—Surgeons may receive a different diagnosis when a breast biopsy is interpreted by a second pathologist. The extent to which diagnostic agreement by the same pathologist varies at two time points is unknown.

Participants and Methods—Pathologists from 8 U.S. states independently interpreted 60 breast specimens, one glass slide per case, on 2 occasions separated by ≥9 months. Reproducibility was assessed by comparing interpretations between the two time points; associations between reproducibility (intra-observer agreement rates) and characteristics of pathologists and cases were determined and also compared with inter-observer agreement of baseline interpretations.

Results—Sixty-five percent of invited, responding pathologists were eligible and consented; 49 interpreted glass slides in both study phases resulting in 2,940 interpretations. Intra-observer agreement rates between the two phases were 92% (95% CI 88%-95%) for invasive breast cancer,

Correspondence to: Sara L. Jackson, MD, MPH, University of Washington, Mailbox 359892, 325 Ninth Avenue, Seattle, WA 98104, Telephone: (206) 744-8946, Fax: (206) 744-8516, sljack@uw.edu.

84% (95% CI 81%-87%) for ductal carcinoma *in situ* (DCIS), 53% (95% CI 47%-59%) for atypia, and 84% (95% CI 81%-86%) for benign without atypia. When comparing all study participants' case interpretations at baseline, inter-observer agreement rates were 89% (95% CI 84%-92%) for invasive cancer, 79% (95% CI 76%-81%) for DCIS, 43% (95% CI 41%-45%) for atypia, and 77% (95% CI 74%-79%) for benign without atypia.

Conclusions—Interpretive agreement between two time points by the same individual pathologists was low for atypia, and similar to observed rates of agreement for atypia between different pathologists. Physicians and patients should be aware of the diagnostic challenges associated with a breast biopsy diagnosis of atypia when considering treatment and surveillance decisions.

Keywords

Breast Pathology Study (B-Path); breast biopsy; breast pathology; breast diseases; breast neoplasms; breast atypia; breast density; ductal carcinoma *in situ* (DCIS); intra-rater agreement

Introduction

Mammography screening has increased the identification of non-invasive lesions such as atypia (including atypical ductal hyperplasia, ADH) and ductal carcinoma *in situ* (DCIS).[1-3] These lesions are associated with increased risk for breast cancer and thus generate anxiety, additional testing, surveillance and treatment. Practice guidelines for women with atypia and DCIS include enhanced annual screening with magnetic resonance imaging (MRI) and pharmacologic risk reduction with selective estrogen-receptor modulators (SERMs) or aromatase inhibitors (AIs).[4] Some women go so far as to request prophylactic bilateral mastectomies.[5,6]

Surgeons need to rely on the pathologic interpretation, the gold standard for breast tissue diagnosis; however, disagreement among pathologists on non-invasive lesions, such as atypia and some forms of DCIS, has been reported.[7-10,11] Concerns about challenges interpreting these biopsy specimens lead many to obtain second opinions before initiating treatment.[12-14]

While established diagnostic criteria exist to guide pathologists in breast tissue interpretation,[15,16] the extent of disagreement among pathologists on diagnoses of atypia led us to question the reproducibility of the diagnoses—i.e., would pathologists diagnose atypia on a case they had previously interpreted as such? Is the underlying cause for variability the pathologist or the case? Few studies assess intra-observer agreement for breast diagnoses such as atypia,[17] thus we studied agreement rates for individual pathologists who interpreted the same cases at different times, hypothesizing greater consistency with their own diagnosis than with interpretations by other breast pathologists. We examined results from 49 pathologists participating in the Breast Pathology Study (B-Path) who interpreted one slide per test case at two points in time separated by at least 9 months (intra-observer agreement). We then compared the levels of intra-observer agreement with inter-observer agreement.

Methods

Study participants

The B-Path study recruited pathologists from eight U.S. states: Alaska, Maine, Minnesota, New Hampshire, New Mexico, Oregon, Vermont, and Washington. Pathologists who interpreted breast specimens within the prior year and planned to continue in the following year were eligible unless they were in training. Other aspects of identification and recruitment have been previously reported.[18] Demographic data, practice characteristics, and interpretive experience of the pathologists were queried using a web-based survey.[13,18]

Test set cases and consensus reference diagnoses

Using a random stratified sampling method, core needle or excisional breast biopsies from the New Hampshire and Vermont breast pathology registries from the National Cancer Institute-sponsored Breast Cancer Surveillance Consortium were selected for the test set of 240 cases as previously described.[19] Cases were stratified to reflect an even distribution of ages (49% aged 40-49 years; 51% aged ≥50 years) and breast density (51% heterogeneously or extremely dense on mammography). Cases of atypia and DCIS were oversampled; among the 240 test cases, 30% were benign without atypia, 30% were atypia, 30% were DCIS, and 10% were invasive carcinoma. Three experienced and internationally recognized breast pathologists interpreted all test cases and assigned a difficulty level for each case.[11] The 240 cases were randomly assigned to 1 of 4 different test sets (60 cases each) that were stratified by the woman's age, breast density, the expert panel consensus reference diagnosis, and the experts' difficulty rating.[11]

Study procedure

In Phase I, participants independently interpreted 60 cases based on one glass slide per case. In Phase II, the same participants reinterpreted the same 60 cases at least 9 months following Phase I. The glass slides in Phase II were randomly reordered and the participants were not told they were reviewing the same cases. After Phase II, pathologists were queried regarding whether they thought any of the cases in the second set (Phase II) were the same as those in the first set (Phase I). Because pathologists were randomly assigned to 1 of 4 test sets of 60 cases each, all 240 test cases contributed interpretive data to the study.

Diagnostic assessments were recorded using an online assessment tool developed for the study, the Breast Pathology Assessment Tool and Hierarchy for Diagnosis (BPATH-Dx).[11,20] Fourteen distinct diagnostic assessments were categorized into four main BPATH-Dx categories: 1) benign without atypia (including non-proliferative and proliferative without atypia); 2) atypia (ADH and intraductal papilloma with atypia); 3) DCIS; and 4) invasive breast carcinoma. For each case, participants could indicate whether the case was borderline; the most severe diagnosis was the one assigned to the case as their primary diagnosis, followed by the secondary diagnosis, until all were listed.

Human research protections

The Institutional Review Boards of Dartmouth College, the Fred Hutchinson Cancer Research Center, Providence Health & Services of Oregon, the University of Vermont, and

the University of Washington approved all study procedures. All participating pathologists signed an informed consent.

Statistical analyses

We compared Phase I versus Phase II categorical diagnoses to determine the proportion of Phase II interpretations that agreed with Phase I. We then repeated the comparison, accounting for cases considered “borderline” between two diagnoses on the second interpretation. If a borderline diagnosis in Phase II included the diagnosis recorded for Phase I, the participant was given credit for interpretive agreement. Next, we compared individual pathologist interpretations in Phase I to interpretations of the same slide by any other pathologist in Phase II, resulting in 33,120 paired comparisons (552 paired pathologists × 60 cases = 33,120 assessments).

Lastly, we assessed participant, case, and Phase I interpretative assessment characteristics associated with diagnostic intra-observer agreement in both study phases. All reported case characteristics were assessed at the time of the Phase I interpretation. Separate logistic regression analyses tested associations between interpretive agreement (yes versus no) of pathologist, case, and Phase I interpretative assessment characteristics. We used generalized estimating equations (GEE) methodology for model fitting, hypothesis testing, and confidence interval construction. This logistic regression methodology was used without covariates (Table 1, Figure 1) to derive confidence intervals for interpretative agreement in Phase II restricted to cases with specific diagnostic interpretations in Phase I. Finally, we examined associations between interpretive agreement and pathologist and case characteristics restricting to cases interpreted as atypia in Phase I. P-values were two-sided. All statistical analyses were performed using SAS 9.4 for Windows.

Role of the funding source

The funding organization had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; or decision to submit the manuscript for publication.

Results

Intra-observer and inter-observer diagnostic agreement rates

Forty-nine pathologists provided a total of 2,940 interpretations for each phase. Agreement rates were 92% (95% CI 88%-95%) for invasive cancer; 84% (95% CI 81%-87%) for DCIS; 53% (95% CI 47%-59%) for atypia; and 84% (95% CI 81%-86%) for benign without atypia (Table 1a). When pathologist interpretations were compared to all other interpretations of the same slide by peer study participants (33,120 paired comparisons), agreement rates were 89% (95% CI 84%-92%) for invasive cancer; 79% (95% CI 76%-81%) for DCIS; 43% (95% CI 41%-45%) for atypia; and 77% (95% CI 74%-79%) for benign without atypia (Table 1b).

Agreement rates were higher when either the primary or borderline diagnosis from Phase II agreed with the Phase I diagnosis. Rates were 97% (95% CI 94%-98%) for invasive

carcinoma; 91% (95% CI 88%-93%) for DCIS; 58% (95% CI 53%-64%) for atypia; and 84% (95% CI 81%-86%) for benign without atypia (Figure 1).

Association of participant and case characteristics with reproducibility

No statistically significantly associations were noted between pathologists' reproducibility of their diagnoses and any of the measured pathologist characteristics. When analysis was limited to cases interpreted as atypia in Phase I, pathologists who reported that their colleagues considered them experts in breast pathology had higher reproducibility (intra-observer agreement rates) than non-experts (65% compared to 50% agreement, p=0.01) (Table 2).

Lower breast density was associated with slightly higher reproducibility (intra-observer agreement rates) when all cases were considered (p=0.007), but not for cases interpreted as atypia in Phase I (Table 3). Agreement was also higher when pathologists assigned fewer diagnoses per case in Phase I (p<0.001), but this association was not demonstrated for atypia cases. For atypia, agreement was higher with tissue obtained by core needle biopsies biopsy compared to excisional biopsy (p=0.037) (Table 3).

Overall, pathologists had greater reproducibility (intra-observer agreement) if they reported higher levels of confidence or less difficulty with the case in Phase I, if the case was not borderline between two diagnoses, or if they did not want a second opinion. This was also observed when restricted to interpretations of atypia in Phase I (Table 4).

Discussion

When pathologists interpreted the same slide from a set of breast biopsy test cases at two points in time, their interpretative agreement varied according to diagnostic category. While reproducibility (intra-observer agreement) was high for invasive breast carcinoma cases, it was lower for DCIS, and for atypia it was just 53%. Pathologists' *intra*-observer agreement was higher than their *inter*-observer agreement with other study pathologists, and for atypia it was 43%. While pathologists are more likely to agree with their own previous diagnoses than with diagnoses by other pathologists, we note concerning findings with regard to the middle diagnostic categories. No pathologist characteristics, such as training or experience, were associated with improved reproducibility. As one would expect, cases that pathologists rated as difficult or borderline between two diagnoses, or where a second opinion was desired, had lower reproducibility. This suggests that pathologists are aware of cases with potentially low diagnostic agreement. Clinical decisions based upon pathologic diagnoses of atypia should be interpreted in light of these results. Breast atypia is not reproducibly identified, even by the same pathologist, calling into question whether clinicians and women should make clinical decisions based upon the pathology report without additional supporting opinions, ancillary diagnostic markers, and taking the full clinical presentation into consideration. It is also possible that similar “indolent lesions of epithelial origin”[21] in other organ systems may lack diagnostic reproducibility, and further study is needed.

Pathologists in this study had consistently low agreement for atypia diagnoses, whether compared to their own prior diagnosis, their peer study participants' diagnoses, or to the

consensus diagnoses of an expert panel of three breast pathologists. A prior analysis compared breast diagnoses of a larger cohort of pathologists interpreting these test cases to diagnoses of an expert reference panel consensus.[11] Compared to experts, diagnostic agreement was 96% for invasive carcinoma, 84% for DCIS, 48% for atypia, and 87% for benign without atypia.[11] Pathologists with higher weekly case volumes or who work in larger or academic practice settings had higher agreement rates with an expert panel;[11] however, these factors were not associated with *intra*-observer diagnostic consistency in the current study, which could be due to the smaller sample size. Thus, the consistently low reproducibility for atypia does not appear to be related to pathologists' training and diagnostic acumen, but is likely due to inherent characteristics of the tissue specimen and an inability to classify these lesions adequately. This may be due to inherent image complexity of microscopic epithelial characteristics of the individual case, or the diagnostic criteria may be more susceptible to subjective interpretation. In addition, the atypia category may encompass greater intrinsic biologic variability, relative to other diagnostic categories, making differences in agreement less likely to be attributable to the interpreter.[20]

The low reproducibility for atypia is particularly problematic because a diagnosis of atypia implies an increased future risk for invasive cancer, can lead to more intensive surveillance and treatment, and can lead to an excisional biopsy if the diagnosis is made on a core biopsy. Wide diagnostic variation for atypia between pathologists has been previously documented.[7-10] One study of atypia diagnoses by nine pathologists found that intra-observer kappa values were higher (0.56 to 0.80) than the inter-observer kappa (0.34). The addition of immunohistochemical stains improves the agreement rate and decreases atypia diagnoses in favor of usual hyperplasia, which would decrease surgical intervention for these lesions.[17,22] Our study presents intra-observer data on a much larger sample of pathologists who work in multiple geographic areas of the U.S., but our methods did not incorporate the option of additional diagnostic test results such as immunohistochemical stains, which might improve observed agreement for atypia.

The statistically significant relationship between intra-observer agreement and fewer diagnoses for a case probably reflects epithelial complexity or overlapping diagnostic features (diagnostic distraction). Similarly, the association of higher breast density with lower reproducibility suggests that inherent characteristics of the breast tissue increase the diagnostic challenge. Our previous studies found that accuracy was slightly higher when pathologists used glass slides, as the current study did, compared to digital whole slide imaging (WSI), an emerging technology for pathology interpretation.[18,23] Although currently understudied, intra-observer variability also has the potential to be greater using WSI.

Strengths of the study include the enrollment of a large number of pathologists from multiple geographic regions in the U.S. who interpreted 60 cases at least 9 months apart. The increased proportion of DCIS and atypia cases allows power for statistical comparisons. When compared to the entire spectrum of breast pathology seen in their own practices, 74% (n=70) of B-Path participants who completed the CME activity (n=94) reported that they either often or always see cases like these, 22% (n=21) reported sometimes seeing cases like these, and 3% (n=3) did not respond to the question. Because the proportion of atypia and

DCIS cases in this study was higher than in typical clinical practice and second consultative opinions were not allowed, agreement rates are lower than would be expected for clinical settings, where the prevalence of these challenging diagnoses is lower and additional evaluation is common. Further, statistically significant associations between pathologist characteristics, case characteristics, and interpretive agreement could be a consequence of multiple statistical comparisons. Lastly, because of testing conditions, pathologists only interpreted one slide per case, without the benefit of additional clinical information (except for age and biopsy type) or supplemental immunohistochemical test results, which differs from clinical practice.

In conclusion, an individual pathologist's agreement with his/her own interpretations of breast biopsies at a second point in time varies; the lowest observed agreement rates were for atypia and the highest were for invasive carcinoma. Pathologists recognize when cases might have low reproducibility. Given the clinical implications of identifying lesions associated with increased risk for breast cancer, such as atypia, the use of second opinion strategies or adjunctive tests that may discriminate between categories should be evaluated as mechanisms to improve reproducibility and reduce overtreatment. Finally, physicians and patients should be aware of the uncertainty of the pathologic diagnosis of atypia when making clinical decisions.

Acknowledgments

The authors appreciate the efforts of the pathologists who participated in this study.

Grant Support: This work was supported by the National Cancer Institute of the National Institutes of Health under award numbers* R01 CA140560, U54 CA163303, KO5 CA104699, and R01 CA172343 and by the National Cancer Institute-funded Breast Cancer Surveillance Consortium award number HHSN261201100031C. The content is solely the responsibility of the authors and does not necessarily represent the views of the National Cancer Institute or the National Institutes of Health. The collection of cancer and vital status data used in this study was supported in part by several state public health departments and cancer registries throughout the U.S. For a full description of these sources, please see: http://www.breastscreening.cancer.gov/work/acknowledgement.html.

References

1. Welch HG, Black WC. Overdiagnosis in cancer. J Natl Cancer Inst. 2010; 102(9):605–613. [PubMed: 20413742]
2. U.S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2009; 151(10):716–726. [PubMed: 19920272]
3. Page DL, Schuyler PA, Dupont WD, Jensen RA, Plummer WD Jr, Simpson JF. Atypical lobular hyperplasia as a unilateral predictor of breast cancer risk: a retrospective cohort study. Lancet. 2003; 361(9352):125–129. [PubMed: 12531579]
4. Hartmann LC, Degnim AC, Santen RJ, Dupont WD, Ghosh K. Atypical Hyperplasia of the Breast — Risk Assessment and Management Options. N Engl J Med. 2015; 372(1):78–89. [PubMed: 25551530]
5. Tuttle TM, Jarosek S, Habermann EB, et al. Increasing rates of contralateral prophylactic mastectomy among patients with ductal carcinoma in situ. J Clin Oncol. 2009; 27(9):1362–1367. [PubMed: 19224844]
6. Arrington AK, Jarosek SL, Virnig BA, Habermann EB, Tuttle TM. Patient and Surgeon Characteristics Associated with Increased Use of Contralateral Prophylactic Mastectomy in Patients with Breast Cancer. Ann Surg Oncol. 2009; 16(10):2697–2704. [PubMed: 19653045]
7. Rosai J. Borderline epithelial lesions of the breast. Am J Surg Pathol. 1991; 15(3):209–221. [PubMed: 1847606]

8. 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(12): 1133–1143. [PubMed: 1463092]

9. Wells WA, Carney PA, Eliassen MS, Tosteson AN, Greenberg ER. Statewide study of diagnostic agreement in breast pathology. J Natl Cancer Inst. 1998; 90(2):142–145. [PubMed: 9450574]

10. Della Mea V, Puglisi F, Bonzanini M, et al. Fine-needle aspiration cytology of the breast: a preliminary report on telepathology through internet multimedia electronic mail. Mod Pathol. 1997; 10(6):636–641. [PubMed: 9195583]

11. Elmore JG, Longton G, Carney PA, et al. Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens. JAMA. 2015; 313(11):1122–1132. [PubMed: 25781441]

12. Elmore JG, Tosteson AN, Pepe MS, et al. Evaluation of 12 strategies for obtaining second opinions to improve interpretation of breast histopathology: simulation study. BMJ. 2016; 353:i3069. [PubMed: 27334105]

13. Geller BM, Nelson HD, Carney PA, et al. Second opinion in breast pathology: policy, practice and perception. J Clin Pathol. 2014; 67(11):955–960. [PubMed: 25053542]

14. Elmore JG, Harris RP. The harms and benefits of modern screening mammography. BMJ. 2014; 348:g3824. [PubMed: 24938686]

15. O'Malley, FP., Pinder, SE., Mulligan, AM. Breast pathology. Philadelphia, PA: Elsevier/Saunders; 2011.

16. Schnitt, SJ., Collins, LC. Biopsy interpretation of the breast. Philadelphia: Wolters Kluwer Health/ Lippincott Williams & Wilkins; 2009.

17. Jain RK, Mehta R, Dimitrov R, et al. Atypical ductal hyperplasia: interobserver and intraobserver variability. Mod Pathol. 2011; 24(7):917–923. [PubMed: 21532546]

18. Onega T, Weaver D, Geller B, et al. Digitized whole slides for breast pathology interpretation: current practices and perceptions. J Digit Imaging. 2014; 27(5):642–648. [PubMed: 24682769]

19. National Cancer Institute. Breast Cancer Surveillance Consortium. [Accessed June 8, 2015] http://breastscreening.cancer.gov/

20. Allison KH, Reisch LM, Carney PA, et al. Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel. Histopathology. 2014; 65(2): 240–251. [PubMed: 24511905]

21. Esserman LJ, Thompson IM, Reid B, et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. Lancet Oncol. 2014; 15(6):e234–242. [PubMed: 24807866]

22. Reyes C, Ikpatt OF, Nadji M, Cote RJ. Intra-observer reproducibility of whole slide imaging for the primary diagnosis of breast needle biopsies. J Pathol Inform. 2014; 5:5. [PubMed: 24741464]

23. Elmore JG, L G, Pepe, et al. A Randomized Study Comparing Reproducibility and Accuracy of Digital Imaging to Traditional Glass Slide Microscopy for Breast Biopsy Interpretation. Under review at Journal of the National Cancer Institute.

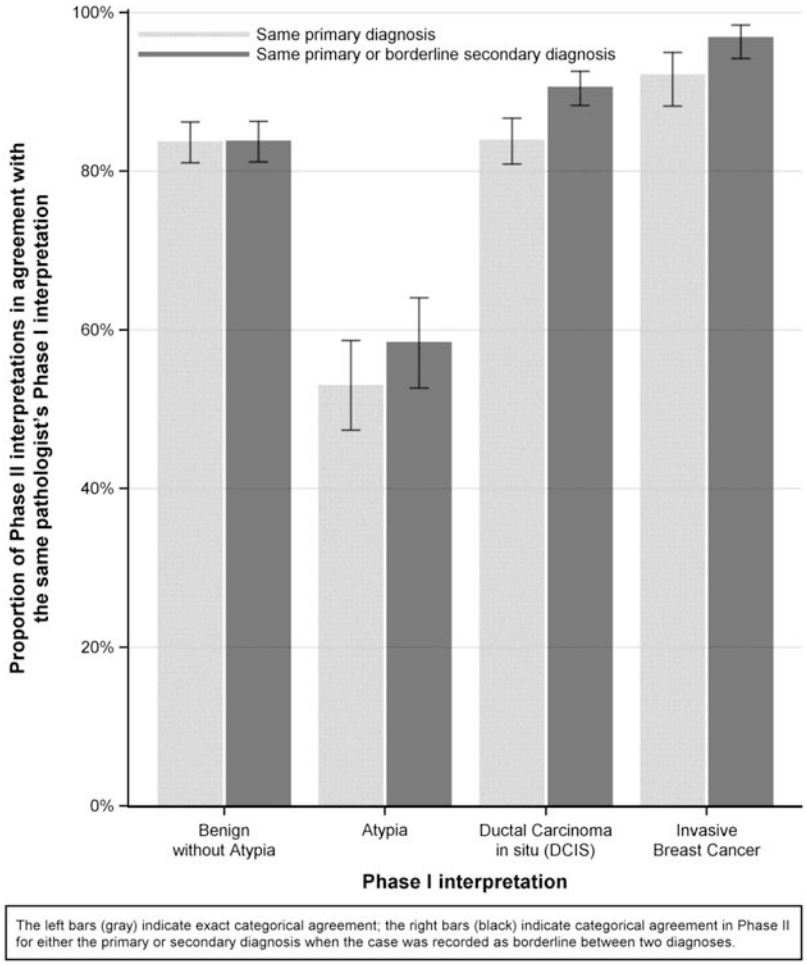

Figure 1. Proportion of interpretations with the same diagnosis in Phase I and Phase II by diagnostic category in Phase I (n=2940 interpretations)

Ann Surg Oncol. Author manuscript; available in PMC 2017 August 01.

Table 1

a. Interpretations of the same breast specimens by 49 participants (intra-observer) at two time points (phase I and phase II, 2,940 paired comparisons). Numbers indicate the numbers of interpretations with agreement highlighted by shading.*

	Phase II Interpretation of Same Individual Pathologist					
Phase I Interpretation of Individual pathologist	**Benign without atypia**	**Atypia**	**DCIS**	**Invasive**	**Total**	**Agreement rates of phase I and II interpretations, % (95% CIs)**
Benign without atypia	**947**	137	41	5	1130	84 (81-86)
Atypia	157	**303**	109	2	571	53 (47-59)
Ductal Carcinoma *in situ* (DCIS)	43	94	**792**	14	943	84 (81-87)
Invasive Breast Cancer	8	4	11	**273**	296	92 (88-95)
Total	1155	538	953	294	2940	79 (77-81)
*The same slide was interpreted on two different occasions separated in time by 9 or more months						

b. Interpretations of the same breast specimens by 49 participants compared all other study pathologist (inter-observer) interpretations of the same slide in phase II (33,120 paired comparisons). Numbers indicate the numbers of interpretations with agreement highlighted by shading

	Phase II Interpretation All Other Study Pathologists					
Phase I Interpretation of Individual pathologist	**Benign without atypia**	**Atypia**	**DCIS**	**Invasive**	**Total**	**Agreement rates of phase I and II interpretations, % (95% CIs)**
Benign without atypia	**9772**	1994	885	76	12727	77 (74-79)
Atypia	2325	**2776**	1306	26	6433	43 (41-45)
Ductal Carcinoma *in situ* (DCIS)	778	1249	**8358**	250	10635	79 (76-81)
Invasive Breast Cancer	125	53	197	**2950**	3325	89 (84-92)
Total	13000	6072	10746	3302	33120	72 (71-73)

Table 2

Association between interpretive agreement and participant characteristics for all cases, and for Atypia cases (phase I diagnosis) only.

Characteristics [a]	Participants	All Phase I interpretations		Interpreted as Atypia in Phase I [b]	
		Agreement with Phase I Interpretation		Agreement with Phase I Interpretation	
	N (%)	Rate (95% CI)	p-value	Rate (95% CI)	p-value
Total	49 (100.0)	0.79 (0.77 - 0.81)	--	0.53 (0.47 - 0.59)	--
Demographics					
Age at Survey (yrs)					
30-39	7 (14.3)	0.75 (0.69 - 0.80)	0.15	0.43 (0.32 - 0.54)	0.30
40-49	20 (40.8)	0.80 (0.77 - 0.83)		0.57 (0.46 - 0.67)	
50-59	16 (32.7)	0.80 (0.76 - 0.83)		0.52 (0.44 - 0.59)	
60+	6 (12.2)	0.76 (0.72 - 0.80)		0.56 (0.44 - 0.68)	
Gender					
Male	27 (55.1)	0.79 (0.76 - 0.81)	0.80	0.55 (0.48 - 0.61)	0.53
Female	22 (44.9)	0.78 (0.76 - 0.81)		0.51 (0.41 - 0.61)	
Training and Experience					
Laboratory group practice size					
< 10 Pathologists	28 (57.1)	0.78 (0.75 - 0.80)	0.25	0.49 (0.42 - 0.56)	0.13
≥ 10 Pathologists	21 (42.9)	0.80 (0.78 - 0.82)		0.58 (0.49 - 0.67)	
Fellowship training in surgical or breast pathology					
No	27 (55.1)	0.80 (0.78 - 0.81)	0.27	0.57 (0.50 - 0.64)	0.11
Yes	22 (44.9)	0.78 (0.74 - 0.81)		0.48 (0.40 - 0.57)	
Affiliation with academic medical center					
No	34 (69.4)	0.78 (0.76 - 0.80)	0.48	0.54 (0.46 - 0.61)	0.92
Yes, adjunct/affiliated clinical faculty	8 (16.3)	0.81 (0.77 - 0.84)		0.51 (0.41 - 0.61)	
Yes, primary appointment	7 (14.3)	0.80 (0.76 - 0.83)		0.52 (0.38 - 0.65)	
Do your colleagues consider you an expert in breast pathology?					
No	38 (77.6)	0.78 (0.76 - 0.80)	0.12	0.50 (0.44 - 0.56)	0.010
Yes	11 (22.4)	0.81 (0.78 - 0.83)		0.65 (0.55 - 0.74)	

Characteristics [a]	Participants	All Phase I interpretations		Interpreted as Atypia in Phase I [b]	
		Agreement with Phase I Interpretation		Agreement with Phase I Interpretation	
	N (%)	Rate (95% CI)	p-value	Rate (95% CI)	p-value
Breast pathology experience (years)					
< 5	7 (14.3)	0.79 (0.72 - 0.84)	0.62	0.47 (0.33 - 0.61)	0.70
5-9	11 (22.4)	0.77 (0.74 - 0.80)		0.52 (0.40 - 0.63)	
10-19	15 (30.6)	0.79 (0.76 - 0.82)		0.54 (0.42 - 0.64)	
≥ 20	16 (32.7)	0.79 (0.76 - 0.83)		0.56 (0.49 - 0.64)	
Breast specimen case load (%)					
< 10	22 (44.9)	0.78 (0.75 - 0.81)	0.14	0.49 (0.41 - 0.57)	0.18
10-24	22 (44.9)	0.79 (0.76 - 0.81)		0.54 (0.45 - 0.63)	
≥ 25	5 (10.2)	0.84 (0.78 - 0.88)		0.66 (0.49 - 0.79)	
Number of breast cases (per week)					
< 5	13 (26.5)	0.76 (0.72 - 0.81)	0.33	0.48 (0.36 - 0.61)	0.46
5-9	18 (36.7)	0.79 (0.76 - 0.81)		0.52 (0.42 - 0.62)	
≥ 10	18 (36.7)	0.80 (0.77 - 0.83)		0.57 (0.49 - 0.65)	

[a] By self-report on baseline survey, 60 biopsy interpretations per pathologist.

[b] Includes all interpretations considered to be atypia by pathologists in Phase I.

Table 3

Association between interpretive agreement and patient and case characteristics for all cases, and for Atypia cases (phase I diagnosis) only.

Case Characteristics	All Phase I interpretations			Interpreted as Atypia in Phase I [a]		
	All Cases	Agreement with Phase I Interpretation		Atypia Cases [a]	Agreement with Phase I Interpretation	
	N (%)	Rate (95% CI)	p-value	N (%)	Rate (95% CI)	p-value
Total	240 (100.0)	0.79 (0.77 - 0.81)	--	50 (100)	0.53 (0.47 - 0.59)	--
Patient Age (yrs) [b]						
40-49	118 (49.2)	0.79 (0.76 - 0.81)	0.34	25 (50.0)	0.52 (0.45 - 0.59)	0.31
50-59	67 (27.9)	0.78 (0.74 - 0.81)		15 (30.0)	0.50 (0.42 - 0.58)	
60-69	29 (12.1)	0.82 (0.78 - 0.86)		5 (10.0)	0.62 (0.51 - 0.72)	
70+	26 (10.8)	0.78 (0.73 - 0.82)		5 (10.0)	0.57 (0.45 - 0.67)	
Breast Density[c]						
Low Density	118 (49.2)	0.81 (0.78 - 0.83)	0.007	21 (42.0)	0.51 (0.45 - 0.58)	0.36
High Density	122 (50.8)	0.77 (0.75 - 0.79)		29 (58.0)	0.55 (0.48 - 0.62)	
Biopsy Type						
Core needle biopsy	138 (57.5)	0.78 (0.76 - 0.81)	0.70	31 (62.0)	0.56 (0.50 - 0.62)	0.037
Excisional biopsy	102 (42.5)	0.79 (0.77 - 0.81)		19 (38.0)	0.48 (0.41 - 0.56)	
Average Number of Diagnoses per Case[d]						
1	14 (5.8)	0.98 (0.95 - 0.99)	<0.001	0 (0.0)	(-)	0.49
>1 and ≤2	194 (80.8)	0.79 (0.77 - 0.81)		39 (78.0)	0.53 (0.47 - 0.59)	
>2 and ≤3	30 (12.5)	0.73 (0.68 - 0.77)		10 (20.0)	0.56 (0.45 - 0.65)	
>3	2 (0.8)	0.42 (0.24 - 0.61)		1 (2.0)	0.33 (0.10 - 0.68)	

[a] Includes all interpretations considered to be Atypia by pathologists in Phase I.

[b] The patient's age and type of biopsy information were provided to participants.

[c] Low density (almost entirely fat or scattered fibroglandular densities) and high density (heterogeneously or extremely dense). Breast density information was not provided to participants.

[d] Represents the number of diagnostic subtypes checked by pathologists in phase I per each interpretation, averaged at the level of the case.

Table 4

Association between intra-observer agreement and participants' perceptions of each case assessment for all diagnoses, and for Atypia cases (phase I diagnosis) only.

Phase I Assessment Characteristics	All Phase I interpretations			Interpreted as Atypia in Phase I [a]		
	All Interpretations	Agreement with Phase I Interpretation		Atypia Interpretations [a]	Agreement with Phase I Interpretation	
	N (%)	Rate (95% CI)	p-value	N (%)	Rate (95% CI)	p-value
Total	2940 (100.0)	0.79 (0.77 - 0.81)	--	571 (100)	0.53 (0.47 - 0.59)	
Confidence in Assessment [b]						
High Confidence	2503 (85.1)	0.82 (0.80 - 0.84)	<0.001	382 (66.9)	0.57 (0.51 - 0.63)	0.03
Low Confidence	437 (14.9)	0.57 (0.50 - 0.65)		189 (33.1)	0.45 (0.35 - 0.55)	
Level of Diagnostic Difficulty of the Case [c]						
Very easy/Easy	2136 (72.7)	0.86 (0.84 - 0.88)	<0.001	238 (41.7)	0.58 (0.51 - 0.65)	0.05
Challenging/Very Challenging	804 (27.3)	0.60 (0.56 - 0.64)		333 (58.3)	0.49 (0.42 - 0.56)	
Do you consider your most advanced diagnosis borderline? [d]						
Yes	746 (25.4)	0.60 (0.56 - 0.64)	<0.001	311 (54.5)	0.50 (0.42 - 0.58)	0.11
No	2194 (74.6)	0.85 (0.83 - 0.87)		260 (45.5)	0.57 (0.51 - 0.63)	
Would you desire a second pathologist's opinion of this case before finalizing?						
Yes	1003 (34.1)	0.63 (0.60 - 0.66)	<0.001	370 (64.8)	0.49 (0.42 - 0.55)	0.03
No	1937 (65.9)	0.87 (0.85 - 0.89)		201 (35.2)	0.61 (0.52 - 0.69)	

[a] Includes all interpretations considered to be atypia by pathologists in Phase I.

[b] From responses on a Likert scale ranging from 1 (highly confident) to 6 (not confident at all), where high confidence was 1, 2 or 3, and low confidence 4, 5 or 6.

[c] From responses on a Likert scale ranging from 0 (very easy) to 5 (very challenging), where increased level of challenge was 4, 5 or 6, and lower level was 0, 1, or 2.

[d] From question, “Do you consider your most advanced diagnosis for this case to be borderline between two diagnoses?”