Published in final edited form as:

Eur J Cancer. 2017 July ; 80: 39–47. doi:10.1016/j.ejca.2017.04.013.

The Diagnostic Challenge of Low-grade Ductal Carcinoma *In Situ*

Tracy Onega, PhD, MA, MS¹, Donald L. Weaver, MD², Paul D. Fredrick, MPH, MBA³, Kimberly H Allison, MD⁴, Anna N. A. Tosteson, ScD⁵, Patricia A. Carney, PhD⁶, Berta M. Geller, EdD⁷, Gary M. Longton, MS⁸, Heidi D. Nelson, MD⁹, Natalia V. Oster, MPH³, Margaret S. Pepe, PhD⁸, and Joann G. Elmore, MD, MPH³

¹Department of Biomedical Data Science, Department of Epidemiology, The Dartmouth Institute for Health Policy and Clinical Practice and Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Lebanon, NH

²Department of Pathology, University of Vermont and UVM Cancer Center, Burlington, VT

³Department of Medicine, University of Washington School of Medicine, Seattle, WA

⁴Department of Pathology, Stanford University School of Medicine, Stanford, CA

⁵Department of Medicine, The Dartmouth Institute for Health Policy and Clinical Practice and Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Lebanon, NH

⁶Department of Family Medicine, Oregon Health & Science University, Portland, OR

⁷Department of Family Medicine, University of Vermont, Burlington, VT 05401

⁸Department of Biostatistics, University of Washington, Seattle, WA 98101

⁹Providence Cancer Center, Providence Health and Services Oregon, and Departments of Medical Informatics and Clinical Epidemiology and Medicine, Oregon Health & Science University, Portland, OR

Abstract

Background—Diagnostic agreement among pathologists is 84% for ductal carcinoma in situ (DCIS). Studies of interpretive variation according to grade are limited.

Methods—A national sample of 115 pathologists interpreted 240 breast pathology test set cases in the Breast Pathology Study and their interpretations were compared to expert consensus interpretations. We assessed agreement of pathologists' interpretations with a consensus reference diagnosis of DCIS dichotomized into low and high grade lesions. Generalized estimating

Corresponding Author: Tracy Onega, PhD, MS, MA, HB 7927 Rubin 8, One Medical Center Dr., Lebanon, NH 03756, 603-653-3671, Tracy.L.Onega@dartmouth.edu.

Conflict of interest statement:

There are no conflicts of interest to or financial disclosures from any authors.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

equations were used in logistic regression models of rates of under-and over-interpretation of DCIS by grade.

Results—We evaluated 2,097 independent interpretations of DCIS (512 low-grade DCIS and 1,585 high-grade DCIS). Agreement with reference diagnoses was 46% (95% CI 42–51) for low-grade DCIS and 83% (95% CI 81–86) for high-grade DCIS. The proportion of reference low-grade DCIS interpretations over-interpreted by pathologists (i.e., categorized as either high-grade DCIS or invasive cancer) was 23% (95% CI 19–28); 30% (95% CI 26–34) were interpreted as a lower diagnostic category (atypia or benign proliferative). Reference high-grade DCIS was under-interpreted in 14% (95% CI 12–16) of observations and only over-interpreted 3% (95% CI 2–4).

Conclusion—Grade is a major factor when examining pathologists' variability in diagnosing DCIS, with much lower agreement for low-grade DCIS cases compared to high-grade. These findings support the hypothesis that low-grade DCIS poses a greater interpretive challenge than high-grade DCIS, which should be considered when developing DCIS management strategies.

Keywords

Ductal Carcinoma in situ; grade; breast pathology; pathology interpretation

INTRODUCTION

Ductal carcinoma in situ (DCIS) is a noninvasive neoplastic breast lesion consisting of cytologically abnormal cells confined to the ducts without extension beyond the basement membrane. Although generally considered a direct precursor to invasive breast carcinoma, DCIS represents a spectrum of abnormality ranging from low-grade lesions with little to no risk for invasive cancer to high-grade lesions with a high likelihood for invasive malignant transformation.¹ Historically, DCIS was detected by palpation and treated with total mastectomy, similar to treatment of invasive carcinoma. The sharp rise in DCIS incidence that occurred in the 1990s was due to detection by screening mammography, which led to discovery of earlier, less aggressive forms of DCIS. However, surgical treatment of DCIS remains remarkably unchanged relative to invasive carcinoma, raising concerns of overtreatment, particularly for lower-grade DCIS.

Although the potential for invasive transformation of DCIS continues to be poorly understood, some immunohistochemical and genetic characteristics are shared between DCIS and invasive cancer.² However, validated prognostic genetic markers for DCIS^{3,4} are not yet in common clinical use. Consequently, pathologists in clinical practice often base their diagnoses on assessment of morphologic features, and these diagnoses guide treatment for patients. DCIS grade is determined using several different pathology classification systems⁵ --most commonly classified into low- and high-grade, or low-, intermediate- and high-grade categories. Distinguishing DCIS grade along the continuum of breast lesions that range from benign precursors, such as atypical ductal hyperplasia (ADH) to invasive cancer, is critical in determining appropriate treatment and reducing potential overtreatment of indolent cancers.⁶

Recent studies indicate discrepancies among pathologists in distinguishing between ADH, DCIS, and invasive cancer. In our previous work in the Breast Pathology Study (B-Path), interpretive agreement was 84% for DCIS, compared to 96% for invasive breast cancer, and 48% for ADH.⁷ Although prior studies have noted similar discrepancies for DCIS interpretation^{8,9}, to our knowledge only one has examined pathologist interpretive variation for DCIS by grade, which assessed three DCIS classification systems.¹⁰ The present study estimates diagnostic agreement and quantifies relative over and under interpretation by comparing pathologists' interpretations of low and high-grade DCIS with a reference standard interpretation. We hypothesized that agreement would be higher with high-grade DCIS compared to low-grade DCIS. This study may inform ongoing debates in management of DCIS.

METHODS

Study Population

The B-Path Study enrolled a geographically diverse sample of pathologists from 8 U.S. states (Alaska, Maine, Minnesota, New Hampshire, New Mexico, Oregon, Vermont, Washington). Eligibility included having completed training, and having interpreted breast specimens for at least one year in practice. Pathologists were identified through membership in professional organizations, calls to pathology facilities, internet searches, or their affiliation with the Breast Cancer Surveillance Consortium (BCSC)¹¹ or Providence Health & Services in Oregon. The Institutional Review Boards at Dartmouth College, Fred Hutchinson Cancer Research Center, Providence Health & Services Oregon, University of Vermont, and University of Washington approved all study activities. Informed consent was obtained from pathologists. Details of the survey design and comparison of the study participants to non-participants are described elsewhere.¹²

Breast Pathology Test Set and Reference Diagnosis

Pathologists were assigned to interpret one of four test sets consisting of 60 breast cases as previously described.¹² The cases were reviewed independently by each pathologist, in random order; no standardized diagnostic definitions were provided. Pathologists recorded their interpretations using a standardized diagnostic form online.¹³ Pathologists received up to 20 free Category 1 Continuing Medical Education Credits¹⁴ for participating in this study, which included receiving individualized feedback comparing their categorical diagnoses with both that of their peers and the consensus panel reference diagnoses.

Development of the test cases has been reported in detail elsewhere.¹² Briefly, a stratified sampling design was used to randomly select breast biopsy cases (240 total; 23 invasive carcinoma, 55 high-grade DCIS, 18 low-grade DCIS, 72 atypical hyperplasia (atypia), 72 benign without atypia) from two BCSC screening registries, excluding mastectomy specimens. The registries collect standardized data on women's age, breast density, and biopsy type. One new slide was made for each case. The test sets included a higher proportion of ADH and DCIS cases than would normally be seen in clinical practice to facilitate calculating reasonably stable estimates of agreement for these diagnoses.

Key Definitions

A consensus panel of three expert breast pathologists defined the reference diagnosis for each case.¹⁵ Independent review followed by a modified Delphi approach was used to achieve a consensus diagnosis within diagnostic categories (benign, ductal carcinoma in situ [DCIS], and invasive carcinoma) for each case. Assessments of nuclear grade and the presence or absence of necrosis (focal or extensive) were recorded for all interpretations of DCIS. We used these two variables to dichotomously classify DCIS into low- and high-grade cohorts for the reference standard and participant pathologist interpretations. Cases with a reference diagnosis of DCIS were divided into “low-grade” and “high-grade” based on the consensus panel’s recorded nuclear grade and necrosis score. To maintain sufficient sample size for our statistical analyses, we applied the following definition: any specimen *without* necrosis and with a low or intermediate nuclear grade was categorized as “low-grade DCIS”, and the remainder were classified as “high-grade DCIS.” Thus, high-grade DCIS included: specimens *with at least focal* necrosis and with a low or intermediate nuclear grade; or specimens with high nuclear grade with or without necrosis. This definition is supported by prior evidence from studies showing that 62% of interpretive disagreement in nuclear grade was in distinguishing between low and intermediate grade nuclei, compared to 34% disagreement in distinguishing between intermediate and high-grade nuclei.¹⁶ Thus, establishing a dichotomous classification for DCIS simplified the challenges associated with three tiers of nuclear grade and aligned the classification with potential clinical treatment pathways for DCIS. The consensus panel was not required to achieve consensus on necrosis and nuclear grade during the modified Delphi work, therefore, we used the grade determination from two of three experts if there was not full agreement (majority opinion).

Our analysis thus describes agreement of pathologists using five diagnostic categories for each case: invasive carcinoma (n=23), high-grade DCIS (n=55), low-grade DCIS (n=18), atypical hyperplasia (atypia) (n=72), benign without atypia (n= 72). *Over-interpretation* was defined as cases classified by the pathologists at a higher diagnostic category relative to the reference; *under-interpretation* was defined as cases classified lower than the reference diagnosis; concordant cases were those with agreement between the pathologists’ and the reference diagnosis.

Statistical Analysis

We summarized counts and frequencies of pathologists by demographic and clinical/practice characteristics. We summarized agreement between the diagnostic categories of the pathologists with that of the reference standard diagnosis. The simple (Cohen’s) kappa statistic was used to calculate inter-observer agreement. Within each diagnostic category, we measured the proportion and standard error of pathologists’ interpretations that agreed with reference diagnostic categories. We modeled rates of over- and under-interpretation according to diagnostic category using generalized estimating equations (GEE) with an independent correlation structure and robust standard errors to account for correlated response data within a logistic regression model. Estimates and standard errors of Least Square-means were computed and transformed back to the original response scale via the inverse-link function to provide misclassification and agreement rates and their 95% confidence intervals. We used Wald confidence intervals and hypothesis testing to estimate

level of significance for group comparisons. We examined the interpretive variability among low- and high-grade DCIS cases at the case level using descriptive statistics. All analyses were performed using SAS, version 9.4.

RESULTS

Of the 115 pathologists, most were between the ages of 40 and 59 (72%), were not affiliated with an academic center (76%), did not have fellowship training in breast pathology (95%), and were not considered experts in breast pathology by their colleagues (78%).⁷

When comparing pathologists' independent interpretations with the reference standard interpretation, pathologists agreed with 46% of the reference-defined low-grade DCIS cases. Pathologists' interpretations of reference-defined low-grade DCIS cases included the following diagnoses: 16% benign without atypia, 14% atypia, 46% low-grade DCIS, 22% high-grade DCIS, 1% invasive (Table 1). In contrast, for high-grade DCIS, 83% of pathologists' interpretations were in agreement (Table 1). Agreement for low-grade DCIS was similar to atypia (48% and 46%, respectively) (Figure 1a). When the reference-defined atypia and low-grade DCIS cases were combined into a single category, agreement within this larger combined categorical grouping was 62%. (Figure 1b)

The proportion of reference low-grade DCIS cases over-interpreted was 23% (95% CI 19–28) and the proportion under-interpreted was 30% (95% CI 26–34) (Table 1). High-grade DCIS was under-interpreted relative to the reference diagnosis for 14% (95% CI 12–16) of interpretations; only 3% (95% CI 2–4) of interpretations were over-interpreted as invasive carcinoma. Low-grade DCIS cases showed greater variability in interpretation, with a majority of pathologists interpreting many cases in categories other than low-grade DCIS (such as lobular in situ lesions or atypical ductal hyperplasia). An example of a case of low-grade DCIS with one of the highest diagnostic agreement rates (66% agreement) is shown in Figure 2A–B and an example of a case with low agreement is shown in Figure 2C–D. In contrast, the vast majority of reference high-grade DCIS cases (83%) were interpreted concordantly by most pathologists. Figure 3A–B shows images from a case of consensus high-grade DCIS with 100% agreement and Figure 3C–D shows images from an unusual case with only 7% agreement with the consensus diagnosis of high-grade DCIS.

DISCUSSION

This study provides a novel assessment of the interpretive variability of DCIS by grade in relation to a spectrum of five possible diagnostic categories ranging from benign to invasive breast cancer. Using a national sample of 115 pathologists and a test-set based approach, we found a high level of diagnostic agreement for high-grade DCIS; however, agreement was markedly lower for low-grade DCIS and relatively similar in magnitude to agreement for atypia. Overall, reference low-grade DCIS cases exhibited extensive categorical diagnostic variation by pathologists within individual cases supporting the hypothesis that the inherent histopathological and morphological challenges in the diagnosis of low-grade DCIS are similar to the known challenges in diagnosing atypical ductal hyperplasia. This finding

underscores the need for better biomarkers or classification methods to distinguish indolent from aggressive forms of DCIS.

This study expands upon our prior work highlighting challenges in breast pathology interpretation using the B-Path Study test set.⁷ We previously reported an overall agreement of 84% for DCIS, which was comparable to agreement for benign without atypia (87%), lower than agreement for invasive carcinoma (96%), and much higher than agreement for atypia (48%).⁷ In our previous analysis, 16% of reference DCIS cases were discordant with the reference standard, with 13% under-interpreted as benign or atypia and 3% over-interpreted as invasive.⁷ Our findings in the current grade-stratified approach to assessing agreement for DCIS suggest that a large proportion of the discordance stems from challenges associated with inter-observer diagnostic agreement at the lower end of the DCIS spectrum.

Our previous work with this test set demonstrated that atypia (atypical hyperplasia) is often over interpreted as DCIS and this current work demonstrates one third of low-grade DCIS is under-interpreted, emphasizing the challenge pathologists experience categorically segregating atypical hyperplasia from low grade DCIS. Historical and contemporary proposals to address this problem include alternative classification terminology for less aggressive or indolent lesions.^{17,18} Our goal was to objectively evaluate from where on the spectrum of DCIS the majority of discordance stems and not necessarily endorse new classifications; however, we acknowledge that combining atypical hyperplasia and low grade DCIS improved categorical diagnostic agreement to 62%. It should be noted that although pathologists recognize intuitively the challenges and subjectivity of differentiating atypical hyperplasia from low grade DCIS, this issue has undergone limited study using robust methods. Shifting the treatment threshold further up the spectrum of DCIS does not wholly resolve the issue. While our observed concordance for high grade DCIS was 83%, low-grade DCIS was over interpreted as high-grade DCIS for 22% of interpretations. If high grade DCIS is treated more aggressively than low grade DCIS, repercussions persist for patients.

Our prior work also showed that other entities, such as lobular carcinoma in situ (LCIS) can be in the differential diagnosis of low-grade DCIS and this contributes to disagreement in the test set setting.¹⁹ In clinical practice, immunohistochemical stains for e-cadherin can assist differentiating LCIS and DCIS. Further refinements in diagnostic distinction between low and high-grades of DCIS may improve patient treatment outcomes and satisfaction. Ultimately, effective clinical stratification may require the addition of biomarkers such as hormone receptor status, Her2 status, or yet to be discovered biomarkers, in addition to DCIS grade.

Hannemann²⁰ and others²¹ note that making treatment decisions for DCIS based on grade is difficult because of the challenges in assessing histological type and grade. Several studies have shown that high-grade DCIS is more likely to recur and/or recur sooner following local therapy compared to low-grade DCIS.^{16–23} In a natural history study, even low-grade DCIS was associated with subsequent invasive breast cancer in 39% of cases when the follow-up was extended to over four decades.²⁴ Molecular profiling of DCIS continues to be investigated to identify gene expression profiles that may be strongly correlated with

prognosis. However, the morphologic overlap between atypia and low-grade DCIS seems to also exist at the molecular level, such that genetic markers do not yet discriminate well between these two diagnostic categories.² As the prognostic value of genetic and morphological features becomes better understood, the inherent challenges of pathologic interpretation of atypia and DCIS may be aided with predictive biomarkers.

Our study suggests that more effort should also be directed at segregating low-grade from high-grade DCIS, particularly when considering implications for clinical management. Wide variation in treatment of DCIS is seen among countries. For example, in a study of 12 countries, 67–90% of women with DCIS had breast conserving surgery, with radiotherapy in 41–100% of those cases.²⁵ Further, 6–59% received sentinel lymph node biopsy (SNLB) and <1–49% had axillary dissection (ALND), both of which were significantly more likely for larger size and higher grade DCIS lesions.²⁵ Because access to screening and technical improvements in mammography increased the incidence of DCIS, it follows that reducing screening would decrease incidence; however, opinions are divided and it has been suggested that the problem is overtreatment not overdiagnosis.²⁷ Different approaches to screening, including risk-based screening, and a willingness to observe more indolent lesions may help address the issues of overdiagnosis and overtreatment. In that regard, a growing recognition of DCIS as a heterogeneous disease, has generated several clinical studies, including the LORD, LORIS, and COMET trials,^{26–30} to test the efficacy of incorporating observation and active monitoring for lower risk DCIS as alternatives to surgical treatment. Thus, diagnostic interpretation of DCIS grade with methods to dichotomously stratify DCIS into treat or observe cohorts is likely to have major implications for management. We will have to wait for outcomes from those clinical trials.

While many studies have examined classification of DCIS by grade based on a variety of characteristics,^{5, 10, 25, 28, 31–34} this study is the largest and the first to use a sample of pathologists from 8 U.S. states to characterize interpretative agreement for low- and high-grade DCIS. This allowed for a more focused evaluation demonstrating the specific end of the spectrum of DCIS that suffers from low diagnostic reproducibility. It should be emphasized that this study only estimates diagnostic variability in a test-set setting and, considering the availability of additional slides, levels, special stains, clinical information and second opinions in clinical practice settings, results of this study are likely an over-estimation of diagnostic variability in actual U.S. practice. However, our methods were designed to more closely mimic real-world practice settings than other studies examining diagnostic variability in breast pathology, many of which use highly selected cases with specific areas on slides highlighted for pathologists. Cases in our study were randomly selected in each diagnostic category (rather than being selected as “good for teaching” or “challenging” cases) and the entire slide was available for review. Our study design was an effort to more closely approximate the range of case material that might be seen in typical practice within each of these categories.

Other strengths of this study include use of a standardized pathology interpretation form and the high number of pathologists and test cases. We also were able to estimate rates of over- and under-interpretation using a statistical approach that accounted for correlation in individual pathologist’s responses.¹⁰ Despite the strengths of this study, we acknowledge

limitations inherent to a test-set based study, such as inclusion of only one representative slide per case, no opportunity to employ immunohistochemical markers, no second opinion available, and no additional clinical history information. Further, we defined a broad classification of low- and high-grade DCIS to explore diagnostic variability at the low end of the DCIS spectrum; we recognize that more granular classification systems, which may have their own variation, are used in practice settings. In addition, within our test set of pathology cases, the overall proportion of DCIS was higher than typically found in clinical practice and outcome data are not available.

This study highlights the challenges associated with diagnosing DCIS by focusing on the differences in diagnostic variability of low-grade versus high-grade DCIS. High-grade DCIS is a more reproducible diagnosis with a low likelihood of over interpretation. However, low-grade DCIS, which accounts for approximately 46% of DCIS lesions,³⁵ is the major contributor to diagnostic variability within DCIS overall. Future work related to diagnosis and treatment of DCIS needs to clearly distinguish between the forms of this disease that are likely indolent and aggressive. While we await development of validated biomarkers that predict short-term progression of potential precursor lesions, such as ADH and low-grade DCIS to invasive breast cancer, we need to focus on improving pathological interpretation of indolent proliferative disease, accepting inherent challenges in diagnostic reproducibility, and examining less aggressive options for clinical management of low-grade DCIS.

Acknowledgments

This work was supported by the National Cancer Institute of the National Institutes of Health under award numbers R01 CA140560, U54 CA163303, R01 CA172343, U01CA86082, U01CA70013 and K05-CA104699 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 sources see: <http://www.breastscreening.cancer.gov/work/acknowledgement.html>

References

1. Virnig BA, Tuttle TM, Shamliyan T, Kane RL. Ductal carcinoma in situ of the breast: a systematic review of incidence, treatment, and outcome. *J Natl Cancer Inst.* 2010; 102:170–178. [PubMed: 20071685]
2. Buerger H, Mommers EC, Littmann R, Simon R, Diallo R, Poremba C, Dockhorn-Dworniczak B, van Diest PJ, Boecker W. Ductal invasive G2 and G3 carcinomas of the breast are the end stages of at least two different lines of genetic evolution. *J Pathol.* 2001; 194:165–170. [PubMed: 11400144]
3. Rakovitch E, Nofech-Mozes S, Hanna W, Baehner FL, Saskin R, Butler SM, et al. A population-based validation study of the DCIS Score predicting recurrence risk in individuals treated by breast-conserving surgery alone. *Breast Cancer Research and Treatment.* 2015; 152(2):389–398. [PubMed: 26119102]
4. Solin LJ, Gray R, Baehner FL, Butler SM, Hughes LL, Yoshizawa C. A multigene expression array to predict local recurrence risk for ductal carcinoma in situ of the breast. *JNCI.* 2013; 105(10):701–710. [PubMed: 23641039]
5. Sloane JP, Amendoeira I, Apostoklikas N, Belloc JP, Bianchi S, Boecker W, Bussolati G, Coleman D, et al. Consistency achieved by 23 European pathologists in categorizing ductal carcinoma in situ of the breast using five classifications. *Human Pathology.* 1998; 29:1056–1062. [PubMed: 9781641]
6. Esserman L, Thompson IM, Reid B, Nelson P, Ransohoff D, Welch HG, Whang S, Berry DA, Kinzler KW, Black WC, Bissell M, Pames H, Srivastava S. Addressing overdiagnosis and

overtreatment in cancer: a prescription for change. *Lancet Oncology*. 2014; 15(6):e234–42. [PubMed: 24807866]

7. Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson ANA, Nelson HD, Pepe MS, Allison KH, Schnitt SJ, O'Malley FP, Weaver DL. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA*. 2015; 313(11):1122–1132. [PubMed: 25781441]
8. Rosai J. Borderline epithelial lesions of the breast. *Am J Surg Pathol*. Mar; 1991 15(3):209–221. [PubMed: 1847606]
9. 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]
10. Wells WA, Carney PA, Eliassen SM, Grove MR, Tosteson ANA. Pathologists' agreement with experts and reproducibility of breast ductal carcinoma-in-situ classification schemes. *The American Journal of Surgical Pathology*. 2000; 24(5):651–659. [PubMed: 10800983]
11. Breast Cancer Surveillance Consortium. [Last accessed 07/23/15] <http://breastscreening.cancer.gov/>
12. Oster NV, Carney PA, Allison KH, Weaver D, Reisch L, Longton G, Onega T, Pepe M, Geller BM, Nelson H, Ross T, Tosteson AN, Elmore JG. Development of a diagnostic test set to assess agreement in breast pathology: Practical application of the Guidelines for Reporting Reliability and Agreement Studies (GRRAS). *BMC Women's Health*. 2013; 13:3. [PubMed: 23379630]
13. Nelson HD, Weerasinghe R, Martel M, Bifulco C, Assur T, Elmore JG, Weaver DL. Development of an electronic breast pathology database in a community health system. *J Pathol Inform*. 2014; 5:26. [PubMed: 25191625]
14. Carney PA, Allison K, Geller B, Oster N, Weaver D. Feasibility and acceptability of a web-based continuous professional development program designed to enhance breast pathology interpretation. In press.
15. Allison KH, Reisch LM, Carney PA, Weaver DL, Schnitt SJ, O'Malley FP, Geller BM, Elmore JG. Understanding Diagnostic Variability in Breast Pathology: Lessons Learned from an Expert Consensus Review Panel. *Histopathology*. 2014; 65(2):240–51. [PubMed: 24511905]
16. Allison KH, Rendi MH, Peacock S, Morgan T, Elmore JG, Weaver DL. Histological features associated with diagnostic agreement in atypical ductal hyperplasia of the breast: illustrative cases from the B-Path study. *Histopathology*. 2016; doi: 10.1111/his.13035.
17. Tavassoli FA. Ductal carcinoma in situ: introduction of the concept of ductal intraepithelial neoplasia. *Mod Pathol*. 1998 Feb; 11(2):140–154. [PubMed: 9504685]
18. Esserman LJ, Thompson IM, Reid B. Overdiagnosis and overtreatment in cancer: an opportunity for improvement. *JAMA*. 2013; 310:797–798. [PubMed: 23896967]
19. Douglas-Jones, Gupta SK, Attanoos RL, Morgan JM, Mansel RE. A critical appraisal of six modern classifications of ductal carcinoma in situ of the breast (DCIS): correlation with grade of associated invasive carcinoma. *Histopathology*. 2003; 29:397–409.
20. Hanneman J, Velds A, Halfwerk JBG, Kreike B, Peterse JL, van de Vijver MJ. Classification of ductal carcinoma in situ by gene expression profiling. *Breast Cancer Research*. 2006; 8:R61. [PubMed: 17069663]
21. Masood S. Why the term 'low-grade ductal carcinoma in situ' should be changed to 'borderline breast disease': diagnostic and clinical implications. *Womens Health*. 2012; 8(1):57–62.
22. Erbas B, Prvenzano E, Armes J, Gertig D. The natural history of ductal carcinoma in situ of the breast: a review. *Breast Cancer Research and Treatment*. 2006; 97:135–144. [PubMed: 16319971]
23. Fisher ER, Dignam J, Tan-Chiu E, Costantino J, Fisher B, Paik S, Wolmark N. Pathologic findings from the National Surgical Adjuvant Breast Project (NSABP) eight-year update of Protocol B-17: intraductal carcinoma [see comments]. *Cancer*. 1999; 86:429–438. [PubMed: 10430251]
24. Ringberg A, Idvall I, Ferno M, Anderson H, Anagnostaki L, Boiesen P, Bondesson L, Holm E, Johansson S, Lindholm K, Ljungberg O, Ostberg G. Ipsilateral local recurrence in relation to therapy and morphological characteristics in patients with ductal carcinoma in situ of the breast. *Eur J Surg Oncol*. 2000; 26:444–451. [PubMed: 11016463]
25. Bijker N, Peterse JL, Duchateau L, Julien JP, Fentiman IS, Duval C, Di Palma S, Simony-Lafontaine J, de Mascarel I, van de Vijver MJ. Risk factors for recurrence and metastasis after

- breast-conserving therapy for ductal carcinoma-in-situ: analysis of European Organization for Research and Treatment of Cancer Trial 10853. *J Clin Oncol*. 2001; 19:2263–2271. [PubMed: 11304780]
26. Ponti A, Lynge E, James T, Majek O, von Euler-Chelpin M, Anttila A, Fitzpatrick P, Mano MP, Kawai M, Scharpantgen A, Fracheboud J, Hofvind S, Vidal C, Asuncion N, Salas D, Bulliard JL, Segnan N, Kerlikowske K, Taplin S. ICSN DCIS Working group. International variation in management of screen-detected ductal carcinoma in situ of the breast. *Eur J Cancer*. 2014; 50(15): 2695–704. [PubMed: 25149183]
 27. Patel TY. It is overtreatment, not overdiagnosis. *Acad Radiol*. 2015; 22:1044–1045. [PubMed: 26100201]
 28. Francis A, Thomas J, Fallowfield L, Wallis M, Bartlett JM, Brookes C, Roberts T, Pirrie S, Gaunt C, Young J, Billingham L, Dodwell D, Hanby A, Pinder SE, Evans A, Reed M, Jenkins V, Matthews L, Wilcox M, Fairbrother P, Bowden S, Rea D. Addressing overtreatment of screen detected DCIS; the LORIS trial. *Eur J Cancer*. 2015 Nov; 51(16):2296–303. Epub 2015 Aug 18. DOI: 10.1016/j.ejca.2015.07.017 [PubMed: 26296293]
 29. Elshof LE, Tryfonidis K, Slaets L, van Leeuwen-Stok AE, Skinner VP, Dif N, Pijnappel RM, Bijker N, Rutgers EJ, Wesseling J. Feasibility of a prospective, randomised, open-label, international multicentre, phase III, non-inferiority trial to assess the safety of active surveillance for low risk ductal carcinoma in situ - The LORD study. *Eur J Cancer*. 2015 Aug; 51(12):1497–510. Epub 2015 May 26. DOI: 10.1016/j.ejca.2015.05.008 [PubMed: 26025767]
 30. Patient-Centered Outcomes Research Institute (PCORI). [Last accessed: June 21, 2016] Comparison of Operative versus Medical Endocrine Therapy for Low Risk DCIS: The COMET Trial. <http://www.pcori.org/research-results/2016/comparison-operative-versus-medical-endocrine-therapy-low-risk-dcis-comet>
 31. Stallard S, Hole DA, Purushotham AD, Hiew LY, Mehanna H, Cordiner C, Dobson H, Mallon EA, George WD. Ductal carcinoma in situ of the breast – among factors predicting for recurrence, distance from the nipple is important. *Eur J Surg Oncol*. 2001; 27:373–377. [PubMed: 11417983]
 32. Sanders ME, Schuyler PA, Dupont WD, Page DL. The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. *Cancer*. 2005; 103:2481–2484. DOI: 10.1002/cncr.21069 [PubMed: 15884091]
 33. Scott MA, Lagios MD, Axelsson K, Rogers LW, Anderson TJ, Page DL. Ductal carcinoma in situ of the breast: Reproducibility of histological subtype analysis. *Human Pathology*. 1997; 28:967–973. [PubMed: 9269834]
 34. Sneige N, Lagios MD, Schwarting R, Colburn W, Atkinson E, Webera D, Sahin A, Kemp B, Hogue A, Risin S, Sabichi A, Boone C, Dhingra K, Kelloff G, Lippman S. Interobserver reproducibility of the Lagios nuclear grading system for ductal carcinoma in situ. *Human Pathology*. 1999; 30:257–262. [PubMed: 10088542]
 35. Kricker A, Goumas C, Armstrong B. Ductal carcinoma in situ of the breast, a population-based study of the epidemiology and pathology. *British Journal of Cancer*. 2004; 90:1382–1385. [PubMed: 15054459]

Highlights

- Low-grade DCIS is subject to much greater variability in interpretation than high-grade DCIS
- A reference standard of low-grade DCIS was almost as likely to be interpreted as a higher diagnostic category than as a lower diagnostic category
- DCIS is heterogeneous and clinical management may need to account for grade, in which case distinguishing low- from high-grade DCIS is very important

Figure 1a

Figure 1b

Figures 1a and 1b.

Proportion of participant interpretations that agreed with the expert consensus diagnosis (n=115 participants)

Figure 2.

A–B) The top two panels show 4x (A) and 20x (B) images from a case of consensus low-grade DCIS that had one of the highest diagnostic agreement rates (66%). This case has uniform involvement of ducts by a proliferation that has clear cytologic monotony and cribriform architecture with polarized spaces, characteristic of low grade DCIS. It measures just over 2 mm (a size threshold frequently used for a low-grade DCIS diagnosis) and is cut across at the edge of the slide. The majority of participants who did not record a low-grade DCIS diagnosis recorded a highest order diagnosis of ADH (29%) or lobular in situ neoplasia (21%) for the case. C–D) The bottom two panels show 2x (C) and 10x (D) images of a case with only 28% diagnostic agreement with the consensus diagnosis of low-grade DCIS. The majority of participants who did not make this diagnosis recorded a highest order diagnosis of LCIS. The solid growth pattern and low grade monotony present can be seen in either of these diagnostic entities. An E-cadherin stain (which was not available to participants) could help resolve this differential in clinical practice.

Figure 3.

A–B) The top two panels show 10x (A) and 20x (B) images from a case with 100% diagnostic agreement with the consensus diagnosis of high grade DCIS. There is obvious central comedo-necrosis present in the center of a duct lined by a proliferation with pleomorphic, hyperchromatic nuclei. C–D) The bottom two panels show 10x (A) and 20x (B) images of an unusual case that had very low diagnostic agreement (7%) with the consensus high-grade DCIS diagnosis. This case lacks necrosis and the nuclei of the proliferation are not as obviously high grade as those shown in A–B. A second duct contains micropapillary structures with a similar cytology (B). The majority of participants recorded a highest order diagnosis of ADH (88%) for this case.

Table 1

Comparison of breast tissue interpretation from 115 pathologists compared to the reference diagnosis

Reference Diagnosis of Ductal Carcinoma <i>in situ</i> (DCIS)	No. of Interpretations	Participating Pathologists' Interpretation					Rate of Over- and Under- Interpretation Compared to the Reference Diagnosis		Overall Agreement Rate Compared to the Reference Diagnosis %
		n (row %)					Over-interpretation	Under- Interpretation	
Low-grade DCIS	512	Benign w/o Atypia	Atypia	Low-grade DCIS	High-grade DCIS	Invasive	23 (19, 28)	30 (26, 34)	46 (42, 51)
		82 (16%)	73 (14%)	238 (46%)	112 (22%)	7 (1%)			
High-grade DCIS	1585	51 (3%)	73 (5%)	93 (6%)	1321 (83%)	47 (3%)	3 (2, 4)	14 (12, 16)	83 (81, 86)

^d Generalized estimating equations (GEE) with an independent correlation structure and robust standard errors were used in a logistic regression model to account for correlated response data.

Estimates and standard errors of LS-means were computed and transformed back to the original response scale via the inverse-link function to provide misclassification and agreement rates and their 95% confidence intervals.