A Framework for Evaluating Diagnostic Discordance in Pathology Discovered During Research Studies

Sherry Feng, BS; Donald L. Weaver, MD; Patricia A. Carney, PhD; Lisa M. Reisch, PhD; Berta M. Geller, EdD; Andrew Goodwin, MD; Mara H. Rendi, MD; Tracy Onega, PhD; Kim H. Allison, MD; Anna N. A. Tosteson, ScD; Heidi D. Nelson, MD, MPH; Gary Longton, MS; Margaret Pepe, PhD; Joann G. Elmore, MD, MPH

• ***Context.***—**Little is known about the frequency of discordant diagnoses identified during research.**

Objective.—**To describe diagnostic discordance identified during research and apply a newly designed research framework for investigating discordance.**

Design.—**Breast biopsy cases (N = 407) from registries in Vermont and New Hampshire were independently reviewed by a breast pathology expert. The following research framework was developed to assess those cases: (1) compare the expert review and study database diagnoses, (2) determine the clinical significance of diagnostic discordance, (3) identify and correct data errors and verify the existence of true diagnostic discrepancies, (4) consider the impact of borderline cases, and (5) determine the notification approach for verified disagreements.**

Results.—**Initial overall discordance between the original diagnosis recorded in our research database and a breast pathology expert was 32.2% (131 of 407). This was reduced to less than 10% after following the 5-step research framework. Detailed review identified 12 cases (2.9%) with data errors (2 in the underlying pathology registry, 3 with incomplete slides sent for expert review, and 7 with data abstraction errors). After excluding the cases with data errors, 38 cases (9.6%) among the remaining 395 had clinically meaningful discordant diagnoses ($\kappa = 0.82$; SE, 0.04; 95% confidence interval, 0.76–0.87). Among these 38 cases, 20 (53%) were considered borderline between 2 diagnoses by either the original pathologist or the expert. We elected to notify the pathology registries and facilities regarding discordant diagnoses.**

Conclusions.—**Understanding the types and sources of diagnostic discordance uncovered in research studies may lead to improved scientific data and better patient care.**

(*Arch Pathol Lab Med.* 2014;138:955–961; doi: 10.5858/arpa.2013-0263-OA)

High-quality medical research depends on precise clinical data, including accurate interpretation of pathologic diagnosis. Studies of breast cancer screening, diagnosis, and treatment, including cancer clinical trials, use pathologists' diagnoses as the gold standard outcome.[1–5] Differences among pathologists' diagnoses occur in clinical practice and have been quantified in studies of interobserver interpretive variability.[6–8] Some cases of diagnostic discordance noted during research activities may have little to no clinical significance for patient care, whereas other cases may be considered medical errors and be associated with adverse patient outcomes or medical-legal consequences.[9]

Surprisingly, little is known about the frequency, severity, or reasons for diagnostic discordance discovered during the course of clinical research. This raises unique ethical issues because those discordant cases may have otherwise remained undiscovered. In addition, although borderline cases have long been considered challenging in pathology practice, the extent to which borderline cases may affect research has not been quantified, to our knowledge, within the context of large-scale studies. Finally, the discovery of discordant pathology diagnoses during a research study requires full review of the underlying quality of the research data because errors can originate at any point during the study from case identification to research data classification.

We are conducting a National Cancer Institute–funded study, to characterize the accuracy of breast pathology

Accepted for publication August 29, 2013.

From the School of Medicine (Ms Feng), the Division of General Internal Medicine (Dr Reisch and Dr Elmore), and the Department of Anatomic Pathology (Dr Rendi), University of Washington, Seattle; the Departments of Pathology, College of Medicine, and the Vermont Cancer Center (Dr Weaver), Family Medicine and Radiology (Dr Geller), and Pathology (Dr Goodwin), University of Vermont, Burlington; the Departments of Family Medicine and Public Health & Preventive Medicine (Dr Carney) and Medical Informatics & Clinical Epidemiology and Medicine (Dr Nelson), Oregon Health and Science University, Portland; the Section of Biostatistics and Epidemiology (Dr Onega), and the Department of Community & Family Medicine (Dr Tosteson), Dartmouth College, Lebanon, New Hampshire; the Department of Pathology, Stanford University, Stanford, California (Dr Allison); Biostatistics Modeling and Methods (Mr Longton) and Biostatistics and Biomathematics (Dr Pepe), Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle.

The authors have no relevant financial interest in the products or companies described in this article.

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 United States. For a full description of those sources, please see: http://www.breastscreening.cancer.gov/work/acknowledgement.html.

Reprints: Joann G. Elmore, MD, MPH, General Internal Medicine Harborview Medical Center, 325 Ninth Ave, Box 359780, Seattle, WA 98104-2499, (e-mail: jelmore@u.washington.edu).

Table 1. Scientific Framework for Identifying and Managing Diagnostic Discordance Among Pathologists Discovered During Research Activities

Step	Research Activity	Assessment Approach	Investigator Action
1	Compare the expert review and database diagnoses.	• Compare data items from different data sources for concordance, including independent review of pathology slides.	• Carefully oversee research staff conducting document comparisons. • Consult expert pathologist (blinded to the original diagnosis) to review actual slides from the case.
2	Determine the clinical significance of diagnostic discordance.	• Evaluate the potential effect on the patient's treatment and clinical care from diagnostic discordance.	• Consult clinical experts as needed to review the potential significance of diagnostic disagreements. This step may narrow the number of cases with potentially significant diagnostic disagreements that will require further detailed review.
3	Identify and correct data errors, and verify the existence of true diagnostic discrepancies.	• Verify data on the diagnostic discordant cases. • Check quality of research data for abstraction or coding errors (blinded panel of pathologists). • Following data corrections, calculate percentage of agreement and κ coefficients.	• Consult clinical experts to assess clinical interpretation of original pathologists' reports. • Create decision rules to be used to modify research data from original source data. • Carefully oversee data error corrections, including documenting all data changes.
4	Consider the impact of borderline cases that cross diagnostic categories.	• Define borderline cases. • Search study documents and databases to identify borderline cases using agreed-upon definitions.	• Consult breast pathology and biostatistical experts on the extent to which borderline diagnostic assessments should be taken into account in case classification and analyses.
5	Determine the notification approach for cases with verified discordant diagnoses.	• Consider the time lag between original clinical care that the patient would have received and research activities that identified significant discordant diagnoses.	• Consult experts in research ethics, pathology, and human subjects to discuss how to proceed. • Consider contacting the laboratory that conducted original interpretation to notify of discordant diagnoses. • Document all methods/notification practices.

interpretation in the United States. The study included developing a test set of breast pathology cases that represents a broad spectrum from benign, nonproliferative findings through invasive breast cancer, with oversampling of cases of atypia and ductal carcinoma in situ (DCIS). As part of the study oversight, we noted an initial rate of potential discordance in diagnoses between our expert pathologist review and the diagnosis recorded in our study database. This prompted an extensive review of all data sources, including gathering new data on the original pathologist's final report and full data from the respective state breast-pathology registries. The purpose of this article is to outline the research framework we developed to identify and characterize the kinds and sources of discordant diagnoses and to describe the findings and outcome from our evaluation.

MATERIALS AND METHODS

We developed a research framework and used it to identify, assess, and manage pathology discrepancies found during our study (Table 1). The 5 framework steps included: (1) compare the expert review and study database diagnoses, (2) determine the clinical significance of the diagnostic discordance, (3) identify and correct the data errors and verify the existence of the true diagnostic discordance, (4) consider the impact of borderline cases that cross diagnostic categories, and (5) determine the notification approach for verified discordant diagnoses. All study activities had Institutional Review Board approval.

Step 1: Compare the Expert Review and Study Database Diagnoses

We identified and reviewed 407 excisional and core breast biopsy specimens for potential inclusion in a larger ongoing study evaluating the effect of interpretive variation on treatment of breast disease in the United States. Details of the primary study have been reported elsewhere.[10] Briefly, cases were selected from the state-based pathology registries in New Hampshire and Vermont. Each registry routinely receives data on benign and malignant breast pathology cases from a variety of settings, including private practices, based in small community hospitals, and university-affiliated practices in tertiary medical centers. Trained medical records abstractors hired by the registries routinely abstract key variables from original pathology reports signed by the pathologist responsible for each registry-patient's diagnostic interpretation.

Cases for the present study were selected using random sampling, stratified on patient age (40–49 years versus 50 years or older), breast density (low versus high), and 5 major breast-pathology diagnostic categories defined as (1) nonproliferative changes, (2) proliferative changes without atypia, (3) atypia (eg, atypical ductal hyperplasia and intraductal papilloma with atypia), (4) DCIS, and (5) invasive breast cancer. Data from the respective pathology registries were abstracted and then classified according to the 5 diagnostic categories.[10] Biopsies from women aged 40–49 years, women with dense breast tissue, and women with diagnoses of atypia or DCIS were oversampled to increase the statistical power for specific aims in the larger, ongoing study. Hereafter, we refer to the registry-derived diagnosis abstracted and then recorded in our study database as the *database diagnosis*.

Pathology laboratories in New Hampshire and Vermont were asked to provide the complete set of original, archived glass slides for all cases identified. A university-based breast pathology expert associated with the study (D.W.) conducted an independent review of the slides, blinded to both the original pathology report and the database diagnosis, and assigned a diagnosis using the 5 diagnostic categories. We refer to this glass slide review diagnosis as the *expert diagnosis*.

For all 407 cases, we compared the database diagnosis to the expert diagnosis to quantify the overall extent of diagnostic discordance. Because it is common for multiple histologic findings to be recorded for a single biopsy (eg, nonproliferative changes in one area and atypia in another), we based our assessment only on the most concerning finding for each case. Thus, if the expert pathologist noted DCIS and the original pathologist noted atypia and DCIS, we considered this to be a concordant case of DCIS because the most concerning diagnoses matched.

Step 2: Determine the Clinical Significance of Diagnostic Discordance Identified

Noted discordance between the expert diagnoses and the database diagnoses that were unlikely to result in meaningful changes in treatment recommendations were defined as *minor discordance* (eg, discordance between the nonproliferative, proliferative without atypia, and atypia categories). We, therefore, merged those 3 diagnostic categories into a single "*benign*" category. We defined *clinically meaningful discordance* as those cases for which the database diagnosis would likely generate one set of treatment recommendations and the expert diagnosis would generate a different set. The expert and database diagnoses for the 3 main diagnostic categories (benign, DCIS, and invasive carcinoma) were then compared again using descriptive statistics with κ coefficients.[11,12]

Step 3: Identify and Correct Data Errors and Verify the Existence of True Diagnostic Discordance

New data were obtained for the subgroup of cases with potentially clinically meaningful diagnostic discordance, including a rereview of the original pathology report and a check for any possible data errors. Copies of the original pathology reports were obtained and deidentified for those cases. Two university-based breast pathology experts (A.G., M.H.R.) independently reviewed the original pathology reports and recorded their own interpretations of the original pathologist's primary diagnosis onto a study-specific histology assessment form that used the 5 diagnostic categories. This review was performed with no access to the original slides and blinded to the database diagnosis and the expert diagnosis. A comparison of the 2 university-based pathologists revealed complete agreement regarding their interpretations of the original pathologist's primary diagnosis for all cases. We refer to their interpretation of the original pathology reports' primary diagnoses as the *original diagnosis*. If the original diagnosis and the database diagnosis differed, this indicated that a data error may have occurred in either the pathology registry (eg, registry staff incorrectly abstracted or entered data from the original pathology reports) or in our research database (eg, if the wrong slides or an incomplete set of slides were sent for expert review).

Step 4: Consider the Impact of Borderline Cases That Cross Diagnostic Categories

Each case was reviewed for information to indicate whether either the original pathologist or the expert considered the case borderline between 2 or more diagnostic categories. When the phrasing of the original pathology report suggested that a finding might be classified in more than one of our 5 diagnostic categories, the 2 reviewing pathologists recorded the borderline diagnostic categories as well as the final diagnosis of the original pathologist. During the expert review, a case was considered borderline if notes on the histology form indicated a differential diagnosis between 2 or more diagnostic categories.

Step 5: Determine the Notification Approach for Cases With Verified, Meaningful Diagnostic Discordance

Once clinically meaningful diagnostic discordance had been identified and carefully verified, investigators needed to consider how they would manage that information. Based on the timing of the clinical breast biopsies and the actual study, our team made the decision to contact the original pathology laboratory directly about all final cases with verified clinically meaningful diagnostic discordance and to contact the pathology registries about the data errors.

RESULTS

Step 1: Compare the Expert Review and Study Database Diagnoses

In the 2 pathology registries, 19 498 breast biopsy cases, representing 13 677 women, met our eligibility requirements. Our sample of 407 breast biopsy cases (2.1%) was identified from those registry cases, and the original glass slides were ultimately obtained from 9 pathology facilities in Vermont (providing 305 cases [74.9%]) and 8 facilities in New Hampshire (providing 102 cases [25.1%]). Fifty-two pathologists provided the original diagnostic interpretations for those 407 cases.

Discordance between the expert diagnosis and the database diagnosis for the 5 diagnostic categories was 32.2% (131 of 407). The Figure shows the overall level of diagnostic discordance after step 1, followed by the outcome of our review after the subsequent steps in our framework, as described below.

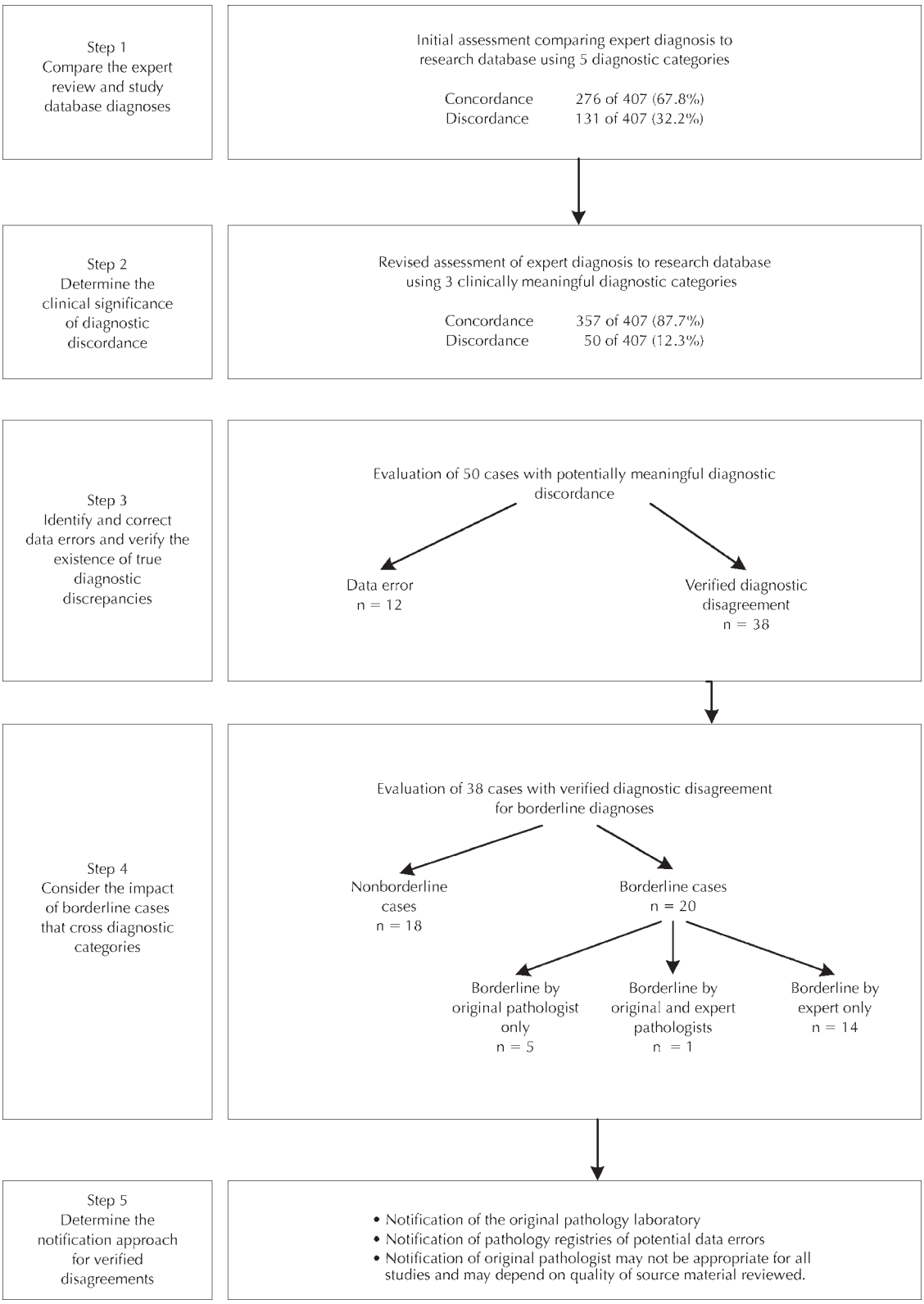
Step 2: Determine the Clinical Significance of Diagnostic Discordance Identified

More than one-half of the study cases (235 of 407, 57.7%) were noted by both the expert diagnosis and the database diagnosis as being within one of the 3 diagnostic categories of nonproliferative, proliferative without atypia, or atypia. As the clinical management of women with these diagnostic assessments is similar, those cases were merged into a single benign category for further comparative review. Among the 235 cases in the benign category, the expert and database diagnoses disagreed on the specific subtype assessment diagnosis category in 81 cases (34.5%); of those 81 cases, most ($n = 55$, 68%) were recorded as *atypia* in the study database and as *proliferative without atypia* by the expert (data not shown).

Discordance between the expert diagnosis and the database diagnosis when only the 3 clinically meaningful diagnostic categories were considered decreased to 12.3% (50 of 407). Discordant diagnoses were almost evenly divided between false-negative and false-positive categories relative to the expert diagnosis.

Step 3: Identify and Correct Data Errors and Verify the Existence of True Diagnostic Discordance

We identified research data errors in 12 of the 407 cases (2.9%). For 3 of the 12 cases (25%), we found that not all original slides were made fully available for expert review. Each of those cases was interpreted as *DCIS* by the original pathologist and as *benign* by the expert. Without a complete slide set, we could not verify whether those 3 cases were concordant. Two of the 12 cases (17%) with research data



Flow chart of review of diagnostic discordance uncovered in a pathology research study and summary of findings.

errors appeared to be due to pathology registry errors. In the first case, the pathology registry diagnosis was *atypia*, whereas both the expert diagnosis and the original diagnosis were invasive breast cancer. In the second case, the pathology registry diagnosis was *DCIS*, whereas both the expert diagnosis and the original diagnosis were *invasive breast cancer* (in that case, the original pathology report noted a "focus of micro-invasion," which pathology registry staff either missed or misinterpreted when they originally abstracted the record). The remaining 7 cases (58%) are the

Table 2. Comparison of Expert Diagnosis to Research Database Diagnosis After Exclusion of 12 Cases With Data Errors				
	Research Database Diagnosis			
Expert Diagnosis	Benign, No. (%)	DCIS, No. (%)	Invasive, No. (%)	Total, No. (%)
Benign	235[a] (59.5)	22 (5.6)	0 (0.0)	257 (65.1)
DCIS	8 (2.0)	78 (19.7)	1 (0.3)	87 (22.0)
Invasive	3 (0.8)	4 (1.0)	44 (11.1)	51 (12.9)
Total	**246** (62.3)	**104** (26.3)	**45** (11.4)	**395** (100.0)

Abbreviation: DCIS, ductal carcinoma in situ.

[a] Of the 235 cases classified as concordant in the combined benign category, 154 (65.5%) were categorically concordant within the 3 subcategories in the benign assessment category, and 81 (34.5%) were considered to have minor discordance between the expert diagnosis and the research database diagnosis.

result of women having multiple breast biopsies on or near the index date, with incorrect slides sent for the expert review, and abstraction mistakes when interpreting free text from the pathology registries into study-specific diagnostic categories.

The revised comparison using only 3 diagnostic categories and excluding the 12 cases with research errors was associated with substantial overall concordance as shown in Table 2. Percent agreement was 90.4% (357 of 395), with an unweighted κ coefficient of 0.82 (SE, 0.04; 95% bootstrap confidence interval [CI], 0.76–0.87). The weighted κ coefficient was 0.85 (SE, 0.03; 95% CI, 0.80–0.90).

Step 4: Consider the Impact of Borderline Cases That Cross Diagnostic Categories

Among the remaining cases (n = 38) with potentially significant discordant diagnoses, 20 (53%) were considered "borderline" between 2 differential diagnoses: 14 of the 20 (70%) by the expert reviewer only, 5 (25%) by the original pathologist only, and 1 (5%) by both the expert and the original pathologist. For these borderline cases, the final assessment was usually the less-severe diagnosis. Of the 14 cases considered borderline only by the expert, 3 (21%) were assigned the more-severe diagnosis, and 11 (79%) were given the less-severe diagnosis. Of the 5 cases considered borderline by the original pathologist, all 5 (100%) were assigned the less-severe diagnosis. We noted one case where the original pathologist identified it as borderline between atypia, DCIS, and invasive cancer, and provided a final diagnosis of atypia; the expert diagnosis for that case was DCIS with pseudoinvasion. For the single case considered borderline by both the expert and the original pathologist, the expert considered the case invasive (borderline with DCIS) and the original pathologist DCIS (borderline with invasive).

In 16 of the 20 borderline cases (80%), the diagnoses of the expert and the original pathologist could be considered to be in "partial agreement" (eg, the original diagnosis was *DCIS* and the expert diagnosis was *atypia with a differential diagnosis of DCIS*). In 3 of the 4 remaining cases (75%), partial agreement could not be determined because the expert pathologist reported the case was *borderline* without stating the differential diagnosis. The final case remained discordant because the original pathologist's diagnosis was *DCIS borderline with invasive cancer* and the expert diagnosis was *atypia*.

Step 5: Determine the Notification Approach for Cases With Verified Discordant Diagnoses

We decided to notify the original laboratories for each of the 38 cases with a verified clinically meaningful discordant diagnosis noted between the study expert and the original pathologist. We also contacted the pathology registry about the data errors noted in their system, and their databases were corrected accordingly.

DISCUSSION

Researchers rarely acknowledge the extent to which potential medical errors emerge during the course of their investigations or how identifying these potential errors should be handled. Researchers may also neglect careful data-cleaning efforts, allowing medical abstraction errors and/or methodological errors to adversely alter registry or research databases, possibly inflating estimates of diagnostic discordance. The present study is an example of the identification and subsequent evaluation of diagnostic discordance noted within a large National Institutes of Health–sponsored breast-pathology study. Our work resulted in the development of a research framework, which we used to identify, assess, and manage pathology discordance. Importantly, the framework used mixed methods (quantitative and qualitative) that proved invaluable in elucidating the types and severity of the diagnostic discordance noted in research dependent on histopathology.

Although the primary National Institutes of Health–supported study[10] will not be affected by the diagnostic disagreements identified, we were compelled to investigate cases with potentially clinically meaningful diagnostic discordance to further understand their genesis. Our study revealed that most discordance involves benign diagnoses that likely have minor consequences for clinical treatment. In addition, pathology registry errors were an infrequent contributor to discordance, an encouraging finding. After full review, the research framework we applied verified unintended, yet potentially clinically meaningful, diagnostic discordance in 38 of 395 cases (9.6%).

Our initial step 1 comparison of the expert diagnosis to the database diagnosis using 5 diagnostic categories generated significant diagnostic discordance that, at first, appeared alarmingly high at 32.2%, or 1 out of every 3 of the study cases. In many of those cases, the discordance involved differences in categorization between nonproliferative, proliferative disease without atypia, and atypia. After merging those 3 assessment categories into a single benign category and applying the step 2 filter of potential clinical significance, the number of meaningful disagreements was reduced to 12.3%. Although that filter is vital to identifying cases with potentially significant disagreements for further detailed review, that does not imply that pathologists should ignore interobserver variation in cases with benign diagnoses. In fact, assessment of proliferative disease with atypia

affects cancer risk assessment and can modify breast cancer screening recommendations.[13] Additionally, in many institutions, atypia identified in a diagnostic core biopsy often requires excision of additional tissue for complete histologic evaluation to exclude more-advanced disease. However, major treatments, such as chemotherapy and radiation, are not determined at those diagnostic thresholds. Clearly, examining disagreements according to their clinical significance is dependent on the objective clinical decision for which the study has been conducted.

The third step in our research framework involved evaluating data quality. In our study, that required collection of the original pathology reports on each patient and a more-cumbersome qualitative investigation to determine whether a case had been incorrectly abstracted or coded in the pathology registry or in the study research database. If there were no errors related to the data quality, we considered that verification of diagnostic discordance.

We discovered that many apparent disagreements arise from challenging cases that manifest differential diagnoses crossing treatment recommendations. Thus, a fourth step in our framework was to evaluate borderline cases. Borderline cases exist in a poorly understood gray zone between 2 diagnostic categories, resulting from subjective rules imposed on a continuum of disease. Strategies to address discordance in research activities would differ for the various underlying causes of apparent disagreement. Thus, our framework for evaluating diagnostic discordance is critical for investigators who design and analyze studies, as well as for clinicians who apply research findings to medical practice.

Interestingly, both the expert and the original pathologists used the less-severe diagnosis as their final assessment in most borderline cases, raising consideration of the more-severe diagnosis in a comment section of the report. Of the borderline cases, atypia and DCIS were the most challenging diagnoses to differentiate. However, we were surprised to see that the original pathologist and the expert pathologist did not identify the same cases as borderline. Of 20 cases considered borderline by either the expert or the original pathologist, only one case (5%) was identified as *borderline* by both. Notably, the expert pathologist identified almost 3 times the number of borderline cases as the original pathologists did. That is likely due to the expert's acquired appreciation of the types of lesions referred for consultative opinion in clinical practice and the extra attention given to what each case might produce when additional sections were made from the original paraffin-embedded tissue block for the overall study. Although educational efforts may reduce diagnostic disagreement at the thresholds between categoric groupings, the borderline phenomenon in pathology is unlikely to be eliminated. We believe borderline cases require special attention in research studies and should be the subject of additional investigation and possibly separate analysis in research involving diagnosis and outcome.

Once clinically significant diagnostic disagreements are identified and verified during research activities, the fifth and final step is the management of that information. Although some research groups do not share that information with the original pathologists, providing information on the diagnostic disagreements to the original pathologist may be important to quality improvement within our field. In many instances, providing that information to the original pathologist may have little to no effect on patient care because of the usual time lag between research activities and clinical care. Alternatively, providing that information to the original laboratory or pathologist allows the opportunity for education and/or consultation with risk management professionals. The original laboratory is in the best position to determine whether a discordant diagnosis has already been identified through other review mechanisms or clinical follow-up, whether the impact of a potential misdiagnosis may already have manifested, or whether there may still exist the opportunity for improved care for a particular patient.

Our study does have some limitations. Our initial sample of 407 cases may be considered small, and our study population was limited to data recorded in 2 pathology registries. Thus, our findings may be biased by regional differences. Nevertheless, each pathology registry included a wide variety of clinical settings, from small private practices to large academic centers, such that regional differences may be less significant. We also deliberately oversampled 2 challenging diagnoses, atypical ductal hyperplasia and DCIS, and have not corrected for their population-based prevalence in this analysis. Another limitation, common among many studies in pathology, was that the database, expert, and original diagnosis for each case was forced into a standardized classification scheme for the study, the Breast Pathology Assessment Tool and Hierarchy for Diagnosis (B-PATH-Dx). That approach may not mirror clinical practice, where final diagnostic reports often allow pathologists to write detailed text that captures their thoughts and diagnostic differentials. Research activities usually require such standardization, and both clinicians and patients often look for an unambiguous diagnoses that lead to recommendations. Although we acknowledge that histologic interpretation is a complicated art, we contend that the communication of pathologic findings would be improved by being simplified and organized, as research shows that clinicians who read pathology reports misinterpret the intended diagnosis 30% of the time.[14,15]

In summary, diagnostic discordance is noted within research activities in pathology and should be considered carefully. We describe a multistep approach for evaluating those diagnostic discrepancies, determining whether they are due to data errors or are clinically meaningful and we discuss the effect those errors might have on clinical care and research. Inadequate evaluation of diagnostic discordance might lead to an incorrect and highly inflated estimate of diagnostic errors and, if those inflated estimates of diagnostic errors were published, it might leave pathologists understandably leery of participating in future research activities. A better understanding of the types of diagnostic discordance and the sources from which those apparent disagreements arise during research may lead to improved scientific data, more accurate conclusions, and ultimately improved patient care.

This work was supported by grants R01 CA140560 and KO5 CA104699 from the National Cancer Institute and grants U01CA86082, U01CA70013, U01CA69976, HHSN261201100031C from the National Cancer Institute funded Breast Cancer Surveillance Consortium. It was also funded under NCATS Grant TL1 TR 000422. We thank Yasman Moshiri, Vignesh Raghunath, Tom Morgan, Natalia Oster, Sara Jackson, and Brenda Shinosky for their assistance.

References

1. Feinstein AR. A bibliography of publications on observer variability. *J Chronic Dis*. 1985;38(8):619–632.

2. Elmore JG, Feinstein AR. A bibliography of publications on observer variability (final installment). *J Clin Epidemiol*. 1992;45(6):567–580.

3. Fenton JJ, Abraham L, Taplin SH, et al; Breast Cancer Surveillance Consortium. Effectiveness of computer-aided detection in community mammography practice. *J Natl Cancer Inst*. 2011;103(15):1152–1161.

4. Rorke LB. Pathologic diagnosis as the gold standard. *Cancer*. 1997;79(4):665–667.

5. Pisano ED, Hendrick RE, Yaffe MJ, et al; DMIST Investigators Group. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. *Radiology*. 2008;246(2):376–383.

6. Renshaw AA, Gould EW. Comparison of disagreement and error rates for three types of interdepartmental consultations. *Am J Clin Pathol*. 2005;124(6):878–882.

7. Lurkin A, Ducimetiere F, Vince DR, et al. Epidemiological evaluation of concordance between initial diagnosis and central pathology review in a comprehensive and prospective series of sarcoma patients in the Rhone-Alpes region. *BMC Cancer*. 2010;10:150. doi: 10.1186/1471-2407-10-150.

8. 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.

9. Troxel DB. Diagnostic errors in surgical pathology uncovered by a review of malpractice claims. Part III. Breast biopsies. *Int J Surg Pathol*. 2000;8(4):335–337.

10. Oster NV, Carney PA, Allison KH, et al. 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 Womens Health*. 2013;13:3. doi: 10.1186/1472-6874-13-3.

11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.

12. Fleiss JL. *Statistical Methods for Rates and Proportions*. 2nd ed. New York, NY: Wiley-Interscience; 1981.

13. Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Ann Intern Med*. 2011;155(1):10–20.

14. Buckley JM, Coopey SB, Sharko J, et al. The feasibility of using natural language processing to extract clinical information from breast pathology reports. *J Pathol Inform*. 2012;3:23. doi: 10.4103/2153–3539.97788.

15. Powsner SM, Costa J, Homer RJ. Clinicians are from Mars and pathologists are from Venus. *Arch Pathol Lab Med*. 2000;124(7):1040–1046.