Published in final edited form as:

Med Decis Making. 2017 January ; 37(1): 91–100. doi:10.1177/0272989X16638326.

The Influence of Disease Severity of Preceding Clinical Cases on Pathologists' Medical Decision Making

Paul D. Frederick, MPH, MBA¹, Heidi D. Nelson, MD, MPH², Patricia A. Carney, PhD³, Tad T. Brunyé⁴, Kimberly H. Allison, MD⁵, Donald L. Weaver, MD⁶, and Joann G. Elmore, MD, MPH¹

¹Department of Medicine, University of Washington School of Medicine, Seattle, WA

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

³Department of Family Medicine, Oregon Health & Science University, Portland, OR

⁴Center for Applied Brain & Cognitive Sciences, Tufts University

⁵Department of Pathology, Stanford University School of Medicine, Stanford, CA

⁶Department of Pathology, University of Vermont and UVM Cancer Center, Burlington, VT

Abstract

Background—Medical decision making may be influenced by contextual factors. We evaluated whether pathologists are influenced by disease severity of recently observed cases.

Methods—Pathologists independently interpreted 60 breast biopsies (one slide per case; 240 total cases in the study) in a prospective randomized observational study. Pathologists interpreted the same cases in two phases, separated by a washout period of > 6 months. Participants were not informed that the cases were identical in each phase and the sequence was reordered randomly for each pathologist and between phases. A consensus reference diagnosis was established for each case by three experienced breast pathologists. Ordered logit models examined the effect the pathologists' diagnosis on the preceding case or the five preceding cases had on their diagnosis for the subsequent index case.

Results—Among 152 pathologists, 49 provided interpretive data in both phase I and II; 66 from only phase I and 37 from phase II only. In Phase I, pathologists were more likely to indicate a more severe diagnosis than the reference diagnosis when the preceding case was diagnosed as ductal carcinoma in situ (DCIS) or invasive cancer (proportional odds ratio [POR] 1.28, 95% confidence interval [CI] 1.15-1.42). Results were similar when considering the preceding 5 cases,

Correspondence to: Paul Frederick, Department of Medicine, Phone: 206-744-8372, paulfred@u.washington.edu.

Data Availability: The authors confirm that, for approved reasons, some access restrictions apply to the data underlying the findings. Public deposition of the data may threaten participant identification. Data are available from request, and requests may be sent to: Mr. Tom Morgan, Research Manager, General Internal Medicine, University of Washington, 325 Ninth Ave., Seattle, WA 98104, trolf@u.washington.edu.

and for the pathologists in phase II who interpreted the same cases in a different order compared to phase I (POR 1.17, 95% CI 1.05-1.31).

Conclusion—Physicians appear to be influenced by the severity of previously interpreted test cases. Understanding types and sources of diagnostic bias may lead to improved assessment of accuracy and better patient care.

Keywords

breast; biopsy; cancer; diagnosis; interpretation; bias; sequential context effects

Introduction

The underlying prevalence of disease varies across and within physicians' clinical practices and may influence their medical decision-making. From clinical practice it seems that when physicians diagnose a patient with a severe medical condition, such as cancer, that this experience influences their evaluation and assessment of subsequent patients. While this influencing effect has not been directly examined, physicians' interpretations have been noted to shift toward more severe diagnoses when the overall prevalence of disease is higher than expected by chance.¹ This effect, known as context bias, has been reported among radiologists in testing situations where the prevalence of disease was notably higher than in clinical practice.¹

Although diagnostic criteria exist for many disciplines,²⁻⁴ their application involves human judgment.⁵ Thus, the medical decision-making process may be influenced by perceptual, cognitive, social, and emotional factors⁶ and decisions based on rational choice may be biased by beliefs, expectations, and context.^{7,8} Context bias may be particularly influential for diagnoses known to have high levels of diagnostic variability or for cases considered borderline between two diagnostic entities or categories.

While context bias has been described with respect to the potential bias resulting from the overall prevalence of disease within test sets, it is unclear whether the immediately preceding clinical cases exert a biasing effect on a subsequent medical decision. Empirical and theoretical work in judgment and decision-making provides a strong foundation for examining this possibility. In general, when making concrete (i.e., size) or abstract (i.e., price) judgments, observers tend to consider an item's immediate spatial and temporal context in arriving at an interpretation.⁹ In some cases judgments can be biased toward a preceding stimulus or response, whereas in other cases it may be biased away; these judgment patterns are respectively termed assimilation and contrast, terms borrowed from the perceptual psychophysics literature.^{9,10} Though assimilation has proven to be a generally more common pattern,¹¹ the extent to which someone assimilates or contrasts a response relative to a preceding event is influenced by a range of participant-level and context-level parameters, such as whether feedback is provided, a participant is an expert observer, or the alternation rate of successive events.^{9,12} In general, the judgment and decision making literature has provided substantial evidence that an immediately preceding judgment serves as a reference point (or "anchor"¹³) for arriving at a subsequent judgment.

The present study examines relatively complex, naturalistic judgments made on behalf of pathologists examining breast biopsy specimens. In clinical practice, pathologists interpreting an extended sequence of cases are expected to rely solely on the perceptible features of the current case. However given the background literature reviewed, the hypothesis is that pathologists' diagnoses will be influenced by the temporal context of preceding case judgments, possibly demonstrating assimilation of diagnoses across cases. Specifically, assimilation would predict over-diagnosis following high diagnostic severity cases, and under-diagnosis following low severity cases. To examine this possibility we tested whether diagnoses of breast biopsy specimens, in a controlled situation where we varied the order of case presentation, are affected by the diagnostic severity of immediately preceding cases.

Methods

The methods for this two-phase prospective randomized observational study are described in detail elsewhere.¹⁴ Briefly, we developed four test sets of breast tissue specimens, each of which contained 60 breast biopsies (240 total cases, 1 slide per case). A consensus panel of three experienced breast pathologists established a reference diagnoses for each case.¹⁵ For this study the diagnostic mapping shown in the appendix (eTable 1) was used and resulted in test cases that included 23 (10%) cases of invasive breast cancer, 58 (24%) high-grade ductal carcinoma in situ (DCIS), 18 (7%) low-grade ductal carcinoma in situ (DCIS), 81 (34%) with atypical hyperplasia (atypia), and 60 (25%) benign cases without atypia.

Each participant interpreted 60 cases. Atypical ductal hyperplasia (ADH) and ductal carcinoma in situ (DCIS), categories known to have high diagnostic variability and to be particularly challenging to diagnose, were over represented. All cases were obtained from women aged 40 and older. 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.

Pathologists who interpret breast specimens in clinical practice in eight U.S. states were invited to participate in the main study and 65% enrolled. Among the 252 pathologists who completed baseline surveys and were randomly stratified, 44 (17%) declined test set interpretation. The remaining 208 pathologists interpreted either glass slides or digital whole slide images (WSI). Data from pathologists' interpretation on the digital whole slide images are not included in this study. The 152 pathologists who interpreted glass slides in either phase I or in phase II or in both phases were included in this analysis (Figure 1), this included 66 pathologists who provided data in only phase I, 37 pathologists who provided data from phase II only and 49 who provided interpretive data in both phase I and II. Pathologists were not informed that the cases were identical in each phase or that the sequence of cases was ordered differently at random in each phase and for each individual pathologist. A washout period of at least 6 months was included between phases.

Interpretations were performed independently by pathologists, and diagnoses were reported using a Breast Pathology Assessment Tool and Hierarchy for Diagnosis (BPATH-Dx) form.¹⁶ Participant's highest order diagnosis was used to classify their case diagnosis into

one of the five diagnostic categories (see eTable 1). Pathologists were instructed to interpret the cases using their own clinical practice criteria.

Data were structured for each pathologist into an overlapping two-event sequence based on the order of their case interpretations (Figure 2). The unit of analysis consisted of the chronologically ordered pair of cases where the first event was the exposure (e.g., the pathologists' diagnoses on preceding cases), and the second event, the outcome (e.g., the pathologists' diagnosis on the subsequent index case). Cross-tabulations of first event and second event were constructed to describe the dependencies occurring in the interactive sequences of diagnoses. Strength and direction of associations between first and second events were examined for pathologists who participated within each phase as well as for those who participated in both phases.

The severity of either the preceding case or the 5 preceding contiguous cases defined the first event (exposure). When considering the preceding five cases, the measure of severity was based on the proportion of cases falling into one or more specified diagnostic classifications (e.g., at least 3 of the preceding 5 cases were diagnosed by the participant as either DCIS or invasive carcinoma) and for the one preceding case, severity was described dichotomously (e.g., 'invasive carcinoma' versus 'not invasive carcinoma'). The participants' diagnosis on the index case was then "anchored" to the reference standard diagnosis and the actual change from the reference diagnosis comprised the second event (outcome). As each case had an expert consensus reference diagnosis (ranging from 1, benign without atypia, to 5, invasive carcinoma) and each participant's diagnosis on the index case used the same scale (ranging from 1 to 5), the change between the participants' diagnosis and the expert consensus reference diagnosis could range from -4 to +4. The calculated change between participant and the reference diagnoses was then collapsed and recoded as -1 (participant index case diagnosis was less severe than the reference), 0 (participant diagnosis agrees with the reference diagnosis), and +1 (participants case diagnosis was more severe than the reference diagnosis).

For the intra-observer analysis, data from the 49 pathologists who interpreted identical test cases in both Phase I and Phase II were used. The change in the severity of the preceding case in Phase II relative to the severity of the preceding case in Phase I defined the first event (exposure) and the change in the severity of the identical index cases defined the second event (outcome). In the first event (exposure), the random reordering of cases in their presentation in Phase II provided dynamically different scenarios with regard to severity of the preceding case. The Phase I interpretations for first and second events were considered the anchor diagnosis when calculating the actual change between diagnostic ratings. The coding structure for both first and second events was similarly collapsed to -1 (participants' diagnosis on the Phase II case was less severe than the diagnosis given to the Phase I case), 0 (No change in participant's diagnosis of the case in Phase II compared to the case in Phase I), and +1 (participants' diagnosis on the Phase II case was more severe than the diagnosis given to the Phase I case).

The funding source had no role in the study design, the analysis and interpretation of data, writing the report, or the decision to submit for publication.

Statistical Methods

Contingency tables examined the associations between first events (exposure to preceding cases or changes in exposure to preceding cases) and second events (the change in the index case diagnosis). Summary statistics included frequencies and cell percentages. Proportional odds logit models were fit from a GEE (Generalized estimating equations) approach using PROC GENMOD in SAS[®] software and an independent working correlation structure to describe the strength, direction and significance of the comparisons made in the contingency tables while accounting for the dependency between repeated measurements within pathologist. In addition to the primary independent variable, all models were adjusted for the anchor diagnosis used to calculate the change in the index case diagnosis (the outcome). Wald statistics for type 3 GEE representing the effect of single indicator variables or the overall effect of multiple indicator variables tested the significance of each variable and generated 95% confidence intervals.

Since the ordinal outcome tested has three levels, two separate odds ratios representing the two cut points (1 vs 0, -1 and 1, 0 vs -1) of the ordinal response could be reported. However, for each proportional odds model, a constant summary odds ratio representing the two odds ratios was reported under the condition that the two odds ratios are equal, which cannot be assumed without statistical testing. To test this assumption of proportional odds, we used PROC SURVEYLOGISTIC and the default Taylor series method for the variance in conjunction with the CLUSTER statement to account for correlated responses by pathologists. The score chi-square test tested the validity of the proportional odds assumption but the test itself tends to overstate the evidence against the null hypothesis (i.e. the *P* value is artificially too small), of equivalent slopes across all logits, especially with small sample sizes.¹⁷ Therefore, if the null hypothesis of the proportional odds assumption was rejected, then 1) the proportional odds model was run without adjustment for the anchor diagnosis and the validity of the proportional odds assumption retested, 2) comparison of logit slopes were assessed first visually by plotting the empirical cumulative logit curves against the primary independent variable by using the observed counts¹⁷ overlaid separately for individual pathologists and for all pathologists combined (ignoring clustered responses), and 3) results reported from proportional odds model were compared to equivalent binary logistic models when the dependent ordinal variable was dichotomized using the two cut points of the ordinal outcome. For binary logistic models, the basic structure of the GEE ordinal regression models remained the same except a logit function was applied, albeit with a loss of statistical power.^{18,19}

A valid proportional odds model was assumed if the null hypothesis of proportional odds failed to be rejected or if the logit curves were approximately parallel for the primary independent variables in the models. Sensitivity analyses were performed to determine if our results were robust in both phases and for the diagnostic schema used to classify pathologists' interpretations. All statistical analyses were performed using SAS 9.4 (SAS Institute, Cary, NC). Tests of statistical significance were based on two-sided probability ($p < 0.05$).

Results

Data from the 152 pathologists who interpreted glass slide cases in Phase I (n=115) or in Phase II (n=86) were included in this study with 49 participating in both Phase I & II (Figure 1). Each pathologist interpreted 60 cases resulting in 6,900 individual case assessments in Phase I and 5,160 case assessments in Phase II. When the first event (exposure to previous cases) consisted of a single case, 6,785 and 5,074 pairs of sequential events comprised the study population for Phase I and II, respectively. For the intra-observer analysis of the pathologists who provided data in both phases, 2,842 matched unit pairs were available (Figure 2). Most participants were between 40 and 59 years of age, male, were not affiliated with an academic medical center, had more than 10 years of experience in breast pathology, and did not report that their colleagues considered them a breast pathology expert. Approximately half of the participants completed a fellowship program (Table 1).

Among the 6,785 pairs of diagnostic events in phase I, 71.5% (4,852) of index case interpretations were in overall agreement with the reference diagnosis with 14.3% (970) and 14.2% (963) of index interpretations more or less severe than the reference standard, respectively (eTable 2). However, when the preceding case had more severe disease, a higher percentage of index case interpretations were more severe than the reference diagnosis compared to when the preceding case had a less severe diagnosis (eTable 2). For example, when the preceding case was DCIS or invasive carcinoma, 15.9% (487 of 3,069 event pairs) of index cases were given a higher diagnostic rating than the reference diagnosis compared to 12.7% (391 of 3,069 event pairs) of index cases given a lower diagnostic rating than the reference ($p<0.001$). The converse also held true when the preceding case was less severe (e.g., not DCIS or invasive).

Based on the significance of the score chi-square statistic, the proportional odds assumption for equal slopes failed in models that included as a covariate the anchor diagnosis used to calculate the dependent variable. The proportional odds model assumption, however, held for all models when adjustment did not include the anchor diagnosis. Graphical assessment of the proportional odds assumption revealed that all primary independent variables met the test for parallelism with the exception of invasive disease where visually the slopes of the two cumulative logit curves appear to cross (eFigure 1a, 1b, 1c). In addition, our results from the method of dichotomizing the ordinal dependent variables at the upper and lower cut points supported our findings.

The odds of pathologists' moving to a more severe interpretation than the reference diagnosis on the index case was greater than one for all definitions of preceding case severity except when the previous case was benign without atypia only, and all associations were statistically significant except when preceding case severity included invasive carcinoma only (Figure 3). For example, the odds of moving to a more severe index diagnosis was 1.28 times greater when the preceding case was diagnosed as either DCIS or invasive carcinoma (proportional odds ratio, POR 1.28, 95% confidence interval (CI), 1.15-1.42, P value <0.001) than if the preceding case diagnosis was either benign without atypia or atypia (reference). When the preceding case included benign without atypia, pathologists were 23% less likely to provide a more severe diagnosis of the index case (POR

0.77, 95% CI, 0.69-0.86, P value <0.001) than if the preceding case was diagnosed as either atypia, DCIS, or invasive disease (reference). A similar effect was noted when each distinct diagnostic class of the immediately preceding cases was entered in a multivariable model (P value <0.001).

When considering the five preceding cases, as the number of these cases diagnosed as DCIS or invasive carcinoma increased, the diagnosis of the subsequent index cases shifted to a more severe category relative to the reference diagnosis. Table 2 shows results for Phase I. For example, when a pathologist diagnosed at least 3 of the 5 preceding cases as DCIS or invasive disease in Phase I, the proportional odds that the index case would be diagnosed as more severe than the reference diagnosis was 1.78 times greater (POR, 1.78, 95% CIs 1.32-2.41; P value <0.001) than the proportional odds when the 5 preceding cases had no DCIS or invasive carcinoma (reference). Phase II data produced either slightly attenuated estimates when considering the influence of severity in the immediately preceding case (eTable 3) and inconsistent results in the effect of proportional severity in the preceding five cases on change between diagnostic ratings (eTable 4).

Data for the participants who interpreted identical cases in both Phase I and II are shown in Table 3. Participants were 17% more likely to shift their diagnosis of the index case in Phase II to a more severe diagnosis relative to their diagnosis of the identical case in Phase I when the severity of the preceding cases in Phase II shifted to a higher diagnosis (POR 1.17, 95% CI, 1.05-1.31, P value = 0.009).

Discussion

This is the first study, to our knowledge, to examine the influence of immediately viewed cases on pathologist decision making for subsequent index case interpretations. We found that pathologists were more likely to over-interpret an index case when the preceding cases were invasive carcinoma or ductal carcinoma in situ (DCIS). Similarly, pathologists were more likely to under-interpret an index case when the preceding cases were benign. We noted similar effects when considering the diagnostic severity of the 5 cases preceding the index case. We also noted similar effects when we analyzed data from pathologists who interpreted the same index case in two different interpretive rounds; the pathologists were more likely to indicate a more severe diagnosis for the same case if it was preceded by a case with higher disease severity, e.g., DCIS or invasive carcinoma than if preceded by a benign case.

Our findings are consistent with a small study among six radiologists¹ who, when interpreting a test set of arteriograms that contained more cases of pulmonary embolism than would have occurred in clinical practice, shifted more of their diagnoses toward higher suspicion compared to a group of radiologists who interpreted a test set with many fewer cases of pulmonary embolism. While this previous study considered the overall prevalence of disease severity within a testing situation, our study was able to evaluate the impact of the immediately preceding case or cases.

We noted that the influence of preceding cases lessened and the influence of proportional severity on direction of change from the reference diagnosis was reversed during Phase II. It may be that pathologists were influenced by participating in Phase I and adjusted their expectations for the prevalence of disease in the test set. It is also possible that, even though participants were not told that they were going to be shown the same test cases in Phase II, they may have recalled the diagnosis they made during Phase I. Even though the participants were not informed about the prevalence of risk-associated lesions in the test sets, it is also likely that many participants recognized that the test sets had more DCIS and atypia than would be expected for 60 randomly selected breast biopsies.

The present results contribute to the judgment and decision making literature in three primary ways. First, we extend prior empirical research to a naturalistic domain that is both exceedingly complex and carries high-stakes outcomes for treatment, prognosis, and patient quality of life. This complements extant research with radiologists, nurse practitioners, and cytopathologists interpreting cervical smears,^{11,12,20} bridging the gap between traditional laboratory judgment and decision making research and relatively real-world tasks and demands. Second, we support converging evidence from perceptual psychophysics²¹ and judgment and decision-making,¹¹ demonstrating the prevalence of assimilation in biasing human judgment. Indeed pathologists showed higher diagnostic classifications following relatively severe (DCIS, invasive) cases, and lower classifications following benign cases, providing strong bidirectional support for assimilation influences on medical decision making. Finally, we also demonstrate that expert pathologists are not invulnerable to the influence of assimilation during diagnostic interpretation, highlighting the potential scope and clinical impact of our findings.

In clinical practice, very subtle differences often exist between diagnostic classification of medical data. Within the field of breast pathology, the classification of benign non-atypical diagnoses and atypical ductal hyperplasia (ADH) and between ADH and minimal extent low grade DCIS, can be challenging. Since the diagnostic differences are subtle, it is possible that the findings on recently reviewed cases may influence the diagnosis on a subsequent index case, especially when a clinician is “on the fence.” It is also true that pathologists worry about malpractice as part of their interpretations, as was recently reported in a study that included the same pathologists as the current study.²² In that study, pathologists reported ordering additional stains, recommending additional surgical sampling, obtaining second reviews, or choosing the more severe diagnosis for borderline cases because of their concerns about malpractice. Such concerns could arise on cases subsequent to those that follow a particularly challenging case. As benign disease is the most common diagnosis for a breast biopsy, our increased prevalence of atypia and DCIS relative to the population prevalence may have enhanced our ability to detect bias in our study.

A strength of this study is that it included a high number of pathologists (115 in Phase I and 86 in Phase II) from diverse clinical practices. Another strength is the intra-observer matched design data, with 49 pathologists interpreting the same cases in both phases. Our methods of regressing “change on change” for these 49 pathologists’ data therefore controlled for internal and external time-invariant confounders.²³

Limitations of our study need to be considered. While our results have applicability to testing situations, translation of our findings to clinical practice is unclear. Pathologists interpreted only one slide per case within a testing situation enhanced with a higher overall prevalence of biopsies with atypia and DCIS. With a higher prevalence of benign cases in clinical practice, fewer opportunities exist for severe cases to influence subsequent case diagnoses than the converse. Pathologists also share cases and consult one another prior to issuing a final clinical diagnosis. The consultant pathologist would not have experienced the same contextual bias and their second opinion may moderate in clinical practice the effects we observed in our study situation. Special stains and additional sections, frequently obtained in clinical practice, provide adjunct diagnostic information and introduce a temporal delay that may additionally alter context.

As participants were able to interpret these test cases at times convenient to their own schedules, it is possible that there were starts and stops in timing of interpretations. The date and time stamps recorded for each case in our web-based data collection system show that in Phase I, 96% of pathologists conducted their assessments without extensive time-delays between preceding case and subsequent index cases, i.e., 4% of interpretations had delays that exceeded 30 minutes (data not shown). It was felt that including all unit pairs regardless of the time-delays would, in part, lend itself to the day-to-day reality of biopsy interpretations experienced by pathologists.

In conclusion, cognitive biases may play a role in pathologists' review and diagnoses of biopsy cases in a testing situation. It is unclear if these findings would apply to actual clinical practice, but if they do, understanding when, how and why these sources of diagnostic bias occur in the medical decision making process may lead to improved accuracy and better patient care.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research reported in this publication was entirely supported by the National Cancer Institute of the National Institutes of Health under award numbers R01 CA140560, U54 CA163303, R01 CA172343, and KO5 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 funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. 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. Eggleston TK, Feinstein AR. Context bias. A problem in diagnostic radiology. *JAMA*. 1996; 276(21): 1752–1755. [PubMed: 8940325]
2. Page DL, Rogers LW. Combined histologic and cytologic criteria for the diagnosis of mammary atypical ductal hyperplasia. *Human pathology*. 1992; 23(10):1095–1097. [PubMed: 1328030]
3. Harris, Jay, R.; Lippman, ME.; Osborne, K.; Morrow, M. *Diseases of the Breast*. Lippincott Williams & Wilkins; 2012.

4. O'Malley, FP.; Pinder, SE. Breast Pathology. Elsevier; 2006.
5. Ricci S, Celani M, Righetti E. Development of clinical guidelines: methodological and practical issues. *Neurological Sciences*. 2006; 27(3):s228–s230. [PubMed: 16752054]
6. Adolphs R. The neurobiology of social cognition. *Curr Opin Neurobiol*. 2001; 11(2):231–239. [PubMed: 11301245]
7. Pronin E. Perception and misperception of bias in human judgment. *Trends Cogn Sci*. 2007; 11(1): 37–43. [PubMed: 17129749]
8. Gorini A, Pravettoni G. An overview on cognitive aspects implicated in medical decisions. *European journal of internal medicine*. 2011; 22(6):547–553. [PubMed: 22075278]
9. Matthews WJ, Stewart N. Psychophysics and the judgment of price: Judging complex objects on a non-physical dimension elicits sequential effects like those in perceptual tasks. *Judgment and Decision Making*. 2009; 4(1):64–81.
10. Vlaev I, Chater N. Context effects in games: Local versus global sequential effects on choice in the prisoner's dilemma game. *Judgment and Decision Making*. 2007; 2(5):380–389.
11. Beckstead JW. Modeling sequential context effects in judgment analysis: A time series approach. *Judgment and Decision Making*. 2008; 3(7):570–584.
12. Beckstead JW, Boutis K, Pecaric MR, Pusic MV. Stimulus Sequence Features Influence Physicians' Response Tendencies in Radiological Image Interpretation. *Applied Cognitive Psychology*. 2013; 27(5):625–632.
13. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science*. 1974; 185(4157):1124–1131. [PubMed: 17835457]
14. 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(1):3. [PubMed: 23379630]
15. 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]
16. Elmore JG, Longton G, Carney PA, et al. Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens. *JAMA*. 2015; 313(11):1122–1132. [PubMed: 25781441]
17. Harrell, J. Springer Series in Statistics. New York: Springer-Verlag; 2001. FE 2001. Regression Modeling Strategies.
18. Bender R, Grouven U. Using binary logistic regression models for ordinal data with non-proportional odds. *Journal of clinical epidemiology*. 1998; 51(10):809–816. [PubMed: 9762873]
19. Scott SC, Goldberg MS, Mayo NE. Statistical assessment of ordinal outcomes in comparative studies. *Journal of clinical epidemiology*. 1997; 50(1):45–55. [PubMed: 9048689]
20. Laming D. Screening cervical smears. *British Journal of Psychology*. 1995; 86(4):507–516. [PubMed: 8542199]
21. DeCarlo LT, Cross DV. Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General*. 1990; 119(4):375.
22. Reisch LM, Carney P, Oster N, et al. Medical Malpractice Concerns and Defensive Medicine: A Nation-wide Survey of Breast Pathologists. *AJCP*. 2015 In press.
23. Shahar E. Evaluating the effect of change on change: a different viewpoint. *Journal of evaluation in clinical practice*. 2009; 15(1):204–207. [PubMed: 19239603]

Figure 1.

Figure 2.

Figure 3.

The odds of moving to a more or less severe interpretation than the reference diagnosis on the index case based on severity of the preceding case Proportional odds ratios (POR) and 95% confidence intervals (CIs). All models adjusted for 'anchor' diagnosis (reference diagnosis) of index case. *Severity of diagnosis is determined by the participant. [†]Each indicator variable is entered as a single variable in the model. [‡]Multiple indicator variables are entered into the model with the exception of benign (reference).

Table 1

Characteristics of participating pathologists, by study phase in which they interpreted breast biopsy test cases.

Participant Characteristics	Participants Phase I n (%)	Participants Phase II n (%)	Participants Both Phases n (%)
Total	115 (100%)	86 (100%)	49 (100%)
Demographics			
Age at Survey (years)			
30–39	16 (14)	12 (14)	7 (14)
40–49	41 (36)	30 (35)	20 (41)
50–59	42 (37)	29 (34)	16 (33)
60+	16 (14)	15 (17)	6 (12)
Gender			
Men	69 (60)	54 (63)	27 (55)
Women	46 (40)	32 (37)	22 (45)
Clinical Practice and Breast Pathology Expertise			
Fellowship training in breast pathology			
No	109 (95)	82 (95)	48 (98)
Yes	6 (5)	4 (5)	1 (2)
Affiliation with academic medical center			
No	87 (76)	55 (64)	34 (69)
Yes, adjunct/affiliated	17 (15)	20 (23)	8 (16)
Yes, primary appointment	11 (10)	11 (13)	7 (14)
Years interpreting breast pathology cases (not including residency/fellowship training)			
< 5	22 (19)	13 (15)	7 (14)
5–9	23 (20)	15 (17)	11 (22)
10–19	34 (30)	28 (33)	15 (31)
20	36 (31)	30 (35)	16 (33)
Considered an expert in breast pathology by colleagues			
No	90 (78)	67 (78)	38 (78)
Yes	25 (22)	19 (22)	11 (22)
Test Set			
Randomization			
A	29 (25)	22 (26)	12 (24)
B	27 (23)	19 (22)	12 (24)
C	30 (26)	24 (28)	12 (24)
D	29 (25)	21 (24)	13 (27)

Shifts in interpretation of the index case based on the proportional severity of diagnoses recorded for the preceding 5 cases (Phase I)

Table 2

Number of severe diagnoses recorded for 5 cases Immediately Preceding Index Case *	Total Assessments n (col %)	Direction of Pathologist's Classification of the Index Case Compared to Reference Standard			Odds of using a more severe diagnosis on Index Case Relative to Reference Standard †	
		Less severe diagnoses than reference standard	Agree with reference standard	More severe diagnosis than reference standard	Proportional Logit Model	P value
		n (row %)				
n	6325 (100.0)	889 (14.1)	4528 (71.6)	908 (14.4)		
Number of Preceding Cases with either DCIS ‡ or Invasive Carcinoma recorded for diagnosis						
0 of 5 (0%)	325 (5.1)	70 (21.5)	214 (65.8)	41 (12.6)	1.0 (ref.)	<0.001
1 of 5 (20%)	1315 (20.8)	212 (16.1)	970 (73.8)	133 (10.1)	1.13 (0.85, 1.52)	
2 of 5 (40%)	2060 (32.6)	297 (14.4)	1470 (71.4)	293 (14.2)	1.42 (1.06, 1.91)	
3 of 5	2625 (41.5)	310 (11.8)	1874 (71.4)	441 (16.8)	1.78 (1.32, 2.41)	

Note: POR proportional odds ratio

* Each unit pair consists of six total assessments, five assessments define the first event (exposure) and one subsequent assessment defines the second event (outcome).

† Covariates included in the model were the primary indicator of interest, proportional severity, in addition to covariate adjustment for reference diagnosis of the index case.

‡ DCIS includes high grade DCIS and low grade DCIS

Table 3

Intra-observer data demonstrating the relationship between change in diagnostic severity of the preceding cases and the change in diagnostic severity of the same index case between Phase I and Phase II

Severity of Pathologist's Preceding Case Diagnosis in Phase II Compared to Severity of Preceding Case Diagnosis in Phase I *	Directional Shift of Pathologist's Index Case Diagnosis in Phase II Compared to their Diagnosis for the Identical Case in Phase I			Odds of Upward Shift of Index Case Diagnosis in Phase II relative to Phase I Diagnosis †	
	Downward Shift	No change in diagnosis	Upward Shift	POR (95% CI)	P Value
n	2842 (100.0)	2115 (74.4)	353 (12.4)	1.17 (1.05, 1.31)	0.009
Phase II case less severe than Phase I case	1090 (38.4)	816 (74.9)	111 (10.2)		
Phase II case as severe as Phase I case	711 (25.0)	530 (74.5)	93 (13.1)		
Phase II case more severe than Phase I case	1041 (36.6)	769 (73.9)	149 (14.3)		

Note: POR proportional odds ratio

* Cases are reordered randomly in Phase II

† Assuming the effect of moving from "Phase II case more severe than Phase I case" to "Phase II case as severe as Phase I case" is the same as that from "Phase II case as severe as Phase I case" to "Phase II case more severe than Phase I case", i.e., exposure is treated as linear. Covariates included in the model were the primary indicator of interest, change in severity, in addition to independent adjustment for participant's diagnosis of the Phase I index case.