Amended Reports

Development and Validation of a Taxonomy of Defects

Frederick A. Meier, MD, CM,[1] Richard J. Zarbo, MD, DMD,[1] Ruan C. Varney, CT,[1] Mona Bonsal, MD,[1] Daniel S. Schultz, MD,[1] Colleen M. Vrbin,[2] Dana M. Grzybicki, MD, PhD,[3] and Stephen S. Raab, MD[3]

Key Words: Amended reports; Defects; Pathology specimens; Pathology reports; Misidentification; Misinterpretation; Pathology errors

DOI: 10.1309/9UPELFVQU5WLCUHX

Abstract

Amended pathology reports produce rework, confusion, and distrust. To develop a reproducible amendment taxonomy we derived a classification from 141 amended reports, then validated it with 130 new cases before 4 observers independently reviewed 430 cases measuring agreement (κ). Next, agreement in classifying 30 other amended reports in 7 institutions was measured. We further tracked amendment rates, defect categories, defect discoverers, and discovery mechanisms.

In the 430-case validation set agreement was excellent (κ = 0.8780 [range, 0.8416-0.9144]). Among the 7 institutions, agreement was good (κ = 0.6235 [range, 0.3105-0.8975]). Amendment rates ranged from 2.6 to 4.8 per 1,000 reports. Misinterpretation fractions varied least (23%-29%). Misidentification fractions ranged more widely (20%-38%). Specimen defects were least frequent (4%-10%) and report defects most frequent (29%-48%). Misidentifications and report defects inversely correlated. Pathologists discovered most misinterpretations, and clinicians found most misidentifications. Conference review revealed 40% to 80% of misinterpretations. This taxonomy produced excellent reproducibility and good agreement across institutions.

Diagnostic anatomic pathology turns tissues and cells into information. Need for information from specimens drives the diagnostic process. The sum of implicit information available in submitted samples constrains it. Surgical pathology and cytopathology reports record the information actually extracted from the specimens.[1] Amendments of pathology reports document defects in the diagnostic process that lose information or add misinformation.

Comparisons of unwanted variation in diagnostic anatomic pathology can be inadequate for 3 reasons. First, regarding 2 monitors of unwanted variation that are mandated,[2,3] criteria used to evaluate discrepancies in frozen section/permanent section diagnoses[4-10] and cytopathologic-histopathologic correlations[11-16] are inconsistent from one institution to another.[17] The inconsistency makes it difficult to regard as indices of relative quality different practitioners' or institutions' frozen section/permanent section correlation rates or different institutions' cytologic-histologic diagnostic agreement rates.[18-21]

Second, defect detection may be active or passive. Examples of active monitors are the double reading of a fraction of all cytology slides (an exercise mandated by Federal regulations),[22-29] double review of surgical pathology cases that diagnose new malignancies, and consultations required on cases that fall into specific specimen types or that yield specific diagnoses.[30-38] Passive detection methods include discovery of discrepancies in pathology reports by clinician readers or participants at conferences, like tumor boards, which compare and contrast pathologic information with other sources of diagnostic data.[39] At present, it is not at all clear whether active and passive monitors detect similar or different sorts of variation, nor is it clear how much active

and passive methods contribute to quality improvement.[1-4,11,17,21-24]

Third, variables other than the supposed indices of quality being measured may distort comparisons. These other variables include whether review is carried out before or after issuing a report ("sign out") and whether review is systematic, ie, involves all of a particular genre of case or specimen type, or is selective, ie, regards only "positive" or "high-risk" cases. Still other variables involve active reviewers themselves. They may be nonspecialized, eg, local colleagues without special competence in a particular area, or specialized, eg, either intramural colleagues with subspecialty interests or extramural expert consultants. The reviewers may also be individuals (eg, the traditional consultant) or a panel; finally, review may involve one or several institutions.[36-38,40]

These 3 considerations prompted us to focus on amended reports as potential comparative indices of pathology defects.[41,42] Released (published—signed out) reports summarize information extracted from specimens during the diagnostic process. Amendments record loss of information or introduction of misinformation during the process. These defects are mixed in at the beginning or—in our experience more often—added during the process. Amended reports reflect defects in all anatomic pathology activities, so we hypothesized that a consistent method of classification of amendments would help in the study of unwanted variation that troubles the process across its full range of function.

Amendments, strictly defined as *changes,* not additions, to information, that occur *after*, not before, release (publication—signing out) of reports seem to us to index system defects in a quartet of ways. First, they are sources of confusion or "noise" in a system that fail to transmit, muddle, or contradict the intended, accurate message. Second, they sow distrust among receivers of the "signal" sent by the system, raising suspicions that the system's product is likely to be inaccurate in general or in specific contexts. Third, amendments produce rework, as they are created and transmitted and as their reception is documented. Finally, amendments suggest a distribution of defects, of unwanted variation, within the anatomic pathology process. In this context, an explicit and validated classification, a taxonomy, of amended report defects, that is reasonably complete and reproducible, may serve as a draft for a more comprehensive classification of anatomic pathology defects across the entire system.

When summarizing the background for this investigation, we found that past studies of pathology defects have been assembled from incommensurate sources. The completeness and consistency of classifications that use such data seemed questionable.[3,4,11,14,17-24] In particular, they failed to define ranges of events or domains that they studied, or they focused on a narrow spectrum of function. Finally, they did not apply explicit and validated classifications that could be used reproducibly across different practice settings. The effort reported herein looks at a system-wide domain, amended reports; uses a consistent classification; produces commensurate results, defect rates, and fractions; and applies them in intradepartmental and interdepartmental contexts, without active surveillance methods or specialized reviewers.

Materials and Methods

Standard Terms

During the study, standard terms for alterations to anatomic pathology reports were used. The previous nonstandard terminology had included 6 not specifically defined categories: addendum, addition, amendment, correction, revision, and supplement. This variety was replaced by a 2-category regimen: the term *amendment* applied to all changes that were not purely additions of the information to the case; the designation *addendum* was reserved for reports that purely added information to the case without alteration of previously reported material. The other terms (addition, correction, revision, and supplement) were removed from use.

Study-Categorize-Apply-Revise-Test Derivation Method

First, the 4 senior authors (F.A.M., R.J.Z., R.C.V., and M.B.) studied 141 amended reports in a derivation set, a collection of examples used to develop an initial classification. Second, by grouping similar amendments, they divided defects that led to amendments into 4 categories: misinterpretations, misidentifications, specimen defects, and report defects. Next, they applied these categories prospectively and independently to a training set, a collection of examples used to improve a proposed classification, of 131 amendments.

The reviewers aimed to classify all amendments in one and only one category. They examined all circumstances in which a report reviewer was unable to classify the reason for an amendment into 1 of the 4 categories. They also examined all instances in which independent observers classified a defect into different categories. As a result of the ensuing case discussions, definitions of defect categories were specified and revised. These revised definitions appear in the "Results" section.

After agreeing on the revisions (of study definitions), the reviewers independently assigned the reasons for amendments in a new 430-case validation set. For assessing taxonomic agreement both in this validation set and the later interinstitutional study, we prospectively adopted the following interpretation scheme: a κ statistic less than 2 was "poor"; 2 through 4 was "fair"; more than 4 through 6 was "good"; more than 6 through 8 was "very good"; and more than 8 was "excellent." Classifier agreement in the 430-case validation set is the first κ statistic reported in the "Taxonomic Agreement" section of

DOI: 10.1309/9UPELFVQU5WLCUHX

the "Results" section.

Quality and Other Variables

During the years of the development and validation of the taxonomy, 2 quality measures were tallied in an ongoing manner: (1) *amendment rates,* defined as the annual number of amended reports per reported cases for the years involved (2001-2004), and (2) amended report *error fractions* divided, for the same years, into the defect types of misinterpretations, misidentifications, specimen defects, and report defects. The fractions in these categories were expressed as percentages of all amended reports for that year. We applied the Pearson test to the year-to-year changes in the defect fractions to examine whether increases or decreases in the fractions of specific pairs of defect categories varied in connected ways or not.

In 2004, two other variables were also collected for each amended report: *defect discoverer* and the *mechanism of defect discovery.* The defect discoverers were pathologist, clinician, other staff, or unknown. The discovery mechanisms were pathologist review, review based on new information, conference review, review triggered by clinician call, and unknown mechanism of review. "Pathologist review" involved unprompted discovery of a report defect in the course of routine practice by the initial pathologist or another pathologist in circumstances in which the next 3 mechanisms of review were excluded. The second mechanism, "new information," entailed reviews triggered by information provided from further pathologic specimens or other sources of information, excluding conference review and a specific clinician call. "Conference review" entailed the discovery of defects during the preparation, presentation, or follow-up of information developed around correlation conferences like tumor boards. "Clinician call" indicated defects discovered after a clinician called to express surprise or doubt about a report's content.

Interinstitutional Testing

To assess reproducibility across institutions, 15 unselected amended reports from files of the practice in which the taxonomy had been developed and another 15 from files of a second large hospital practice were deidentified, and these 30 cases were circulated between the 2 institutions and among 5 other practices, 2 in the Northeast, 1 in the Southeast, and 2 in the Midwest. Classifiers all received the same page of classification directions and definitions. The κ statistic was then used to measure the degree of agreement among the participating classifiers; this is the second κ statistic reported in the "Taxonomic Agreement" subsection of the "Results" section.

Results

Defect Categories

The study validated 4 defect categories: misinterpretations, misidentifications, defective specimens, and defective reports.

Misinterpretations divided into 3 *subtypes* that occurred in relation to 2 *levels* of diagnostic information. Misinterpretations in the first subtype were diagnostic conclusions that added inaccurate information—false-positives or overcalls. Those in the second subtype failed to recognize or lost accurate information from the specimen—false-negatives or undercalls. These misinterpretations occurred at primary and secondary *levels* of diagnosis. Primary level diagnoses were those that changed between positive and negative status or between malignant and benign interpretations. Secondary level diagnostic information involved features that affected the clinical context or prognostic implications of a pathologic diagnosis, eg, the grade, stage, state of surgical margins, or lymph node status of specimens resected for malignancy. The third subtype, misclassifications, confused similar diagnostic categories, eg, the names of a soft tissue sarcoma, but these confusions neither added nor subtracted primary diagnostic implications or secondary diagnostic information's modifying impact, eg, the differently labeled sarcoma behaved biologically with the same degree of aggressiveness and same pattern of spread—and was treated the same way—under either name. Misclassifications thus failed to change clinical consequences and so had no impact on either diagnostic level.

Misidentifications, the second defect category, contained 4 subtypes: *patient* identification could be lacking or wrong; *tissue* designation could be faulty, eg, lung confused with liver or stomach confused with colon; *laterality* specification could be reversed, right vs left or left vs right; and *anatomic localization* could be wrong, eg, skin of thigh misidentified as skin of shoulder.

Specimen defects included 5 subtypes: *lost* specimens, specimens with *inadequate* sample *volume* or *size,* samples with *absent* or *discrepant* measurements, samples that were initially affected by *inadequately representative* sampling, and samples that had initially escaped with *absent* or *inappropriate* ancillary studies when the latter were necessary for an accurate diagnosis.

Report defects of 3 subtypes were observed. First, they included *missing* or *erroneous nondiagnostic* information about practitioners involved in the case, specification of the procedure in which the specimen was collected, dates of specimen collection, and codes regarding the patient, procedure, or diagnosis, etc. Second, they covered *dictation* or *transcription slips*, typographical errors in the strict sense. Third, they encompassed *failures* or *aberrations* of *electronic formats* or *transmission* of the information in the reports. Defects in all 3 report subtypes were limited to those that affected *nondiagnostic* information.

Table 1 outlines the classification.

Taxonomic Agreement

DOI: 10.1309/9UPELFVQU5WLCUHX

In the validation set of 430 independently reviewed cases at the institution where the taxonomy was developed, the classifier agreement was "excellent": median κ of 0.8780 with a range of 0.8416 to 0.9144. The median κ for the 7-institution, 30-case comparison was "very good" at a median κ of 0.6235 with a range of 0.3105 to 0.8975. **Table 2** summarizes the 2 κ comparisons.

Quality Variables

Amended Report Rates

Table 3 gives the amended report rates for the 4 years during which the taxonomy was developed and validated in the practice initiating the taxonomy. During that period, the number of amended reports fell slightly from year 1 to year 2 and then rose more appreciably in years 3 and 4. During the same period, the number of surgical cases rose very slightly between year 1 and year 2 and then fell more strikingly during years 3 and 4. The amended report rate fell to a low of 2.6 per 1,000 in year 2 and rose to a high of 4.8 per 1,000 in year 4. This is the range of amended report frequencies against which the taxonomy was developed.

Amended Report Defect Types

Table 4 summarizes the fractions of amended report defects for the same 4 years. Misinterpretations provided the steadiest fraction of defects. They followed the pattern of the amended report frequency over the first 3 years, falling slightly in year 2, rising in year 3, and falling again in year 4 when the amendation rate was highest, but changing in only a narrow 6% range between 23% and 29% of defects during the entire period. Misidentifications were more variable and followed a different pattern. They rose sharply from year 1 to year 2 and remained similar between years 2 and 3, then fell strikingly in year 4, in an 18% range from 20% to 38% of defects. Specimen defects made up the smallest category in all years: "ping-ponging" from 4% in year 1 to 10% in year 2 back to 6% in year 3 and then back up to 9% in year 4, for an overall 6% range from 4% to 10%. The pattern of report defects was strikingly different: report defects accounted for the largest proportion of amended reports in year 1, then fell successively in years 2 and 3, but they rose again, to their highest level, in the last year of the study, for a 20% swing during the 4 years from 28% to 48%.

We also calculated Pearson coefficients of variation to compare year-to-year changes in defect types. Misinterpretations and specimen defects did not change together, either with one another or with misidentifications or report defects. Misidentifications and report defects were, in contrast, according to the Pearson test, linked. Report defects rose as identification defects fell. We suggest these variables were connected in this way because increased attention to reports in the study situation simultaneously caught more prepublication misidentifications, and prevented them, but provoked observation of postpublication report defects that, before the study, had

Table 1
Classification Defects in Amended Pathology Reports

Misinterpretations
Subtypes
False-positives (overcalls)
False-negatives (undercalls)
Misclassifications
Levels
Primary (benign/malignant; positive/negative)
Secondary (grade, stage, margin, and lymph node status)
Misidentifications
Patient
Tissue
Laterality
Anatomic location
Specimen defects
Lost
Inadequate sample volume or size
Absent or discrepant measurements
Nonrepresentative sampling
Absent or inappropriate ancillary studies
Report defects
Missing or wrong nondiagnostic information (re, eg, practitioners, procedures, dates, diagnostic codes)
Dictation or transcription slips (typographical errors in strict sense)
Aberrations of electronic formats or transmission

Table 2
Taxonomic Agreement

No. of Cases	No. of Reviewers	Median κ	κ Range	Rating
430	4	0.8780	0.8416-0.9144	Excellent
30	7	0.6235	0.3105-0.8975	Very good

Table 3
Amended Report Rates by Year

	2001	2002	2003	2004
No. of amended reports	141	131	158	225
No. of surgical cases	50,317	50,398	47,153	46,468
Amended report frequency per 1,000 cases	2.8	2.6	3.4	4.8

DOI: 10.1309/9UPELFVQU5WLCUHX

escaped notice.

Stratifying Variables

Defect Type by Discoverer

Table 5 lays out the distribution of the 4 defect types among the 4 categories of defect discoverers for year 4 of the study. Pathologists discovered misinterpretations and report defects most frequently, clinicians found misidentifications most often, and the discoverers of report defects were usually not known. The small numbers of specimen defects were scattered through the different groups of discoverers.

Defect Type by Discovery Mechanism

Table 6 summarizes the distribution of the defect types by the mechanism of discovery during year 4. Conference review was the most fruitful mechanism for discovering misinterpretations. Clinician calls were most fruitful for detecting misidentifications. Report defects were detected about as often by pathologist review, clinician calls, and conference reviews as in unknown ways. Again, the few specimen defects were scattered among multiple mechanisms.

Usefulness of Conference Review

Table 7 summarizes the fraction of all amended reports detected during conference reviews (usually tumor boards). In particular, it specifies the fraction of misinterpretations discovered in this way. During the 4-year study, confer-

Table 4
Amended Report Defect Types by Year*

	2001†	2002†	2003†	2004
Total defects detected	144	144	165	225
Defect type				
Misinterpretation	36 (25.0)	33 (22.9)	46 (27.9)	53 (24.0)
Misidentification	39 (27.1)	50 (34.7)	63 (38.2)	44 (19.6)
Specimen	6 (4.2)	15 (10.4)	10 (6.1)	20 (8.9)
Report	63 (43.8)	46 (31.9)	46 (27.9)	108 (48.0)

* Data are given as number (percentage).
† For 2001-2003 data, some amended reports contained more than 1 defect (3 in 2001, 13 in 2002, 7 in 2003).

Table 5
Defect Type by Discoverer (Year 4; N = 225)

Discoverer/No. Discovered	Misinterpretation	Misidentification	Specimen Defect	Report Defect
Pathologist/104	39	6	12	47
Clinician/69	5	29	2	33
Other staff/1	0	0	0	1
Unknown/51	9	9	6	27
Type total (% of overall total)	53 (23.6)	44 (19.6)	20 (8.9)	108 (48.0)

Table 6
Defect Type by Discovery Mechanism (Year 4; N = 225)

Discovery Mechanism/No. Discovered	Misinterpretation	Misidentification	Specimen Defect	Report Defect
Pathologist review/34	2	2	1	29
New information/31	12	2	9	8
Conference review/44	26	3	3	12
Clinician call/69	4	29	3	33
Unknown/47	9	8	4	26
Type total (% of overall total)	53 (23.6)	44 (19.6)	20 (8.9)	108 (48.0)

Table 7
Conference Review

Year	2001	2002	2003	2004
No. of amended reports from conference review/total (%) amended reports	22/141 (15.6)	14/131 (10.7)	24/158 (15.2)	44/225 (19.6)
Conference review fraction (%) of misinterpretations	22/36 (61.1)	14/33 (42.4)	24/46 (52.2)	44/53 (83.0)

DOI: 10.1309/9UPELFVQU5WLCUHX

ence review detected a relatively steady 15% to 19% of all amended reports. It accounted, however, for the majority of discoveries of misinterpretations, with similar levels in year 1 and in year 4 (61% and 83%). The lowest level was in year 2 (42%), with just a bare majority in year 3 (52%).

Discussion

This report describes how we used derivation, training, and validation sets to develop, improve, and demonstrate the reproducibility of a classification of amended report defects. The classification then demonstrated excellent interobserver agreement in a prospective validation set at the developing institution and very good agreement in a small set of deidentified examples circulated among 7 institutions, 6 with little training and experience in applying the classification (Table 2).

As observed in the introduction to this article, pathology defects can usefully be thought of as variations in the diagnostic process that add misinformation or lose (or fail to notice) available, accurate information. As we summarized in the introduction, much effort has been spent in the past, indeed mandated by federal regulation and required by accrediting agencies, to ferret out discrepancies that arise within the process, such as cytologic-histologic noncorrelation and frozen section diagnosis–permanent section diagnosis noncorrelation. We also pointed out further attempts to detect mischievous variation by various patterns of double reading of slides (a practice that is also mandated in cytology) or sending cases for consultation.

We argued that such a variety of assessment methods leads to inconsistency, tends to produce incommensurate results, and resists generalization. The study of amended reports, in contrast, can be done with a single approach, consistently applied, to produce comparable results that allows generalization between institutions and over time.

The authors' previous experience in the College of American Pathologists Q-Probes study program also brought them to appreciate a core difficulty in anatomic pathology defect detection, that of separating *quality* variables, comparable differences in the diagnostic process that facilitate or hinder the obtaining of accurate information, from *stratifying* variables, differences in practice circumstances that facilitate or hinder the comparison of variation among local processes. Among these other, stratifying variables, some influence the sort of review that turns up some discrepancies; others influence the kind of reviewer who judges whether a discrepancy is present. In this study, we tried to identify and measure these variables.

This plan of study initially presented a problem of definition. Amendments, as indices of defects, the authors came to understand, are *changes,* not additions to diagnostic reports, produced *after* the reports have been released (signed out), not *before* the report had been issued or published. This definition, we found, had to be applied strictly in a semantic field that included a variety of other terms: addenda, additions, corrections, revisions, and supplements.

As readers of pathology reports have occasion to observe, the alternative designations are sometimes used in self-serving ways. We encountered the following examples. An addendum changed "benign prostate tissue" to "prostatic adenocarcinoma." In our taxonomy, this would be amendment owing to misinterpretation. An addition altered "intradermal nevus" to "eccrine poroma." In the taxonomy, this amendment was due to misidentified patient skin biopsy specimens. A revision adjusted "mild dysplasia (cervical intraepithelial neoplasia [CIN] 1)" to "focal severe squamous dysplasia (CIN 3)." This change is amendment, prompted by obtaining of deeper levels, after a telephone call from a gynecologist questioning the initial report, ie, an amendment for a specimen defect owing to inadequate sampling prompted by a clinician call.

We advocate listing corrections of misinterpretations, misidentifications, and specimen defects like the examples just given, as well as report defects involving nondiagnostic information, under the unequivocal designation of amendments. Based on the experience of developing this taxonomy, we urge that most report changes made after release be recognized as amendments. Only additions of information should be designated *addenda,* and other semantic variations ought to be banished as merely confusing the evaluation of the product of our practice.

In this study, we examined the range of amended report defects (from a derivation set), categorized them, applied the initial categories (in a training set), revised them, and then tested them (in a validation set). This scheme for developing a classification has previously been used successfully, eg, to build clinically useful classifications for assessing the likelihood of the presence or absence of streptococcal pharyngitis in clinical sore throat.[43-49] We found that this scheme produced a reasonably complete and consistent, reproducible classification of amendments.

While developing the taxonomy we also measured 2 quantitative indices of quality, the rates of amendment and the fraction of amendments in each of the 4 categories. We further measured 2 stratifying variables that modulated, in our perception, amendment-causing defects: who found them (the defect discoverer) and how they were found (the mechanism of discovery).

The amended report rates that we found, 2.6 to 4.8/1,000, we now take to be preintervention rates of amendments against which we can measure the impact of various amendment reduction strategies. Regarding the distribution of

DOI: 10.1309/9UPELFVQU5WLCUHX

amended report defect types, we were struck by the relative stability of the contribution from misinterpretation, that ranged from a little less than a fourth to a little less than a third of amended reports. The more prevalent misidentifications and report defects may be the "low hanging fruit" for process improvements to reduce amendment report rates. Specimen defects, in contrast, contributed the smallest fractions from year to year during our study.

Regarding the stratifying variables, defect discoverer and mechanism of discovery, we were interested and pleased to determine that pathologists themselves are the most frequent discovers of defects in interpretations and reports. We were also interested to find anecdotal experience confirmed by documentation that clinicians are the most frequent discoverers of misidentifications. Study of the mechanism of discovery highlighted the important role of conference review (in our institutions, mostly tumor boards) as a steady source of defect detection. Conferences during the period of the study accounted for detection of 10% to 20% of all defects that led to amendments. In particular, conference review was, in all years, the most efficient way to detect interpretation errors, detecting in 3 out of 4 years more than half of the errors in the misinterpretation group and in year 4 more than 80% of misinterpretations.

We are now attempting to apply lessons learned in this study. At Henry Ford Hospital, Detroit, MI, we now use the validated taxonomy as a computerized "amended reports dictionary" in real time, throughout our practice, with one of us (R.C.V.) acting as the controlling editor of the amended procedure. Defects are recorded with documentation specific enough to permit accurate classification of errors and complete enough to provide triggering information for root cause analysis.

At Henry Ford Hospital and Shadyside Hospitals, Pittsburgh, PA, we adopted Toyota Production System principles of process improvement to reduce to the point of elimination the systematic sources of errors that require report amendment. As we began this process, in the postvalidation year (2005) we saw report defect rates rise sharply. At Henry Ford Hospital, this rise was also associated with the introduction of a new anatomic pathology computer system that had the untoward effect of distributing preparation of final reports among 16 pathologists. Previously, 4 transcriptionists had prepared the final reports. Our ongoing examination of defect discovery also stresses the role that clinicians, their office staff members, accessioning clerks, pathology assistants, and residents performing gross examination of specimens all have in avoiding and routing out patient and specimen misidentifications. Efforts to prevent these defects have moved upstream to specimen collection, transport accessioning, and gross examination steps in the testing process. Changes in processes at each of these steps have been necessary to drive down the rate of misidentification.

In the surgical pathology division at Henry Ford Hospital, interpretation and specimen defects have been the focus of senior staff conferences. In this venue, interpretive nonreproducible diagnostic distinctions, distracting histologic backgrounds, contrasting histologic vs cytologic emphases in interpretation, presence of critical histologic appearances at central vs peripheral locations on slides, and the impact of differing knowledge of clinical context recur as themes in interpretive discrepancies. None of these sources of misinterpretation appears amenable to simple interventions. Their definition provides a jumping-off point for a long expedition toward improvement.

Conclusion

A validated taxonomy of amended report defects in anatomic pathology divides these defects into misinterpretations, misidentifications, specimen defects, and report defects. With practice, independent observers can apply this taxonomy with excellent agreement within a single large practice. Very good agreement was obtained across multiple institutions, without much experience with the taxonomy. The latter finding suggests that the classification is also robust to interpractice differences. For the taxonomy to work, an important prerequisite is strict adherence to the definition of amendments vs addenda with the exclusion of other categories of report changes. Besides amended report rates, the distribution of amendments among the 4 defect categories may be of practical value in planning and assessing improvement initiatives. The study also suggests that who discovers the defect and how the defect is discovered may also be useful, practice-specific stratifying variables. Our findings add to the consensus regarding the value of conference review for detecting especially misinterpretation defects in anatomic pathology reports.

The taxonomy may be of further research value. We are currently testing its usefulness in the study of discrepancies in anatomic pathology—frozen section vs permanent section noncorrelation and cytologic-histologic noncorrelation. In the original domain of amended reports, we have also linked the classification of these defects to measures of their clinical influence, attempting to develop a harm severity scale. In this effort, consistent classification will serve as the foundation for reproducible measurement of unintended variation.

From the [1]Department of Pathology and Laboratory Medicine, Henry Ford Hospital, Detroit, MI; [2]Pathology Department, UPMC Shadyside, Pittsburgh, PA; and [3]Department of Pathology, University of Colorado, Anschutz Medical Campus, Aurora.

Supported by grant HS 13321-01 from the Agency For Healthcare Research and Quality, Rockville, MD.

DOI: 10.1309/9UPELFVQU5WLCUHX

Address reprint requests to Dr Meier: Dept of Pathology and Laboratory Medicine, Henry Ford Hospital, 2799 W Grand Blvd, Detroit, MI 48202.

References

1. Nakhleh RE. What is quality in surgical pathology? *J Clin Pathol.* 2006;59:669-672.
2. Nakhleh R, Coffin C, Cooper K; for the Association of Directors of Anatomic and Surgical Pathology. Recommendations for quality assurance and improvement in surgical and autopsy pathology. *Am J Clin Pathol.* 2006;126:337-340.
3. Raab SS, Nakhleh RE, Ruby SG. Patient safety in anatomic pathology: measuring discrepancy frequencies and causes. *Arch Pathol Lab Med.* 2005;129:459-466.
4. Raab SS, Tworek JA, Souers R, et al. The value of monitoring frozen section–permanent section correlation over time. *Arch Pathol Lab Med.* 2006;130:337-342.
5. Coffin CM, Spilker K, Zhou H, et al. Frozen section diagnosis in pediatric surgical pathology: a decade's experience in a children's hospital. *Arch Pathol Lab Med.* 2005;129:1619-1625.
6. Kucera E, Kainz C, Reinthaller A, et al. Accuracy of intraoperative frozen-section diagnosis in stage 1 endometrial adenocarcinoma. *Gynecol Obstet Invest.* 2000;49:62-66.
7. Houck K, Nikrui N, Duska L, et al. Borderline tumors of the ovary: correlation of frozen and permanent histopathologic diagnosis. *Obstet Gynecol.* 2000;95:839-843.
8. Udelsmann R, Westra WH, Donovan PI, et al. Randomized prospective evaluation of frozen-section analysis for follicular neoplasms of the thyroid. *Ann Surg.* 2001;233:716-722.
9. Tanis PJ, Boon RPA, Koops HS, et al. Frozen section investigation of the sentinel node in malignant melanoma and breast cancer. *Ann Surg Oncol.* 2001;8:222-226.
10. Maia DM. The reliability of frozen-section diagnosis in the pathologic evaluation of Hirschsprung's disease. *Am J Surg Pathol.* 2000;24:1675-1677.
11. Vrbin CM, Grzybicki DM, Zaleski MS, et al. Variability in cytologic-histologic correlation practices and implications for patient safety. *Arch Pathol Lab Med.* 2005;129:893-898.
12. Chute DJ, Covell J, Pambuccian SE, et al. Cytologic-histologic correlation of screening and diagnostic Papanicolaou tests. *Diagn Cytopathol.* 2006;34:503-506.
13. Benedet JL, Matisec JP, Bertrand MA. An analysis of 84,244 patients from the British Columbia cytology-colposcopy program. *Gynecol Oncol.* 2004;92:127-134.
14. Ohori NP, Schoedel KE, Rajendram S. Cytologic-histologic correlation of nongynecologic cytopathology cases: separation of determinate from indeterminate cytologic diagnosis for analysis and monitoring of laboratory performance. *Diagn Cytopathol.* 2003;28:127-134.
15. Higgins RV, Matkins JF, Marroum MC. Comparison of fine-needle aspiration cytology findings of ovarian cysts with ovarian histologic findings. *Am J Obstet Gynecol.* 1999;180:550-553.
16. Ylagan LR, Farkas T, Dehner LP. Fine-needle aspiration of the thyroid: a cytohistologic correlation and study of discrepant cases. *Thyroid.* 2004;14:35-41.
17. Renshaw AA. Comparing methods to measure error in gynecologic cytology and surgical pathology. *Arch Pathol Lab Med.* 2006;130:626-629.
18. Adams KC, Absher KJ, Brill YM, et al. Reproducibility of subclassification of squamous intraepithelial lesions: conventional vs ThinPrep Paps. *J Low Genit Tract Dis.* 2003;7:203-208.
19. Stoler MH, Schiffman M. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA.* 2001;285:1500-1505.
20. Yeoh GPS, Chan KW. The accuracy of Papanicolaou smear predictions: cytohistological correlations of 283 cases. *Hong Kong Med J.* 1997;3:373-376.
21. Raab SS. Improving patient safety through quality assurance. *Arch Pathol Lab Med.* 2006;130:633-637.
22. Renshaw AA. Rescreening in cervical cytology for quality control: when bad data is worse than no data or what works, what doesn't, and why. *Clin Lab Med.* 2003;23:695-708.
23. Zarbo RJ, Meier FA, Raab SS. Error detection in anatomic pathology. *Arch Pathol Lab Med.* 2005;129:1237-1245.
24. Frable WJ. Surgical pathology: second reviews, institutional reviews, audits, and correlations: what's out there? error or diagnostic variation? *Arch Pathol Lab Med.* 2006;130:620-625.
25. Raab SS, Stone CH, Jensen CS, et al. Double slide viewing as a cytology quality improvement initiative. *Am J Clin Pathol.* 2006;125:526-533.
26. Whigham P, Ilaro MJM, Flanagan MB, et al. Discrepancy analysis, communication, and feedback for cytotechnologist quality improvement of nongynecologic cytopathology. *Diagn Cytopathol.* 2006;34:265-269.
27. Wright RG, Halford JA, Ditchmen EJ. Detection of false-negative Papanicolaou smears by rapid rescreening in a large routine cervical cytology laboratory. *Pathology.* 1999;31:379-381.
28. Shield PW, Cox NC. The sensitivity of rapid (partial) review of cervical smears. *Cytopathology.* 1998;9:84-92.
29. Nanda K, McCrory DC, Myers ER, et al. Accuracy of the Papanicolaou test in screening for follow-up of cervical cytologic abnormalities: a systematic review. *Ann Intern Med.* 2000;132:810-819.
30. Kronz JD, Westra WH, Epstein JI. Mandatory second opinion surgical pathology at a large referral hospital. *Cancer.* 1999;86:2426-2435.
31. Selman AE, Niemann TH, Fowler JM, et al. Quality assurance of second opinion pathology in gynecologic oncology. *Obstet Gynecol.* 1999;94:302-306.
32. Tsung JS. Institutional pathology consultation. *Am J Surg Pathol.* 2004;28:399-402.
33. Kronz JD, Westra WH. The role of second opinion pathology in the management of lesions of the head and neck. *Curr Opin Otolaryngol Head Neck Surg.* 2005;13:81-84.
34. Staradub VL, Messenger KA, Hao N, et al. Changes in breast cancer therapy because of pathology second opinions. *Ann Surg Oncol.* 2002;9:982-987.
35. Novis DA. Routine review of surgical pathology cases as a method by which to reduce diagnostic errors in a community hospital. *Pathol Case Rev.* 2005;10:63-67.
36. Zardawi IM, Bennett G, Jain S, et al. Internal quality assurance of a surgical pathology department in an Australian teaching hospital. *J Clin Pathol.* 1998;51:695-699.
37. Coffin C. Pediatric pathology: pitfalls and strategies for error prevention. *Arch Pathol Lab Med.* 2006;130:610-612.

DOI: 10.1309/9UPELFVQU5WLCUHX

38. Azam M, Nakhleh R. Surgical pathology extradepartmental consultation practices: a College of American Pathologists Q-Probes study of 2746 consultations from 180 laboratories. *Arch Pathol Lab Med.* 2002;126:405-412.
39. Gross GE. The role of tumor board in a community hospital. *CA Cancer J Clin.* 1987;37:88-92.
40. Gupta D, Layfield LJ. Prevalence of inter-institutional anatomic pathology slide review: a survey of current practice. *Am J Surg Pathol.* 2000;24:280-284.
41. Stastny JF, Geisinger KR, Michael CW, et al. Another quality assurance issue: amended reports: what do we really know about them? *Diagn Cytopathol.* 1998;18:67-70.
42. Nakhleh RE, Zarbo RJ. Amended reports in surgical pathology and implications for diagnostic error detection and avoidance: a College of American Pathologists Q-Probes study of 1,667,547 accessioned cases in 359 laboratories. *Arch Pathol Lab Med.* 1998;122:303-309.
43. Centor RM, Witherspoon JM, Dalton HP, et al. The diagnosis of strep throat in adults in the emergency room. *Med Decis Making.* 1981;1:239-246.
44. Wigton RS, Connor JL, Centor RM. Transportability of a decision rule for the diagnosis of streptococcal pharyngitis. *Ann Intern Med.* 1986;146:81-83.
45. Poses RM, Cebal RD, Collins M, et al. The importance of disease prevalence in transporting clinical prediction rules: the case of streptococcal pharyngitis. *Arch Intern Med.* 1986;105:586-591.
46. McIsaac WJ, Goel V, To T, et al. The validity of a sore throat score in family practice. *CMAJ.* 2000;163:811-815.
47. McIsaac WJ, Kellner JD, Aufricht P, et al. Empirical validation of guidelines for the management of pharyngitis in children and adults. *JAMA.* 2004;291:1587-1595.
48. Gerber MA, Shulman ST. Rapid diagnosis of pharyngitis caused by group A streptococci. *Clin Microbiol Rev.* 2004;17:571-580.
49. Centor RM, Allison JJ, Cohen SJ. Pharyngitis management: defining the controversy. *J Gen Intern Med.* 2007;22:127-130.

DOI: 10.1309/9UPELFVQU5WLCUHX