Quality and Patient Safety
In Anatomic Pathology

Practical Solutions:
Studying Amended Reports
To Evaluate and Improve
Surgical Pathology Processes
Learning Objectives/Disclosure

- Define amended reports
- Classify types of amended reports
- Illustrate amended report monitoring

*No commercial interest in amended reports*
To Evaluate and Improve:

- Characterize amendments
- Relate frequencies & fractions to process
- Monitor process changes to test how frequencies & fractions change
Topics For Discussion

- How should addenda vs. amendments be sorted?
- Do amendments mean something regarding quality vs. safety?
- How should ‘old cases’ be handled when errors are discovered?
Diagnostic Anatomic Pathology

- Turns tissues and cells into information
- Records information from specimens in reports
- Reveals loss of information, addition of misinformation in amended reports
Amendments: Definition

- Changes in information
- Not additions to information
- After, not before release*

*Release = ‘signing out’, publication
Amendments:
Indices of Anatomic Pathology Process Defects

- Sources of confusion ("noise")
- Causes of distrust (among "receivers" of "signal")
- Instances of rework (producing otherwise unnecessary "retransmissions" of "signal")
- Suggestions of distribution of unwanted variation ("noise") in pathology process ("system")
Potential Utility of Amended Report Monitoring

- Observes a system-wide domain: Reports
- Permits consistent classification: Taxonomy
- Generates defect rates and fractions: Quantification
- Assesses interventions: Process change
Problem of Non-Standard Terms For Altered Reports

- Addendum
- Addition
- Amendment
- Correction
- Revision
- Supplement
Standard Terminology

- **Amendment**: All changes that do not only add information
- **Addendum**: Only changes that just add information
- **All other terms** removed from use
Taxonomy Development

- Review of **derivation set**: 141 changed reports
- Division of changed reports into **four categories**:
  - misinterpretations
  - misidentifications
  - specimen defects
  - report defects
- Application of categories to **training set**: 131 changed reports
Categorical Imperatives

- Classification of each defect in one-and-only one category
- Agreement of (4) independent observers on each defect’s class
- Revision of study definitions
- Application of revised categories to validation set: 430 changed reports
- Assessment of classifier agreement using kappa statistic
- Comparison of classifier agreement within vs. among institutions
Inter-institution Comparison

- Thirty unselected amendments
  (15 HFH; 15 UPMC Shadyside)
- Circulated among seven practices
  (3 in NE, 1 in SE, 3 in MW)
- Agreement among classifiers measured using kappa statistic
## Taxonomic Agreement

<table>
<thead>
<tr>
<th>Cases</th>
<th>Reviewers</th>
<th>Median Kappa</th>
<th>Kappa Range</th>
<th>Kappa Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>430</td>
<td>4</td>
<td>0.8780</td>
<td>0.8416-0.9144</td>
<td>excellent</td>
</tr>
<tr>
<td>(HFH)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>7</td>
<td>0.6235</td>
<td>0.3105-0.8975</td>
<td>very good</td>
</tr>
<tr>
<td>(7 Hosp)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Quality Measures

Amendment Rates: Annual number of amendments per annual reported cases

Error Fractions*: Misinterpretation, mis-ID, specimen defects, report defects

*Changes measured by Pierson test
Stratifying Variables

- **Defect Discoverer:** Pathologist, clinician, other staff, unknown.
- **Mechanism of Discovery:** Clinician call, conference review, pathology review, other.
Four Defect Categories

- Misinterpretations (Mis-IP)
- Misidentifications (Mis-ID)
- Specimen Defect (SD)
- Report Defect (RD)
## Misinterpretation: Three Subtypes

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Added Misinformation</td>
<td>False positives; over calls</td>
</tr>
<tr>
<td>Missed, lost Information</td>
<td>False negatives; under calls</td>
</tr>
<tr>
<td>Misclassified Information</td>
<td>Confusion of similar ‘weight’ categories</td>
</tr>
</tbody>
</table>
## Misinterpretation: Two Levels

<table>
<thead>
<tr>
<th>Primary Level</th>
<th>Changes between positive/negative; benign/malignant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondary Level</td>
<td>Changes in grade, stage, margin(s), node status</td>
</tr>
</tbody>
</table>
Misidentification

- Patient
- Tissue
- Laterality
- Anatomic Localization
Specimen Defects

- Lost
- Inadequate size, volume
- Absent, discrepant measurement
- Inadequate representation
- Absent, inappropriate ancillary studies
Report Defects*

- Missing, erroneous **non-diagnostic** information (practitioner, procedure, date, codes)
- Dictation, transcription slips (typographical errors in strict sense)
- Failures or aberrations of e-formats, e-transmissions

*Report defect category is considered only after other three categories are definitely excluded*
## Range of Amended Report Rates Against Which Taxonomy Developed

<table>
<thead>
<tr>
<th>Year</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year One</td>
<td>2.8/1000</td>
</tr>
<tr>
<td>Year Two</td>
<td>2.6/1000</td>
</tr>
<tr>
<td>Year Three</td>
<td>3.4/1000</td>
</tr>
<tr>
<td>Year Four</td>
<td>4.8/1000</td>
</tr>
</tbody>
</table>

*Amended report rates rise with real time monitoring*
Amended Report Defect Types During Taxonomy Development

<table>
<thead>
<tr>
<th>Misinterpretation</th>
<th>Steady fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>23% - 29% of defects</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mis-ID and Report Defects</th>
<th>Linked fractions:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mis-ID fell as RD’s rose*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specimen Defects</th>
<th>‘Ping-pong’ pattern:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-10% of defects</td>
</tr>
</tbody>
</table>

*Hypothesis: Increased attention simultaneously caught more pre-publication mis-IDs more post-publication RDs
<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of Case Reports</td>
<td>46,468</td>
</tr>
<tr>
<td>Total Amendments</td>
<td>225</td>
</tr>
<tr>
<td>Amendment Rate</td>
<td>4.8/1000</td>
</tr>
<tr>
<td>Misinterpretations</td>
<td>53 (24%)</td>
</tr>
<tr>
<td>Mis-IDs</td>
<td>44 (19%)</td>
</tr>
<tr>
<td>Specimen Defects</td>
<td>20 (9%)</td>
</tr>
<tr>
<td>Report Defects</td>
<td>108 (48%)</td>
</tr>
</tbody>
</table>
### Stratifying Variables - ‘Baseline’

<table>
<thead>
<tr>
<th></th>
<th>Defect Discovers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pathologists</strong></td>
<td>Most misinterpretations (74%) Plurality of report defects (44%)</td>
</tr>
<tr>
<td><strong>Clinicians</strong></td>
<td>Most misidentifications (66%)</td>
</tr>
</tbody>
</table>

### Discovery Mechanism

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conference Review</strong></td>
<td>Most misinterpretations (83%)</td>
</tr>
<tr>
<td><strong>Clinician Calls</strong></td>
<td>Most misidentifications (66%)</td>
</tr>
</tbody>
</table>
‘Sands of Pathology’*

Taxonomy’s *intra* institutional agreement ‘excellent’; *inter* institutional agreement ‘good’

Taxonomy must be consistently applied to provide comparable results among sites (and their components) over time.

*A pearl forms around a nidus of irritating sand*
A Gram of Sand (or Salt): The Need For An Independent Editor Using Objective Criteria

Self serving
‘addenda’

Misapplied Criteria

Failure to Find
Root Cause

Without editorial monitoring, amendments tend to default to report defects

“Benign prostate tissue” to “prostatic adenoca”

“Intradermal nevus” to “eccrine poroma” specimen mis-ID not report defect

“Mild dysplasia CIN1” to “Focal severe dyspl. CIN3” not false negative (undercall) but inadequate sampling (specimen defect)
Contrast of ‘Baseline Rates’

Passive Monitoring 2.8/1000

Active Monitoring 4.8/1000
Importance of Conference Reviews

- Detected 10% - 20% of all defects requiring amendment at baseline

- Discovered 50% of mis-interpretations during passive monitoring, 80% of mis-interpretations during active monitoring
Domain: Surgical Pathology Process

- Patient and specimen identification
- Specimen handling
- Diagnostic interpretation
- Result reporting
Measures: Applying Taxonomy

- Total amendments
- Amendments per thousand cases
- Amendment fractions by defect type
Events During Monitoring

- Pathologist production of reports’04
- Amendment process editing began ’05
- Production System introduced ’06
- HFPS fully operational ‘07
## Monitoring Results 1: Defect Rates

<table>
<thead>
<tr>
<th>Year</th>
<th>Amendments</th>
<th>Cases</th>
<th>Amendments/1000 Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>225</td>
<td>46,468</td>
<td>4.8</td>
</tr>
<tr>
<td>2005</td>
<td>475</td>
<td>46,880</td>
<td>10.1</td>
</tr>
<tr>
<td>2006</td>
<td>374</td>
<td>48,010</td>
<td>7.8</td>
</tr>
<tr>
<td>2007</td>
<td>306</td>
<td>48,422</td>
<td>6.3</td>
</tr>
</tbody>
</table>
Observations: Defect Rates

- Introducing edit function doubled defect detection
- Defects fell one-fifth with HFPS introduction
- They fell another 15% with full implementation
### Results II: Monitoring Defect Fractions

<table>
<thead>
<tr>
<th>Year</th>
<th>Misinterp</th>
<th>MisIDs</th>
<th>Spec. Defects</th>
<th>Rep. Defect</th>
<th>Spec. Defects</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>53 (23.5%)</td>
<td>44 (19.5)</td>
<td>20 (9.0)</td>
<td>108 (48)</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>87 (18.3%)</td>
<td>74 (15.6)</td>
<td>9 (1.9)</td>
<td>305 (64.2)</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>59 (15.8%)</td>
<td>46 (12.3)</td>
<td>16 (4.3)</td>
<td>253 (67.6)</td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>22 (7.0%)</td>
<td>38 (12.0)</td>
<td>33 (11.0)</td>
<td>213 (70.0)</td>
<td></td>
</tr>
</tbody>
</table>
Observations: Defect Fractions

- Mis-interpretation fraction fell during monitoring
- Fall steeper after HFPS’s full implementation
- Mis ID fraction fell in first 2 years, then leveled off
- Specimen defects produced few amendments
- Report defects fraction were largest, varied most, increased most
<table>
<thead>
<tr>
<th>Measures</th>
<th>Interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification defects</td>
<td>Goal-directed clinician education</td>
</tr>
<tr>
<td>Specimen defects</td>
<td>Redesign of specimen accession</td>
</tr>
<tr>
<td>Interpretation defects</td>
<td>Double reading- breast, prostate*</td>
</tr>
<tr>
<td></td>
<td>*2% of cases: ~800/47,500</td>
</tr>
</tbody>
</table>
## Monitoring Amendments ’06-’08

<table>
<thead>
<tr>
<th></th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Amendments</strong></td>
<td>374</td>
<td>306</td>
<td>271</td>
</tr>
<tr>
<td><strong>Amendments per 1000 reports</strong></td>
<td>7.8</td>
<td>6.3</td>
<td>5.5</td>
</tr>
</tbody>
</table>
## Fractions of Amendment Subtypes

<table>
<thead>
<tr>
<th>Category</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misinterpretations</td>
<td>16%</td>
<td>7%</td>
<td>3%</td>
</tr>
<tr>
<td>Misidentifications</td>
<td>12%</td>
<td>12%</td>
<td>9%</td>
</tr>
<tr>
<td>Specimen defects</td>
<td>4%</td>
<td>11%</td>
<td>4%</td>
</tr>
<tr>
<td>Report Defects</td>
<td>68%</td>
<td>70%</td>
<td>84%</td>
</tr>
</tbody>
</table>
Focused double-review of a small fraction of cases associated with fall in misinterpretations*

Goal directed clinician education on specimen ID had only modest impact on mis-ID

Accession re-design associated with increase, then decrease in specimen defect fraction

* “Undetected confounding factors cannot be excluded in real practice”
Take Home Messages
From On-Going Monitoring

- Pathologist “finalling” increases report defects
- Editing (introducing systematic classification) increases defect detection
- LEAN initiative (HFPS) reduced defects
- HFPS influence grew with ongoing, widening integration into practice
Interpretation, Identification, Specimen Defects ’07-’09

- Less frequent (16%)
- More worrisome
# Interpretation Defects ’07-’09

<table>
<thead>
<tr>
<th></th>
<th>Year A [’07-’08]</th>
<th>Year B [’08-’09]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amendment rate</td>
<td>7.1/1000</td>
<td>4.8/1000</td>
</tr>
<tr>
<td>Primary Dx</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>Secondary Dx</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Reclassification</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Year A ['07-'08]</td>
<td>Year B ['08-'09]</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Patients</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>Anatomic site</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Laterality</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Tissue</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>29</td>
</tr>
</tbody>
</table>
## Specimen Defects ’07-’09

<table>
<thead>
<tr>
<th></th>
<th>Year A ['07-'08]</th>
<th>Year B ['08-'09]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ancillary Testing</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Nonrepresentativeness</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Other defects</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>6</td>
</tr>
</tbody>
</table>
## Interpretative Defects: Diagnosis Involved ’07-’09

<table>
<thead>
<tr>
<th>Year A</th>
<th>Year B</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI</td>
<td>4</td>
</tr>
<tr>
<td>Skin</td>
<td>4</td>
</tr>
<tr>
<td>Cervix</td>
<td>2</td>
</tr>
<tr>
<td>Prostate</td>
<td>2</td>
</tr>
<tr>
<td>Breast</td>
<td>1</td>
</tr>
<tr>
<td>Lung</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>14</td>
</tr>
</tbody>
</table>

*an additional oral pathology diagnostic revision*
## Detection Mechanism ‘07-’09

<table>
<thead>
<tr>
<th>Review Type</th>
<th>Year A</th>
<th>Year B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinician Requested Review</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Tumor Board Review</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Other Post-Sign-out Review</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>18</strong></td>
<td><strong>6</strong></td>
</tr>
</tbody>
</table>
Observations ’07–’09

- Pre-sign out double review may decrease primary interpretation defects.
- Interpretation defects are most frequent in high volume specimen types [GI, skin, cervix].
- Identification defects are relatively resistant to interventions.
Need for Audit and Editing ‘09

- 2 months’ audit
- 7846 cases reported
- 181 addenda reviewed
- 2.3 addenda/1000 cases
Addenda [ N=181]

- 78 (43%) added GI diagnoses (mostly \textit{H. pylori} status)
- 34 (19%) added GI diagnoses (no reason supplied)
- 25 (14%) added reports of molecular testing
- 19 non-GI added diagnoses (no reason supplied)
- 6 others [non-GI] added diagnoses (reason supplied)
- 10 changed diagnoses*
  *fitting amendment definition
Observations From Audit

- In a substantial fraction of addenda (29%) no reason for addition was supplied.
- Two common causes account for 57% of addenda:
  - H. pylori reports (43%)
  - Molecular pathology reports (14%)
- Five and half percent (10/181) of ‘addenda’ were really amendments.

*There were more of these misclassified diagnostic changes during the audit period than there were reported amendments.*
Conclusions From Audit

- Independent editor of both Amendments and Addenda
- Maintains system integrity
Intervention: Amended Reports Dictionary

- In real time
- Throughout practice
- One controlling editor
- Consistent classification
- Capturing information for root cause analysis
Impact on Process Improvement

- Observed defects rise, then fall
- Untoward increase in report defects
- Multiple observed causes of mis-ID’s: (collection, transport, accessioning, grossing, cutting)
- Possible decrease in mis-IPs
Sources of “Our” Interpretation And Specimen Defects

- Non-reproducible diagnostic distinctions
- Distracting histologic backgrounds
- Contrasting histo- vs. cytologic emphasis
- Central vs. peripheral location of critical diagnostic findings
- Differing knowledge of clinical context
“Get It While You Can”

(Janice Joplin – Nicknamed ‘Pearl’)

- Defects that cause amendments divide into:
  - Misinterpretations  Misidentifications
  - Specimen defects  Report defects
- Excellent taxonomic agreement obtained within a practice, good agreement among practices
- Classification integrity requires real time monitoring and independent editing *
- Defect fractions assess interventions; amendment rates assess success

* Strict adherence to definitions of amendments vs. addenda, elimination of other ‘change’ categories
References

Thanks!

Discussion Topics

- Monitor both addenda and amendments?
- Amended reports measure quality or safety?
- Amend ‘old cases’ when errors discovered?

Collaborators
Richard J. Zarbo, Ruan Varney, Stephen Raab, Colleen Vrbin
CAP Companion Meeting at USCAP 2010
Quality and Patient Safety in Anatomic Pathology: Practical Solutions

Studying Amended Reports to Evaluate and Improve Surgical Pathology Processes

Frederick A. Meier, MD, CM
Henry Ford Health System
Detroit, MI 48202
March 20, 2010

Take Home Messages:

Defects that cause amendments divide into:
* Misinterpretations  * Misidentifications
* Specimen defects  * Report defects

Excellent taxonomic agreement can be obtained within a practice, good agreement among practices

Classification integrity requires real time monitoring and independent editing*

Defect fractions assess interventions; amendment rates assess success

* Strict adherence to definitions of amendments vs. addenda, elimination of other ‘change’ categories

Syllabus

Introduction: In discussions of quality and patient safety in anatomic pathology, amended reports often appear as potential material for evaluating surgical pathology processes. This presentation proposes a consistent definition of amended reports and a classification of the types of amendments. It shows how this taxonomy has been used to monitor amended reports over several years. The presentation also offers some conclusions that have emerged from this monitoring.

The monitoring characterizes amendments into four classes, relates their frequencies to process attributes, and links process changes to these frequencies. The monitoring experience also introduces three topics for discussion: how should the distinction between addenda and amendments be maintained? What, realistically, do amendments tell about the quality of surgical pathology processes; in particular, do they say anything about patient
safety? Finally, how should ‘old cases’ be handled when errors are discovered months or years after the results have been released – should they produce belated amendments?

From the experience that this presentation describes, we do think that amended reports are indeed sources of evaluation and improvement. Diagnostic anatomic pathology (surgical pathology and cytopathology), fundamentally, turns tissues and cells into information. This product is recorded in reports, so amendments to these reports reveal loss of information or addition of misinformation, which is damaged product. In every production system, responsible producers should attend to their product’s defects.

**Definitions:** Amendments are changes in information – not just additional information – that are made after report are released, that is, ‘signed out’ or published, not before.

Information theory presents the case for amendments as indices of process defects. They are sources of confusion that revised previous information: in information theory this confusion is “noise”. Noise breeds distrust among people ("receivers") who wonder how much they should rely on the information (“signal”) that they receive. These “receivers” are presented with the question: should I trust “signal” from this “source”? The answer to this question depends on how often the “source” generates "noise" or transmits through “noise”: this is the “signal to noise ratio”. Beyond the distrust, amendments generate – indeed they are – rework; they otherwise unnecessary “retransmissions of signal”. Finally, in an information system, “noise” patterns may be instructive manifestations of unwanted variation (“interference” in the information theory "system"). In our case, the information system is the diagnostic pathology process.

Reports are also a system-wise domain; all parts of the pathology enterprise, as their raison d’être, produce reports. We hope to show in this presentation, that a valid taxonomy can classify amendments consistently. Such consistent classification, in turn, can be used to generate two kinds of quantified monitoring indices: defect rates and defect fractions. We have tried to use the defect rates to monitor our product and its “signal to noise ratio”. We have tried to use the defect fractions to monitor interventions to decrease defects, that is, interventions that decrease “interference”.

The initial hurdle to get over in the study of amended reports is the ambiguity of nonstandard terms for altered reports: addendum, addition, amendment, correction, and revision – presented here in alphabetical order – were all descriptive designation used in our practice without clear distinctions separating one from another.

We replaced this vague situation with a two-state terminology. Post- release changes sort into one of two alternative bins: Either amendments – all changes except those that only add information, or addenda – only changes that just add information; addenda may in no way
change information. To avoid confusions about the confusion that post-release changes cause - all other terms must disappear.

**Taxonomy Development:** We developed and validated our classification for amended reports in the following way. First, we examined 141 changed reports, collected over a year. Next, we sorted the changed reports into 4 categories: misinterpretations, misidentifications (mis-IDs), specimen defects, and report defects. We then applied these categories to a new set of a different series of 131 changed reports. We did this with two objectives: classify each defect into one-and-only-one category and achieve agreement among 4 independent observers as to the class in which each amendment belonged. To achieve these two objectives we revised the study definitions. With the revised categories in mind, the four independent observers classified another new set of amendments, this time a series of 430 reports not previously classified. To the observers’ classifications in this validation round we applied the kappa statistic for measuring agreement.

At the time we were doing this, our laboratory was participating in a multi-institutional study of unwanted variation in pathology and cytology organized by Doctors Stephen Raab and Dana Grzybicki. This connection, and the help of the study’s capable staff, permitted us to circulate another set of new cases: fifteen from our practice and fifteen from the pathology practice at University of Pittsburgh Medical Center Shadyside. Practitioners from seven practices evaluated these 30 cases – three of the practices were in the Northeast, one in the Southeast, and three in the Midwest. Again we used the kappa statistic for measuring agreement.

In the validation set (N= 430 cases, all from Henry Ford Hospital), the four initial reviewers demonstrated excellent agreement by kappa measurement: a median kappa of 0.8780 (range: 0.8416 – 0.9144) in this prospective test of their classifying consistency. In the circulation set (N = 30 cases – 15 from Ford, 15 from Shadyside) seven reviewers (including one of the original four) showed very good kappa agreement: a median of 0.6235 (range: 0.3108 – 0.8975) across the different practices.

The validation of taxonomic agreement permits us to use of a pair of quality measures that can follow the ebb and flow of amendments over time: amendment rates – annual number of amendments per total number of reported cases – and amendment fractions – the percentages of amendments that fall into misinterpretation, misidentification, specimen defect, and – if no other applies – report defect categories. The interactions of these fractions – whether they vary together or not – can be assessed statistically by the Pierson correlation test.

While they do not track product quality as the sources of "interference" (the frequency and distribution of report defects), two other measures are valuable as well. These two measures are valuable as “stratifying variables” – findings that help in the investigation of amendments’
root causes. The measures are “defect discoverers” and “mechanisms of defect discovery”. There are four sorts of defect discoverers: pathologist, clinician, other staff, and unknown, as well as four mechanisms of discovery: clinician call, conference review, pathology review, and other.

**Taxonomic Classes:** Before presenting what the monitoring of amended reports found, we should summarize the major characteristics of the four categories of misinterpretation, misidentification, specimen defect, and report defect.

**Misinterpretations** come in three subtypes. False-positives or “over-calls” add misinformation. False-negatives or “under-calls” miss or lose information. Sometimes diagnoses are rearranged without “over-call” or “under-call”, that is, on the same plane of importance or implication. Changing the name of, for example, a soft tissue sarcoma from that of one entity to that of another when the two names indicate entities of the same degree of malignancy or aggressive behavior would be an example of ‘misclassified’ but neither false-positive nor false-negative interpretation. These three types of misinterpretation, it turns out, appear on two different ‘levels’ of diagnosis. The primary level is that on which changes appear between positive and negative, benign and malignant. The secondary level is home to modifiers of prognosis like changes in grade (histologic appearance), stage (extent of involvement), margin involvement (or lack of such involvement), and the number and extent of involvement of lymph nodes in sections examined.

**Misidentifications** can also happen at different levels: patients can be confused, as can tissues (e.g. stomach vs. colon biopsies), and the laterality of specimens from paired organs (e.g. right vs. left breast) as well as anatomic location within an organ (e.g. skin of shoulder vs. skin of thigh).

**Specimen defects** come in five kinds. Specimens can be completely lost; they can be too small; their descriptions may lack critical measurements; they may not have been adequately sampled; or critical ancillary studies have not been done or the wrong studies were selected.

**Report defects:** If defects that cause amendments do not fall into one (and only one) of the three categories just described, then the sub-type of these report defects is one of three kinds: (i) missing, erroneous, non-diagnostic information; examples of non-diagnostic changes that stimulate amendments are missing or wrong practitioner, procedure, date, or code (physician, diagnostic, or test codes); (ii) dictation or transcriptions that are typographical errors in the strict sense, (iii) failures or aberrations of electronic (computer) formats or electronic transmissions. Please note that changes in diagnosis – those “above the line” – are, by definition, excluded from this default category.
**Monitoring Experience - Development:** With these classifications in mind, we can follow the behavior of amendments over time.

During the taxonomy’s development, retrospective rates (years 1 and 2) proved to be lower (2.6 – 2.8/1,000) than prospective, “real time” amendment rates (3.4 - 4.8/1,000). We point this out as an important finding: amended report rates in systems in which interventions are yet to be undertaken rise with real time monitoring.

Over the period of the classification’s development, among the four defect categories, misinterpretations made up the most consistent fractions of 23% -29% of defects. The Pierson test found misidentifications and report defects to be reciprocally linked: misidentifications fell as report defects rose. Perhaps, this connection is due – like the rise in “real time” amendment rates - to the increased attention we paid to amendments. Regarding the reciprocal relationships, increased attention may have led to both more pre-publication detection of misidentifications, lowering that rate, and more post-publication awareness of report defects, leading to their amendment. Specimen defects followed a “ping pong” pattern that has continued over all the years that we have followed them: lower one year then higher the next year, then lower again; this lack of a consistent downward trend has been frustrating, but specimen defects make up the smallest of the four fractions, they “bounced” between 4% and 10% during the initial phase of our study.

Before we look at the impact of various interventions on these indices, we should summarize the ‘baseline state’ from which we set out on our ‘quality journey’. At that point, the amendment rate (225/46,468) was 4.8/1000 with 24% (53) misinterpretations, 19% (44) misidentifications, and 9% (20 specimen defects and almost half, 48% (108) report defects.

Regarding the stratifying variables of defect discoverer, at baseline pathologists discovered almost three-quarters (74%) of misinterpretations and 44% of report defects (more than any other discoverer category). Clinicians discovered two-thirds (66%) of misidentifications. Regarding the stratifying variable of defect discovery mechanism, most interpretations (83%) were discovered preparing for, at, or after clinical pathological conferences – a point to which we will return. Most (two-thirds) of misidentifications came to light during or following calls from clinicians.

Also before launching into our account of amended report monitoring with the validated taxonomy, we should stress four points that we learned from the classification development project. First, excellent agreement within institutions and good agreement among different institutions can be attained, but the classification must be consistently applied to obtain comparable results among sites and over time.

Second, achieving consistent application of the classification requires an independent editor using objective criteria to prevent, for example, self-serving ‘addenda’ in which ‘benign
prostate tissue’ becomes ‘prostatic adenocarcinoma’ without the change being recognized as a misinterpretation. Besides being ignored, classification criteria can be misapplied: when samples have been confused, a report amended from ‘intradermal nevus’ to ‘eccrine poroma’ is not due to a report defect but due to misidentification. Failure to dig down to the amendment’s root cause can lead to the change from ‘mild dysplasia CIN1’ to ‘focal severe dysplasia CIN3’ being misattributed to defective interpretation when the change proves due, in actuality, to a specimen issue, inadequate sampling: the higher grade lesion appeared on deeper levels.

Third, there is a big difference between the lower level of amendments that appears with passive monitoring (not quite 3/1000) and the higher rate that appears with active monitoring (almost 5/1,000).

Finally, during the taxonomy’s development phase, conference reviews appeared as a very important mechanism for amendment discovery in the baseline state: in our experience getting ready for conferences, presenting at conferences (like tumor boards), and following up on issues that conference discussions raised detected one to twenty percent of all defects, half of misinterpretations during passive monitoring period and four-out-of-five misinterpretations during active monitoring. Institutions following in our footsteps should keep these four points in mind: consistency can be achieved, that it requires an independent editor using objective criteria, that there is a big difference between passive and active monitoring in terms of detection rates, and that conference review is the most important amendment discovery mechanism in the baseline state.

**Monitoring Experience – Application:** The four defect categories apply to a sequence of events that make up the surgical pathology process. First in the sequence is patient and specimen identification, monitored by the defect category of misidentification. Second in the sequence is specimen handling, monitored by specimen defects. Third is diagnostic interpretation, monitored by the category of misinterpretation. Last in the sequence is result reporting, monitored by report defects.

To this sequence we applied the measures of total amendments per year and the rates of amendments per thousand cases to assess trends in amendment frequency, and the fractions of the four defect types divided by the amendment totals to assess the distribution of amendment types.

The sequence of events in the surgical pathology process is always a work in progress, always undergoing modification. The years in which we applied the amended report taxonomy to the surgical pathology process at Ford, we saw three important modifications that affected the sequence. Amendment editing as a real time application of the classification began in 2005; the LEAN principles of the Henry Ford Production System
(HFPS) were introduced, including interventions into the first two steps in the sequence, patient and specimen identification and specimen processing in 2006; the HFPS had informed the whole sequence by 2007. The process changes were associated with interesting changes in the measure of amendment rates. (See the table tallying these changes from the Power Point presentation accompanying this syllabus: Monitoring Results I: Defect Totals.)

The introduction of amendment process editing was associated with a doubling of observed amendments per thousand cases between the baseline year of 2004 and 2005, from 4.8/1,000 to 10.1/1,000. During the first year of the Production System’s changes specifying and standardizing the steps in the process to reduce batching and cut cycle terms, defects fell 20%, from 10.1/1,000 to 7.8/1,000. As the HFPS permeated the whole process, producing a redesign of specimen and information flow, focus on decreasing process defects and reducing re-work, the amendment rate fell another 15%, from 7.8/1,000 to 6.3/1,000 between 2006 and 2007.

Over the same period, changes in the fractions of defects in different categories were also of interest. These changes are summarized in a Table entitled “Results II: Monitoring Defect Fractions in the accompanying Power Point presentation.

During the four year period, the misinterpretation fraction fell 70%, from almost a quarter of all defects (23.5%) in 2004 to 7% in 2007. This fall was steeper after the HFPS’s full implementation in 2007: the decline in the fraction between 2006 and 2007 was from 16% to 7%, so the fractional contribution of misinterpretation more than halved. Misidentification fractions fell 4% between 2004 and 2005, and 3% between 2005 and 2006, but then leveled off in 2006 and 2007 at 12% of all defects. Specimen defects were the fewest in the first three years: they declined steeply, from 9% to 2% between 2004 and 2005, then rebounded to 4.3% in 2006, and rose to its highest fraction (11%) in 2007, the year when the misinterpretation fraction became the smallest fraction of defects. As the other fractions fell, report defects increased from 48% to 64% to 68% to 70% at the application period’s end.

In these changes we see that the HFPS’s interventions had more impact on identification and interpretation defects and less impact on specimen defects. The report defection fraction, one should remember, is the “default category” that increases as the other fractions decrease.

Monitoring Experience: Second Phase: As we continued monitoring into 2008, we focused on the three major initiatives of the HFPS in the surgical pathology process. These entailed, first ‘goal-directed’ education of the clinical staff (nurses, office staff, and physicians) to change their specimen labeling process to eliminate misidentifications. In the process phase of specimen handling we redesigned and change our own specimen accession...
(identifying, describing, sampling, cassetting, and fixing) process. In the third phase of diagnostic interpretation, the Ford surgical pathologists adopted a practice of pre-sign-out double reading of breast and prostate specimens were introduced. This double review involved about 800 of about 47,500 cases a year, so 2% of all cases.

Over the period that we monitored these three interventions, total amendments fell steadily from 374 to 261 a year, that is, from an amendment rate of 7.8 to 5.5/1,000. Focused double reviews of a small fraction of cases were associated with an 88% decrease in the amendment fraction: from 16% in 2006, through 7% in 2007, to 2% in 2008. We hasten to add here that this is a real world association not the conclusion of a well-controlled experiment. Undetected confounding factors may have also contributed to this fall.

The “front end” effort to improve patient and specimen identification has no initial impact on the misidentification factor – which was 12% in both 2006 and 2007 – but it was associated - perhaps rewarded with a striking decrease, from 12% to 5%, between 2007 and 2008. Accession redesign was associated with an initial rise in specimen defects, from 4% to 11% between 2006 and 2007, then a return to 3% in 2008 (still not as low as the 2% recorded in 2005). This set of observations, illustrates how the amendment totals and rates can be used to check the overall trend associated with improvement initiatives, while defect fractions can be used to examine the specific impact of particular interventions.

Connection of Monitoring to Process: If one looks at the relation between amendment rates and fractions between 2004 and 2008 and the simultaneous changes in our surgical pathology processes, one finds four “take home messages” from on-going amended report monitoring.

First, pathologists “finalling” reports themselves (rather than with the help of transcriptionists) greatly increased report defects (missing or erroneous non-diagnostic information: practitioner, procedure, date, various codes, etc.) at the beginning of the monitoring period. By the end of the period--we have recounted-- defects in this category accounted for 90% of report revisions. Second, editing – introducing systematic amendment classification –increases defect detection. Third, the Henry Ford Production System LEAN initiatives did indeed reduce defects, but the impact was greater on some defect types (misidentification and misinterpretation) than others (specimen defects). Fourth, this impact grew with ongoing, widening integration of the LEAN principles into daily practice. Relevant to the last point, the impact of efforts to standardize clinicians’ specimen collection and labeling practices on misidentifications took two years before a decline appeared in the fraction of this sort of defect.
Finally, some challenges are difficult. The specimen processing re-design led initially to an increase in recorded defects then a decrease to the previous level but – at this point in the monitoring – without achieving a major quality breakthrough.

**Monitoring Experience - Continued Application:** Over a more recent two year space (2007-2009), we have focused attention in greater detail on the 10% of amendments due to interpretation, identification, and specimen defects.

First, we compared amendments between mid 2007 and mid 2008 and mid 2008 and mid 2009. Amendment rates continued to fall, over this time span from 7/1000 to 5/1000. This decline was reflected in two of the categories. For interpretation defects the decline was mostly due to a decrease in amendments of primary diagnoses. Among identification defects, patient misidentification was the most frequent defect in both years but the most prominent finding was rise, almost threefold, in laterality confusions. The decrease in specimen defects which finally appear to be trending down - can be traced to elimination of defects associated with ancillary testing.

In both years, GI and skin specimens remained the most frequent sources of interpretation defect with cervix and breast specimens also appearing in both years, but prostate, kidney, lung, and thyroid specimens coming and going from the list.

In the second year both clinical and conference review were less frequent discovery mechanisms.

As we continue to monitor amendments, four observations seem to be holding up. Pre-sign out double review of breast and prostate specimens appears worthwhile. Identification defects, at their new lower level, have again become relatively impervious to continuous improvement by our current intervention. Interpretation defects are more frequent in high volume (GI, skin, and cervix) specimen types.

**Importance of the Edit Function:** In early 2009 we again had to make the case for an independent editor. During a 60 day period – over which 7,846 cases were reported – 181 addenda were added (2.3 addenda/1000 cases). The biggest fraction of these additions (43%) was added GI diagnoses for which a reason was supplied (mostly statements of *H. pylori* immunostain results). Another relatively large group (19%) was also made up of GI diagnoses, but no reason for the additions was obvious in these cases. Fourteen percent of addenda were added molecular pathology test reports, so – *H. pylori* immunotesting and molecular testing account for over half (57%) of addenda. It is troubling, however, that no reason for an addendum was supplied in not only the 34 GI cases mentioned above but also in another 19 cases, so a large fraction (29%) of addenda was in instances in which the reason for amending could not be determined in retrospect. Even more troubling was the turning up of 10 cases in
which amendments had masqueraded as addenda: 5.5% of addenda should have been classed as amendments. In fact, there were more of these misclassified diagnostic changes during the period of our addenda audit than there were reported amendments. We took the finding of this audit to indicate that the independent editors need to review both amendments and addenda to maintain system integrity.

**Conclusion:** We have found value in the application of an amended reports dictionary in real time, throughout the practice, by one controlling editor, using a consistent classification to capture information for root cause analysis. Studying amended reports in this way during a period of LEAN process improvement saw observed defects first rise, then fall, recorded an untoward increase in report defects, and found multiple observed causes of misidentification at collection, transport, accessioning, grossing and cutting steps in the process. It also witnessed a decrease in misinterpretations that may indeed be real and substantial.

**Misinterpretations: An Interesting But Difficult Problem:** Interpretation defects may, however, be relatively difficult to approach systematically. There are five themes that emerge from their study as they surface in amendments that suggest how this might be true. The first is that some of the diagnostic distinctions that pathologists use, especially to describe the lowest grade borderlands of neoplasia, are not reproducible among observers. The second theme is that distracting histologic backgrounds increase the risk of misinterpretation. The third theme is that observers’ different emphasis on histological versus cytological appearances can lead to differences in diagnostic interpretations. Fourth, if an observer discovers a finding in a peripheral zone of a slide, he or she may neglect another finding in a central zone, or vice versa. Finally, in some situations, only differing knowledge of clinical contexts seems to drive diagnostic discrepancies.

**Pathology Pearls:** The string of four pearls that we can offer from our seven years study of amended reports are: defects which cause amendments divide into misinterpretations, misidentifications, specimen defects, and report defects; excellent taxonomic agreement in the application of this classification can be obtained within a practice and good agreement among several practices; consistent classification requires real time monitoring of both amendments and addenda by an independent editor. Finally, defect fractions assess interventions and defect rates assess success.

On the last slide in the Power Point presentation appears -- along with thanks for interest in the topic-- three questions for its audience to consider. Are you persuaded by our argument that both addenda and amendments require monitoring? Do amended report rates and distributions measure only product quality or do they also index safety? Should cases published years before be amended in the context of much later diagnostic developments?
The last slide also acknowledges long term collaborators in the study of amended reports: Richard Zarbo is the Chairman of Pathology and Laboratory Medicine at Ford and the enlivening spirit of our HFPS approach to improvement. Ruan Varney has managed the amended report dictionary from the beginning. Stephen Raab joined in our initial development of the taxonomy and – with his research assistant from Pittsburgh days, Colleen Vrbin - carried out the inter-practice comparison.