Error Reduction in Cytopathology: Working Smarter, Not Harder

Stephen S. Raab, M.D.
Department of Pathology, University of Colorado Denver, USA
American Society of Cytopathology 56th Annual Scientific Meeting, Orlando, Florida
November 9, 2008
Overview

• Quality improvement in non-gynecologic cytology
• Business/engineering models
• Steps in performing a quality improvement QI initiative
• Example in thyroid gland fine needle aspiration
• Questions
Challenges in QI

• Medicine has shifted from an individual model of providing healthcare to a team-oriented approach
• Lack of organizational commitment
• Turf issues
• Training
• Long term evaluation is problematic
Challenges in QI

• Middle management not engaged
• Up-front costs
• Cultural model (top down versus bottom up)
• Punitive history difficult to eradicate
• Disincentives for improvement
• Difficulties in linking improvement with outcome
Where do we start?

- Literature
- Existing quality improvement methods
- Communication with colleagues
- Utilize methods in hospital/medicine
- Utilize methods outside of medicine
Business/engineering model of work

- Three components: quality, efficiency, and cost
- Work is composed of numerous steps
- Components of work:
  - Activities
  - Pathways
  - Connections
- Improvement must be monitored and evaluated for each of the three components
Steps in QI

- Define the current condition
- Determine the causes of failure (where, when, how, and why)
- Design an intervention to alter a step in the process
- Implement the intervention
- Evaluate the success of the intervention
Defining the current condition

• Example: thyroid gland fine needle aspiration

• Define the steps in the process
  – Decision to perform FNA (pre-pre-analytic)
  – Performance of FNA (pre-analytic)
  – Analysis of FNA (analytic)
  – FNA result reporting (post-analytic)
  – Post-FNA decision making (post-post analytic)

• Measure failures in the steps of this process
Practice variability

- Less than optimal care results from the wide variability in practices
- Data indicate that there is wide variability in the practice of thyroid gland FNA
- Variability may be seen in all testing phases
- Variability in one testing phase affects the variability in following testing phases
Thyroid gland fine needle aspiration

• Examine cases in which the FNA diagnosis was benign and the patient had a neoplastic lesion on surgical excision (false negative cases)

• Examine cases in which the FNA diagnosis was indeterminate (e.g., atypical, suspicious), neoplastic or malignant and the patient had a non-neoplastic lesion on surgical excision (false positive cases)
Root cause analysis

• Traditional cytologic histologic correlation (sampling, screening, or interpretation)
• Other methods of root cause analysis may lead to targeting different error types and developing different types of best practice
  – Failure modes and effect analysis
  – Eindhoven method
  – Toyota method
Part A. Quality of specimen

Amount of tumor

A. Poor quality specimen
   No tumor identified

B. Excellent quality specimen
   No tumor identified

C. Poor quality specimen
   Tumor identified

D. Excellent quality specimen
   Tumor identified
Categorization of 30 thyroid gland false negatives

<table>
<thead>
<tr>
<th>PQ, NT</th>
<th>GQ, NT</th>
</tr>
</thead>
<tbody>
<tr>
<td>PQ, T</td>
<td>GQ, T</td>
</tr>
</tbody>
</table>
Categorization of 30 thyroid gland false positives

<table>
<thead>
<tr>
<th></th>
<th>PQ, NT</th>
<th>GQ, NT</th>
<th>PQ, T</th>
<th>GQ, T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image" alt="PQ, NT" /></td>
<td><img src="image" alt="GQ, NT" /></td>
<td><img src="image" alt="PQ, T" /></td>
<td><img src="image" alt="GQ, T" /></td>
</tr>
</tbody>
</table>


Initial impressions

- All failures have a cognitive and a system component
- We tend to over and under interpret poor samples
- Rarely, we miss lesions (FNs) that retrospectively we think could have been diagnosed
- For some FNs, patient-related or operator causes may have contributed
<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Latent errors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical</td>
<td>Errors in physical items, such as equipment, physical installations, software, materials, labels and forms</td>
<td></td>
</tr>
<tr>
<td>External</td>
<td>Technical failures beyond the control and responsibility of the investigating organization</td>
<td>TEX</td>
</tr>
<tr>
<td>Design</td>
<td>Failures due to poor design of equipment, software, labels, or forms</td>
<td>TD</td>
</tr>
<tr>
<td>Construction</td>
<td>Correct design was not followed accurately during construction</td>
<td>TC</td>
</tr>
<tr>
<td>Materials</td>
<td>Material defects not classified under TD or TC</td>
<td>TM</td>
</tr>
<tr>
<td>Organizational</td>
<td>Failures at an organizational level beyond the control and responsibility of the investigating organization</td>
<td>OEX</td>
</tr>
<tr>
<td>Transfer of knowledge</td>
<td>Failures resulting from inadequate measures taken to ensure that situational or domain-specific knowledge or information is transferred to all new or inexperienced staff</td>
<td>OK</td>
</tr>
<tr>
<td>Protocols/procedures</td>
<td>Failures related to the quality and availability of the protocols within the department</td>
<td>OP</td>
</tr>
<tr>
<td>Management priorities</td>
<td>Internal management decisions in which safety is relegated to an inferior position in the face of conflicting demands or objectives. This error results from a conflict between production needs and safety (e.g., decisions about staffing levels)</td>
<td>OM</td>
</tr>
<tr>
<td>Culture</td>
<td>Failures resulting from collective approach to risk and attendant modes of behaviour in the investigating organization</td>
<td>OC</td>
</tr>
<tr>
<td><strong>Active errors (human)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>External</td>
<td>Errors or failures resulting from human behaviour</td>
<td>HEX</td>
</tr>
<tr>
<td>Knowledge-based behaviours/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>knowledge-based errors</td>
<td>Human failures originating beyond the control and responsibility of the investigating organization</td>
<td></td>
</tr>
<tr>
<td>Rule-based behaviours</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualifications</td>
<td>Incorrect fit between an individual's qualifications, training, or education and a particular task</td>
<td>HRQ</td>
</tr>
<tr>
<td>Coordination</td>
<td>Lack of task coordination within a healthcare team in an organization</td>
<td>HRC</td>
</tr>
<tr>
<td>Verification</td>
<td>Failures in the correct and complete assessment of a situation, including relevant patient conditions and materials to be used before the intervention</td>
<td>HRV</td>
</tr>
<tr>
<td>Intervention</td>
<td>Failures that result from faulty task planning (selecting the wrong protocol) and/or execution (selecting the right protocol but carrying it out incorrectly)</td>
<td>HRI</td>
</tr>
<tr>
<td>Monitoring</td>
<td>Failures during monitoring of process or patient status during or after intervention</td>
<td>HRM</td>
</tr>
<tr>
<td>Skill-based behaviours</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slips</td>
<td>Failures in performance of fine motor skills</td>
<td>HSS</td>
</tr>
<tr>
<td>Tripping</td>
<td>Failures in whole-body movements</td>
<td>HST</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient-related factor</td>
<td>Failures related to patient characteristics or conditions that influence treatment and are beyond the control of staff</td>
<td>PRF</td>
</tr>
<tr>
<td>Unclassifiable</td>
<td>Failures that cannot be classified in any other category</td>
<td>X</td>
</tr>
</tbody>
</table>
Root cause analysis

• Organizational latent errors
  – OP : Quality and availability of protocols that are too complicated, inaccurate, unrealistic, absent or poorly presented
  – OK: Failures resulting from inadequate measures taken to ensure that situational or site-specific knowledge or information is transferred to all new or inexperienced staff
Root cause analysis

• Rule based behavior errors
  – HRK: The incorrect fit between an individual’s qualification, training, or education and a particular task
  – HRC: A lack of task coordination within a health care team in an organization
  – HRV: The incorrect or incomplete assessment of a situation, including related conditions of the patient/donor and materials to be used before beginning the task
Patients with a non neoplastic thyroid nodule that is interpreted as neoplastic or non-definitive

Unsatisfactory Sample

- extremely busy service
  - overinterpretation of specimen as neoplastic instead of unsatisfactory
    - clinician pressure to diagnosis as adequate
      - OM
  - overinterpretation of specimen as neoplastic instead of unsatisfactory
    - clinician pressure to diagnosis as adequate
      - OC
- lack of immediate interpretation service
  - overinterpretation of specimen as neoplastic instead of unsatisfactory
    - lack of experience of cytologist
      - PRF and OK
- patient related factors
  - overinterpretation of specimen as neoplastic instead of unsatisfactory
    - rule-based or knowledge based error
      - HRQ or HKK

Adequate Sample

- interpretation of specimen as neoplastic
  - rule-based or knowledge based error
    - HRQ or HKK
Thyroid gland FNA failures

- Root cause analysis for active causes:
  - Poor quality specimens
  - Over diagnosis of poor quality specimens
  - Poor clinician-pathologist-cytotechnologist communication
  - Variable use and meaning of diagnostic categories
  - Lack of individual training or experience
Thyroid gland FNA failures

- Root cause analysis for latent causes:
  - Lack of organizational commitment for improvement
  - Lack of protocols
  - Practitioners hurried because of heavy work loads
  - Lack of cost incentives
  - Cultural acceptance of current failure rate
Work assessment of failures

• Poor quality specimen
  – Lack of training in procurement (activity, pathway)
  – Lack of immediate feedback loops (activity, connection)
  – Over diagnosis of unsatisfactory specimens (activity)
  – Poor clinician-pathologist communication (activity, pathway, connection)
  – Variable use and meaning of diagnostic categories (activity, connection)
  – Lack of individual training or experience (activity)
Areas of improvement

• Change system processes
  – Standardize work (pathways, connections, activities)
  – Improve efficiency (pathways and activities)
  – Improve communication
  – Create error checks (e.g., redundancy)

• Focus on cognition
  – Training and education
Service redesign

• Standardize terminology
  – Create specimen adequacy scale
  – Use standardized diagnostic terminology

• Improve pathology-clinical communication
  – Create a pathologist immediate interpretation service for radiologic-guided FNAs
## Likelihood ratios for diagnoses

<table>
<thead>
<tr>
<th>Pathol</th>
<th>Unsat</th>
<th>Benign</th>
<th>Atypical</th>
<th>Malignant</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.73</td>
<td>0.52</td>
<td>0.62</td>
<td>14.1</td>
</tr>
<tr>
<td>Inexperienced</td>
<td>∞</td>
<td>0.81</td>
<td>1.0</td>
<td>3.0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.67</td>
<td>0.75</td>
<td>∞</td>
</tr>
<tr>
<td>2</td>
<td>0.34</td>
<td>0.50</td>
<td>0.69</td>
<td>4.1</td>
</tr>
<tr>
<td>3</td>
<td>0.98</td>
<td>0.53</td>
<td>0.33</td>
<td>∞</td>
</tr>
</tbody>
</table>
Adequacy

• Three criteria:
  – Background grade
  – Cellularity grade
  – Preservation grade

• Each criterion was graded on a 0-3 scale

• Specimens with a summed adequacy score of $\leq 5$ were classified as *non-specific*
Criteria

Background grade

0 - Bloody background, little or no colloid
1 - Bloody background, little or no colloid, macrophages and/or inflammatory elements
2 - Definite areas of sharply demarcated watery colloid with atypical wrinkles/crackled features on at least one slide, +/- bloody background, +/- macrophages and/or inflammatory elements
3 - Abundant sharply demarcated watery colloid with typical wrinkles/crackled features on many slides. +/- macrophages and or inflammatory elements. Minimal or no blood.
Criteria

Cellularity grade

0 – Less than 8 to 10 cellular fragments (at least 10 cells) on 2 slides
1 - > or = 8 to 10 cellular fragments (at least 10 cells) on 2 slides
2 – Cellular fragments (at least 50 cells) on 50% slides reviewed
3 – Cellular fragments (at least 50 cells) on 100% slides reviewed
## Number of FNAs in pre and post-interventions categories

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Standardization</th>
<th>Immediate interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>Unsatisfactory</td>
<td>89</td>
<td>78</td>
</tr>
<tr>
<td>Non-specific</td>
<td>0</td>
<td>155</td>
</tr>
<tr>
<td>Benign</td>
<td>1,130</td>
<td>744</td>
</tr>
<tr>
<td>Atypical</td>
<td>126</td>
<td>43</td>
</tr>
<tr>
<td>Neoplasm</td>
<td>133</td>
<td>90</td>
</tr>
<tr>
<td>Malignant</td>
<td>46</td>
<td>49</td>
</tr>
</tbody>
</table>
Results of implementation

- Standardized thyroid gland fine needle aspiration (FNA) services
- Sensitivity: Pre-intervention: 70.2%; post intervention: 92.3% (P < 0.001)
- False negative rate: Pre-intervention: 41.8%; post intervention: 18.2% (P = 0.006)
- Repeat FNA rate: Pre-intervention: 12.7%; post intervention (with immediate interpretation): 3.7% (P = 0.001)
- Cost savings $167,000 per 100 patients ($501 million annually in USA)
Other examples

• Immediate interpretation, standardization of terminology, feedback loops in lung FNA – improved sensitivity and specificity
• Immediate feedback loops in breast FNA – decreased time to treatment
• Double viewing (redundancy) in bronchial washing/brushing – decreased false negative rate
• Reprocessing voided urine cytology specimens diagnosed by cytotechnologist as atypical – improved sensitivity and specificity
Other examples

• Diagnostic standardization practices
  – Most improvement through face-to-face, frequent microscope sessions
  – Least improvement with use of picture images
  – Smaller groups tend to be more standardized
  – Groups with respected, strong cytotechnologist or cytopathologist tend to be most standardized
Summary

• Standardization of diagnostic categories is a form of developing a best practice
• Standardization must be incorporated into a larger program of evaluation of services
• Standardization must be evaluated in terms of sustainability, reliability, implementability, root cause analysis, and other means to assess utility
• Highest levels of improvement result from targeting poor specimens within and outside of pathology lab