Research in Digital Pathology

Mark S. Roberts, MD, MPP
Professor of Medicine, Health Policy and Management and Industrial Engineering
University of Pittsburgh Schools of Medicine, Public Health and Engineering
Outline of Talk

• Purpose and need for evaluation of new technology
• Review of Telepathology evaluation
• Components of current evaluation plans
Need for Evaluation

- New technology does not always improve diagnosis and therapy
- With all innovation, there are expected and unintended consequences
- One interesting historical example....
Technology Examples: Radiology

- W.C. Roentgen discovers X-rays in 1895
  - There are hundreds of reports of their use in hospitals by mid 1896
  - X-ray machines are ubiquitous by 1905
- Innovation that
  - “worked”
  - Had a clear benefit
- But there are downsides….
In the early 1900’s there was a virtual “epidemic” of childhood chest malignancy (thymomas):

Thousands (*that’s really thousands*) of children had normal thymus glands removed because no-one knew that the normal thymus in an 8-year old is about 10 times as big as it is in an adult.
What should the test of accuracy be?

- When telepathology is compared to standard reading methods, what is the appropriate comparator?
  - The same pathologists light microscopy reading?
  - A consensus “gold standard” light microscopy reading?
Comparison to Teleradiology

- Digital x-rays were not known to be equivalent to plain films
  - Radiologists objected to not be able to pick up the film, hot light an area, step back, etc
- Evaluation studies demonstrated virtual equivalence in diagnoses
  - (as equivalent as radiological diagnosis can be)
Concordance in standard methods imperfect

There is not 100% concordance in the diagnosis (in fact there may be n+1 readings where n is the number of readers)

SECTION OF DECISION SCIENCES AND CLINICAL SYSTEMS MODELING
Concordance in standard methods imperfect

• Even when the observer is the same….  

• There is not 100 percent concordance in the diagnosis
Concordance in standard methods imperfect

There is not 100% concordance in the diagnosis
Telepathology: evaluation examples:

- Dermatology: telepathology for dermatologic diagnosis is reliable but not perfect (Piccolo, Arch Derm, 2002)


- Renal histopathology: found no difference, but insufficient sample size to declare equivalence (Furness P. Histopathology, 2007)
Telepathology: current evaluation state

- Typical evaluation methodology: blinded single slide reading

Telepathology: current evaluation state

- Diagnostic disagreement is not one-sided

<table>
<thead>
<tr>
<th>Level of discordance</th>
<th>Reason for discordance</th>
<th>VM diagnosis</th>
<th>OM diagnosis</th>
<th>Review consensus diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>II</td>
<td>Chronic gastritis</td>
<td>Acute and chronic gastritis</td>
<td>YM</td>
</tr>
<tr>
<td>C</td>
<td>III</td>
<td>Gastric adenoma</td>
<td>Gastric hyperplastic polyp</td>
<td>YM</td>
</tr>
<tr>
<td>C</td>
<td>I</td>
<td>Ulceration ventriculi</td>
<td>Ulcerated, well differentiated adenocarcinoma</td>
<td>OM</td>
</tr>
<tr>
<td>C</td>
<td>II</td>
<td>Chronic atrophic gastritis</td>
<td>Chronic gastritis with low grade dysplasia</td>
<td>OM</td>
</tr>
<tr>
<td>B</td>
<td>II</td>
<td>Hyperplastic polyp</td>
<td>Hyperplastic mucosa</td>
<td>OM</td>
</tr>
<tr>
<td>C</td>
<td>II</td>
<td>Mild colitis</td>
<td>Normal colon</td>
<td>OM</td>
</tr>
<tr>
<td>C</td>
<td>III</td>
<td>Colitis ulcerosa with dysplasia</td>
<td>Colitis ulcerosa</td>
<td>YM</td>
</tr>
<tr>
<td>C</td>
<td>III</td>
<td>Adenoma papillare</td>
<td>Hyperplastic polyp</td>
<td>OM</td>
</tr>
</tbody>
</table>

Discrepancy: B, clinically not important; C, clinically important. Reason for discrepancy: I, image quality; II, interpretation; III, insufficient clinical information.

OM, optical microscopy; VM, virtual microscopy.
Telepathology-what to evaluate

• Primary evaluation should be:
  – “Do we arrive at the same diagnosis?”

• Secondary evaluation components:
  – Does it improve workflow/workload distribution
  – Does it improve/enhance collaboration/education
  – Does the “system” work:
    • Ease of use
    • Quality of image
Ideal evaluation scenarios

- Live pathologist
- Whole Slide Imaging
- Robotic microscope

**SECTION OF DECISION SCIENCES AND CLINICAL SYSTEMS MODELING**
Live pathologist

Pathology slide

Whole Slide Imaging

Robotic microscope

Diagnosis A

Diagnosis B

Diagnosis C
Evaluation Methodology

Pathology slides

Live pathologists “Gold Standard”

Robotic microscope

Pathology slides

Live pathologists “Gold Standard”

Whole Slide Imaging

SECTION OF DECISION SCIENCES AND CLINICAL SYSTEMS MODELING
Evaluation outcome: diagnosis

<table>
<thead>
<tr>
<th>Pathological Diagnosis</th>
<th>Telepathology Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis concordant w/light microscopy Reading</td>
<td>Robotic Microscopy</td>
</tr>
<tr>
<td>Diagnosis different from light microscopy reading</td>
<td>Whole Slide Imaging</td>
</tr>
<tr>
<td>$X_1$</td>
<td>$X_2$</td>
</tr>
<tr>
<td>$N_1-X_1$</td>
<td>$N_2-X_2$</td>
</tr>
</tbody>
</table>
Evaluation Outcomes: process measures

- Pathologist feedback
  - Ease of use
  - Surety of diagnosis
  - Quality of image
Future Possibilities

- Recall the progression of teleradiology
- Image processing capability improved the actual process of radiological diagnosis
  - Bone windows
  - Lung windows
  - Contrast enhancement
  - Magnification
  - Border identification
Telepathology

- Potential for image processing remains early and undeveloped
- Potential for virtual consults
- May need to overcome standard habits