Uncertainty in Medical Imaging

It’s all over the chain!

Ioannis Sechopoulos, Ph.D., DABR
Associate Professor
Advanced X-ray Tomographic Imaging Lab
Department of Radiology and Nuclear Medicine
Radboud University Medical Center and
LRCB – Dutch Reference Centre for Screening
Nijmegen, the Netherlands
The old imaging chain...

http://apachetechnology.in/KC/Multimedia/DIP/Medical/DigitalXRay.aspx
...but today, the longer chain

Biology  Engineering  Perception
Physics  Mathematics  Epidemiology
...but today, the longer chain
UNCERTAINTY: BIOLOGY
What is cancer?

Hanahan and Weinberg, Cell, 100(1), 57–70, 2000
What is cancer?

Emerging Hallmarks:
- Deregulating cellular energetics
- Avoiding immune destruction
- Genome instability and mutation
- Tumor-promoting Inflammation

Enabling Characteristics

Hanahan and Weinberg, Cell, 144(5), 2011, 646–674
Tumor development
When should we detect cancer?
When should we TREAT cancer?
Disease growth and detection
Ductal Carcinoma In Situ (DCIS)
In situ cancers reported separately

<table>
<thead>
<tr>
<th></th>
<th>Detected DBT total</th>
<th>Detected DBT alone</th>
<th>Detected DM total</th>
<th>Detected DM alone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>67</td>
<td>21</td>
<td>47</td>
<td>1</td>
</tr>
<tr>
<td>Total invasive cancers</td>
<td>58</td>
<td>17</td>
<td>41</td>
<td>-</td>
</tr>
<tr>
<td>Total in situ cancers</td>
<td>9</td>
<td>4</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

Over-detection?

Imaging
Physical exam
Genetic testing
Over-diagnosis?

Pathology
Over-treatment?

Lumpectomy or Mastectomy
Radiotherapy
Chemotherapy
Co-morbidities

<table>
<thead>
<tr>
<th>Death due to</th>
<th>HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other coronary heart diseases</td>
<td>1.14</td>
</tr>
<tr>
<td>Heart failure</td>
<td>1.29</td>
</tr>
<tr>
<td>Other heart disease</td>
<td>1.24</td>
</tr>
<tr>
<td>Gastrointestinal disease</td>
<td>1.1</td>
</tr>
<tr>
<td>External causes</td>
<td>1.14</td>
</tr>
<tr>
<td>Suicide</td>
<td>1.39</td>
</tr>
<tr>
<td>Symptoms</td>
<td>1.09</td>
</tr>
<tr>
<td>Diabetes</td>
<td>1.18</td>
</tr>
<tr>
<td>Pulmonary circulation</td>
<td>1.51</td>
</tr>
<tr>
<td>Urinary system disease</td>
<td>1.16</td>
</tr>
<tr>
<td>Other bacterial disease</td>
<td>1.2</td>
</tr>
</tbody>
</table>
Estimating over-diagnosis

<table>
<thead>
<tr>
<th>Location</th>
<th>Percentage (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malmö (55–69 years)</td>
<td>10.5 (8.4–12.7)</td>
</tr>
<tr>
<td>Canada I</td>
<td>12.4 (9.9–14.9)</td>
</tr>
<tr>
<td>Canada II</td>
<td>9.7 (7.5–11.9)</td>
</tr>
<tr>
<td>Overall ($I^2=22.3%, p=0.276$)</td>
<td>10.7 (9.3–12.2)</td>
</tr>
</tbody>
</table>
To be continued...
So don’t over-treat...

Which ones?
The LORIS trial: 932 patients

- Patient presents with screen detected or incidental microcalcification
- Local diagnosis of non-high grade DCIS
- Randomise

**Surgery Arm**
- Surgery
- Annual mammograms for 10 years

**Active Monitoring Arm**
- New ipsilateral abnormality detected - follow investigation algorithm
  - No invasion or grade migration
  - Invasive disease or grade migration
- Annual mammograms for 10 years

References:
The LORD trial

1240 women aged
≥ 45 years
asymptomatic, screen-detected, pure low-grade DCIS microcalcifications only.
The LORD trial

Standard treatment + annual mammo vs. Annual mammo only

Primary end-point: 10-year ipsilateral invasive breast cancer free percentage
UNCERTAINTY: Biology

Different types and stages of (breast) cancer
UNCERTAINTY: Biology

Some tumors will never kill you
UNCERTAINTY: Biology

Need to discriminate which ones
UNCERTAINTY: Biology

How should they be dealt with?
UNCERTAINTY: ENGINEERING
Indirect digital detector

- Contact leads for readout electronics
- Contact fingers
- Amorphous silicon array
- Thin-film transistor
- Glass substrate
- Scintillator (CsI:Tl)

Courtesy of GE Medical Systems
Indirect digital detector
How do you simulate that?

Freed et al, Medical Physics 36(11), 2009.
Point Spread Functions

Freed et al, Medical Physics 36(11), 2009.
Contrast-Detail Phantom
Contrast-Detail Phantom

Detectability Threshold
Contrast-Detail Curve

- Contrast
- Size (detail)

Worse Detectability
Better Detectability
Point Spread Functions

Freed et al, Medical Physics 36(11), 2009.
Contrast-Detail Phantom
Example deblurring study

Jannetta et al, PMB, 49(21), 2004, 4997
<table>
<thead>
<tr>
<th>Mode</th>
<th>Original</th>
<th>MEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.8BF</td>
<td>51</td>
<td>77</td>
</tr>
<tr>
<td>1.8FF</td>
<td>54.5</td>
<td>81</td>
</tr>
<tr>
<td>3.0FF</td>
<td>79</td>
<td>104</td>
</tr>
</tbody>
</table>

*Jannetta et al, PMB, 49(21), 2004, 4997*
Best metric

Receiver operating characteristics
But...

Expensive
Contrast-detail analysis

Pattern effect
Detection task (to be continued)
2. Image criteria related to detector performance, exposure parameters and patient movement
   2.1 Clear reproduction of glandular tissue
   2.2 Clear reproduction of fibrous strands in fat tissue
   2.3 Clear reproduction of vascular structures in fat tissue
   2.4 Clear reproduction of pectoral muscle margin
   2.5 Clear reproduction of calcifications, when present
   2.6 Acceptable noise level in the reproduction of the pectoral muscle
3. Overview judgement related to either screening or diagnostic procedures
   3.1 Is image quality sufficient for early breast cancer detection?
JAFROC Results

Zanca et al, BJR, 85 (2012), 1233–1241
VGA Results

Siemens OpView v. 2 vs IMS Raffello

$S = 2.55$

Zanca et al, BJR, 85 (2012), 1233–1241
UNCERTAINTY: Engineering

Imperfect detector response
UNCERTAINTY: Engineering

PSF modeled experimentally and simulation
UNCERTAINTY: Engineering

True PSF(s) unknown
UNCERTAINTY: Engineering

Single PSF is an approximation
UNCERTAINTY: Engineering

Blind deconvolution (?)
UNCERTAINTY: Engineering

How should we evaluate image quality?
UNCERTAINTY: Engineering

What should we aim to improve?
UNCERTAINTY: PERCEPTION
Radiologists are not perfect...
...and their imperfections vary
Inter-reader variability

Mammography alone, 17 readers, 164 cases

Drukker et al, Academic Radiology, Vol 20, No 7, July 2013
Inter-reader variability

Readers

Normals  Cancers

Mammo + ABUS, 17 readers, 164 cases

Drukker et al, Academic Radiology, Vol 20, No 7, July 2013
Inter-reader variability

Inter-reader $\kappa$ statistics for all cases

Drukker et al, Academic Radiology, Vol 20, No 7, July 2013
Inter-reader variability

Mammography
8 readers, 45 cases
2 reads per reader per case

Ciccone et al, Eur J Cancer, 28A(6/7), 1054, 1992
Inter-reader variability

Inter-reader kappa: 0.45 (first read)
0.44 (second read)
(moderate)

Ciccone et al, Eur J Cancer, 28A(6/7), 1054, 1992
But also...
Intra-reader variability

Intra-reader kappa: 0.35 – 0.67
mean: 0.56
(moderate)

Ciccone et al, Eur J Cancer, 28A(6/7), 1054, 1992
What does it take to detect a lesion?
One of the radiologists later admitted, “I just didn’t see it.”
Three steps to detection

Search
Recognition
Decision
Search: Did the eyes pass by the lesion?
Search
Recognition: Was the lesion recognized as such?
Wasting time looking at irrelevant areas...

<table>
<thead>
<tr>
<th></th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevant</td>
<td>96.25%</td>
<td>99.58%</td>
<td>99.17%</td>
<td>100%</td>
</tr>
<tr>
<td>Not relevant</td>
<td>3.75%</td>
<td>0.42%</td>
<td>0.83%</td>
<td>0%</td>
</tr>
</tbody>
</table>

*Krupinski et al, Human Pathology, Volume 44, Issue 3, March 2013, Pages 357–364*
Decision: Is what I’m looking at a cancer?
# Reasons for false negatives

<table>
<thead>
<tr>
<th></th>
<th>Experienced</th>
<th>Inexperienced</th>
</tr>
</thead>
<tbody>
<tr>
<td>False negatives</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td>Search</td>
<td>24%</td>
<td>29%</td>
</tr>
<tr>
<td>Recognition</td>
<td>24%</td>
<td>42%</td>
</tr>
<tr>
<td>Decision</td>
<td><strong>52%</strong></td>
<td>29%</td>
</tr>
</tbody>
</table>

Krupinski, Acad Radiol 1996;3:137-144
## False negatives: mammo vs tomo

<table>
<thead>
<tr>
<th></th>
<th>Tomo + Mammo -</th>
<th>Tomo - Mammo +</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visibility</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>Radiographic appearance</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Interpretative error</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

Lång et al, Br J Radiol 2014;87:20140080
Don’t make lesions just more apparent so they can be seen....
...make them clearer so they can be characterized...
UNCERTAINTY: Perception
(for now?) Humans read radiology images
UNCERTAINTY: Perception

Mistakes happen (differently!)
UNCERTAINTY: Perception

Our research should pay attention to these perception results
UNCERTAINTY: EPIDEMIOLOGY
Screening for breast cancer works... 

....right?
Sure!

People live longer!
"The five-year survival rate for all cancers improved from 51 percent in the early 1980s to almost 60 percent in the early 1990s. . . . since the 1971 National Cancer Act, much of the research into early cancer detection and treatment has paid off."
Take one cancer patient...
Take another cancer patient...
Lead time bias
X-year survival can’t be used for determining screening effectiveness
In some cases two cohorts are available

Comparison pre-screening

Female breast cancer in England: incidence and mortality

Rate per 100,000

Incidence

Mortality

Year

Source: Office for National Statistics
but...

Female breast cancer in England: incidence and mortality

Incidence

Mortality

Source: Office for National Statistics

Breast Cancer in England, 2010 Release
Also...
Screening or therapy?

Important and challenging

Incidence must be the same (or corrected for)
Example:
Modeling incidence to estimate over-diagnosis
“in 2008, breast cancer was overdiagnosed in more than 70,000 women...31% of all breast cancers diagnosed.”
# Underlying breast cancer risk

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Incidence increase (%/yr)</th>
<th>Overdiagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base case</td>
<td>constant</td>
<td></td>
</tr>
<tr>
<td>Best guess</td>
<td>0.25%</td>
<td>31%</td>
</tr>
<tr>
<td>Extreme assumption</td>
<td>0.50%</td>
<td>26%</td>
</tr>
<tr>
<td>Very extreme assumption</td>
<td>0.5% + highest</td>
<td></td>
</tr>
<tr>
<td></td>
<td>baseline</td>
<td>22%</td>
</tr>
</tbody>
</table>

Underlying breast cancer risk

- Incidence US
- ‘extreme’ (26%)
- ‘best guess’ (31%)

Courtesy of Mireille Broeders and Nicolien van Ravesteijn
Underlying breast cancer risk

- Incidence US
- 'extreme' (26%)
- 'best guess' (31%)

Courtesy of Mireille Broeders and Nicolien van Ravesteijn
Underlying breast cancer risk

![Graph showing breast cancer incidence and APC model predictions from 1950 to 2000.](image)

- **incidence US**
- **APC model (10%)**
- **'extreme' (26%)**
- **'best guess' (31%)**

Courtesy of Mireille Broeders and Nicolien van Ravesteijn
Estimating over-diagnosis

Puliti et al, J Med Screen 2012
UNCERTAINTY: Epidemiology

Survival rate misleading for screening effectiveness
UNCERTAINTY: Epidemiology

Determining true improvement in outcome involves modeling
UNCERTAINTY: Epidemiology

Modeling introduces uncertainty
In short...

Earliest possible moment to detect/diagnose/treat cancer might not be optimal
In short...

We could estimate and deblur due to PSF on average, but not for each pixel.
In short...

How do we determine image quality improvement?
In short...

We need to improve image quality in a way that would address where the errors are made most often.
In short...

The impact of medical imaging is not always that easy to quantify!
Thanks for listening!

Any Questions?

No?

SUPER!