TAKING DIAGNOSTIC REASONING TO VEGAS: APPLICATION AND EVALUATION OF PROBABILISTIC THINKING USING A NOVEL EDUCATIONAL GAME

Nathan Stehouwer, MD
Lukasz Weiner, MD
Collin Swafford, MD
Keith Ponitz, MD
Andrea Scioscia, MD
Michael Dell, MD
Goals of this workshop:

• List at least three educational benefits of teaching formats which encourage learners to consider probabilistic reasoning.

• Describe and apply a scoring system for probabilistic reasoning which rewards accurate appraisals of uncertainty.

• Implement a game of diagnostic reasoning, in which learners rate the probability of each item on their differential diagnosis, at didactic conferences at their home institutions.
Agenda

• Introduction to probabilistic reasoning
• Discuss Brier scoring and its application to differential diagnosis
• Participate in diagnostic reasoning game
• Debrief: Discuss benefits, challenges, and applications of this methodology
Decision-Making Model of Diagnostic Reasoning

Don’t treat
Additional testing needed
Start treatment

0% Probability $P_{test}$ $P_{rx}$ 100% Probability
How well do physicians think probabilistically?

- Unconscious probabilistic reasoning underlies “intuition” (Kreiter, 2017)
- Our usual preferred terms (probable, possible, unlikely, frequent) are unclear and may be interpreted differently by colleagues and by patients (Lurie, 1999)
- In one survey, only 2 out of 300 clinicians surveyed reported thinking in terms of numerical probability (Grimes, 2005)
How well do physicians think

What is the probability (in percents) in your opinion, that the patient has:
• Acute coronary syndrome
• Dissecting aortic aneurysm
• Reflux esophagitis
• Biliary colic
• Anxiety disorder?
• Others

Source: Cahan, 2003

Figure 2. Frequency distribution of the total probabilities assigned by participants. The mean total probability was 136.7% (± 53.9%). Sixty-five percent of participants had a total probability >100% (i.e. exhibited subadditivity).
Possible benefits of thinking probabilistically

- Appropriately determine when testing or treatment are indicated
- Improved interpretation of diagnostic test results (Bayesian inference)
  - Accurate pretest probability assessment largely negates the effect of false positive and negative results (Diamond, 1980)
- Improved risk/benefit assessments (Lurie, 1999)
- Make clinical uncertainty explicit and acceptable
  - Physicians may overstate diagnostic certainty, which appears to be negatively correlated with experience (Schoenherr, 2018)
- Encourage metacognition – corrective to System 1 biases
Can we teach probabilistic thinking?

- Decision tools
  - Students taught a prediction rule for estimating pretest probability of coronary artery disease (CAD) outperformed students given more traditional reading on CAD (Hickam, 1987).
  - Such tools limited by being:
    - Binary
    - Tied to prevalence in past population studied
    - Specific to one question at a time
- Bayesian inference supports sound clinical reasoning (Kreiter, 2017) but is challenging to objectively assess.
FiveThirtyEight

Sign in using Google or Facebook to save your selections, and check back next week to see how well you did against FiveThirtyEight and everybody else playing the game.

Your picks  Your results  Leaderboards

Week 20

2/2 forecasts completed

Sunday, Jan. 20 at 3:05 p.m.

L.A. Rams : 81%
Away

New Orleans : 64%
Home

FiveThirtyEight's forecast

You'll lose 81.2 points if the L.A. Rams win,
You'll gain 42.8 points if New Orleans wins.

Sunday, Jan. 20 at 6:40 p.m.

New England : 55%

Kansas City : 6.1%

FiveThirtyEight's forecast

You'll lose 10.4 points if New England wins,
You'll gain 9.6 points if Kansas City wins.

Source: FiveThirtyEight.com
Key Innovation – Application of Brier Scoring to Differential Diagnosis

- Sum of the squares of the difference between predicted and actual outcome

\[ BS = \frac{1}{N} \sum_{t=1}^{N} \sum_{i=1}^{R} (f_{ti} - o_{ti})^2 \]

- Score is affected by:
  - Uncertainty: Is the event easy to predict \textit{a priori} (the child is very unlikely to have coronary artery disease) or difficult to predict (the child might or might not have pneumonia)
  - Reliability: Do I predict bronchiolitis at an overall rate similar to what is observed?
  - Resolution: How well are predictions adjusted to particular circumstances?
Scaled Brier Scores, By Assigned Probability

-5
-4
-3
-2
-1
0
1
2
3
4
5

0 20 40 60 80 100

Probability Assessed To a Given Diagnosis

Scaled Brier Score

SCORE IF PATIENT HAS DIAGNOSIS
SCORE IF PATIENT DOES NOT HAVE DIAGNOSIS
The score is a function of BOTH picking the right diagnosis, AND assessing appropriate level of confidence.

<table>
<thead>
<tr>
<th>Accuracy in diagnosis</th>
<th>Accuracy in appraising level of confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smart, but overconfident</td>
<td>Master clinician</td>
</tr>
<tr>
<td>Dangerous</td>
<td>Appropriately cautious, developing clinical acumen</td>
</tr>
</tbody>
</table>
## Example: Card Drawing Predictions

<table>
<thead>
<tr>
<th>Probability of...</th>
<th>Result</th>
<th>Player 1</th>
<th>Score</th>
<th>Player 2</th>
<th>Score</th>
<th>Player 3</th>
<th>Score</th>
<th>Player 4</th>
<th>Score</th>
<th>True probability</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red card in 1 draw</td>
<td>YES</td>
<td>50%</td>
<td>2.5</td>
<td>50%</td>
<td>2.5</td>
<td>50%</td>
<td>2.5</td>
<td>50%</td>
<td>2.5</td>
<td>50%</td>
<td>2.5</td>
</tr>
<tr>
<td>One 2, 3, or 4 card in 1 draw</td>
<td>YES</td>
<td>20%</td>
<td>-1.4</td>
<td>23%</td>
<td>-0.9</td>
<td>20%</td>
<td>-1.4</td>
<td>22%</td>
<td>-1.1</td>
<td>23%</td>
<td>-0.9</td>
</tr>
<tr>
<td>At least 1 face card in 1 draw</td>
<td>NO</td>
<td>50%</td>
<td>2.5</td>
<td>23%</td>
<td>4.5</td>
<td>40%</td>
<td>3.4</td>
<td>45%</td>
<td>3.0</td>
<td>23%</td>
<td>4.5</td>
</tr>
<tr>
<td>2 Face Cards in 2 draws</td>
<td>NO</td>
<td>12%</td>
<td>4.9</td>
<td>5%</td>
<td>5.0</td>
<td>20%</td>
<td>4.6</td>
<td>8%</td>
<td>4.9</td>
<td>5%</td>
<td>5.0</td>
</tr>
<tr>
<td>2 Black Cards in 2 draws</td>
<td>NO</td>
<td>25%</td>
<td>4.4</td>
<td>25%</td>
<td>4.4</td>
<td>25%</td>
<td>4.4</td>
<td>25%</td>
<td>4.4</td>
<td>25%</td>
<td>4.4</td>
</tr>
<tr>
<td>One black, one red in 2 draws</td>
<td>YES</td>
<td>25%</td>
<td>-0.6</td>
<td>25%</td>
<td>-0.6</td>
<td>20%</td>
<td>-1.4</td>
<td>50%</td>
<td>2.5</td>
<td>51%</td>
<td>2.6</td>
</tr>
<tr>
<td>exactly one red card in 3 draws</td>
<td>YES</td>
<td>12%</td>
<td>-2.7</td>
<td>50%</td>
<td>2.5</td>
<td>60%</td>
<td>3.4</td>
<td>25%</td>
<td>-0.6</td>
<td>38%</td>
<td>1.2</td>
</tr>
<tr>
<td>At least one 5 in 3 draws</td>
<td>NO</td>
<td>20%</td>
<td>4.6</td>
<td>8%</td>
<td>4.9</td>
<td>10%</td>
<td>4.9</td>
<td>10%</td>
<td>4.9</td>
<td>22%</td>
<td>4.5</td>
</tr>
<tr>
<td>One pair in 5 card draw</td>
<td>YES</td>
<td>28%</td>
<td>-0.2</td>
<td>6%</td>
<td>-3.8</td>
<td>1%</td>
<td>-4.8</td>
<td>15%</td>
<td>-2.2</td>
<td>42%</td>
<td>1.6</td>
</tr>
<tr>
<td>Two pair in 5 card draw</td>
<td>NO</td>
<td>10%</td>
<td>4.9</td>
<td>0.5%</td>
<td>5.0</td>
<td>0.5%</td>
<td>5.0</td>
<td>5%</td>
<td>5.0</td>
<td>5%</td>
<td>5.0</td>
</tr>
<tr>
<td>No face cards or aces</td>
<td>NO</td>
<td>3%</td>
<td>5.0</td>
<td>75%</td>
<td>-0.6</td>
<td>1%</td>
<td>5.0</td>
<td>3%</td>
<td>5.0</td>
<td>15%</td>
<td>4.8</td>
</tr>
</tbody>
</table>

Player 1 Total: 23.8  Player 2 Total: 22.7  Player 3 Total: 25.6  Player 4 Total: 28.2  Score: 35.1
On to the fun stuff!

• Get in teams of 5-6 people
• Designate one team leader who will enter answers
• Use computer (ideal) or phone and access the Google sheet which was sent out in advance
• You will be given a short description of an actual case
• You will have only about 2 minutes per case – please discuss, then enter estimated probabilities of each diagnosis in percentage.
• In these cases, only one diagnosis is correct, so probabilities should add up to 100%
• “Other” may well be the correct diagnosis
CASES

- Attendees will participate in a game of probabilistic diagnostic reasoning
- Cases will not be distributed in advance and are not included in this powerpoint
Benefits of this teaching tool?
Applications
Keys to to the probabilistic reasoning game conference

- Use actual case series – not invented cases
- Choose representative cases - do not select for unexpected twists (not morning report cases”)
- Need a high volume of cases for scoring to be valid – quick discussion on each, try to pare down case presentation to minimum necessary information
- Low/high probability events require a greater number of cases
- Diagnostic categories need to be broad enough that one can reasonably narrow down to 4-5 likely categories (i.e., “pneumonia” rather than “pneumococcal pneumonia.”)
Future questions

• Does this exercise improve assessment of confidence?
• What methods (for example, emphasizing epidemiology, or teaching “anchor and adjust” method) can improve assessment of pretest probability?
• Can ability to accurately describe probabilities translate into improved clinical decision making?
“Though some have argued that doctors are not good gamblers, by more clearly identifying the issues and sources of uncertainty and using probabilistic thinking to guide reasoned bets, physicians can become better consumers of data” (Jenssen et al, 2015)
References

• Hickam DH, Sox HC. Teaching Medical Students to Estimate Probability of Coronary Artery Disease. JGIM 1987; 2:73-77.
• Upshur REG. Two Techniques for Teaching the Estimation of Prior Probabilities. Teach Learn Med 2000; 12(3): 141-4