Work-based Assessment and Outcomes-based Systems

Eric Holmboe
Early Principles: CBME

• World Health Organization (1978):
  • “The intended output of a competency-based programme is a health professional who can practise medicine at a defined level of proficiency, in accord with local conditions, to meet local needs.”

Work-based Assessment

- The essential element in a competency-based system
- What matters most to patients and healthcare systems
- Messy yes, but a poor excuse for not doing it and performing the research to do it better
  - Standardized testing was once a young science too
- What we’ve been doing for decades in assessment no longer sufficient if we are to meet societal needs
The Ubiquitous and Iconic Miller “Pyramid (1990)

Does (action)

Shows How (performance)

Knows How (competence)

Knows (knowledge)

Performance in Practice/Multi-source feedback/Direct Observation

Standardized Patients/Simulation

Diagnostic Reasoning using clinical vignettes or CSR

Multiple choice Questions

Dominated the last century
<table>
<thead>
<tr>
<th>Measure Focus</th>
<th>Measure Name/Description</th>
<th>Baseline Rate</th>
<th>Most Recent Rate</th>
<th>Aspirational Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin Use</td>
<td>Outpatient visits at which adults with cardiovascular disease are prescribed/maintained on aspirin</td>
<td>47%&lt;sup&gt;13&lt;/sup&gt;</td>
<td>53%&lt;sup&gt;14&lt;/sup&gt;</td>
<td>Increase to 65% by 2017</td>
</tr>
<tr>
<td>Blood Pressure Control</td>
<td>Adults with hypertension who have adequately controlled blood pressure</td>
<td>46%&lt;sup&gt;15&lt;/sup&gt;</td>
<td>53%&lt;sup&gt;16&lt;/sup&gt;</td>
<td>Increase to 65% by 2017</td>
</tr>
<tr>
<td>Cholesterol Management</td>
<td>Adults with high cholesterol who have adequate control</td>
<td>33%&lt;sup&gt;17&lt;/sup&gt;</td>
<td>32%&lt;sup&gt;18&lt;/sup&gt;</td>
<td>Increase to 65% by 2017</td>
</tr>
<tr>
<td>Smoking Cessation</td>
<td>Outpatient visits at which current tobacco users received tobacco cessation counseling or cessation medications</td>
<td>23%&lt;sup&gt;19&lt;/sup&gt;</td>
<td>22%&lt;sup&gt;20&lt;/sup&gt;</td>
<td>Increase to 65% by 2017</td>
</tr>
</tbody>
</table>
## Faculty OSCE Clinical Skills

<table>
<thead>
<tr>
<th>Competency</th>
<th>Mean (SD)</th>
<th>Range</th>
<th>Generalizability</th>
</tr>
</thead>
<tbody>
<tr>
<td>History Taking</td>
<td>65.5% (9.6%)</td>
<td>34% - 79%</td>
<td>0.80</td>
</tr>
<tr>
<td>Physical Exam</td>
<td>78.9% (13.6%)</td>
<td>36% - 100%</td>
<td>0.52</td>
</tr>
<tr>
<td>Counseling</td>
<td>77.1% (7.8%)</td>
<td>60% - 93%</td>
<td>0.33</td>
</tr>
<tr>
<td>Patient Satisfaction¹</td>
<td>5.62 (0.48)</td>
<td>4.43 – 6.63</td>
<td>0.60</td>
</tr>
</tbody>
</table>

¹On 7-point scale

N=44

The Problem of Uncertainty

From Lineberry et.al. critique of Script Concordance Testing:

*The former [expert opinion] is a unique aspect of SCTs unaccounted for by classical test theory (CTT)*...

We are unaware of any psychometric theory sophisticated enough to guide estimation of the adequacy or inadequacy of sampling from ‘any population frequency distribution that may be observed across a 5-point scale’

Really, isn’t uncertainty something MDs do everyday?!
Clinical Questions Raised by Clinicians at the Point of Care

Systematic Review (11 articles):

- Mean frequency of questions $= 0.57$ per patient visit
  - GIM Physician: Roughly 90 visits/week
  - Assuming conservative estimate of 46 weeks of work/year
    - $\# \text{ of questions} \approx 2,228$ per year or $\approx 22,000+ \text{ over 10 years}$

- Physicians pursued 51% (range 36-66%) of questions
  - Found answers for 78% (range 67-88%)

*No amount of testing is going to “fix this”*

Comprehensive Care Project

- 236 GIM practices across US
- 46 total performance measures
- Patient demographics, characteristics and scope of practice varied widely

Table 2: Mean Patient Characteristics Measured at the Physician Level (Chart Review of 22,526 Patients among 236 Physicians)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean (SD)</th>
<th>Range across Physicians</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>60 (6.0)</td>
<td>44–77</td>
</tr>
<tr>
<td>Patients with age ≥ 65</td>
<td>44% (0.16)</td>
<td>5–89%</td>
</tr>
<tr>
<td>Male gender</td>
<td>40% (0.12)</td>
<td>10–75%</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic white</td>
<td>37% (0.32)</td>
<td>0–99%</td>
</tr>
<tr>
<td>Non-Hispanic black</td>
<td>9% (0.18)</td>
<td>0–100%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>8% (0.18)</td>
<td>0–100%</td>
</tr>
<tr>
<td>Unknown race</td>
<td>46% (0.37)</td>
<td>0–100%</td>
</tr>
<tr>
<td>Drug/alcohol abuse</td>
<td>2% (0.02)</td>
<td>0–13%</td>
</tr>
<tr>
<td>Psychiatric/cognitive impairment</td>
<td>15% (0.15)</td>
<td>0–60%</td>
</tr>
<tr>
<td>Problems with adherence to treatment</td>
<td>5% (0.08)</td>
<td>0–52%</td>
</tr>
<tr>
<td>Social factors limiting self-care</td>
<td>13% (0.18)</td>
<td>0–100%</td>
</tr>
<tr>
<td>Comorbidity index</td>
<td>1.0 (0.55)</td>
<td>0.02–3.5</td>
</tr>
<tr>
<td>Patients with comorbidity index &gt; 2</td>
<td>11% (0.09)</td>
<td>0–48%</td>
</tr>
</tbody>
</table>
Results

- Wide range of performance across measures and composites (shown to left)
- At best, performance on the MOC exam only explained 4.9% of the variance of performance on the composite measures
- If “does” (performance) matters most to patients, then perhaps this is where we need more effort

High Stakes Testing as Normal Science

Thomas Kuhn (1962):

“Normal science, the activity in which most scientists inevitably spend almost all of their time, is predicated on the assumption that the scientific community *knows* what the world is like. Much of the success of the enterprise derives from the community’s willingness to defend that assumption, if necessary at considerable cost”

Cambridge Model: “Righting” the Pyramid

System related influences

Performance

Competence

Individual related influences

Work-based assessment has to be a larger part of our future if we are to make meaningful gains in quality and safety

Rethans, Norcini, et al, 2002
Final Thoughts

“Seek out new knowledge and be willing to engage in dialogue about disconfirming data that challenges past beliefs. Learn and listen from others, most importantly from physicians and patients who sit at the core of the education and care experience. Be open to bringing the best ideas home for testing, trial and adaptation for improvement. All organizations and programs, including the regulatory entities, must embrace the obligation to investigate what they do, admit limitations, and improve for the benefit of the public.”

From: Holmboe ES and Batalden PB. Achieving the Desired Transformation: Thoughts on Next Steps for Outcomes-based Medical Education. Acad Med. 2015; in press.