Assessing individual clinical performance - what the research says

Ian Scott MBBS, FRACP, MHA, MEd

Director of Internal Medicine and Clinical Epidemiology
Princess Alexandra Hospital
Associate Professor of Medicine
University of Queensland
Brisbane

Queensland Clinical Senate
31/10/14
Differentiating performance from competence

Individual related factors

Competence
What clinicians can do in professional practice

Performance
What clinicians actually do in professional practice

System related factors

Consultant
Advanced trainee
Basic trainee

Adapted from Rethans 2002
For what purpose?

- Maximise high value care
- Promote professional development
- Determine remuneration
- Recertify or revalidate
- Identify poor performers

TEAM

Increasing requirement for valid, specific, reliable performance measures
More resource intense
More frequent

INDIVIDUAL
What to measure?

- Communication
- Diagnostic accuracy
- Care appropriateness
- Professional conduct
How to measure?

• **Implicit vs explicit methods**
  - Intuitive ratings or impressions
    vs
  - Structured data to which specific criteria or standards applied
    • Explicit reputed to be more objective and reliable, esp if multiple reviewers

• **Direct vs indirect**
  - Evaluating real practice/behaviour
    vs
  - Assessment of skills and behaviour in arbitrary or contrived test environments
    • High fidelity testing - in-cognito SPs, clinical vignettes - chart reviews
## How to measure?

<table>
<thead>
<tr>
<th>Measurement instrument</th>
<th>Performance attributes being measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chart audits</td>
<td>Clinical expertise, reasoning</td>
</tr>
<tr>
<td>Peer practice reviews</td>
<td>All attributes</td>
</tr>
<tr>
<td>Mini-CEX</td>
<td>Clinical expertise, reasoning, communication, professionalism</td>
</tr>
<tr>
<td>Multisource feedback</td>
<td>All attributes</td>
</tr>
<tr>
<td>Significant incident analysis</td>
<td>Clinical expertise, decision-making, professionalism, learning/teaching</td>
</tr>
<tr>
<td>Practice logs, QA activities, self-assessments</td>
<td>Clinical expertise, reasoning, teaching and learning</td>
</tr>
<tr>
<td>High fidelity procedural simulations</td>
<td>Technical expertise</td>
</tr>
<tr>
<td>Patient satisfaction surveys</td>
<td>Communication, health advocacy, professionalism</td>
</tr>
<tr>
<td>Administrative data (deaths, readmissions, complications, LOS)</td>
<td>Clinical expertise</td>
</tr>
</tbody>
</table>
How good are the measures?

• Measurement standards
  - Clinical relevance
  - Scientific soundness
    • Based on process-outcome relationships
      - Expert consensus
      - Clinical trials
    • Good psychometric properties
      - Validity, reliability
    • Adequate sampling
    • Adjustment for patient/environmental confounders
  - Attribution accuracy and controllability
  - Timeliness
  - Metric balance
  - Ease of use, feasibility, sustainability
  - No unintended adverse consequences
How good are the measures?

- 55 tools for directly assessing clinical competence¹
  - 21 tools for students; 32 for residents/fellows
  - Rater training described for 26 tools
  - Only 11 tools had validity evidence based on internal structure and relationship to other variables
  - Self-assessed changes in KSA (n = 9) or objectively measured change in KS (n = 5)
  - *Strongest validity evidence for Mini-CEX*

- 30 tools for directly assessing psychomotor skills²
  - Validity evidence identified for 24 tools
  - Internal consistency in 14, test-retest reliability in 5, inter-rater reliability in 20
  - Effect on KSA reported for 5 tools
  - *Strongest evidence for 7-item Global Rating Scale and the Procedure-Based Assessment*

1. Kogan et al JAMA 2009
How good are the measures?

- 43 articles on MSF: high reliability, validity, feasibility
  
  - In general minimum of 8 medical colleagues, 8 co-workers, 25 patients to achieve reliability ($\alpha \geq 0.90$) and generalizability ($Ep2 \geq 0.80$)
  
  - Positive ratings: patients - co-workers - resident peers - faculty - consultant evaluators
  
  - Variance in measures of four competencies - patient management, clinical assessment, communication, professional development - varied according to specialty
    - 73% for internal medicine physicians
    - 70% for psychiatrists
    - 68% for pediatricians

- Procedural competence for surgeons require additional assessment methods

1. Donnon et al Acad Med 2014
## How good are the measures?

<table>
<thead>
<tr>
<th>Measurement instrument</th>
<th>Measurement properties</th>
<th>Overall utility</th>
<th>Performance attributes being measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chart audits</td>
<td>+++</td>
<td>++</td>
<td>Clinical expertise, reasoning</td>
</tr>
<tr>
<td>Peer practice reviews</td>
<td>++</td>
<td>+</td>
<td>All attributes</td>
</tr>
<tr>
<td>Mini-CEX</td>
<td>+++</td>
<td>+++</td>
<td>Clinical expertise, reasoning, communication, professionalism</td>
</tr>
<tr>
<td>Multisource feedback</td>
<td>+++</td>
<td>++</td>
<td>All attributes</td>
</tr>
<tr>
<td>Significant incident analysis</td>
<td>+</td>
<td>+</td>
<td>Clinical expertise, decision-making, professionalism, learning/teaching</td>
</tr>
<tr>
<td>Practice logs, QA activities, self-assessments</td>
<td>++</td>
<td>++</td>
<td>Clinical expertise, reasoning, teaching and learning</td>
</tr>
<tr>
<td>High fidelity procedural simulations</td>
<td>++</td>
<td>++</td>
<td>Technical expertise</td>
</tr>
<tr>
<td>Patient satisfaction surveys</td>
<td>++</td>
<td>+</td>
<td>Communication, health advocacy, professionalism</td>
</tr>
<tr>
<td>Administrative data</td>
<td>++</td>
<td>++</td>
<td>Clinical expertise</td>
</tr>
</tbody>
</table>
What are the benefits of performance assessment?

Overeem et al Med Educ 2007

- 64 studies of 6 different methods
  - simulated patients, video observation, direct observation, peer assessments, chart audits, portfolio

- 21 studies assessed effectiveness
  - 8 - improved learner satisfaction
  - 4 - encouraged learning objectives, CPD plans
  - 12 - self-reported changes or intended changes in practice
  - 2 - improved referral letters, chart documentation

- No studies of effects on patient care or outcomes
What are the benefits of performance assessment?

Miller & Archer BMJ 2010

- 16 studies
  - MSF (8), mini-CEX (4), direct observation of procedural skills (1), multiple methods (3)
  - Self-reported outcomes, mostly educational impact

- MSF: strong evidence of educational impact; mixed results for performance change
  - More so for family physicians, less so for surgeons
  - More so with credible feedback associated with coaching

- Remaining studies: positive educational value but no objective evidence of improved performance

- Multiple assessment methods
  - Time consuming; administrative workload; neutral or negative impact on training for surgeons, positive impact for physicians
What are the benefits of performance assessment?

• Audits and feedback
  • Jamveldt et al Cochrane Database Syst Rev 2012

• Surgical registries
  – Improve surgeon performance and more appropriate patient selection
    • NY cardiac surgeons

• Profiling of minority of clinicians at high risk of recurrent medico-legal complaints
  • Bismark et al BMJ Qual Safe 2013
What are the limitations of performance measures?

• Few studies of long-term clinical impact and effectiveness of individual performance assessment
  - Observed changes in performance small to moderate
  - No study has assessed effect on patient outcomes

• Few measures of
  - diagnostic error
  - overuse or inappropriate care
  - ability to manage complex, multi-morbid patients with psychosocial issues

• Most measures do not capture other factors that impact outcomes
  - Teamwork, leadership, innovation, cultures

• Thresholds for determining adequate performance levels for process measures differ according to frequency of clinical condition, evidence-base, psychometric properties, sampling rates
What are the limitations of performance measures?
What are limitations of performance measures?

• Validity and reliability of many measures questionable
  - May not measure effectiveness or appropriateness of actual care
  - Some process measures have no relation to patient outcomes
  - Most measures sample very small fraction of clinical actions (~10-15%)
  - Measures may not adequately reflect task complexity and context
  - Issues around data accuracy - will EMR save us?
  - Evidence for 'pay for performance' very mixed
  - Non-MSF measures may not identify poor performers

• Unintended effects
  - Gaming, patient exclusions, treating to the measure, neglect of non-measured areas of practice, surveillance bias
  - Crowds out intrinsic motivation - esp for high level executive tasks
    • Strongest negative impact: measures contingent on very specific task performance (micromanagement) or associated with unrealistic targets, deadlines, sanctions
  - Opportunity costs

• Competing measures - catch 22
  - Conflicting aims and targets, good care vs financial bottom-line
Recommendations for individual performance measures

• Develop a positive culture
  - Constructive not punitive

• Be clear about purpose of performance measurement
  - Means for assessing and potentially improving patient care
  - Identifying very small minority of poorly performing individuals

• Ensure clinicians involved in choosing assessment methods, adequately trained in the use of assessment methods, fully aware of their limitations

• Professional attributes regarded as important must be targets for assessment
Recommendations for individual performance measures

- Multiple assessment methods using multiple data sources preferred
  - Overcomes content (or skill) specificity and bias or inaccuracy involving data sources

- High sampling rates for multiple less structured assessment methods gives best picture of overall performance

- Sufficient resources and physician time to allow adequate collection and analysis of data, feedback and debriefing

- Combination of measurement feedback, reflection and mentoring helps - insightful practice
Future directions

• Move from measuring processes to outcomes

• Use performance measures strategically and dynamically
  - To address contemporary care challenges and encourage innovation
  - BUT give incentives/resources and make sure targets are SMART

• Measure performance at the level of the microsystem rather than individual clinician
  - Take KPIs out of individual contracts and into microsystem accountable care contracts

• Use aggregate measures that facilitate peer comparisons and maximise positive impact of professional competitiveness

• Use performance measurement to promote rapid-learning healthcare organisations (big data, BI, EMR)

• Evaluate effects, applications, ROI and adverse effects of performance measures

• Standardise performance measures and methods of data collection, analysis and reporting across every HHS