Meaningful measurement of individual performance

Ian Scott

Director of Internal Medicine and Clinical Epidemiology
Princess Alexandra Hospital
Associate Professor of Medicine
University of Queensland
Brisbane
Adjunct Assoc Professor of Medicine
Monash University
Melbourne

Quantum Leap Conference
Sydney 24/9/12
Definitions

- **Competence**: ‘what the doctor has been trained to do’
  - involves acquiring and maintaining the requisite knowledge, skills and behaviours to perform at or above the minimum standard

- **Performance**: ‘what the doctor actually does from day to day’
  - amenable to ongoing improvement through the development of both technical and non-technical knowledge, skills and behaviours of the medical practitioner over the course of their professional career
Differentiating performance from competence

Competence
What clinicians can do in professional practice

Performance
What clinicians actually do in professional practice

Individual related factors

System related factors

Consultant
Advanced trainee
Basic trainee

Adapted from Rethans 2002
“It is no longer enough to do a job to the best of one’s ability”

Quality Measures and the Individual Physician
Danielle Ofri, M.D., Ph.D.
N Engl J Med 2010; 363 (7): 606-7

Balancing “No Blame” with Accountability in Patient Safety
Robert M. Wachter, M.D., and Peter J. Pronovost, M.D., Ph.D.

What are the main influences on individual behaviour?
• Strong belief in the efficacy of our personal actions
• What our peers are doing

Wakefield et al Qual Saf Health Care 2011
Why the need for performance assessment

- Identify under-performers
  - Impaired
    - Depression, anxiety, substance abuse, physical illness, cognitive impairment
    - Problem doctors: antisocial or disruptive behaviour
  - Incompetent
    - Less than acceptable level of knowledge or skill

- Guide professional development and helping all physicians to continuously improve

- Facilitate external validation for healthcare stakeholders
Prevalence of under-performers

• 17% physicians aware of impaired/incompetent colleague

• 8% to 15% suffer substance abuse

• 3% to 5% are 'problem doctors'

• 10% restrict practice for a prolonged period due to disabling physical illness

• 33% will experience, at some point in their career, a period during which they have a condition which impairs their ability to practice safely

• 5% to 12% will demonstrate significant deficiencies in knowledge or skills at some point in their career
Main domain problems

- Communication
- Diagnostic accuracy
- Care appropriateness
- Professional conduct
Decline in performance over time

• Systematic review 62 studies
  - Most found declines in physician performance with time
    • Decreasing medical knowledge
    • Less adherence to standards of appropriate diagnosis, screening, preventive care, therapy
    • Attrition of clinical skills
    • Worse health outcomes

Choudhry et al Ann Intern Med 2005
Self-assessment alone does not work

- Physicians’ ability to independently and accurately assess and evaluate their own performance is poor
  - Only 7/20 (35%) comparisons between self- and external assessment showed positive correlation
  - Worst self-assessment accuracy among least skilled or most confident
    - Davis et al JAMA 2006

- Physicians typically overestimate their adherence to quality standards of care
  - Holmboe et al J Contin Educ Health Prof 2006

- Fewer than 30% of physicians undertake clinical audits or performance reviews of their own practice
  - Audet et al Health Aff 2005

- Voluntary performance assessment systems demonstrate low participation rates and fail to produce desired practice change
External regulation is not enough

• No evidence that externally regulated performance systems
  - assure competence of individual practitioners
  - foster reflection or CQI
  - prevent widely publicised failures that prompted their implementation

• Why not?
  - Clinical performance influenced by multiple domains of practice which interact to result in standard of care
  - Content specific knowledge and skills not sufficient to guarantee performance
  - Non-technical areas of performance, when less developed, negatively influence the standard of practice
Improving performance is difficult

- Traditional CPD does not guarantee improved clinician knowledge and performance
  - Mazmanian & Davis JAMA 2002
  - Davis et al JAMA 1995

- Clinicians tend to pursue education around topics they are already good at while avoiding areas in which there may be room for improvement

- Clinical audits and feedback improve guideline concordant care by no more than 10%
  - Jamtvedt et al Cochrane Database Syst Rev. 2011

- Providing patients, managers and others with data on clinician performance does not significantly impact on their perceptions or choice of clinician
Avoiding the problem

• Very few healthcare organisations or medical societies systematically monitor physician performance, have formal programs to identify serious under-performers, or provide remediation programs
  » Neff Am Coll Phys Exec 2000
  » Humphrey J Continu Educ Health Prof 2010

• Hospital credentialling and disciplinary processes lack granularity and standardisation

• Many physicians are reluctant to act when faced with impaired or incompetent colleagues
  - US study: 33% to 36%
    » DesRoches et al JAMA 2010
  - NZ study: 20%
    » Raniga et al NZ Med J 2005
  - UK study: 6%
    » Cooke et al. Med Educ 2001
Measuring individual performance is challenging

- Evidence- or consensus-based
- Agreed standards for satisfactory performance
- Sound measurement properties
  - Validity
  - Reliability
    - Internal consistency, test-retest reliability (or reproducibility), intra-rater reliability, inter-rater reliability
  - Discrimination
  - Response process
  - Standardised specifications
  - Adequate sampling
  - Adjustment for confounding patient/system factors
- Attribution accuracy
- Timeliness
- Feasibility and ease of use
- Little potential for unintended adverse consequences

Scott et al Intern Med J 2011
Measuring individual performance is challenging

Effects of casemix on clinicians quality rankings

n=181 PCPs
9 quality measures
Breast/colorectal cancer screening, HbA1c, lipids
Patient xistics
age, co-morbidity, SES, ethnicity

Hong et al JAMA 2010
Choice of assessment methods

Implicit vs explicit measures
• Implicit judgements rely on global ratings or impressions
• Explicit measures based on structured data with explicit criteria

Direct vs indirect assessment
• **Direct performance measures**
  - Audit of medical charts or clinical registries or administrative datasets
  - Standardised (incognito) patients
  - Mini-CEX
  - Multi-source feedback (360° appraisals)
  - Sentinel (or significant) event analysis
  - Direct observation of clinical interactions, procedural skills
  - Video-recordings of clinical interactions
  - Chart-stimulated recall interviews
  - Patient satisfaction interviews and surveys
  - Evidence-based practice logs or portfolios

• **Indirect performance measures**
  - Clinical vignettes (paper or computer based)
  - Objective structured clinical examinations (OSCEs)
  - Oral viva examinations (long and short cases)
  - High-fidelity simulation exercises
  - Professional self-appraisal tools
  - Interactive, case-based small group discussions

Scott et al
Intern Med J 2011
Choice of assessment methods

- Multiple assessment methods using multiple data sources are preferred to single or a small number of methods and/or data sources – in order to overcome content (or skill) specificity and bias or inaccuracy involving data sources

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

- Professional attributes regarded as important by clinicians must be the targets for assessment – even though this may pose methodological challenges

Scott et al Intern Med J 2011
# Utility of specific methods

<table>
<thead>
<tr>
<th>Tool</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Performance attributes assessed</th>
</tr>
</thead>
</table>
| **Mini-CEX** | Standardised scoring template which takes 30 minutes to complete  
Can be used repeatedly in different situations and settings  
Ability to assess different professional attributes within one exercise  
Provides immediate feedback | May be confounded by Hawthorne effects and observer bias  
Needs 6 to 10 exercises per physician to generate consistent ratings of performance due to inter-rater variability  
Limited content specificity  
Takes time | All except health advocacy; teaching and learning; leadership and management |
| **MSF** | Ability to assess different professional attributes from several colleagues from different disciplines  
Based on collective observations over an extended period of time  
Ease and rapidity of use  
Does not require large numbers of surveys | May be confounded by recall and observer bias  
Open to vindictive ratings  
Ideally requires up to 10 observers  
High levels of inter-rater variability may be seen for specific performance attributes | Virtually all |

Scott et al Intern Med J 2011
Evidence for performance assessment and feedback

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
Evidence for performance assessment and feedback

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
Four Modular Components

1. Clinician Profile and Outcomes
2. Multisource Feedback (360 Peer Review)
3. Continuing Education & Professional Development
4. Annual Mandatory Training Compliance Check
Module One: Clinician Profile & Outcomes

- Qualifications & AHPRA
- Credentials and Scope of Clinical Practice
- Benchmarking (VLAD, HRT, Clinical Audit etc.)
- Customised clinical review (DRGs, LOS, Mortality, Complications, etc.)
- Formal Complaints
- Medico-Legal Issues
- Critical Incidents
- Morbidity & Mortality Attendance
Module Two: Multisource Feedback

- Adapted from a validated tool
- Tailored to individual specialty after consultation
- Confidential & anonymous
- Includes reviews by nominated colleagues (peers, registrars, senior nursing staff, MD team)
- Scores are aggregated

* 1993 Ramsey et al, Use of peer ratings to evaluate physician performance, JAMA 1993; 269: 1655-1660
Module Three: Professional Development Plan

- Contribution to Training and Education
- Professional Development
- Research Activities
- Supervisor comments
- Clinician comments
Module Four: Mandatory Training

- Mandatory training disc developed at PAH
- Record of compliance as per QH policy
  - Corporate Induction
  - Code of Conduct
  - Harassment & Bullying
  - ATSI Cultural Awareness
  - Aggressive Behaviour Management
  - Child Safety
  - Fire & Safety
  - Manual Handling
  - Fatigue Management
Outcomes*

- 702 reviews completed
- 13,170 individual team and peer reviews distributed
- Informed performance management processes for 9 clinicians
- 3 clinicians referred to Cognitive Institute communication programs
- No breaches of confidentiality
- No grievances / appeals / complaints
- 30 DMS’s attended Training Workshop

*November 2011
Individual Clinician Value

Formal evaluation demonstrating:

- Increased SMO PAD compliance from 2% to 98%
- Most respondents agreed process was meaningful 60%
- Most respondents agreed SMPR produced valid information 61%
- Most respondents agreed feedback would inform their practice 68%
Organisational & Professional/Personal Outcomes

- Identification of a small group of clinicians who scored statistically significant lesser outcomes than peers at the 99% confidence level.
- Identification of significant communication and teamwork concerns requiring 3rd party involvement.
- Opportunity to discuss academic, managerial and clinical governance involvement by SMO’s/VMO’s
Key features

• Develop a positive culture
  • Constructive not punitive

• Be clear about purpose of performance measurement

• Clearly express any desired behaviours

• When using MSF
  • Keep the number of items to 12
  • Keep the scale simple and fit for purpose
  • Use 6 to 10 raters
  • Compare results with self-assessment
  • Train those giving feedback
  • Involve assesses
Need for more insightful practice

- Insightful practice: engagement, insight and action in response to suite of feedback
- Quality improvement: meeting personal objectives
- Professional support: facilitated feedback appraisal
- Independent feedback on performance: suite of contextualised feedback
- Reflection on credible evidence from practice: setting personal objectives for improvement
- Responsibility and accountability
## Insightful practice

<table>
<thead>
<tr>
<th>Tool</th>
<th>Source</th>
<th>Prepared by</th>
</tr>
</thead>
</table>
| Multi-source feedback (MSF)*              | General Medical Council (GMC) colleague survey\[^1\]^\[^8\]^\[^9\]  
                                          | 2Q MSF\[^18\]^\[^20\]                   | GMC                               |
| Patient satisfaction questionnaires*      | GMC patient survey\[^18\]^\[^19\]          | Developed by study author         |
|                                           | Consultation and relational empathy\[^16\]^\[^21\] | GMC                               |
|                                           |                                             | Developed by study authors        |
| Open book self-assessed knowledge test    | Consisted of 60 items focusing on chronic disease management, referral issues and prescribing | Royal College of General Practitioners (RCGP Scotland) |
| Prescribing safety data feedback†         | 12 measures of undesirable co-prescriptions\[^18\]^\[^22\] | Developed for study               |
| Quality of care data feedback            | Single area of interest selected for each participant’s practice by an external assessor\[^18\] | Web-based report                  |
| Patient complaints                        | -                                           | As received                       |

*For the purpose of the research study programme, participants collected and reflected on output from two patient satisfaction questionnaires and two MSF questionnaires, both on two occasions, in order to test the reliabilities of individual tools. In any real system, only one tool would be used and the collection of data would likely be spread over a longer period of time. The reliabilities of individual tools are not reported here.

†These data on 12 undesirable co-prescriptions were developed for the purpose of this study.\[^18\]^\[^22\] Other tools used are available to GPs to include when considering data for current appraisal submission.

GP, general practitioner.

---

* Murphy et al BMJ Qual Safe 2012
### Insightful practice

<table>
<thead>
<tr>
<th>Question</th>
<th>Rating scale</th>
<th>Completed by</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reflection template</strong></td>
<td>Likert 1–7*</td>
<td>GP participant</td>
</tr>
<tr>
<td>Source of feedback highlighted</td>
<td></td>
<td>Face-to-face appraiser (preappraisal)</td>
</tr>
<tr>
<td>1. Important issues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Concern in performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Led to planned change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Gave valuable feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Assessment of insightful practice template</strong></td>
<td>Likert 1–7*</td>
<td>Face-to-face appraiser (postappraisal)</td>
</tr>
<tr>
<td>Doctor demonstrated</td>
<td></td>
<td>Anonymous assessor (postappraisal)</td>
</tr>
<tr>
<td>1. Satisfactory engagement with the TIPP process</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Insight into the feedback provided on performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Plans for appropriate action where applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Engagement, insight and action (global rating of insightful practice)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Suitability for recommendation as on track for revalidation without further opinion</td>
<td>Binary yes/no</td>
<td>▶ Face-to-face appraiser (postappraisal)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>▶ Anonymous assessor (postappraisal)</td>
</tr>
</tbody>
</table>

*Likert scale descriptors (1–7): (1) strongly disagree; (3) disagree; (5) agree; (7) strongly agree.

TIPP, Tayside In-Practice Portfolio.
Insightful practice

- Significant difference between face-to-face and anonymous assessment in:
  - Mean scores of global AIP*: former scored more highly
  - Mean difference 1.07, 95% CI 0.73 to 1.41, p<0.001.
- Dichotomous judgment on GPs’ suitability for revalidation

- No portfolio was considered unsatisfactory at face-to-face assessment vs
  42/180 (23.3%) of the three anonymous markings of each of the 60 portfolios were considered unsatisfactory (p<0.001).

- Face-to-face appraisal did not discriminate between GPs and therefore could not be classed as reliable

- High reliability demonstrated by anonymous global assessment by three assessors (G= 0.85) of GPs’ insightful practice

- Recommendation on GPs’ suitability for revalidation was also highly reliable by four assessors (G=0.83.)

*Assessment of insightful practice
Insightful practice

• Combination of feedback from multiple sources, reflection and mentoring is consistent with the call for innovation in assessing professional competence
  - Answers the call for innovation in measuring professionalism to cover previously poorly tested areas of insightfully seeking and responding to feedback and results of audit

• Shows how assessment instruments might be used together to promote performance improvement
Closing comments

• Performance assessment - means for assessing and potentially improving patient care - not aimed solely at a very small minority of poorly performing individuals

• Clinicians need to be actively involved in choosing assessment methods, be adequately trained in the use of assessment methods, and be fully aware of their limitations

• Professional attributes regarded as important must be targets for assessment

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

• Currently available MSF methods should be used for formative purposes (professional development and improvement) rather than summative purposes (recertification)