Measuring and Improving Performance in Surgical Training

AAMC November 10, 2014
Gary Dunnington MD
Jay L Grosfeld Professor and Chair of Surgery
Indiana University School of Medicine
No Disclosures
The Reality of Resident Evaluation

- 20-25% of residents have performance problems identified during training
- Most problems are identified early but with delays in addressing problems
- Poor evaluation strategies yield poor diagnosis of problems
- Remediation is often “penicillin for everything”
- Most residents complete training with the same problems identified early in training
Problems in Resident Evaluation

- Inadequate sampling of performance (number of ratings, raters and skills sampled)
- Inaccuracies due to overreliance on memory
- Hidden performance deficits
- Lack of meaningful benchmarks
- Hesitancy to act
- Systematic rater error

Forecasting Residents’ Performance – Partly Cloudy
Williams, Dunnington, Klamen, Acad Med, May, 2005
Assessment of Incoming Fellows: Survey of Fellowship Council Program Directors

- 43% of all PDs felt that incoming fellows were unable to independently perform 30 minutes of a major procedure.

- 30% felt that new fellows could not independently perform basic operations, such as a laparoscopic cholecystectomy.

Mattar et al, Annals of Surgery, Sept., 2013
Presidential Address

Why Johnny cannot operate

Richard H. Bell, Jr, MD, Philadelphia, PA

I consider the performance of surgical operations to be the most complex psychomotor activity that human beings are called upon to perform. In the arts, athletics, games, and other realms of human activity, I have found nothing that matches the difficulty of surgery. Maximum return on that investment, it is critical that the procedures be done well, resulting in the best possible functional outcomes and the minimum need to intervene. As it stands now, however, 13% of general and vascular surgical patients in private sector hospitals in the United States experience complications that require re-operation.
Why Johnny (or Joanna) Can’t Operate

- Inadequate experience in residency training
- Absence of formal assessment of operative skills
Surgery Milestones PC3 and MK2

• **PC3:** Assesses proficiency in basic surgical skills to proficiency in “essential operations”

• **MK2:** Assesses knowledge for common “essential operations” to comprehensive knowledge of the pre, intra, and postoperative spectrum of surgical care
Four Problems with Using Milestones

• Evidence base supporting each item
• Human memory and information processing
• Faculty do not think about performance in multipoint scales
• Deconstruction of competence into measurable components

Placing Constraints on the Use of the ACGME Milestones
Williams, Dunnington, Mellinger, Klamen
Acad Med, in press
Evidence Base Supporting each Item

• 6/28 components of six competencies typically observed by faculty (Dunnington, Williams, 2006)

• 13 minutes direct faculty observation in 9 hour ED shift (Chisholm, 2004)

• Raters will provide a rating, with or without basis
Human Memory and Information Processing

- Using factor analysis, faculty only “see” two performance facets; clinical performance and professional behavior (Verhulst, Williams, 1986)
- In describing outstanding, average and poor residents, faculty begin with most defining attribute then elevate or discount others to “fit” (Ginsberg, 2010)
- Assessments can’t reliably measure six competencies as independent constructs (Lurie, 2009)
Faculty Do Not Think in Multipoint Rating Scales

• Faculty judgments are more categorical, dichotomous: competent or not (Gengerich, 2011)

• Since 2006 we have used rating system modeled after Consumer Reports for autos – “identify problem areas” (clinical performance, professional behavior and serious problems)

• Milestones committee “struggled” to develop satisfactory anchors for multipoint scale
Deconstruction of Competence to Measurable Components

“... the ACGME’s demand for objectivity and measurement in assessment of competence ineluctably leads to the assessment of pieces of performance in the fragmentary fashion that we have argued cannot be presumed to add up to the kind of competence we are actually interested in.”

Huddle, Heudebert, Acad Med, 82; 2007
Measuring Milestones: Two Solutions

- Abandon the milestones
- Increase workplace assessment based on direct observation over global ratings (OSATS VOP, PAME, OPRS, CAMEO)
IU Surgical Skills Laboratory
Deliberate Practice

- Well defined task
- Appropriate difficulty level for the particular individual
- Informative feedback
- Opportunities for repetition and correction of errors

Ericsson, Psychological Review, 1993
From Practice Arena to Performance Arena
Rationale for Practice Arena before Performance Arena

• Avoids exposing patients to the sharpest slope of the surgical skills learning curve
• Enhances the quality of OR teaching
• Potential to significantly decrease the number of operative cases needed for competency
• Hour for hour, it may be more efficient than OR training for early trainees (Darzi)
ACS/APDS Surgical Skills Project Phases

- **Phase I**  Basic/core skills and tasks
- **Phase II**  Advanced procedures
- **Phase III**  Team based skills
Verification of Proficiency Modules

- Knot tying
- Basic suturing
- Chest tube placement
- Emergency surgical airway
- Basic laparoscopic skills
- Laparoscopic cholecystectomy
- Central venous access
- Bowel anastomosis
- Arterial anastomosis
- EGD
- Colonoscopy
Verification of Proficiency Modules

• Review of video of expert performance
• Guided practice until performance within time standards
• Video of performance
• Blinded review by expert faculty with “pass” or “needs more practice”
• OR performance ONLY after verification (proficiency = OR ready)
Fundamentals of Endoscopic Surgery (FES)

- Measures proficiency of EGD and Colonoscopy
- Vigorous validation methodology (like FLS)
Development of the Operative Performance Rating System

• Identify commonly performed or technically challenging procedures
• Literature review to determine five most common technical factors impacting outcome
• Faculty consensus to establish rating anchors
• Addition of five validated general performance items from Toronto work
• Piloting the new instrument

Larson, Williams, Dunnington, Surg Oct 2005
Specific OPRS Case Evaluation Instruments

- Excisional biopsy
- Open inguinal hernia
- Lap chole
- Dialysis graft
- Colon resection
- Mastectomy/partial
- Thyroidectomy
- Parathyroidectomy
- Lap inguinal hernia
- Lap colectomy
- Ventral hernia
- Lap ventral hernia
- Lap appendectomy
The ABS Validation Project

- Gold standard surgeons evaluated video performances (7 per procedure)

- Procedures: lap chole, open inguinal hernia, lap ventral hernia, thyroidectomy
Resident Score Report By Half-Year

Score Categories:
- Poor
- Fair
- Good
- Very Good
- Excellent

PGY Levels:
- PGY 3.1
- PGY 3.2
- PGY 4.1
- PGY 4.2
- PGY 5.1
- PGY 5.2

Performance Levels:
- 1
- 2
- 3
- 4
- 5

Graph showing performance scores for different PGY levels from Poor to Excellent.
Personal Best: Top athletes and singers have coaches. Should You? - Atul Gawande

"No matter how well trained people are, few can sustain their best performance on their own. That's where coaching comes in." (p. 1)

"Professional athletes use coaches to make sure they are as good as they can be." (p. 2)
I Phone App for Personal Best Surgery

Laparoscopic Appendectomy Rating Form

Coach:

Performing Surgeon:

Please rate this resident's performance during this operative procedure. The caption above each item provides descriptive anchors for 3 of the 5 points on the rating scale. "N/A" (Not applicable) should only be selected when the resident did not perform that part of the procedure.
The Northwestern PASS Autonomy Scale

- Show and tell
- Active assist
- Passive assist
- Supervision only
SIMPL: System for Improving and Measuring Procedural Learning

- Combines Personal Best with Zwisch rating
- Three click system
- Case difficulty, overall quality, autonomy
- Long evaluation instrument used for remediation
Key Elements for an Operative Performance Evaluation System

• Validated instruments with technical items based on literature review
• Faculty friendly for high compliance
• Provides quality feedback to trainees
• Provides norms to Program Director and trainees
• Ten raters/resident per year
• 20 ratings/resident per year
• Ratings completed within 72 hours
Measuring Operative Performance for Surgeons in Practice: The Future of Maintenance of Certification?

Video review of 20 bariatric surgical procedures by peers assessing technical skills with correlation to patient outcomes

Bottom quartile in surgical skills associated with:

- Higher complication rate (14.5% vs 5.2%, p<0.001)
- Higher mortality (0.26% vs 0.05%, p<0.01)
- Longer operations (137 vs 98 minutes, p<0.001)
- Higher reoperation (3.4% vs 1.6%, p<0.01)
- Higher readmission (6.3% vs 2.7%, p<0.001)

Birkmeyer et al, NEJM, October 10, 2013
“If you cannot measure it, you cannot improve it.”

Lord Kelvin