Public Reporting of Surgical Outcomes

Benjamin D. Kozower, MD, MPH
Associate Professor of Surgery & Public Health Sciences
University of Virginia
Charlottesville, VA
August 12, 2015
Disclosures

Chair, STS General Thoracic Surgery Database

Member, STS Quality Measurement Taskforce
Need a new knee? Heart valve? Back surgery? This Web site could help you find the top surgeons near you.
<table>
<thead>
<tr>
<th>Select to</th>
<th>Name</th>
<th>Better Outcomes</th>
<th>Recommended Most By Surveyed Doctors</th>
<th>Board Certifications</th>
<th>Surgeon's Hospitals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adams, Reid B.</td>
<td>★★★</td>
<td>---</td>
<td>---</td>
<td>Surgery</td>
</tr>
<tr>
<td></td>
<td>Faulkenberry, William L.</td>
<td>---</td>
<td>★★★</td>
<td>---</td>
<td>Surgery</td>
</tr>
<tr>
<td></td>
<td>Branum, Gene D.</td>
<td>---</td>
<td>★★★</td>
<td>---</td>
<td>Surgery</td>
</tr>
<tr>
<td></td>
<td>Garwood, Robert Allen</td>
<td>---</td>
<td>★★★</td>
<td>---</td>
<td>Surgery</td>
</tr>
<tr>
<td></td>
<td>Oates, Thomas M.</td>
<td>---</td>
<td>★★★</td>
<td>---</td>
<td>Surgery</td>
</tr>
</tbody>
</table>

How confidently we say the surgeon's results are better-than-average: Very confidently = ★★★★★★ Confidently = ★★★★★ Probably = ★★★
Public Reporting of Surgical Outcomes

• Routine expectation of stakeholders
  – Patients, payers, legislators, policy makers

• Ethical responsibility
  – Accountability
  – Affirm patient autonomy

• Goals
  – Education of consumers
  – Improve patient care (2 mechanisms)
  – Improve resource utilization
Public Reporting of Surgical Outcomes

• Opportunity
  – Currently being done with poor methodology and administrative data
  – Need more relevant, patient-centered and long-term outcomes
  – Need robust methodologies to compare outcomes, “quality”, and performance

• Need for transparency
Problems with Current Methods for Ranking Surgeons and Hospitals …..

"I have just heard that 50% of UK cardiac surgeons are below average,...... this has got to stop!"

attributed to: The Rt Hon Frank Dobson
Former Labor Secretary for Health

“As this statement illustrates, many interested parties, including surgeons, politicians, journalists and consumers, do not understand the nuances and limitations of risk models and focus on bottom-line mortality rates with little regard for their accuracy or statistical validity”.

Shahian et al. Ann Thorac Surg 2001;72
Barriers to Public Reporting: Requirements for Comparing Institutions

- We don’t have robust, accurate data
- We don’t have relevant outcomes measures
- We need sufficient statistical power to identify meaningful differences
- We don’t know how to communicate effectively to consumers

Lansky D, Health Affairs April, 2012
Gaynor JW, With Permission 2015
Barriers to Public Reporting: Requirements for Comparing Institutions

- We don’t have robust, accurate data
- We don’t have relevant outcomes measures
- We need sufficient statistical power to identify meaningful differences
- We don’t know how to communicate effectively to consumers

Lansky D, Health Affairs April, 2012
Gaynor JW, With Permission 2015
Impact of Case Ascertainment Methodology on Hospital Rankings

Impact of Case Ascertainment Methodology on Hospital Rankings

Predicted Risk of Mortality Models: Surgeons Need to Understand Limitations of the University HealthSystem Consortium Models

Benjamin D Kozower, MD, MPH, FACS, Gorav Ailawadi, MD, FACS, David R Jones, MD, FACS, Robert D Pates, PhD, Christine L Lau, MD, FACS, Irving L Kron, MD, FACS, George J Stukenberg, PhD

BACKGROUND: The University HealthSystem Consortium (UHC) mortality risk adjustment models are increasingly being used as benchmarks for quality assessment. But these administrative database models may include postoperative complications in their adjustments for preoperative risk. The purpose of this study was to compare the performance of the UHC with the Society of Thoracic Surgeons (STS) risk-adjusted mortality models for adult cardiac surgery and evaluate the contribution of postoperative complications on model performance.

STUDY DESIGN: We identified adult cardiac surgery patients with mortality risk estimates in both the UHC and Society of Thoracic Surgeons databases. We compared the predictive performance and calibration of estimates from both models. We then reestimated both models using only patients without any postoperative complications to determine the relative contribution of adjustments for postoperative events on model performance.

RESULTS: In the study population of 2,171 patients, the UHC model explained more variability (27% versus 13%, p < 0.001) and achieved better discrimination (C statistic = 0.88 versus 0.81, p < 0.001). But when applied in the population of patients without complications, the UHC model performance declined severely. The C statistic decreased from 0.88 to 0.49, a level of discrimination equivalent to random chance. The discrimination of the Society of Thoracic Surgeons model was unchanged (C statistic of 0.79 versus 0.81).

CONCLUSIONS: Although the UHC model demonstrated better performance in the total study population, this difference in performance reflects adjustments for conditions that are postoperative complications. The current UHC models should not be used for quality benchmarks. (J Am Coll Surg 2009;209:551–556. © 2009 by the American College of Surgeons)
Comparing Administrative & Clinical Data: Apples & Oranges

• The University HealthSystem Consortium (UHC) uses decile rankings
  – Includes all patients in the cardiovascular service line
  – Cardiac surgery, cardiology, interventional cardiology, EP, etc.…

• The Society of Thoracic Surgeons (STS) uses tercile (3-star) ratings
  – Includes surgical patients operated on by cardiac surgeons
  – Only risk-adjusts for common procedures
Comparing Administrative & Clinical Data

- 28% difference in cardiac surgery patient populations
- STS appropriately uses preoperative risk factors
- UHC includes all diagnoses, including postoperative complications
Predicted Risk of Mortality

Kozower BD et al. JACS. 2009
STS National Database

The STS National Database was established in 1989 as an initiative for quality improvement and patient safety among cardiothoracic surgeons. There are three components to the STS National Database, each focusing on a different area of cardiothoracic surgery—Adult Cardiac, General Thoracic, and Congenital Heart Surgery, with the availability of Anesthesiology participation within the Congenital Heart Surgery Database. The Database has grown exponentially over the years, both in terms of participation and stature. Learn more in the STS National Database Brochure.

- Adult Cardiac, General Thoracic, Congenital
- 95% data accuracy, externally audited
- Predictive risk models for common procedures
Barriers to Public Reporting: Requirements for Comparing Institutions

- We don’t have robust, accurate data
- We don’t have relevant outcomes measures
- We need sufficient statistical power to identify meaningful differences
- We don’t know how to communicate effectively to consumers

Lansky D, Health Affairs April, 2012
Gaynor JW, With Permission 2015
The Swedish Approach to Hip Replacement

- Reported health gain (EQ-5D index gain) after one year
- Patient satisfaction at one year
- Complications at two years
- Ten-year implant survival
- None of these measures are currently reported in the US
- 90% follow up recorded in Sweden, Publicly Reported

Swedish Hip Arthroplasty Register 2012
Relevant Thoracic Surgery Outcomes:
Measuring Hospital Performance

STS Database Risk Models: Predictors of Mortality and Major Morbidity for Lung Cancer Resection

Benjamin D. Kozower, MD, MPH, Shubin Sheng, PhD, Sean M. O’Brien, PhD, Michael J. Liptay, MD, Christine L. Lau, MD, David R. Jones, MD, David M. Shahian, MD, and Cameron D. Wright, MD

Departments of Surgery & Public Health Sciences, University of Virginia Health System, Charlottesville, Virginia; Duke Clinical Research Institute, Duke University, Durham, North Carolina; Department of Cardiovascular and Thoracic Surgery, Rush University, Chicago, Illinois; and Division of Thoracic Surgery and Center for Quality and Safety, Massachusetts General Hospital, Boston, Massachusetts

Objectives

1. To create models for perioperative risk of lung cancer resection using the STS General Thoracic Surgery Database

2. To develop risk models that measure variation in hospital performance for quality improvement purposes
Results: Patient Population

19,026 lung cancer surgery pts (100%)

- Benign disease (n=50)
- Extrapleural pneumonectomy (n=34)
- Missing age, gender or mortality (n=142)

18,800 at 111 centers (98.8%)
## Mortality and Morbidity (30 Day or In-Hospital)

<table>
<thead>
<tr>
<th>Category</th>
<th>Count (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td>413 (2.2%)</td>
</tr>
<tr>
<td>Major complications</td>
<td>1,491 (7.9%)</td>
</tr>
<tr>
<td>Length of stay</td>
<td>5 days</td>
</tr>
</tbody>
</table>
Hospital Performance Variation:
Mortality or Major Morbidity – STS Database
18,800 lung cancer resections from 111 hospitals

Barriers to Public Reporting: Requirements for Comparing Institutions

- We don’t have robust, accurate data
- We don’t have relevant outcomes measures
- We need sufficient statistical power to identify meaningful differences
- We don’t know how to communicate effectively to consumers

Lansky D, Health Affairs April, 2012
Gaynor JW, With Permission 2015
Surgical Mortality as an Indicator of Hospital Quality: The Problem with Sample Size

Only CABG has the case volumes to reach the threshold of 138 cases to detect a doubling of the mortality rate.

Dimick et al JAMA 2004;292:847
• Medicare data 2009-2012
• 90-day mortality, readmission, complications
• Risk-adjustment methodology is unknown
• 5 categories, only report top three for now
Postoperative Mortality Is an Inadequate Quality Indicator for Lung Cancer Resection

Yinin Hu, MD, Timothy L. McMurry, PhD, Kristen M. Wells, PhD, James M. Isbell, MD, MSCI, George J. Stukenberg, PhD, and Benjamin D. Kozower, MD, MPH

Department of Surgery, Division of Thoracic Surgery, and Department of Public Health Sciences, University of Virginia Health System, Charlottesville, Virginia

• Objectives
  – To compare operative mortality definitions
  – To determine if mortality could be used as a quality metric

• Methods
  – SEER-Medicare data 2006-2010
  – Hierarchical GLE models to estimate 30 and 90-day mortality
Results

11,787 patients from 686 hospitals
Provider volumes range from 1-383
33% of hospitals perform fewer than 5 resections
Results

<table>
<thead>
<tr>
<th>Mortality</th>
<th>n</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-hospital</td>
<td>355</td>
<td>(3.01)</td>
</tr>
<tr>
<td>30 day</td>
<td>435</td>
<td>(3.69)</td>
</tr>
<tr>
<td>Perioperative</td>
<td>482</td>
<td>(4.09)</td>
</tr>
<tr>
<td>90 day</td>
<td>812</td>
<td>(6.89)</td>
</tr>
</tbody>
</table>
Results: Provider Rankings

• Calculated observed/expected 90-day mortality estimates for 686 hospitals

• Hospital variation
  – 3% of hospitals (20/686) had an O/E 90-day mortality different than 1
    • 18 hospitals had O/E > 1
    • 2 hospitals had O/E < 1

• 90-day mortality is not a useful metric for evaluating hospitals following lung cancer resection
Increasing Sample Size: Composite Measures

Mortality or Major Morbidity – STS Database
18,800 lung cancer resections from 111 hospitals

Barriers to Public Reporting: Requirements for Comparing Institutions

- We don’t have robust, accurate data
- We don’t have relevant outcomes measures
- We need sufficient statistical power to identify meaningful differences
- We don’t know how to communicate effectively to consumers

Lansky D, Health Affairs April, 2012
Gaynor JW, With Permission 2015
Communicating Outcomes with Consumers

• 60 million Americans use Yelp and dry cleaner ratings every month
• Healthcare is a very common search, but far from the top 10
  – 66% of internet users performed a healthcare search in the past year

<table>
<thead>
<tr>
<th>Rank</th>
<th>Search Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Robin Williams</td>
</tr>
<tr>
<td>2</td>
<td>World Cup</td>
</tr>
<tr>
<td>3</td>
<td>Ebola</td>
</tr>
<tr>
<td>4</td>
<td>Malaysia Airlines</td>
</tr>
<tr>
<td>5</td>
<td>ALS Ice Bucket Challenge</td>
</tr>
<tr>
<td>6</td>
<td>Flappy Bird</td>
</tr>
<tr>
<td>7</td>
<td>Conchita Wurst</td>
</tr>
<tr>
<td>8</td>
<td>ISIS</td>
</tr>
<tr>
<td>9</td>
<td>Frozen</td>
</tr>
<tr>
<td>10</td>
<td>Sochi Olympics</td>
</tr>
</tbody>
</table>
National Hospital Ratings Systems Share Few Common Scores And May Generate Confusion Instead Of Clarity

Leapfrog Group, Consumer Reports Health, US News & World Report’s, HealthGrades

- Examined 844 hospitals
- Four Rating Systems: Leapfrog, Consumer Reports, US News and World Report’s, HealthGrades
- Objectives were to compare the criteria used in ratings systems and to determine the consistency of hospital ratings
No hospital was rated as a high performer by all 4 rating systems.

Each system has its own rating methods, a different focus to its ratings, and stresses different measures.
Examined 4 different public reporting displays
  – 3 of 4 are used by publically available state reports
Displays shown to 337 adults
Respondents asked to identify which surgeon they were most likely to choose and surveyed on the importance of public reporting
This State Display Was the Most Accurately Interpreted: Which Surgeon Would You Choose?

<table>
<thead>
<tr>
<th>Surgeon</th>
<th>Cases</th>
<th>Mortality</th>
<th>Readmissions</th>
<th>Length of Stay</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>In-Hospital</td>
<td>30-Day</td>
<td>7-Day</td>
</tr>
<tr>
<td>1</td>
<td>92</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>3</td>
<td>89</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>4</td>
<td>91</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
</tbody>
</table>

- ○ Lower than expected (meaning that the hospital or surgeon had fewer deaths than expected after accounting for how sick the patients were in that hospital)
- ● Same as expected (meaning that the hospital or surgeon had as many deaths as expected after accounting for how sick the patients were in that hospital)
- ● Higher than expected (meaning that the hospital or surgeon had more deaths than expected after for how sick the patients were in that hospital)

Fig 2. Chart C—expected patient mortality. This chart shows the percent of expected deaths among patients who have CABG surgery, based on the average of deaths at several hospitals. The symbols in this report represent the results of how well hospitals and surgeons performs surgery and cared for the patient. A statistical test is done to determine whether differences in the results are simply due to chance or random variation. A difference is called “significant” when we are 95% sure that the difference is not likely to result from chance due to random variation.

Donelan K et al. Ann Thorac Surg. 2011;91
Personal Importance of Public Ratings

Fig 3. Personal importance of public ratings: “If you have to select a surgeon, how important is it to you to have data like these available to you?”

Donelan K et al. Ann Thorac Surg. 2011;91
Surgeon Preference by Display Format

The public does not understand current displays of reported outcomes

The “best surgeon” was correctly identified between 16% and 66% of the time, depending on the display

Donelan K et al. Ann Thorac Surg. 2011;91
Public Reporting of Surgical Outcomes

STS Public Reporting Online

Heart Surgery Outcomes - Public Access

STS believes the public has a right to know the quality of surgical outcomes and considers public reporting an ethical responsibility of the specialty. STS Public Reporting Online enables STS National Database participants to voluntarily report to each other and the public their heart surgery composite star ratings and the component ratings from which those are derived.

Isolated CABG:
Search or browse star ratings for surgery groups
Search or browse star ratings for hospitals

Isolated AVR:
Search or browse star ratings for surgery groups
Search or browse star ratings for hospitals
The Society of Thoracic Surgeons Composite Score for Lobectomy for Lung Cancer


STS Quality Measurement Task Force
STS General Thoracic Surgery Database Task Force
Duke Clinical Research Institute

Public Reporting of Surgical Outcomes

• Develop a platform for thoracic surgery
  – Start with lobectomy for lung cancer
  – Most common procedure
  – Use 3 years of rolling data

• Two domain composite
  – Operative mortality
  – Major complications
Methods: Estimation of Composite Scores & Star Ratings

• Composite score = (1-risk adjusted mortality rate) + (1-risk adjusted complication rate)

• Outcomes were weighted inversely by their respective standard deviations

• Star ratings (3 categories)
  – High (3-star) and low (1-star) performers
    • 95% credible intervals outside average scores
  – Easy for the public to comprehend

Results: Patient Population

22,362 lobectomies at 234 programs

- Non-elective status (1,211)
- Missing data (n=281)
- Benign disease (n=122)

20,657 at 231 centers (92.4%)
## Composite Outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operative mortality</td>
<td>1.5%</td>
</tr>
<tr>
<td>Major complications</td>
<td>9.6%</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>4.3%</td>
</tr>
<tr>
<td>Unexpected return to OR</td>
<td>3.9%</td>
</tr>
<tr>
<td>Reintubation</td>
<td>3.4%</td>
</tr>
<tr>
<td>Pulmonary embolus</td>
<td>0.5%</td>
</tr>
<tr>
<td>Initial Vent Support &gt;48 Hours</td>
<td>0.5%</td>
</tr>
<tr>
<td>Bronchopleural fistula</td>
<td>0.4%</td>
</tr>
<tr>
<td>Tracheostomy</td>
<td>1.0%</td>
</tr>
<tr>
<td>ARDS</td>
<td>0.7%</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>0.4%</td>
</tr>
<tr>
<td>Length of stay (median)</td>
<td>4 days, IQR (3,7)</td>
</tr>
</tbody>
</table>
## Composite Score Weights

<table>
<thead>
<tr>
<th></th>
<th>Mortality</th>
<th>Major Complications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviation</td>
<td>0.84</td>
<td>3.42</td>
</tr>
<tr>
<td>Weight</td>
<td>0.80</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Mortality is weighted approximately 4 times that of a major complication in the quality measure.
Reliability of Star Ratings Based on Program Volume (Proportion of Signal to Noise Variation)

<table>
<thead>
<tr>
<th>Number of participants</th>
<th>No Minimum</th>
<th>≥30 cases</th>
<th>≥50 cases</th>
<th>≥100 cases</th>
<th>≥150 cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>231</td>
<td></td>
<td>172</td>
<td>136</td>
<td>80</td>
<td>40</td>
</tr>
<tr>
<td>Reliability (95% CrI)</td>
<td></td>
<td>42% [33%-50%]</td>
<td>51% [41%-60%]</td>
<td>55% [45%-64%]</td>
<td>60% [48%-70%]</td>
</tr>
</tbody>
</table>
### Number of STS Participants Identified as Different from STS Average

<table>
<thead>
<tr>
<th>Criterion for categorizing participants</th>
<th>Below STS-Average (1-star)</th>
<th>STS-Average (2-star)</th>
<th>Above STS-Average (3-star)</th>
</tr>
</thead>
<tbody>
<tr>
<td>98% Credible interval</td>
<td>5 (2.9%)</td>
<td>160 (93.0%)</td>
<td>7 (4.1%)</td>
</tr>
<tr>
<td>95% Credible interval</td>
<td>8 (4.7%)</td>
<td>152 (88.4%)</td>
<td>12 (7.0%)</td>
</tr>
<tr>
<td>90% Credible interval</td>
<td>12 (7.0%)</td>
<td>142 (82.6%)</td>
<td>18 (10.5%)</td>
</tr>
<tr>
<td>80% Credible interval</td>
<td>16 (9.3%)</td>
<td>128 (74.4%)</td>
<td>28 (16.3%)</td>
</tr>
</tbody>
</table>
Distribution of Composite Scores

Participants Sorted in Order of Increasing Composite Score
### Construct Validity:
Complication Rates Vary Across Star Ratings

<table>
<thead>
<tr>
<th></th>
<th>1- Star</th>
<th>2- Star</th>
<th>3- Star</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operative Mortality (95% CrI)</td>
<td>3.2% [1.6% - 5.9%]</td>
<td>1.6% [0.6% - 3.5%]</td>
<td>0.9% [0.4% - 1.6%]</td>
</tr>
<tr>
<td>Major Complications (95% CrI)</td>
<td>17.1% [11.3% - 24.2%]</td>
<td>10.1% [5.1% - 16.9%]</td>
<td>6.5% [3.7% - 9.6%]</td>
</tr>
</tbody>
</table>
Risk Adjusted Outcomes: Star Ratings

STS General Thoracic Surgery Database
Composite Quality Rating
Participant 99999
STS Period Ending 12/31/2013

### Quality Domain
- **Participant Score (85% CI)**
- **STS Mean Participant Score**
- **Participant Rating**

#### Jan 2011 - Dec 2013
- **Overall**
  - 93.8% (92.5, 95.0)
  - 94.6%
  - ![★ ★]

#### Jan 2011 - Dec 2013
- **Absence of Mortality**
  - 96.8% (96.7, 97.7)
  - 97.2%
  - ![★ ★]

#### Jan 2011 - Dec 2013
- **Absence of Major Complication**²
  - 82.6% (79.3, 85.5)
  - 84.4%
  - ![★ ★]

### Quality Domain
<table>
<thead>
<tr>
<th>Eligible Procedures</th>
<th>Detail</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>Mortality</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>Major complication</td>
<td>70</td>
<td>25.7%</td>
</tr>
<tr>
<td></td>
<td>Pneumonia</td>
<td>18</td>
<td>5.7%</td>
</tr>
<tr>
<td></td>
<td>ARDS</td>
<td>2</td>
<td>0.6%</td>
</tr>
<tr>
<td></td>
<td>Bronchopleural fistula</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td></td>
<td>Pulmonary Embolus</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td></td>
<td>Initial Vent Support &gt;48hrs</td>
<td>2</td>
<td>0.6%</td>
</tr>
<tr>
<td></td>
<td>Reintubation</td>
<td>10</td>
<td>3.1%</td>
</tr>
<tr>
<td></td>
<td>Tracheotomy</td>
<td>6</td>
<td>1.8%</td>
</tr>
<tr>
<td></td>
<td>Myocardial infarction</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td></td>
<td>Unexpected return to OR</td>
<td>16</td>
<td>4.9%</td>
</tr>
<tr>
<td></td>
<td>Multiple complications (more than one of the above)</td>
<td>13</td>
<td>3.7%</td>
</tr>
</tbody>
</table>

¹Percentages represent the proportion that the specific complication contributed to the total number of patients with major complication.

This information is intended to facilitate and focus process and quality improvement initiatives by providers.
Limitations

• Generalizability
  – STS outcomes are better than national benchmarks
  – Fewer than 50% of lobectomies in the US

• Short term outcome measure
  – We added vital status out to 5-years
  – Lack patient reported outcomes
STS Study Conclusions

• Developed a two-domain quality measure for lobectomy for lung cancer
• Identified 12% of participating programs as outliers
• Will present risk-adjusted outcomes and star ratings in the public reporting effort
• Encourage participation in voluntary public reporting
Unintended Consequences of Public Reporting

- Misclassification of surgeons or hospitals
- Risk aversion for high risk patients with reduced access to care
- Confusion for the public and payers
- Potential for serious financial consequences
Distribution of Composite Scores:
Potential for misclassification
Comparison to the mean, not to nearest neighbor
The Centers for Medicare & Medicaid Services (CMS) issued a national coverage decision in 2006 that limited coverage of weight loss surgery to centers of excellence (COEs).

Adjusting for patient factors, changes in procedure type, and preexisting time trends toward improved outcomes, there were no statistically significant improvements in outcomes after (vs. before) implementation of the CMS national coverage decision for any complication, serious complications, and reoperation.

Dimick JB. et al. JAMA. 2013;309(8)
A CMS policy restricting care to COEs was associated with a relative decline in the proportion of nonwhite Medicare patients receiving bariatric surgery.

A policy intended to improve patient safety may have been associated with the unintended consequence of reduced use of bariatric surgery by minority patients.
Most surgeons (62%) refused to operate on at least one high-risk CABG patient over the prior year, primarily because of public reporting.
“We must address the underlying incentives for “case-selection creep” by improving risk adjustment methods for the highest-risk patients and by highlighting centers and physicians who undertake high-risk procedures in appropriate patients.”
Public Reporting of Surgical Outcomes: Conclusions

• The goal is to provide the optimal care for patients (lowest mortality, complications and resource utilization)
• Current public reporting initiatives have serious limitations
• Need accurate data, important and relevant outcomes, reliable and transparent methodology
• Need collaboration between key stakeholders
• Need to minimize potential consequences
The Search for Quality
Acknowledgements

• University of Virginia Surgical Outcomes Research Center
  – George Stukenborg, PhD
  – Tim McMurry, PhD
• Society of Thoracic Surgeons
• J William Gaynor, MD