Rankings & Ratings: An Update on U.S. News Hospital Quality Reporting

Ben Harder & Colleagues

OCTOBER 20, 2015
WASHINGTON, D.C.
LEAD INVESTIGATORS

Avery Comarow

Geoff Dougherty, MPH

Murrey Olmsted, Ph.D.

Ben Harder
Our Group

• U.S. News, which employs A.C., G.D. and B.H. and is the sole sponsor of our analysis, receives revenues from multiple advertisers including health systems

Ben Harder

• Anticipates serving as a part-time Senior Fellow at GuideStar
• Wife is a MedStar Health-employed physician
• Sister is a Brigham & Women’s-employed physician

Geoff Dougherty

• Part-time employee, fellowship recipient, and Ph.D. student in the Johns Hopkins Bloomberg School of Public Health Department of Epidemiology

Avery Comarow

• None

Murrey Olmsted

• None
Learning Objectives

• The U.S. News perspective on public reporting
• Update on our ratings & rankings
  – *Transparency's pivotal role*
• Changes under consideration
• Future directions:
  – **Appropriateness & avoidance of low-value care**
  – Best Regional Hospitals ( => Best Regional Health Systems )
  – Post-acute care quality reporting
  – Physician quality reporting
WHY WE REPORT ON HOSPITAL QUALITY

Louise H. Batz (r), 1943–2009
with family on Easter 2009
“WHERE ARE THE GOAL POSTS?”
THE TRIPLE AIM
Our Mission
To provide decision support to patients, families and referring physicians.

Any other societal benefits (quality-improvement, public accountability, guidance on pay-for-performance), while important, are secondary.
Potential for Unintended Consequences
Public reporting, like any intervention, may have unintended consequences. Potential hazards:
• Risk aversion
• Misclassification and its consequences
• Inefficient or perverse resource allocation
  – E.g., diversion of resources from clinical care, board overattentive to imperfect indicators

We are:
• Mindful of potential consequences when making analytical decisions
• Watchful for evidence of actual unintended consequences
Engagement with Patients

Approximately 135,000 people per day use our provider-lookup tools – Best Hospitals, Doctor Finder and Best Nursing Homes.

• 4 million unique users per month

Additionally, tens of thousands of consumers per day read our patient advice articles, evidence-based diet information, health news and more.

We owe them good guidance -- which requires good data.
Engagement with Researchers & Clinicians

Hospital Quality Reporting by US News & World Report
Why, How, and What's Ahead

Clinically significant variation exists in the quality of care delivered in US hospitals. To make informed choices among hospitals, as well as among health systems and insurance networks of which they are a part, the public must have access to meaningful data that facilitate actionable comparisons. For 25 years, US News & World Report has published indices of hospital quality with the objective of providing health care consumers with highly intuitive decision support. Using methods based on the Donabedian model of structure, process, and outcomes, US News annually ranks hospitals in 16 adult and 10 pediatric specialties. The explicit focus of these rankings is high-acuity care, which is often delivered in the cannot be measured with validity in a direct manner must be omitted from a model or included by proxy, which results in inherently imperfect measurement. In addition, the use of any metric as a publicly reported performance measure may have negative unintended consequences, particularly if it leads to risk-averse behavior among physicians or health care organizations. A model designed to identify clinically significant differences among hospitals may prioritize either statistical sensitivity or specificity, but cannot maximize both. To balance such trade-offs, US News has taken an iterative approach to constructing and subsequently refining its methods, which have been described.
UPDATE ON METHODS
Best Hospitals for Common Care

• Debuted in May 2015
• Target audience: Typical patient facing a routine elective procedure or managing a chronic condition that may episodically require admission
• Patient cohorts
  – Hip replacement
  – Knee replacement
  – CABG
  – CHF
  – COPD
• Classifications (ratings)
  – High performing
  – Average
  – Below average
  – Low volume (non-outliers with <25 FFS cases over 3 years)
• No reputational component
Scorecard: Heart bypass surgery

Rating

75% of hospitals rated in this procedure were average. Average hospitals met expected standards of care.

What goes into this rating:

**Survival**
Survival 30 days after admission following coronary artery bypass graft surgery (CABG), adjusted for patient risk.

As expected
Score: 7
As expected: 7 out of 7

**Readmissions for any reason within 30 days**
Success in preventing unplanned returns of CABG patients to the hospital for any cause within 30 days of discharge.

As expected
Score: 7
As expected: 7 out of 7

**Readmissions for any reason within 7 days**
Success in preventing unplanned returns of CABG patients to the hospital for any cause within 7 days of discharge.

As expected
Score: 7
As expected: 7 out of 7

**Patient volume**
Number of Medicare Inpatients who had heart bypass surgery with or without valve replacement or repair in 2010-2012.

**High**
High: At least 2/3rds of patients participating in the Hospital Outcomes Data Collection (HODC) Program met standard.

**Significantly worse than expected**
Score: 3
Significantly worse than expected: 1-5 out of 13

**Hospital-acquired infections**
Success in minimizing six types of infections as reported by the Centers for Disease Control and Prevention. Very high or low numbers of one or more types of infections may significantly affect overall success. Data are available online at data.medicare.gov.

**Nurse staffing**
Relative number of nurses caring for patients hospitalwide.

**Nurse Magnet recognition**
Accredited for high nursing standards by American Nurses Credentialing Center as of Feb. 1, 2015.

**Cardiac ICU**
Intensive-care unit exclusively for critically ill heart patients.

**Intensivist on staff**
At least one intensive-care unit staffed by physician with subspecialty certification or fellowship training in care of ICU patients.

**Patient experience**
How Inpatients rated their hospital stay across seven aspects of their care, from a quarterly survey used by nearly 4,000 hospitals.

**Transparency in STS quality measures**
Most teams of cardiac surgeons participate in a quality-measurement program (registry) run by the Society of Thoracic Surgeons. Some hospitals permit STS to reveal their performance to the public.

As of the data analysis was completed, this hospital’s CABG performance data was not available on STS.org.

Common Care report card: CABG
Common Care: Quality Measures

- SES- & risk-adjusted readmissions (7- & 30-day, LDS SAF)
- Risk-adjusted mortality (30-day, derived from analysis of LDS SAF data)
- Infection rates, per NHSN
- Staffing factors (intensivist staffing, nurse staffing, Magnet status)
- Volume (LDS SAF)
- Patient experience measures, per CMS
- Joint complication rates, per CMS
- Joint revision rates (1-year, LDS SAF)
- Composite CABG rating, per STS
- Transparency on clinical outcomes, via STS
Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors

TECHNICAL REPORT
August 15, 2014

This report is funded by the Department of Health and Human Services under contract HHSM-300-2012-00003I task order 7.

The Medicare Hospital Readmissions Reduction Program
Time for Reform

Andrew S. Boozary, MD, MPP

Joseph Manchin III
The United States Senate (D-West Virginia), Washington, DC.

Roger F. Wicker, JD
The United States Senate (R-Mississippi), Washington, DC.
Data Limitations

• Retrospective analysis cannot adjust for all potential confounders
  – Claims data may obscure variations in risk factors, coding practices, diagnostic accuracy and appropriateness of care

• Lagging performance may not predict current performance.
  – CMS recently accelerated LDS release; next public reporting will add 2013 & 2014

• Differences in setting (inpatient vs outpatient) may affect results
  – We anticipate incorporating outpatient claims in 2017

• Analysis is limited primarily to Medicare FFS patients
  – We will expand a voluntary program to use audited, system-submitted all-payer data

• Important outcomes, e.g. functional status, unmeasurable from claims

• Proxy measures such as volume & staffing may correlate imperfectly to outcomes of interest
The Antidote for Imperfect Data: Better Data

New clinical registries are rapidly emerging to meet federal requirements for physician quality reporting.

Mature registries have a track record of bringing more-perfect data into the public domain.

• Exhibit A: the Society of Thoracic Surgeons (STS)
Obstacles to Transparency

- Patient privacy concerns
- Health system anxieties
  - Provider privacy (e.g. longtime injunction on Medicare physician billing data)
  - Provider buy-in
  - Potential loss of competitive edge in market
  - Medical-legal considerations

- Risk aversion & denial of access (societal anxieties)
  - Experiences from New York, Pennsylvania and Massachusetts
  - This is of particular concern in the context of claims-based risk adjustment
  - Complete transparency (i.e., risk coefficients) could facilitate strategic risk arbitrage
  - Out of an abundance of caution, we have not published risk coefficients

- Our standard, for ourselves & others: Determine the therapeutic window and administer the maximum safe dose of transparency
Will Other Registries Choose Transparency, Like STS?

• One observer thinks not:
  “the medical priesthood [says]: ‘You are not worthy to judge us.’ ”

• We’re more optimistic. Laypeople will use the tools available to them, and medical professionals will insist on those tools being valid.
Heart Bypass Outcomes, by STS Reporting Status

- We compared 431 hospitals practicing voluntary reporting via STS.org (as of Feb. 2015) to 757 nonreporters that perform CABG.
- Hospitals that were not voluntarily transparent had inferior outcomes.

<table>
<thead>
<tr>
<th></th>
<th>30-day mortality</th>
<th>7-day readmissions</th>
<th>30-day readmissions</th>
<th>Avg. volume (Medicare FFS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Voluntarily reporting hospitals</strong> (n=97,751 cases)</td>
<td><strong>17% lower risk</strong></td>
<td><strong>8% lower risk</strong></td>
<td><strong>7% lower risk</strong></td>
<td><strong>75.6 cases/year</strong></td>
</tr>
<tr>
<td><strong>STS nonreporters</strong> (n=118,546 cases)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td><strong>52.2 cases/year</strong></td>
</tr>
</tbody>
</table>
Heart Bypass Outcomes, by STS Status & Composite

- Nonreporting hospitals were inferior to 3-star and 2-star reporters.
- Important implications for U.S. News patient decision-support

<table>
<thead>
<tr>
<th>STS Status</th>
<th>30-day Mortality</th>
<th>7-day Readmissions</th>
<th>30-day Readmissions</th>
<th>Avg. Volume (Medicare FFS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>STS 3-star (n=99 hospitals)</td>
<td>24% lower risk</td>
<td>6% lower risk</td>
<td>5% lower risk</td>
<td>110.3 cases/year</td>
</tr>
<tr>
<td>STS 2-star (n=319 hospitals)</td>
<td>11% lower risk</td>
<td>6% lower risk</td>
<td>5% lower risk</td>
<td>65.6 cases/year</td>
</tr>
<tr>
<td>STS 1-star (n=11 hospitals)</td>
<td>lower risk (n.s.)</td>
<td>0% difference</td>
<td>6% higher risk (n.s.)</td>
<td>54.0 cases/year</td>
</tr>
<tr>
<td>Nonreporters (n=757 hospitals)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>52.2 cases/year</td>
</tr>
</tbody>
</table>
“STS programs that voluntarily participate in public reporting have significantly higher volumes and performance. No evidence of risk aversion was found.”
“No evidence of risk aversion was found.”

- That’s reassuring, considering experiences in NY/PA/MA, where public reporting of CABG and PCI led to reduced access
- Your clinicians may still harbor residual “transparency anxiety”
  - It’s understandable to have concerns about excessive or premature transparency
- Please bring them along. Gently. But quickly.

If your hospital isn't voluntarily transparent today...
What is it waiting for?
Best Hospitals (complex specialty care)

- Published annually since 1990
- Target audience: Patients facing a complex dx or difficult procedure
- Covers 16 specialties, 12 of them data-driven
- Classifications:
  - Ranked nationally #1 to #50
  - High performing (top decile)
  - Unranked (not a negative composite)
- Past methodology refinements:
  - Replaced inpatient mortality with 30-day mortality (2007)
  - Introduced select AHRQ PSIs (2009)
  - Reduced weight assigned to Reputation (2014)
  - Expanded sample of surveyed physicians (2014)
# Cardiology & Heart Surgery Scorecard

**Overall Score**: 54.5 / 100

## What goes into this score:

### Reputation with specialists

- **Score**: 2.3%
- **Modeled**: 0.1% - 2.9%

### Survival

- **Score**: 7
- **Better than expected**: 7 or 8

### Patient safety

- **Score**: 1
- **Limited**: 1 out of 5

### Patient volume

- **Highest**: 4,205 discharges
- **Highest**: 4,900 or above

### Nursing intensity

- **High**: 1.9
- **High**: 1.7 - 1.9

### Nurse Magnet recognition

Certified for compliance with high nursing standards by American Nurses Credentialing Center.

### Advanced technologies

- **Highest**: 8 out of 8

### Patient services

- **High**: 6 out of 7
Best Hospitals: Quality Measures

• Risk-adjusted mortality

• Patient safety score
  – A composite of 7 AHRQ PSIs
  – Not equivalent to PSI-90 – different constituent PSIs, different weights

• Volume of relevant complex MS-DRGs

• Staffing factors (intensivist staffing, nurse staffing, Magnet status)

• Other structural measures

• Program reputation (expert medical opinion)
  – Survey of board-certified physicians in relevant specialties
  – RTI conducts the survey, analysis and weighting
  – Two sampling frames: (1) non-members of Doximity; (2) members of Doximity
  – Proportional weighting to ensure results are nationally representative
Reputation: a (Modest) Role for Expert Clinical Opinion

• Reputational data may capture clinician-possessed information about program quality that’s not apparent in objective data
  – The essence of “expert opinion”
  – We have characterized reputation as a measure of “process” in the Donabedian paradigm of structure, process and outcomes. Obviously, others use term differently.

• Reputation acts primarily as a differentiator among exceptionally high-performing centers

• It has relatively modest effects on which hospitals are ranked
The Reputation Myth

**Second Opinion**

Inside the health rankings.

**The Reputation Myth and the U.S. News Hospital Rankings**

On objective quality measures, 98% of ranked hospitals scored at or above the 98th percentile.

- If reputation were removed from our model:
  - 98% of Best Hospitals would be in the top 100 of ~5,000 hospitals evaluated
  - 84% would be in the top 50
- That is: The best hospitals in a reputation-less ranking methodology would have 84% overlap with the published Best Hospitals
How We Report: Pediatric Care

Best Children’s Hospitals

- Debuted in 2007
- Target audience: Families of children with complex or rare diagnoses
- Covers 10 pediatric specialties
- Classifications:
  - Ranked nationally #1 to #50
  - Unranked

- Past methodology refinements:
  - Entirely reputational through 2006
  - Began data collection from pediatric programs – now an 1,800-item inventory (2007)
  - Data-driven rankings published in 6 specialties (2008)
  - Steady increase in clinical measures, e.g. outcomes, best practices (2008–)
  - Reduced reputational component to 25% (2013)
  - Reduced reputational component to 16.7% (2014)
Pediatrics: Orthopedics Scorecard

Rank

448 in the nation

Overall Score

68.0 / 100

What goes into this score:

Reputation with specialists
% of doctors in this specialty responding to surveys in 2013, 2014 and 2015 who named hospital as among best for very challenging patients.

Modest
Score: 1.7%
Modest: 0.1%-2.9%

Speed and success with complex fractures
Efficient care of complex fractures.

Average
Score: 4
Average: 3 or 4 out of 6

Preventing surgical complications
Reflects % of patients without complications after scoliosis surgery; additional credit for low % of infections after spinal fusion.

Superior
Score: 18
Superior: 16, 17 or 18 out of 18

Use of infection-preventing measures
Success in minimizing infections through hand hygiene, vaccination and other proven measures.

Very good
Score: 24
Very good: 18 to 25 out of 29

Patient volume
Relative number of patients in past year with certain specific disorders.

Low
Score: 20
Low: 1 to 20 out of 57

Procedure volume
Relative number of tests and surgical procedures.

Average
Score: 15
Best Children’s Hospitals: Quality Measures

- Outcomes measures – at least 1 per specialty (e.g. complex heart surgery mortality, 3-year cancer survival, rate of ICU infection)
- Best practices based on patient population (e.g. management of CF and diabetes patients)
- Inpatient and outpatient volumes
- Staffing factors (e.g. intensivist staffing, nurse staffing, Magnet status)
- 37 structural measures including key technologies (e.g. ECMO in Neonatology)
- Program’s reputation among pediatric specialists and subspecialists
U.S. News Dashboard

• Submit feedback, questions & changes
• Request/confirm name changes to reflect hospital rebranding
• Access embargoed data and draft methodologies
• [https://hospitaldashboard.usnews.com](https://hospitaldashboard.usnews.com)
• For VPMAs, CQOs, CEOs, PIOs
• Registration & use are free
Concerns about Validity, Transparency & Purpose

September 14, 2015

Patrick Conway, M.D.
Principal Deputy Administrator
Centers for Medicare & Medicaid Services
Department of Health and Human Services
Attention: CMS-1461-P
P.O. Box 801
Baltimore, MD 21244-8013

Re: Comments on the Hospital Star Ratings Methodology Report

Dear Dr. Conway:

On behalf of our nearly 5,000 hospitals, health systems and other health care organizations, the American Hospital Association (AHA) appreciates the opportunity to comment on the draft methodology for creating an overall star rating system for hospital quality as delineated in the report of CMS’ Technical Expert Panel (TEP) and Aug. 13 webinar. Our members are dedicated to providing useful quality and safety information to our patients, the communities we serve and other stakeholders. America’s hospitals were instrumental in the creation of Hospital Compare more than a decade ago, and remain strongly committed to sharing meaningful, accurate hospital quality information with the patients they serve.
Concerns about Discordance


**Abstract**

Attempts to assess the quality and safety of hospitals have proliferated, including a growing number of consumer-directed hospital rating systems. However, relatively little is known about what these rating systems reveal. To better understand differences in hospital ratings, we compared four national rating systems. We designated “high” and “low” performers for each rating system and examined the overlap among rating systems and how hospital characteristics corresponded with performance on each. No hospital was rated as a high performer by all four national rating systems. Only 10 percent of the 844 hospitals rated as a high performer by one rating system were rated as a high performer by any of the other rating systems. The lack of agreement among the national hospital rating systems is likely explained by the fact that each system uses its own rating methods, has a different focus to its ratings, and stresses different measures of performance.
Concerns about Discordance

Austin et al. found high degree of apparent discordance across 4 public reporting programs. Authors suggest consumers may be confused.

• Possible sources of discordance:
  (1) Raters measuring different, complementary dimensions of care
      – i.e. U.S. News complex specialty care, Leapfrog patient safety
  (2) Raters measuring same dimension with different degrees of validity

• Investigational opportunity: comparison on a single dimension of care:
  – Readmissions of hip/knee patients, U.S. News (with SES) vs CMS (without SES)

• Diversity in study design advances the state of knowledge
  – Discordance is a promising way to expose perversities in P4P programs
Concerns about Measure Validity: PSIs & NHSN Infections

Concerns about Measure Validity
Paradoxically, Rajaram et al. paradoxically found hospitals penalized by HAC Reduction:
• More likely to be teaching hospitals
• Less likely to be Joint Commission-accredited
• Less likely to participate in QI registries
• Had more complex case mix
• Had lower volume
• Achieved superior performance on CMS outcome & process measures
Concerns about Measure Validity

Rajaram et al. raises doubts about the validity of PSI-90 & the 2 NHSN measures used in P4P.
• Past work by this group factored in our 2014 decision to exclude PSI-12 (VTE).

Our Patient Safety composite score, used in Best Hospitals, overlaps ~36% with PSI-90.

We are currently conducting independent analyses of PSI and NHSN data to understand whether modifications to our methods may be warranted.
CHANGES UNDER CONSIDERATION
Common Care: Anticipated Expansion

- Common Care ratings to be updated in Spring 2016
- Analysis based on 2012-2014 inpatient claims
- Inclusion of outpatient claims in 2017
- Additional cohorts under development
  - Aortic valve surgery
  - Abdominal aortic aneurysm (AAA) surgery
  - Lobectomy
  - Colectomy
  - Obstetrics/perinatal care (after-lunch discussion)
  - Others for 2017 and beyond
Common Care: Modifications Under Consideration

- **Cohort-specific measures**
  - Remove readmissions from medical cohorts (CHF, COPD)
  - Address overlapping outcome definitions (7-day and 30-day readmissions)
  - Include process measures (e.g. % of HF patients given discharge instructions)
  - Measure excessive use of open vs endovascular AAA repair

- **HAI measures**
  - NHSN measures are hospitalwide, not cohort-specific
  - NHSN data may be confounded (internal analysis, published research)
  - Claims-based HAIs may have higher validity for hospital comparisons

- **Intensivist definition**
  - Staff vs contract physicians

- **Nursing quality measures**
  - Nurse Magnet: beyond reach for some hospitals, where NDNQI be informative
Hospital Outcomes Data Disclosure (HODD) Program

• Medicare FFS provides an incomplete view of care in Medicare-age population.
• Some health systems are more invisible in FFS than others.
• We piloted a voluntary program for submission of audited all-payer data.
  – Pilot participants: Kaiser Permanente, Intermountain
  – Participation may increase likelihood that U.S. News can identify true outliers
  – Modeling was designed to neither advantage nor disadvantage participants
• We now invite all interested systems to join HODD program
  – Fall 2015 enrollment for spring 2016 Common Care ratings
  – Inclusion in Best Hospitals rankings is not planned for 2016
FUTURE DIRECTIONS
Best Regional Hospitals => Best Regional Health Systems

• Debuted in 2011
• Serves as starting point for consumers comparing provider organizations within a market
• Methodology counts performance in all 5 Common Care service lines and all 12 data-driven adult specialties
  – Preventive health, behavioral health & other important areas are not yet included
  – Attribution is currently at the hospital level
  – Attribution will migrate toward the regional system level
• With rise of narrow networks, consumers increasingly ‘choose’ their systems when they select a health plan
  – Decision support on health system choice will influence decision support on health plan (or ACO) choice
  – We already publish basic ratings of Medicare Advantage plans
  – Ultimately, measurement of systems and plans may converge
Volume-Mortality Relationships (Best Hospitals for Common Care)
• Hip & Knee
Volume-Mortality Relationships (Best Hospitals for Common Care)

• CABG and COPD

More Cases, Fewer Deaths: Heart Bypass Surgery

More Cases, Fewer Deaths: COPD
Risks Are High at Low-Volume Hospitals

Patients at thousands of hospitals face greater risks from common operations, simply because the surgical teams don't get enough practice.

By Steve Sternberg and Geoff Dougherty  |  May 18, 2015 | 12:01 a.m. EDT

Like other hospitals in thinly populated areas, Sterling Regional Medical Center does a bit of everything. The 25-bed Colorado hospital has its own heliport, delivers about 200 babies a year and admits more than 1,200 patients for a variety of conditions and procedures. Replacing worn and painful hips and knees is among them. To patients, the surgery may seem perfectly routine.

Joint replacements are anything but routine at hospitals that don't do many of them, a new U.S. News analysis shows. Sterling is among thousands of U.S. medical centers whose patients face a greater risk of death and complications because their surgical teams do too few procedures, even common ones, for doctors, nurses and technicians to maintain their skills.

These large numbers of low-volume hospitals, the analysis found, continue to put patients at higher risk even after three decades of published research have demonstrated that patients are more likely to die or suffer complications when treated by doctors who only occasionally see similar patients rather than by experienced teams at hospitals with more patients and established protocols.
Hospitals Move to Limit Low-Volume Surgeries

Three of nation's leading hospital systems say they will limit low-volume surgeries.

By Steve Sternberg | May 19, 2015 | 12:01 a.m. EDT

Three of the nation's top academic medical systems – Dartmouth-Hitchcock Medical Center, Johns Hopkins Medicine, and the University of Michigan – say they are planning to impose minimum-volume standards that will bar hospitals in their systems from performing certain procedures unless both the hospitals and their surgeons do them often enough to keep their skill level up.

The move comes in the wake of a story released Monday by U.S. News showing that hospitals that do small numbers of common procedures place patients at far greater risk than those that do lots of them.

The voluntary standards, which are expected to go into effect before the end of the year, represent the first coordinated effort to place limits on hospitals and on surgeons, who traditionally have been allowed to perform virtually every procedure within the scope of their specialty training, even if only once a year.
“It's a promising, bold move. I hope other hospitals across the country follow.”

– Leah Binder, The Leapfrog Group

“Low-volume hobbyists are bad for patients and we have to stop them.”

– Dr. John Birkmeyer, Dartmouth-Hitchcock
Point/Counterpoint: Why We Need Minimum Volume Standards

Having a threshold for surgeons and hospitals is a common-sense way to improve outcomes for thousands of patients each year.

By John Birkmeyer and Peter Pronovost

Point/Counterpoint: A Simple Surgical Threshold Won’t Suffice

It’s tempting to try to link quality to quantity. But the real solution is far more complicated.

By David C. Chang and Keith D. Lillemoe | June 24, 2015 | 10:51 a.m. EDT
Hospital Networks and Value-Based Payment
Fertile Ground for Regionalizing High-Risk Surgery

Recently, 3 major medical centers—Dartmouth, Johns Hopkins, and the University of Michigan—declared a “volume pledge” that restricts performance of 10 selected procedures to surgeons and hospitals that meet volume criteria. This followed an announcement by US News & World Report of a new analysis demonstrating poorer outcomes at low-volume hospitals. Both of these are derived from decades of research showing that high surgical volumes are associated with better outcomes, especially in high-risk procedures.

Stakeholders have responded to the volume-outcome literature by calling for selective referral to high-volume facilities and in some cases enforced this by only reimbursing facilities that meet certain volume criteria.

Pledging to Eliminate Low-Volume Surgery
David R. Urbach, M.D.

On May 18, 2015, leaders at three hospital systems—Dartmouth-Hitchcock Medical Center, the Johns Hopkins Hospital and Health System, and the University of Michigan Health System—publicly announced a “Take the Volume Pledge” campaign to prevent certain surgical procedures from being performed by their surgeons who perform relatively few of them or at their hospitals where relatively few such procedures are performed. The Pledge, promulgated by long-time advocates of quality improvement such as John Birnkrant and Peter Pronovost, challenges other mon Care rankings, was predictably hostile — and completely out of proportion to the modest ambition of the Pledge. Of all the possible approaches to restricting surgical care to high-volume hospitals, perhaps the least controversial ought to be a decision by a large metropolitan academic hospital system that its most complex elective surgery should be performed by the providers and hospitals that do the most of a given procedure. (The Pledge is silent on the question of performing complex surgery in independent small and rural hospitals.) If volume-based distribution of


FUTURE DIRECTIONS

Hospital Networks and Value-Based Payment
Fertile Ground for Regionalizing High-Risk Surgery

Karan R. Chhabra, AB
Rutgers Robert Wood Johnson Medical School, New Brunswick, New Jersey.

Justin B. Dimick, MD, MPH
Center for Healthcare Outcomes and Policy, University of Michigan, Ann Arbor; and Department of Surgery, University of Michigan, Ann Arbor.

Recently, 3 major medical centers—Dartmouth, Johns Hopkins, and the University of Michigan—declared a “volume pledge” that restricts performance of 10 selected procedures to surgeons and hospitals that meet volume criteria. This followed an announcement by US News & World Report of a new analysis demonstrating poorer outcomes at low-volume hospitals. Both of these are derived from decades of research showing that high surgical volumes are associated with better outcomes, especially in high-risk procedures.

Stakeholders have responded to the volume-outcome literature by calling for selective referral to high-volume facilities and in some cases enforcing this by only reimbursing facilities that meet certain volume criteria.

Pledging to Eliminate Low-Volume Surgery
David R. Urbach, M.D.

On May 18, 2015, leaders at three hospital systems — Dartmouth–Hitchcock Medical Center, the Johns Hopkins Hospital and Health System, and the University of Michigan Health System — publicly announced a “Take the Volume Pledge” campaign to prevent certain surgical procedures from being performed by their surgeons who perform relatively few of them or at their hospitals where relatively few such procedures are performed. The Pledge, promoted by long-time advocates of quality improvement such as John Birkmeyer and Peter Pronovost, challenges other


Regionalization of Care & the Health System of Tomorrow

As integrated systems approach “maturity” – e.g. using hub-and-spokes delivery models around centers of excellence – we anticipate changes in:

• Referral patterns
• Volume at individual sites (hub up, spokes down)
• Systemwide outcomes
Developing Quality Measures to Address Overuse

Jason S. Mathias, MD, MS
David W. Baker, MD, MPH

Medical societies and organizations have worked for many years to develop quality measures to assess the quality of health care services (e.g., platelet therapy for patients with acute coronary syndromes). More recently, organizations have focused on addressing unnecessary tests and treatments to improve quality of care by developing "overuse measures." More information is available at http://www.jama.com. The use of a service that is unlikely to improve outcomes or for which potential harms exceed potential benefits is considered overuse.

The standards for developing and updating overuse measures are not clearly defined and are controversial among experts. Two key issues in the development, implementation, and updating of overuse measures are (1) the definition of overuse and (2) the level of evidence required for underuse measures.

Level of Evidence Required for Overuse Measures

Original Investigation

Early Trends Among Seven Recommendations From the Choosing Wisely Campaign

Alan Rosenberg, MD; Aby Aggarwal, PhD; Marc Gottlieb, MPA; John Barron, PharmD; Peter Brady, MBA; Ying Liu, MS; Cindy Li, MSc; Andrea DeVries, PhD

**IMPORTANCE** The Choosing Wisely campaign consists of more than 70 lists produced by specialty societies of medical practices or procedures of minimal clinical benefit to patients in most situations, with recommendations regarding judicious use.

**OBJECTIVE** To quantify the frequency and trends of some of the earliest Choosing Wisely recommendations using nationwide commercial health plan population-level data.

**DESIGN, SETTING, AND PARTICIPANTS** Retrospective analysis of claims data for members of Anthem-affiliated commercial health plans. The low-value services selected were (1) imaging tests for uncomplicated headache; (2) cardiac imaging without history of cardiac conditions; (3) low back pain imaging without red-flag conditions; (4) preoperative chest X-rays with unremarkable history and physical examination results; (5) human papillomavirus testing for women younger than 30 years; (6) use of antibiotics for acute sinusitis, and (7) use of prescription nonsteroidal anti-inflammatory drugs (NSAIDs) for members with hypertension, heart failure, or chronic kidney disease.
Our Interest in Appropriateness

- Low-value care and overuse has been tolerated or incentivized for years.
- Yet it exposes patients to avoidable risk and financial toxicity.
- Population management requires massive removal of low-value care...
- ...making appropriateness a window into system performance on the Triple Aim.

A sampling of likely indicators:

- Performance on Choosing Wisely
- Excessive cardiac interventions (elective angioplasty)
- Too much open surgery (when minimally invasive techniques appropriate)
- C-section rates (risk-adjusted), episiotomy & EED
  - Laudable efforts by ACOG, TJC, March of Dimes, Leapfrog & others
  - We intend to add the power of our pulpit to their efforts
**Maternity Care**
- Potentially highly amenable to consumer choice
- Minimal differences in safety or short-term outcomes
- Substantial differences in frequency of interventions with cost and long-term health implications

### Table: Birth of a Ranking: Which Hospitals Are Best at Delivering Babies?

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Births</th>
<th>Observed primary</th>
<th>Expected primary</th>
<th>Observed repeat</th>
<th>Expected repeat</th>
<th>Episiotomy rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Augusta Health</td>
<td>1,111</td>
<td>24.5</td>
<td>26.7</td>
<td>92.7</td>
<td>88.7</td>
<td>9.5</td>
</tr>
<tr>
<td>Bedford Memorial Hospital</td>
<td>207</td>
<td>20.7</td>
<td>16.5</td>
<td>89.5</td>
<td>n/a</td>
<td>4.6</td>
</tr>
<tr>
<td>Bon Secours DePaul Medical Center*</td>
<td>974</td>
<td>14.3</td>
<td>17.8</td>
<td>69.9</td>
<td>84.2</td>
<td>2.5</td>
</tr>
<tr>
<td>Bon Secours Mary Immaculate Hospital</td>
<td>944</td>
<td>20.3</td>
<td>21.4</td>
<td>92.0</td>
<td>86.0</td>
<td>15.2</td>
</tr>
<tr>
<td>Bon Secours Maryview Medical Center</td>
<td>1,012</td>
<td>19.6</td>
<td>18.1</td>
<td>89.7</td>
<td>88.2</td>
<td>11.6</td>
</tr>
<tr>
<td>Bon Secours Memorial Regional Medical Center</td>
<td>1,804</td>
<td>21.0</td>
<td>21.2</td>
<td>83.0</td>
<td>90.6</td>
<td>10.7</td>
</tr>
<tr>
<td>Bon Secours St. Francis Medical Center</td>
<td>1,077</td>
<td>23.3</td>
<td>22.4</td>
<td>89.9</td>
<td>90.8</td>
<td>16.2</td>
</tr>
</tbody>
</table>
U.S. News Maternity Care Ratings

Excessive intervention in healthy childbirth is a longstanding quality gap.

- Major strides have been made against early elective delivery
  - But variability remains; EED rates are public but largely invisible
- Episiotomy rates are stubbornly elevated at some hospitals
  - An NQF-endorsed measure is available & can be calculated from claims in ~20 states
- C-sections
  - Risk-stratified C-section rates can be calculated from discharge data in ~20 states
  - Hospitals increasing report NTSV C-section rates to TJC, Leapfrog, etc.
- U.S. News options:
  - Rely on state discharge data – limited coverage & risk adjustment challenges
  - Solicit measures directly from hospitals – would add reporting burden
  - Team up with (an)other public-reporting organization(s)
FUTURE DIRECTIONS

U.S. News Maternity Care Ratings
Other clinical and family-centered data might include:
• Breastfeeding on discharge +/- considering mother’s preference
• Baby Friendly
• Availability and frequency of VBAC
POST-ACUTE CARE
Post-Acute Care: Our Current Decision Support Tools

- Best Rehabilitation Hospitals
  - Reputation-based
  - Needs outcomes measurement
- Best Nursing Homes
  - Nursing Home Compare-based
  - This is the best CMS can do?
Post-Acute Care: Greater Scrutiny Is Needed
Quality measurement of post-acute care needs more watchdogs, particularly in the era of bundled care.

- NYTimes investigation into SNF gaming of CMS
- WSJ investigation into SNF utilization of ultrahigh therapy
- WSJ analysis of LTAC discharge timing
Rating Post-Acute Care: Skilled Nursing Facilities
We have acquired SNF Inpatient LDS SAF files for future analyses.

Possible applications in U.S. News public reporting:

• Measurement & reporting of SNF performance on outcomes
  – Short stay, possibly by clinical cohort (e.g. post-joint replacement)
  – Long stay

• Adjustment for SNF in calculating hospital readmissions rates and case-level expected risk
PHYSICIAN PERFORMANCE
The Doctor Finder leverages the world’s most comprehensive physician database, Doximity.
U.S. News augments the Doximity data with hospital quality data, Medicare data, etc.

Procedures Dr. Brennan Performs

When you're facing a medical procedure, experience matters. Research indicates that physicians with higher volumes tend to achieve better outcomes. U.S. News has mined Medicare data to determine procedures that this physician performs. Where possible, physicians who perform a high volume of a procedure relative to their peers nationally or in their state are identified as well. More procedures will be added over time. Read more about our methodology and limitations here.

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Medicare Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthroscopic rotator cuff repair</td>
<td>36 in 2012</td>
</tr>
<tr>
<td>Arthroscopic shoulder debridement</td>
<td>37 in 2012</td>
</tr>
<tr>
<td>Complete meniscectomy</td>
<td>39 in 2012</td>
</tr>
<tr>
<td>Joint injections</td>
<td>461 in 2012</td>
</tr>
<tr>
<td>Partial knee replacement</td>
<td>38 in 2012</td>
</tr>
</tbody>
</table>
Surgeon-level Volumes
We recently added surgeon-level, procedure-specific volumes to Doctor Finder.

We also identified surgeons who were high-volume relative to others performing the same procedure.
Public Reporting on Physician Quality
The Volume Pledge is about quality of surgeon as well as hospital. Surgeon quality, in particular, is of great public interest.
Public Reporting on Physician Quality

For better or worse, we have entered an era of claims-based reporting on physicians. In addition, CMS will be judging physicians for P4P.

• How can we ensure patients (& payers) use valid decision-support?

We have concluded that U.S. News should develop physician ratings. We will proceed with caution and ample communication.
“Others have tried and failed. What makes you think you’ll succeed?”
“Others have tried and failed. What makes you think you’ll succeed?”

Because you and your peers want us to succeed.
Public Reporting on Physician Quality
Meaningful comparisons require appropriate assignment to peer group.
• E.g. a joint specialist cannot be compared to a back specialist

Within peer group, important indicators may include:
• Voluntarily reported registry data (QCDRs)
• Hospital-reported patient satisfaction (CAHPS)
• Risk-adjusted outcomes, where they can be reliably measured & attributed
• Surgical volume as an outcome proxy
• Low-value and/or inappropriate care
• Low-value referral decisions or being embedded in low-value network

Let’s look at low-value, unnecessary care…
Are Doctors Exposing Heart Patients to Unnecessary Cardiac Procedures?

A U.S. News analysis finds some doctors may be putting patients at risk - and reaping the benefits in Medicare payments.

By Steve Sternberg and Geoff Dougherty | Feb. 11, 2015 | 12:01 a.m. EST

Dr. Gregory Sampognaro is one of the busiest interventional cardiologists in the United States, clearing out clogged coronary arteries at a high rate. His numbers were the highest in 2012 in the number of these procedures in the heart, based on Medicare data.

What makes these numbers remarkable is not necessarily the large number he performs in New York or Chicago but in Monroe-Woodbury, a small town in one of the poorest congressional districts in the state.

Sampognaro is one of dozens of cardiologists performing catheterization procedures – angioplasties – at higher rates than referral centers, the U.S. News analysis found.

Doctors and Angioplasty Rates: Our Methodology

How U.S. News analysed the Medicare data.

The dataset used by U.S. News covers most inpatient and outpatient angiograms and angioplasties covered under Medicare's fee-for-service insurance.

By Geoff Dougherty | Feb. 11, 2015 | 12:01 a.m. EST

U.S. News analyzed cardiac catheterizations by obtaining Medicare procedure and billing records from 2012, the only year available, to assess where and how often doctors performed diagnostic angiograms and angioplasty procedures and how much the federal government paid for their services.
QUESTIONS, COMMENTS & CONCERNS?
Outlier Frequency on U.S. News and CMS Outcomes (Knee/Joint)

- Range: <2% (mortality) to 40% (30-day readmissions) from LDS SAF
- 4% are outliers on a CMS complications measure

<table>
<thead>
<tr>
<th>KNEE COHORT</th>
<th>USN Knee Composite</th>
<th>USN Joint Mortality</th>
<th>1-year Knee Revisions</th>
<th>30-day Knee Readmissions</th>
<th>CMS Joint Complications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Better than expected</strong></td>
<td>278 (10%)</td>
<td>10 (0.4%)</td>
<td>203 (7.2%)</td>
<td>530 (19%)</td>
<td>61 (2.2%)</td>
</tr>
<tr>
<td><strong>No different</strong></td>
<td>2,301 (82%)</td>
<td>2,767 (98%)</td>
<td>2,545 (91%)</td>
<td>1,725 (61%)</td>
<td>2,635 (96%)</td>
</tr>
<tr>
<td><strong>Worse than expected</strong></td>
<td>235 (8%)</td>
<td>37 (1.3%)</td>
<td>63 (2.2%)</td>
<td>559 (20%)</td>
<td>53 (1.9%)</td>
</tr>
</tbody>
</table>
Aggregation of Indicators

We used a latent-variable modeling approach to construct cohort-specific composites. E.g., for knee replacement:

- Joint mortality
- + 7-day knee readmissions
- + 30-day knee readmissions
- + HAIs
- + Patient experience
- + Joint complications
- + Knee revisions
- + Knee volume
- + **Staffing factors**

Knee composite rating
System Consolidation and Attribution Considerations

- Primary attribution is by AHA ID
  - Some AHA entities have multiple Medicare provider numbers
  - Systems occasionally merge previously distinct AHA entities

- Small but noticeable uptick in other “merger” requests from systems
  
  Drivers:
  - Closer clinical integration across multiple hospitals
  - Possibly also an expectation of “doing better” in the rankings?

- We evaluate each request individually to assess what’s best for patients
  - Disinclined to combine entities unless they achieve closely matched risk-adjusted outcomes
  - Different outcomes suggests sites are not equivalent to patients

- Moreover, internal analysis suggests that merging tends to depress the rankings performance of the better-performing entity
Unnecessary Care
An example of the studies we’re reading is Chen et al. (NEJM, 2015):

“Routine preoperative testing in not recommended for patients undergoing cataract surgery…

“Preoperative testing occurred frequently and was more strongly associated with provider... than patient characteristics.”