The Debate Over How to Rate Doctors and Hospitals

Ben Harder (mod.) @benharder
Donna Cryer @DCpatient
Jerry Lin @DocSpotHealth
Marshall Allen @marshall_allen
Mark Friedberg @MWFriedberg

#hdatapalooza

May 9, 2016
Health Datapalooza
Washington, D.C.
Moderator’s disclosures

• **Employer.** U.S. News receives advertising and other revenue from various healthcare organizations

• **Personal.** Senior fellow at GuideStar, a data-transparency organization serving the nonprofit sector

• **Other.** Wife is a MedStar Health-employed physician; sister is a Brigham & Women’s-employed physician
THE WALL STREET JOURNAL.

POLITICS

U.S. to Delay Release of New Hospital Ratings

Release date pushed back to July amid questions about methodology
RATINGS & REPORT CARDS

By Steven D. Findlay

ANALYSIS & COMMENTARY

Consumers’ Interest In Provider Ratings Grows, And Improved Report Cards And Other Steps Could Accelerate Their Use

Timeline of public reporting of provider performance

<table>
<thead>
<tr>
<th>Year</th>
<th>Source of Initiative</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>New York State Department of Health</td>
<td>Issues report on heart surgery outcomes and provider ratings.</td>
</tr>
<tr>
<td>1990</td>
<td>Agency for Healthcare Research and Quality</td>
<td>Launches development of CAHPS, a survey instrument to measure the patient experience of care.</td>
</tr>
<tr>
<td>1991</td>
<td>Leapfrog Group</td>
<td>Launches ratings of hospitals, with focus on safety issues.</td>
</tr>
<tr>
<td>1993</td>
<td>California and Pennsylvania</td>
<td>Join New York in publishing heart surgery outcomes and provider ratings</td>
</tr>
<tr>
<td>1994</td>
<td>Consumers’ Checkbook</td>
<td>Launches “Guide to Top Doctors” based on surveys of physicians.</td>
</tr>
</tbody>
</table>
Applications of Provider Ratings

- **Patient/consumer decision support.** Ratings facilitate informed decision-making and patient engagement.*
- **Performance improvement.** Measurement can help doctors and hospitals identify opportunities for improvement.
- **Pay-for-performance.** Value-based payment or contracting may incentivize delivery of higher quality at lower cost.
- **Accountability.** Performance measurement can identify unwarranted and/or unacceptable variation.

*Giving consumers decision-support is U.S. News & World Report’s objective. Other entities’ ratings programs may have a different objective, or multiple objectives.*
Limitations of Provider Ratings

• Potential unintended consequences
  • Risk aversion by providers

• Potential misclassification
  • Are the measures valid?
  • Are the detectable differences important?
  • Are important differences escaping detection?

“Not everything that can be counted counts, and not everything that counts can be counted.”
### U.S. News Hospital Ratings (May 2015)

<table>
<thead>
<tr>
<th></th>
<th>COPD</th>
<th>Heart Failure</th>
<th>Heart Bypass Surgery</th>
<th>Hip Replacement Surgery</th>
<th>Knee Replacement Surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Better than expected</strong></td>
<td>301</td>
<td>434</td>
<td>143</td>
<td>224</td>
<td>262</td>
</tr>
<tr>
<td><strong>Not statistically different</strong></td>
<td>3,400 (76%)</td>
<td>3,399 (76%)</td>
<td>792 (67%)</td>
<td>1,809 (55%)</td>
<td>2,301 (68%)</td>
</tr>
<tr>
<td><strong>Worse</strong></td>
<td>232</td>
<td>237</td>
<td>127</td>
<td>197</td>
<td>234</td>
</tr>
<tr>
<td><strong>Insufficient volume</strong></td>
<td>533</td>
<td>422</td>
<td>119</td>
<td>1,044</td>
<td>591</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>4,466</td>
<td>4,492</td>
<td>1,181</td>
<td>3,274</td>
<td>3,388</td>
</tr>
</tbody>
</table>

Results not shown for 16 adult specialties and 10 pediatric specialties.
## ProPublica Surgeon Scorecard (July 2015)

<table>
<thead>
<tr>
<th>n = surgeon count</th>
<th>Prostate removal</th>
<th>Prostate resection</th>
<th>Gall-bladder (lap.)</th>
<th>Spinal fusion (neck)</th>
<th>Spine (lumbar, anterior)</th>
<th>Spine (lumbar, posterior)</th>
<th>Hip</th>
<th>Knee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Better than expected</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not statistically different</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**16,019 surgeons rated**
A Methodological Critique of the ProPublica Surgeon Scorecard

Mark W. Friedberg, Peter J. Pronovost, David M. Shahian, Dana Gelb Safran, Karl Y. Bilimoria, Marc N. Elliott, Cheryl L. Damberg, Justin B. Dimick, Alan M. Zaslavsky

Summary

On July 14, 2015, ProPublica published its Surgeon Scorecard, an online tool that displays “Adjusted Complication Rates” for individual, named surgeons for eight surgical procedures performed in hospitals.

Public reports of provider performance (or, performance reports) have the potential to improve the quality of health care that patients receive. Valid performance reports (i.e., reports that truly measure what they are advertised as measuring) can stimulate providers to make quality improvements and can help patients make better selections when choosing among health care providers. However, performance reports with poor measurement validity and reliability are potentially damaging to all involved. Therefore, it is important to critically examine the methods used to produce any performance report.

Measuring provider performance is challenging, but methods exist that can help ensure that performance reports are valid and display true differences in performance. This methodological critique of the ProPublica Surgeon Scorecard has three goals: to explain methodological issues in the Scorecard, to suggest ways in which the Scorecard can be improved, and to inform the public about these aspects of the Scorecard. An overview of our conclusions with respect to the first two goals follows. The third—to inform the public—exists because the Scorecard is currently available to the public, and, based on our critique, we hope patients who are choosing a surgeon will be better able to decide how much weight to give the data presented in the Scorecard.
PATIENT ENGAGEMENT

By J. Matthew Austin, Ashish K. Jha, Patrick S. Romano, Sara J. Singer, Timothy J. Vogus, Robert M. Wachter, and Peter J. Pronovost

National Hospital Ratings Systems Share Few Common Scores And May Generate Confusion Instead Of Clarity

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.
The Debate Over How to Rate Doctors and Hospitals

Ben Harder (mod.) @benharder
Donna Cryer @DCpatient
Jerry Lin @DocSpotHealth
Marshall Allen @marshall_allen
Mark Friedberg @MWFriedberg

#hdatapalooza

May 9, 2016
Health Datapalooza
Washington, D.C.