2-18-2011

Detecting Hospital Quality with PRIDIT.

Robert D. Lieberthal, PhD
School of Population Health, Thomas Jefferson University

Follow this and additional works at: https://jdc.jefferson.edu/hplectures

Part of the Health and Medical Administration Commons, and the Public Health Commons

Let us know how access to this document benefits you

Recommended Citation
https://jdc.jefferson.edu/hplectures/18

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's Center for Teaching and Learning (CTL). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in College of Population Health Lectures, Presentations, Workshops by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: JeffersonDigitalCommons@jefferson.edu.
Detecting hospital quality with *PRIDIT*

Robert D. Lieberthal

Jefferson School of Population Health

February 18, 2011
I could not have done this alone

- **PRIDIT** code from Richard Derrig and the Automobile Insurance Board of Massachusetts
- Funding from the Agency for Healthcare Research and Quality
- Lots of great input
  - Amol Navathe, Dan Polsky, Rachel Werner, and two anonymous referees at HSR
  - Attendees at the 2008 NRSA trainees conference, 2008 International Conference on Health Policy Statistics
  - Elaine Yuen and the students of PBH609
Outline

Hospital quality data
  Background
  Hospital Compare data

PRIDIT results
  Quality scores
  Bootstrapped standard errors
  Validating PRIDIT
Outline

Hospital quality data
   Background
   Hospital Compare data

PRIDIT results
   Quality scores
   Bootstrapped standard errors
   Validating PRIDIT
Hospital quality measures are hard to aggregate

- Hospital quality is assessed on process and outcome measures
  - Process measures include appropriate antibiotic use, frequent hand washing
  - Outcome measures include 30 day readmission rates, risk adjusted mortality
- It is hard to determine which observed measures of quality are good indicators of high quality hospitals
  - What is the relative importance of different measures?
  - How can we account for hospital characteristics like teaching status and ownership type?
  - Apparent high performance of hospitals could be a result of locating near a healthy population
- Quality should measure how much a hospital can improve a patient’s health, not how healthy she was to begin with
Hospital Compare contains publicly reported hospital process measures

<table>
<thead>
<tr>
<th>Process measure</th>
<th>Average (US)</th>
<th>Average (PA)</th>
<th>Jefferson Adherence</th>
<th>Patients (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antibiotic timing</td>
<td>87%</td>
<td>88%</td>
<td>82%</td>
<td>303</td>
</tr>
<tr>
<td>Correct antibiotic</td>
<td>93%</td>
<td>93%</td>
<td>98%</td>
<td>302</td>
</tr>
</tbody>
</table>

**Table:** Hospital compare sample data, 7/1/2009-12/31/2009

Both measures contain some discretion
I used process measures and hospital characteristics

- 20 process measures of adherence to best practices from 0-100% at a single point in time
  - Heart attack (8 measures)
  - Heart failure (4 measures)
  - Pneumonia (6 measures)
  - Surgical infection prevention (2 measures)
- 5 other demographic variables from American Hospital Association data
  - Acute care or critical access hospital
  - Hospital ownership (govt, nfp, fp)
  - Emergency services
  - Accreditation
  - Teaching intensity (several levels)

Reporting data is optional for some hospitals, mandatory for others (or Medicare would reduce their payments)
My sample included 4,217 hospitals that reported data
Heart attack measures contained lots of variation

<table>
<thead>
<tr>
<th>Measure</th>
<th>Percent reporting</th>
<th>Average adherence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE inhibitor or ARB for LVSD</td>
<td>73%</td>
<td>80%</td>
</tr>
<tr>
<td>Aspirin at arrival</td>
<td>87%</td>
<td>92%</td>
</tr>
<tr>
<td>Aspirin at discharge</td>
<td>85%</td>
<td>89%</td>
</tr>
<tr>
<td>β-blocker at arrival</td>
<td>87%</td>
<td>85%</td>
</tr>
<tr>
<td>β-blocker at discharge</td>
<td>85%</td>
<td>87%</td>
</tr>
<tr>
<td>PCI &lt; 120 minutes post arrival</td>
<td>30%</td>
<td>64%</td>
</tr>
<tr>
<td>Smoking cessation</td>
<td>65%</td>
<td>79%</td>
</tr>
<tr>
<td>Thrombolytics &lt; 30 minutes post arrival</td>
<td>41%</td>
<td>30%</td>
</tr>
</tbody>
</table>
Heart failure quality measures were well reported

<table>
<thead>
<tr>
<th>Measure</th>
<th>Average (%)</th>
<th>reporting (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE inhibitor or ARB for LVSD</td>
<td>80</td>
<td>89</td>
</tr>
<tr>
<td>Assessment of left ventricular function</td>
<td>80</td>
<td>93</td>
</tr>
<tr>
<td>Discharge instructions</td>
<td>52</td>
<td>83</td>
</tr>
<tr>
<td>Smoking cessation</td>
<td>74</td>
<td>81</td>
</tr>
</tbody>
</table>
Pneumonia measures were well reported

<table>
<thead>
<tr>
<th>Measure</th>
<th>Percent reporting</th>
<th>Average adherence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pneumococcal vaccination</td>
<td>94%</td>
<td>56%</td>
</tr>
<tr>
<td>Antibiotic(s) (&lt; 4) hours after arrival</td>
<td>93%</td>
<td>77%</td>
</tr>
<tr>
<td>Oxygenation assessment</td>
<td>94%</td>
<td>99%</td>
</tr>
<tr>
<td>Smoking cessation</td>
<td>83%</td>
<td>71%</td>
</tr>
<tr>
<td>Appropriate antibiotic(s)</td>
<td>84%</td>
<td>78%</td>
</tr>
<tr>
<td>Blood culture before antibiotic</td>
<td>84%</td>
<td>82%</td>
</tr>
</tbody>
</table>
Surgical infection measures were not well reported

<table>
<thead>
<tr>
<th>Measure</th>
<th>Percent reporting</th>
<th>Average adherence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antibiotic 1 hour before incision</td>
<td>35%</td>
<td>74%</td>
</tr>
<tr>
<td>Antibiotic stopped &lt; 24 hours post surgery</td>
<td>35%</td>
<td>67%</td>
</tr>
</tbody>
</table>
Outline

Hospital quality data
  Background
  Hospital Compare data

PRIDIT results
  Quality scores
  Bootstrapped standard errors
  Validating PRIDIT
Result is overall score

- Output on quality of hospitals and value of different variables
  - A relative ranking of all 4,217 hospitals in the dataset
  - A weighting system for the relative importance of quality indicators, demographic variables
- Example: Temple University Hospital scored 0.01419 (national average is 0)
- Example: Heart failure measure patients given assessment of left ventricular function was weighted 0.69731 (maximum score is 1)
- No negative weights
  - All measures were associated with positive quality—sometimes focus on one measure in isolation can hurt overall quality
  - If I had recoded the hospital characteristics, they would have been negative
- Small hospital bias caveats
  - Volume is not included
  - Hospitals did not report measures with N<25 observations
  - I imputed an average value for unreported variables
A few variables accounted for most of the variation in quality

- Patients given beta-blocker at arrival and at discharge
  - Well reported (~85%)
  - Majority but not total adherence (~85%)
- All 4 heart failure measures (esp. assessment of left ventricular function)
- Measures with total adherence not useful for measuring quality
  - Oxygen assessment for pneumonia–99% adherence!
- Surgical measures not well reported and so did not explain much variation
- All process measures positively associated with quality
- More teaching is better–no residency programs < some residency programs < full residency programs < residency and med school program
Heart attack measures were good indicators

Ranked bins from 1 (best) to 5 (worst)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Full data set</th>
<th>Clinical data only</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE inhibitor or ARB for LVSD</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Aspirin at arrival</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Aspirin at discharge</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>$\beta$-blocker at arrival</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\beta$-blocker at discharge</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PCI $&lt; 120$ minutes post arrival</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Smoking cessation</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Thrombolytics $&lt; 30$ minutes post arrival</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>
Heart failure measures were good indicators

<table>
<thead>
<tr>
<th></th>
<th>Full data set</th>
<th>Clinical data only</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE inhibitor or ARB for LVSD</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Assessment of left ventricular function</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Discharge instructions</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Smoking cessation</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Pneumonia and surgery weren’t as good

<table>
<thead>
<tr>
<th></th>
<th>Full data set</th>
<th>Clinical data only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pneumonia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pneumococcal vaccination</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Antibiotic(s) &lt; 4 hours after arrival</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Oxygenation assessment</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Smoking cessation</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Appropriate antibiotic(s)</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Blood culture before antibiotic</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Surgical infection prevention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antibiotic 1 hour before incision</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Antibiotic stopped &lt; 24 hours post surgery</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>
Hospital demographics were good controls

<table>
<thead>
<tr>
<th></th>
<th>All data Importance</th>
<th>Clinical data Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute care hospital</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>Government hospital</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>Private hospital</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>Accredited hospital</td>
<td>4</td>
<td>N/A</td>
</tr>
<tr>
<td>Emergency service available</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>Major teaching</td>
<td>4</td>
<td>N/A</td>
</tr>
<tr>
<td>Significant teaching</td>
<td>3</td>
<td>N/A</td>
</tr>
<tr>
<td>Any teaching</td>
<td>3</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Hospital quality was evenly distributed

Lots of hospitals in the middle, a few “outliers” of high and low quality
How well does *PRIDIT* distinguish between hospitals?

- What is the difference between scoring 0.05 and 0.10?
- What is the value in raising the score from 0.05 to 0.10?
  - Pay for performance and provider networks—are they different enough to pay one more than the other?
- How does the standard error change across the range of *PRIDIT* scores?
- Importance differs by geography
  - Some areas only have one hospital—see if you need to make it better
  - Some areas have lots of hospitals—direct people to the best one
Sampling *PRIDIT* scores with replacement

- 1000 samples with replacement of the universe of scores
- Not bagging—I didn’t recompute the scores in each sample
- U-shaped distribution of standard errors—much more confidence in median than high or low scores
- Quintile cutoffs: -0.012, -0.003, 0.0045, 0.0135
Standard errors are small relative to *PRIDIT* scores

- You can tell many Philadelphia hospitals apart within a 2 sd range (one sd in each direction)
- Hahneman, Presby and Jefferson are all close (Frankford and Nazareth are also only 4 miles apart)
Validating *PRIDIT* with outcome measures in Hospital Compare

- Better scores should correlate with less readmission and mortality
- Readmission
  - Readmission is 30-day readmission rate
  - Theory: they should have fixed you the first time
  - Less than perfect indicators
    - Not a long look back period
    - If people die, they are never readmitted
- Mortality
  - Mortality is risk adjusted
  - Theory: they should have fixed you well enough so you don’t die
  - Less than perfect indicators
    - Not a long look back period
    - It would be better to have raw scores and do the adjustments myself
Using multiple observations from Hospital Compare

- Multiple observations of the same hospital over time
  - Measure the stability of hospital rankings over time
  - Measure the relative importance of each measurement over time
  - More hospitals report more data over time
  - Hospital Compare has more measures now than when I first studied these hospitals
  - Use outcomes measures over time to rank quality and/or validate quality scores

- Measure the bootstrapped confidence intervals against multiple observations
The extensions are works in progress. I have submitted the validation idea as a grant proposal to the Actuarial Foundation. Any feedback would be really useful to me. Thanks!