Applications: Health Care Provider Performance Assessment

Susan Paddock, Ph.D.
RAND Corporation
OPRE Methods Meeting
October 19, 2017
Application: Healthcare provider performance evaluation

• Goal: *Improve value of health care*
  
  – *Improve quality:* Patients receive about 50% of the evidence-based tests and treatments they should (McGlynn et al., 2003)
  
  – *Lower cost:* Growth of healthcare costs has been on ongoing concern (Commonwealth Fund, 2007)

• There are numerous measures of quality and (increasingly) of cost
Performance Measures Summarize Aspects of Care

• Structural characteristics of healthcare providers
  – Use of electronic health records (yes/no)

• Clinical processes of care
  – % females aged 40-69 who had >= 1 mammograms during past 2 years

• Patient outcomes
  – % readmitted to hospital within 30 days of discharge

• Patient satisfaction & experiences
  – In the last 6 months, how often was it easy to get appointments with specialists?
Policy Levers for Improving Health Care Value

• **Public report cards**
  – Providers are motivated to improve performance (Fung et al., 2008)
  – Example: Medicare Hospital Compare
    • [https://www.medicare.gov/hospitalcompare](https://www.medicare.gov/hospitalcompare)
    • Numerous clinical, patient experience, payment, and outcomes measures for most hospitals in the U.S.
  – Examples of many other report cards are available from the Agency for Healthcare Research and Quality (AHRQ)
Policy Levers for Improving Health Care Value

• Value-based Payment (Pay for Performance)
  – Link payment to performance
  – Examples:
    • Medicare Value-based Purchasing Program for acute care hospitals
    • Medicare Advantage Health Plan Star Ratings
    • Numerous public and private payers have such programs

• Oftentimes, public reporting and pay for performance are combined
Challenges of Public Report Cards

• Evidence is mixed of report cards’ effect on consumer selection of providers (Fung et al., 2008)

• Concerns: Poor design, irrelevant content, hard to understand (AHRQ, 2012)

• In response, AHRQ, in collaboration with the Centers for Medicare & Medicaid Services (CMS), supported 17 studies to improve the science of public reporting

• I will present research conducted under one of these 17 studies (Grant R21HS021860, PI: Paddock)
Example: Statistical Benchmarking

• Use data to define a “realistic standard of excellence” (Kiefe et al., 2001). Examples:
  – National Committee for Quality Assurance (NCQA) uses percentiles of performance measures to score health plans
  – Medicare Hospital Value-Based Purchasing: Compare each quality measure to an achievement threshold equal to the 50th percentile. Hospital must exceed this to receive any points toward bonus payment
Examining the Standard Benchmarking Methodology in Hospital Compare

• For a process measure, hospital performance is often estimated as a direct average (numerator / denominator)

• Medicare Hospital Compare: Hospitals scoring at or above the 90th percentile for a measure are described as top hospitals

• O’Brien et al. (2008) found this rule tended to identify lower-volume (i.e., fewer patients) hospitals
  – However, exploratory data analyses showed that higher-volume hospitals had better performance
Example: Low Volume Hospitals with Relatively Extreme Performance Scores

Data: Medicare Hospital Compare (2009) measure PN-6, Initial Antibiotic Selection for Community-Acquired Pneumonia in Immunocompetent Patients

Minimum number of cases per displayed hospital is 11
What About Stabilizing Hospital Performance Estimates for Low Volume Providers?

• O’Brien et al. (2008) examined empirical Bayes (‘shrunken’) estimates of hospital performance from a two-stage hierarchical model (HM)
  – A provider performance estimate under HM is a weighted average of its own (‘direct’) estimate and the average for all hospitals
  – Less weight is given to the hospital’s own performance data for low-volume hospitals

• HM-based provider performance estimates have lower average error than direct estimates (James & Stein 1961) and protect against regression to the mean (Burgess et al., 2000)
**Example of Shrinkage**

Blue lines: Hospitals with direct estimates of 0.45 to 0.60, which are highlighted to compare the effect of shrinkage for smaller versus larger hospitals. Gray lines: All other hospitals.

**Data:** Medicare Hospital Compare (2009) measure PN-6, Initial Antibiotic Selection for Community-Acquired Pneumonia in Immunocompetent Patients.
Should Posterior Means be Used in this Case?

• In O’Brien et al. (2008):
  – Higher volume hospitals were more likely to be top performers if using the 90\textsuperscript{th} percentile of posterior means as a benchmark
  – However, no hospital with <10 cases exceeded the benchmark for several measures

• Could the higher volume hospitals truly be better?
  – Volume-outcome relationship exists (Silber et al., 2010)

• Could provider performance estimates be ‘over-shrunk’ for benchmarking?
What is the Correct Inferential Target for the Distribution of Provider Performance?
(Shen & Louis 1998, Paddock & Louis 2010)

• Posterior means (PMs) are under-dispersed
• Direct estimates are over-dispersed
• The empirical distribution function (EDF) of provider performance parameters from the Bayesian HM is just right
Hospital Compare Benchmarks using Direct Estimates, PMs, and the EDF

• Analysis of two selected performance measures using 2010 data (Paddock, 2014) showed:
  – 95-99% agreement if targeting the achievement threshold of 50%
  – More differences in hospitals selected as ‘top 10%’ across methods:
    • 10-19% classified differently for DIRECT vs. PM
    • 1-11% classified differently for DIRECT vs. EDF
  – As expected, EDF is a compromise between the DIRECT estimates and PMs
Why Did I Decide To Use Bayesian Methods For This Particular Project?

• Non-hierarchical Bayesian estimators featured in the benchmarking literature – shrinkage properties appreciated (Kiefe et al., 2001)

• Hierarchical (empirical) Bayesian approach is familiar in provider performance evaluation field

• However, clarity is needed in the field about the correct target for inference for benchmarks
Challenges in Conducting the Analysis or Disseminating Findings

- **Challenge:** Findings are relatively new to the health services research community. Further education and dissemination are required among researchers.

- **Challenge:** Further assess sensitivity of findings using the EDF versus standard approach
  - Assess trade-off between transparency versus accuracy
Opportunities in Conducting the Analysis or Disseminating Findings

• **Opportunity:** Bayesian methods offer a way to examine posterior probabilities of benchmark exceedances and uncertainty in benchmarks (Paddock and Louis, 2011)
  – Could be useful when developing scoring rules

• **Opportunity:** Further demonstration for report card designers of the importance of focusing on the correct inferential target

• Example: Ranking versus rating providers
Posterior Means versus Posterior Ranks
(Goldstein & Spiegelhalter, 1996)

• Posterior means and posterior ranks can differ
• Provider performance ranks can be very imprecise
• Many report cards bin provider performance relative to average performance
Why Are Bayesian Ideas Receiving More Attention In Healthcare Provider Assessment?

• How to quantify uncertainty?

• Example highlighted in White Paper for CMS, written by the Committee of Presidents of Statistical Societies (COPPS; 2012):
  – Congress requires CMS to penalize hospitals with readmission rates > 1
  – Hospitals with same performance could be classified differently because of different levels of precision
  – COPPS recommendation to CMS: Enforce the penalty only if the posterior probability of rate > 1 exceeds a pre-determined threshold, c
  – Posterior probability could guide the degree of penalty
    • e.g., probability of exceeding = 0.5 could imply receiving half of the full penalty
Why are Bayesian ideas receiving more attention in healthcare provider assessment?

• How to make decisions using model output?

• Center for Medicare and Medicaid Innovation (CMMI)

• Renee Mentnech (2017 Joint Statistical Meetings): Posterior probabilities and decision-making; actuaries find posterior probabilities to be very useful
Why are Bayesian ideas receiving more attention in healthcare provider assessment?

• How much information to use to obtain provider performance estimates?
  – As much as possible, under a Bayesian perspective
  – There can be disagreement among stakeholders about which information to put into the model
    • Example: Medicare Hospital Compare performance estimates are not adjusted for hospital characteristics (like volume) that are associated with performance (Silber et al., 2010, 2016; George et al., 2017)
  – Solutions could blend best statistical practices with policy solutions to minimize unintended consequences (Damberg et al., 2015)
Summary

• Bayesian approach has several important features for health care provider performance assessment
  – Quantify uncertainty
  – Express results in terms of posterior probabilities
  – Build into the analysis all assumptions

• However, standard statistical analysis cautions hold
  – The correct target for inference must be specified
  – Model must be appropriate for the policy question
References


References


