Ranking Hospital Performance Using Administrative Databases

Rui Wang, Ph.D.

Department of Population Medicine
Harvard Pilgrim Health Care Institute and Harvard Medical School
Department of Biostatistics
Harvard T.H. Chan School of Public Health

Feb 18, 2019

BIRS workshop, Statistical Analysis of Large Administrative Health Databases: Emerging Challenges and Strategies
Banff, Canada

Background

- National reimbursement policies seek to align quality and cost and reduce preventable harm, including healthcare-associated infections (HAIs)
- Claims data are commonly used to track HAIs, but they are limited by variable coding practices and the potential influence of changing reimbursement policies
- Federal value-based incentive programs (VBIPs) incorporate HAI rates reported to the National Healthcare Safety Network in determinations of hospital performance
- Goal: To examine differences in hospital rankings computed using claims versus NHSN, focusing on surgical site infection (SSI) following colon surgery

Ranking Hospital Based on Infection Rates

Infection Control & Hospital Epidemiology (2019), **40**, 208–210
doi:10.1017/ice.2018.310

SHEA
The Society for Healthcare Epidemiology of America

Concise Communication

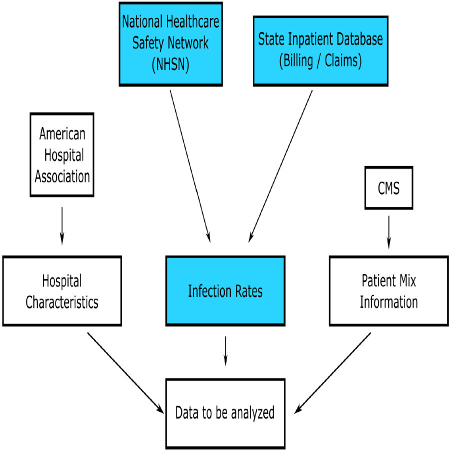
Comparison of hospital surgical site infection rates and rankings using claims versus National Healthcare Safety Network surveillance data

Chanu Rhee MD, MPH[1,2], Rui Wang PhD[1], Maximilian S. Jentzsch MS[1,3], Carly Broadwell BS[1,3], Heather Hsu MD, MPH[1,4], Robert Jin MS[1], Kelly Horan MPH[1] and Grace M. Lee MD, MPH[1,5]

[1]Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, Massachusetts, [2]Division of Infectious Diseases, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, [3]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, [4]Department of Medicine, Boston Children's Hospital, Boston, Massachusetts and [5]Department of Pediatrics, Stanford University School of Medicine, Palo Alto, California

NHSN vs. Billing/Claims

- NHSH: Infection needs to be recognized and reported by infection prevention nurses into the NHSN surveillance system
- Billing: Infection needs to be recognized by the physician, and a claim needs to be submitted to the insurance company for care required for a condition



Study Question: How would ranking differ using two data sources?

Data Source

- Retrospective cohort: adult patients admitted to 155 non-federal acute-care hospitals in 7 states that shared NHSN data through the Preventing Avoidable Infectious Complications by Adjusting Payment (PAICAP) study
- Included admissions in calendar years 2012-2014 from PAICAP hospitals that could be linked to administrative data from the State Inpatient Databases, Healthcare Cost and Utilization Project
- Hospital Characteristics were obtained from the 2011 American Hospital Association Annual Survey
- 6.2 million adult admission, 63,541 colon surgeries
- Reported SSIs: 7.197 (claims) vs. 3,283 (NHSN)

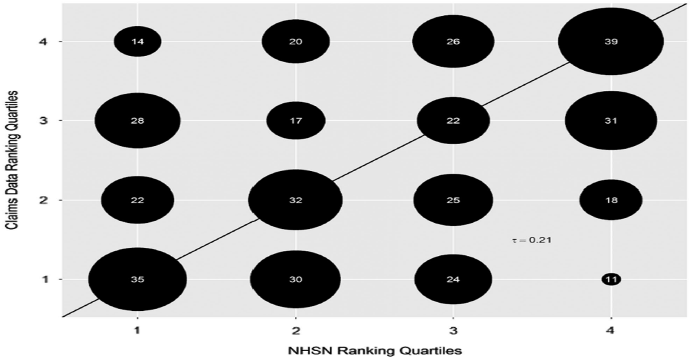
Metrics to Compare Ranking

- Limited available data: ranking were based on reliability adjusted rates by fitting a random effects model with a random hospital-specific intercept
- Concordance correlation coefficient:

$$\rho_c = \frac{2\rho\sigma_x^2\sigma_y^2}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2}$$

- More relevant question: whether hospitals move in- and out- of the worst quantile
- A bubble plot

The Bubble Plot



- 60/99 (60%) of times a hospital ranked in the worst quantile by NHSN ranked out by claims data
- 62/101 (61%) of times a hospital ranked in the worst quantile by claims data ranked out by NHSN

Issues Worth Considering

- Time trends:
  - How to assess stability of ranking over time?
  - Should hospital ranking take into account performance history?
  - How to assess ranking discrepancies over time?
- Patient-mix: current dataset does not contain individual information hence this was not possible

Ranking Hospital Based on Sepsis Mortality

Variation in Identifying Sepsis and Organ Dysfunction Using Administrative Versus Electronic Clinical Data and Impact on Hospital Outcome Comparisons

Chanu Rhee, MD, MPH[1,2]; Maximilian S. Jentzsch, MS[1,3]; Sameer S. Kadri, MD, MS[4]; Christopher W. Seymour, MD, MSc[5]; Derek C. Angus, MD, MPH[5]; David J. Murphy, MD, PhD[6]; Greg S. Martin, MD, MSc[6]; Raymund B. Dantes, MD, MPH[7]; Lauren Epstein, MD, MS[7]; Anthony E. Fiore, MD, MPH[7]; John A. Jernigan, MD, MS[7]; Robert L. Danner, MD[4]; David K. Warren, MD, MPH[8]; Edward J. Septimus, MD[1,9]; Jason Hickok, MBA[10]; Russell E. Poland, PhD[1,10]; Robert Jin, MS[1]; David Fram, BA[11]; Richard Schaaf, SM[11]; Rui Wang, PhD[1]; Michael Klompas, MD, MPH[1,2]; for the Centers for Disease Control and Prevention (CDC) Prevention Epicenters Program

Crit Care Med. 2018 Nov 13. [Epub ahead of print]

Sepsis Burden and the Focus on Quality

- Sepsis is a leading cause of death in U.S. hospitals
- Timely and effective sepsis care can reduce the risk of death
- Sepsis is now the focus of policy initiatives to improve and benchmark the quality of sepsis care
- Claims data have been shown to have low to moderate accuracy for identifying sepsis
- Goal: Can they be used to compare hospital sepsis rates and outcomes for reliable identification of low or high-performing hospitals?

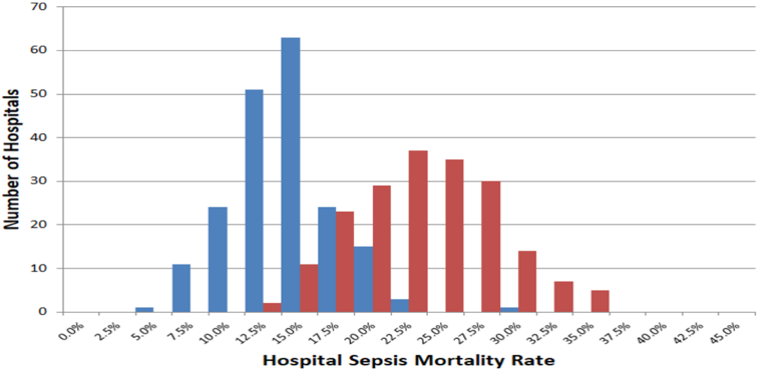
Study Cohort

- Retrospective cohort study of adults hospitalized in 2013 or 2014 at 193 hospitals drawn from 6 datasets:
  – Cerner HealthFacts, Emory, HCA Healthcare, Institute of Health Metrics, UPMC, and Brigham and Women's Hospital
- 4.3 million adult admissions:
  - 117,000 explicit sepsis codes
  - 266,000 EHR clinical sepsis

Hospital Characteristics

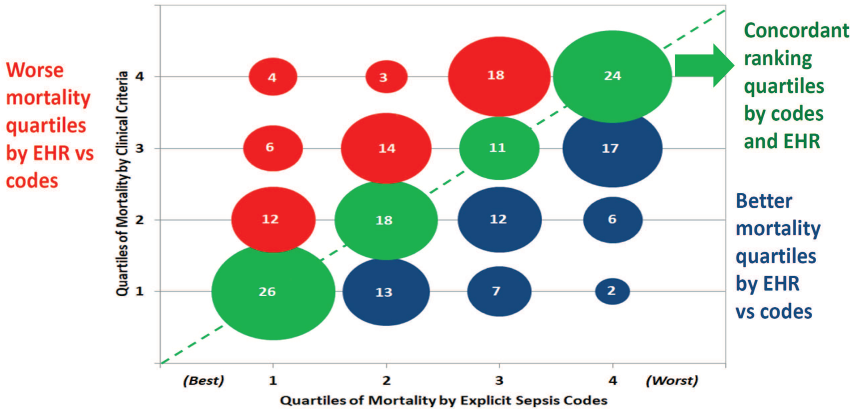
Hospital Characteristic	Distribution Among Hospitals
Region	
Northeast	25(13%)
Midwest	14 (7%)
South	118 (62%)
West	34 (18%)
Teaching Status	
Teaching	64 (33%)
Nonteaching	129 (67%)
Number of Beds	
<200 (Small)	73 (39%)
200-499 (Medium)	98 (52%)
500+ (Large)	17 (9%)

Hospital Sepsis Mortality



Blue: EHR Clinical Sepsis; Red: Explicit Sepsis Codes

Concordance of Hospital Ranking



- 51% ranked by claims in the worst quantile ranked out by EHR
- 51% ranked by EHR in the worst quantile ranked out by claims

Ranking Hospital Performance

"Provide hospital-specific performance metrics for an array of procedures that incorporate the best possible information for each hospital as to how well it performs with its patients in comparison to the outcomes that would be expected if the same patients were to receive care that matched the national norm."

- Best possible information: hospital procedure volume, patient-mix, hospital characteristics
- National norm: how to obtain this?

Ash AS, Feinberg SE, Louis TA, Normand S-LT, Stukel T, Utts J. Statistical Issues in Assessing Hospital Performance. Commissioned by the Committee of Presidents of Statistical Societies for the Centers for Medicare and Medicaid Services (CMS). January 27, 2012.

Clustering Within Hospital

- A random effects model:

$$\text{logit}(\pi_{ij}) = \eta_i + \beta^T X_{ij}, \quad \eta_i \sim N(\eta, \sigma^2).$$

- A marginal model:

$$\text{logit}(\pi_{ij}) = \tilde{\eta} + \tilde{\beta}^T X_{ij}.$$

- Which one should be used as "national norm":
  - $\text{expit}(\eta + \beta^T X_{ij})$
  - $\text{expit}(\tilde{\eta} + \tilde{\beta}^T X_{ij})$

- A stratified model:

$$\text{logit}(\pi_{ij}) = \eta^* + \beta^{*T} X_{ij} + \gamma_i^T \mathbf{H}_{ij}, \quad \mathbf{H}_{ij} = I(\text{in ith hospital}).$$

Adjusting for Patient Mix

- The effect of biomarker may be nonlinear
- Potential interactions
- The need to come up with a parsimonious model for general applicability
- Combining claims data and EHR data may improve adjustment
- Develop and validate sepsis risk-adjustment models using CDC's adult sepsis event criteria and routinely collected EHR data in two large cohort of U.S. hospitals

Data Source

- Primary dataset: all adults admitted to 136 hospitals in the Cerner HealthFacts dataset from 2009-2015
  - 2/3rd used for model development, 1/3rd for internal validation (better way to improve efficiency?)
  - 97,352 patients with CDC sepsis
- Data from adults admitted in 2013-2014 to 137 hospitals in the HCA healthcare network were used for external validation
  - 201,997 patients with CDC sepsis

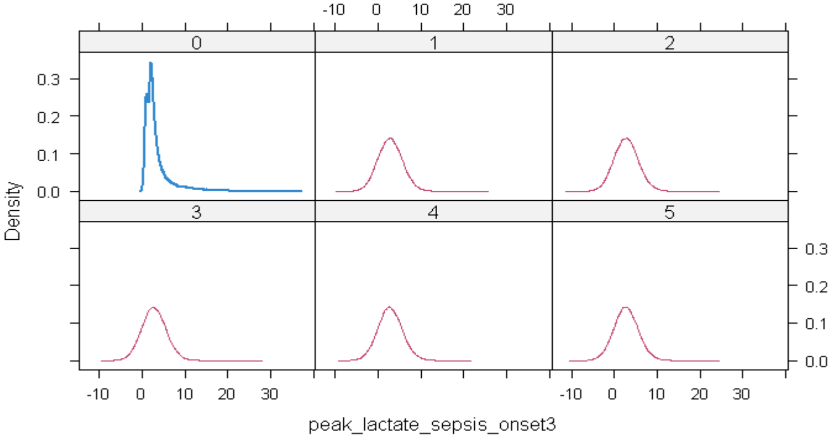
Predictors

Predictor Category	Variables
Demographics	Age, Sex, Race
Comorbidities	Elixhauser Comorbidity Groups + Leukemia, Stem Cell Transplant, Solid Organ Transplant (based on ICD-9 codes)
Infection Site	Pneumonia, Urinary, Intra-abdominal, Skin/Soft Tissue, Septicemia/Bacteremia, Obstetric/Gynecologic, CNS, 2 or more, Unknown/ None (based on ICD-9 codes)
Time to Sepsis	Days from admission to sepsis onset
ICU at Sepsis Onset	Whether patient was in ICU on day of sepsis onset
CDC Adult Sepsis Event Organ Dysfunction Variables	Number of vasopressors, mechanical ventilation, lactate, creatinine, bilirubin, platelet count
Extended Labs	White blood cell count, hematocrit, sodium, anion gap, albumin, AST, INR
Microbiology	Positive blood culture
Vital Signs	Systolic blood pressure, Temperature, Respiratory Rate
Mental Status	Glasgow Coma Scale

Missing Important Covariates

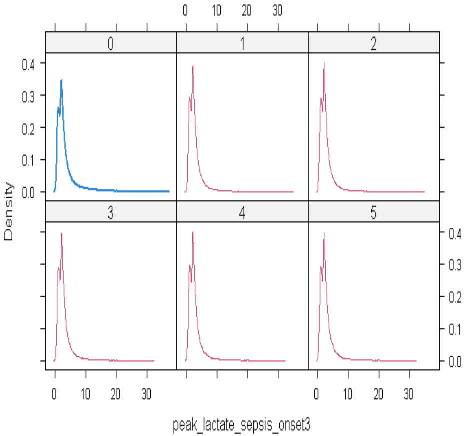
- In the perfect world, 100% of sepsis patients would have a lactate drawn
- In our dataset, 46% were missing
- Truncated linear regression: where the lower limit was set at a serum lactate level 0.1 mmol/L (first percentile) and the upper limit was set at 30.0 mmol/L (99th percentile) (used in Philips et al., 2018. The New York Sepsis Severity Score: development of a risk-adjusted severity model for sepsis. Critical care medicine 46.5: 674-683.)
- Imputed distributions do not resemble the observed distribution

Truncated Linear Regression



Predictive Mean Matching

- Generate predicted values for x for all cases (missing and observed) from posterior predictive distribution.
- For each missing x, identify a set of cases with observed x whose predicted values are close to the predicted values for the case with missing data.
- Then among those, randomly choose one for the missing x.

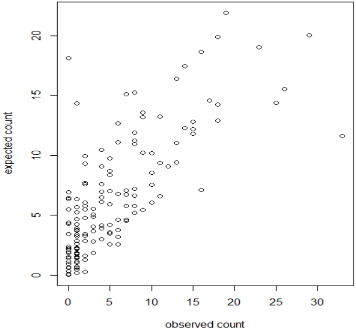


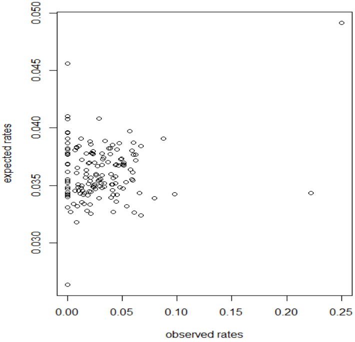
Overview of Models

Predictor Category	Model 1 (Basic Admin. Model)	Model 2 (+Adult Sepsis Event Criteria)	Model 3 (+Extended Labs)	Model 4 (+Blood Culture Results)	Model 5 (Maximal Clinical)	Ford Admin. Model*
Demographics	X	X	X	X	X	
Comorbidities	X	X	X	X	X	
Infection Site	X	X	X	X	X	
Days to Sepsis Onset	X	X	X	X	X	
ICU at Sepsis Onset	X	X	X	X	X	
Adult Sepsis Event Organ Dysfunction Criteria		X	X	X	X	
Extended Labs			X	X	X	
Blood Culture Results				X	X	
Vital Signs and GCS					X	

*Ford model based entirely on administrative data (age, sex, race, early/late mechanical ventilation, shock, hemodialysis, ICU care, comorbidities)

Evaluating Model Performance





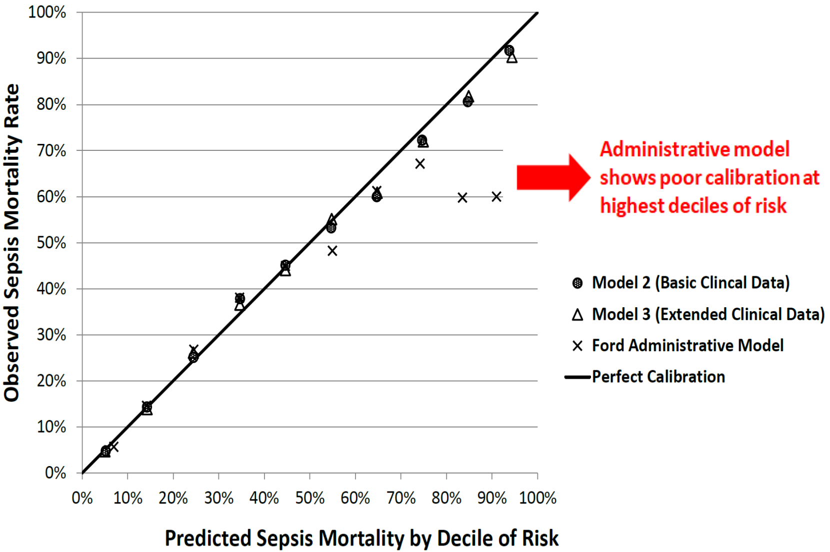
The relationship between expected count and observed count may be confounded by hospital procedure volume.

Model Results

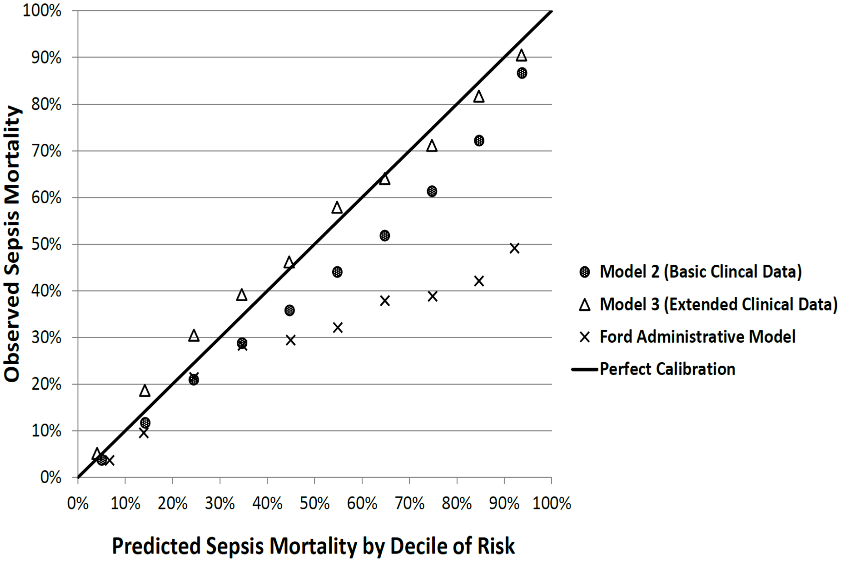
Model Characteristics	Model 1 (Basic Admin. Model)	Model 2 (+CDC Criteria)	Model 3 (+Extended Labs)	Model 5 (Maximal Clinical)	Ford Admin. Model
AUROC Training (Cerner)	0.73	0.82	0.83	0.835	0.78
AUROC Internal Validation (Cerner)	**0.72**	**0.82**	**0.83**	**0.834**	**0.78**
AUROC External Validation (HCA)	**0.71**	**0.82**	**0.83**	**N/A**	**0.77**

***Model 4 (+blood culture results) had no impact on performance → not shown**

Calibration: Cerner Internal Validation



Calibration: HCA External Validation



Summary

- Substantial discrepancies were seen comparing claims database vs. NHSN, and claims vs. EHR
- Incorporating routinely collected EHR data may improve model performance
- Accurate hospital ranking relies on proper adjustment of procedure volume and patient-mix: this can be complicated by missing data, unknown functional form of the covariates, and their interactions
- Unclear the best way to obtain "national norm"
- Small procedure volume introduces large sampling variability within hospital, which further complicates ranking accuracy

Acknowledgements

Study Co-Investigators:

- *Harvard:* Chanu Rhee, Michael Klompas, Maximilian Jentzsch, Robert Jin, Heather Hsu, Kelly Horan, Heather Hsu, Ed Septimus
- *Stanford*: Grace Lee
- *NIH Clinical Center:* Sameer Kadri, Robert Danner
- *CDC:* Raymund Dantes, Lauren Epstein, Anthony Fiore, John Jernigan
- *UPMC:* Chris Seymour, Derek Angus
- *Emory:* David Murphy, Greg Martin
- *Wash U*: David Warren
- *HCA Healthcare:* Jason Hickok, Russell Poland
- *Commonwealth Analytics*: David Fram, Richard Schaaf

Study Support and Supervision

- Richard Platt (Harvard)
- Jonathan Perlin (HCA Healthcare)

Funding:

- CDC Prevention Epicenters (U54CK000484)
- AHRQ K08 (K08HS025008)
- AHRQ 2R01HS018414-06
- Departmental Grant from Harvard Pilgrim Health Care Institute