

Bayesian Hospital Mortality Rate Estimation, Calibration and Standardization for Public Reporting

Edward I. George

Veronika Ročková, Paul R. Rosenbaum, Ville A. Satopää and Jeffrey H. Silber

The Wharton School, University of Pennsylvania

edgerge@upenn.edu

Statistics Conference

In Honor of Aad van der Vaart's 60th Birthday

Leiden, The Netherlands

June 20, 2019

Prologue - Is shrinkage estimation really safe?

- Observe $x_i \mid \mu_i \sim N(\mu_i, v_i), i = 1, \dots, p$ independently.
- Hierarchical model: $\mu_1, \dots, \mu_p \stackrel{\text{iid}}{\sim} N(\mu, v)$.
- eB shrinkage estimator: $\hat{\mu}_j = \frac{v_\mu}{v_i + \hat{v}} x_i + \frac{v_i}{v_i + \hat{v}} \bar{x}$
- $\hat{\mu}_j$ strongly shrinks x_i towards $\bar{x}$ when $v_i \gg \hat{v}$
- Can be great when $\mu_1, \dots, \mu_p$ all have the same μ and v
- Can be a disaster otherwise!

Some History - The National Halothane Study (1969)

- 856,500 surgeries under anesthesia from 1959-1962
- 16,840 deaths within 6 weeks of surgery (about 2%)
- Compared the effect of general anesthetic agents, especially halothane, on postoperative mortality
- Reported standardized mortality rates

THE NATIONAL HALOTHANE STUDY

A STUDY OF THE POSSIBLE ASSOCIATION BETWEEN HALOTHANE ANESTHESIA AND POSTOPERATIVE HEPATIC NECROSIS

Report of
The Subcommittee on the National Halothane Study,
of the Committee on Anesthesia, Division of Medical Sciences,
National Academy of Sciences-National Research Council
Washington, D.C.

Edited by
John P. Bunker, M.D., William H. Forrest, Jr., M.D.
Frederick Mosteller, Ph.D., and Leroy D. Vandam, M.D.

Supported by
Department of Health, Education, and Welfare
Public Health Service, National Institute of General Medical Sciences
Contract PH43-63-65

NAS-ME

APR 25 1969

LIBRARY

National Institutes of Health
National Institute of General Medical Sciences
Bethesda, Md.

1969

Public Reporting of Hospital Mortality Rates Today

- Medicare's web based "Hospital Compare":

To provide the public "with information on how well the hospitals in your area care for all their adult patients with certain medical conditions" such as heart attacks.

(U.S. Department of Health and Human Services, 2007)

- Available at <http://www.hospitalcompare.hhs.gov>
- Hospital Compare's approach is being copied as the "Gold Standard" for general performance comparisons.

A Typical Medicare Mortality Rate Report

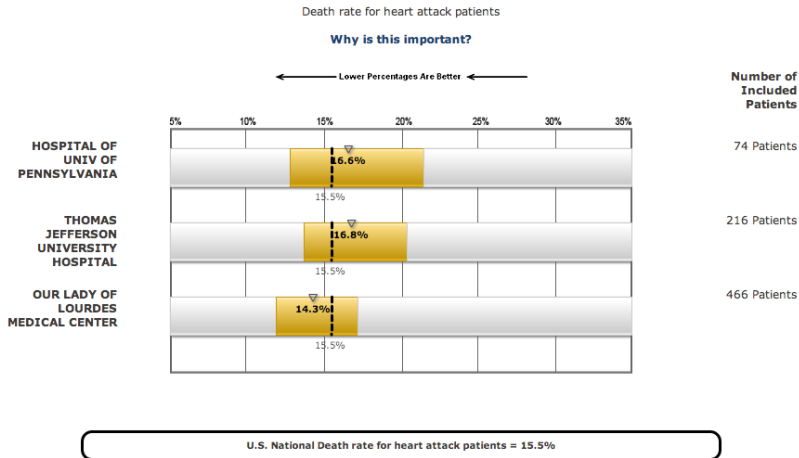


Figure: Comparing heart attack mortality rates with Hospital Compare

Medicare Public Reporting

- Hospital Compare 2008: Out of 4311 hospitals, 4302 of them (99.8%) are “no different than U.S. National rate” and zero hospitals are “worse than U.S. National rate”.
- How did Hospital Compare reach these conclusions?
 - The smaller the hospital volume, the more its mortality rate estimate is “shrunk” to the overall mean.
 - Medicare’s justification: Estimates for small volume hospitals rely on the pooled data of all hospitals: “this pooling affords borrowing of statistical strength that provides more confidence in the results.”
- UH-OH! Hospital Compare’s estimates contradict the conventional wisdom that mortality rates are higher at low volume hospitals!!!

Administrative Data from Medicare Billing Records

- Medicare data on AMI (Acute Myocardial Infarction) cases from July 1, 2009 to December 31, 2011.
- 377,615 AMI patients admitted to 4,289 hospitals.
- 56,567 deaths within 30 days of admission (about 15%)
- 27 patient characteristics (e.g. age, heart failure, hypertension etc)
- 4 hospital characteristics (volume, resident-to-bed ratio, nurse-to-bed ratio, PCI)
- **Training:** first 2 years. **Validation:** remaining 6 months.

Hospital Compare's Reported Mortality Rates by Volume

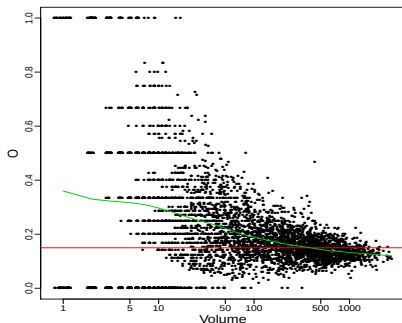


Figure: Observed Hospital Rates

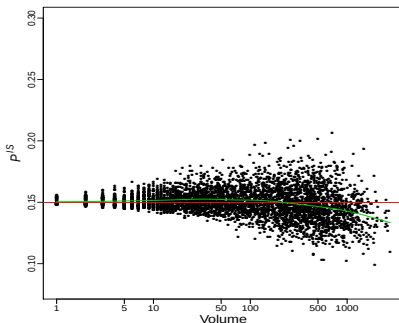


Figure: Reported Hospital Rates

Two Major Steps:

- 1 Hospital Compare begins with a log linear random effects model to predict hospital mortality rates.
- 2 These rates are then standardized (indirectly) to adjust for patient case-mix differences.

Hospital Compare's Random Effects Model

$$\log \left(\frac{p_{hj}}{1 - p_{hj}} \right) = \alpha_h + \mathbf{x}'_{hj}\boldsymbol{\beta},$$

where

- $p_{hj} = P(Y_{hj} = 1)$: 30-day mortality rate at hospital h for patient j , ($h = 1, 2, \dots, H$ and $j = 1, 2, \dots, n_h$).
- $\alpha_h \sim N(\mu, \sigma^2)$: hospital random effects.
- $\mathbf{x}'_{hj}\boldsymbol{\beta}$: patient fixed effects (based on patient characteristics $\mathbf{x}_{hj}$).
- Fit using PROC GLIMMIX in SAS, Krumholz et al. (2006ab).

Hospital Compare Model Estimates

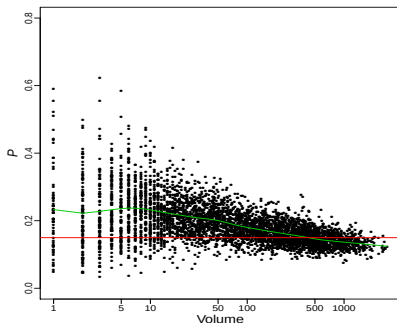


Figure: P_h Hospital Mortality Rates

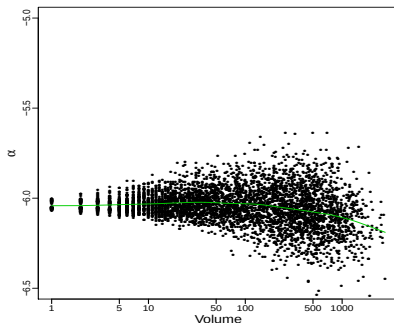


Figure: α_h Hospital Effects

- $P_h = \frac{1}{n_h} \sum_{j=1}^{n_h} p_{hj}$ have shrunk the raw observed mortality rates.
- Strong shrinkage of the α_h 's for small volume hospitals.
- Under this model, mortality rate P_h variation at small volume hospitals driven primarily by patient case-mix differences.

Shouldn't Hospital Characteristics be in the Model?

- Hospital effects have been modeled as $\alpha_h \sim N(\mu, \sigma^2)$, completely random with the same mean and variance!
- This assumption is leading to the strong shrinkage of the α_h 's for small volume hospitals!
- Available hospital characteristics have been left out!
- Proponents of the Hospital Compare approach argue that including hospital characteristics in the model would be “unfair” to the hospitals.
- But isn't excluding hospital characteristics “unfair” to the people seeking accurate information?

Will Adding Hospital Characteristics
Make a Difference?

Suppose we elaborate $\alpha_h \sim N(\mu, \sigma^2)$ to $\alpha_h \sim N(\mu_h, \sigma_h^2)$

- μ_h and σ_h^2 can now be functions of hospital characteristics!

Model	μ_h	σ_h^2
(C,C)	Constant	Constant
(L,C)	Linear(log vol _h)	Constant
(S,L)	Spline(log vol _h)	Log-Linear(vol _h)
(SL,L)	Spline(log vol _h)+Linear(ptca _h , ntbr _h , rtbr _h)	Log-Linear(vol _h)
(SLI,L)	(SL,L) + (log vol _h × age _{hj}) Interaction	

- (C,C) is equivalent to Hospital Compare.
- Each subsequent model nests the previous one.
- Fully Bayesian implementations with non-influential, vague priors.
- MCMC calculations via posterior augmentation with Pólya-Gamma latent variables (Polson, Scott, and Windle 2013).

Emancipating the Means and Variances with Volume

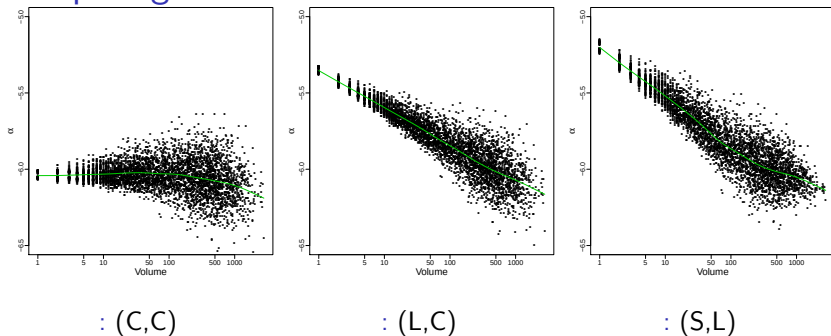


Figure: Posterior means of α_h vs vol_h

- Dramatic improvements over the (C,C) model.
- Data speak clearly because simpler models are nested.
- Higher mortality rates at low volume hospitals.

Adding More Hospital Characteristics and an Interaction

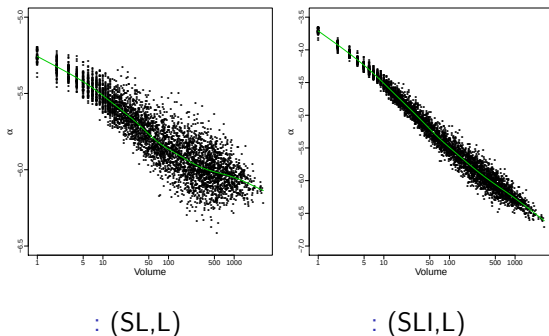


Figure: Posterior means of α_h vs vol_h

- Refinements continue to support higher mortality rates at low volume hospitals

Have Predictions Really Improved?

Model Comparisons via Predictive Bayes Factors

- In-sample Bayes Factors are unreliable with vague priors.
- Instead, use out-of-sample Predictive Bayes factors versus (C,C)

$$BF_{\mathcal{M}_i/\mathcal{M}_{CC}} = \frac{\mathbb{P}(y_{out} \mid y_{in}, \mathcal{M}_i)}{\mathbb{P}(y_{out} \mid y_{in}, \mathcal{M}_{CC})}$$

Model	(L,C)	(S,L)	(SL,L)	(SLI,L)
$\log BF$	27.54	32.13	35.46	37.96

- Vast successive improvements. (SLI,L) clearly best.

Matched Sample Comparisons of Model Predictions

- Predictive BF's gauge overall model fit.
- But what about the calibration of model predictions with future mortality rates on particular segments of patients?
- For this purpose, we compared each model's predictions to out-of-sample mortality rates on two sets of patients:
 - LV: Patients at low-volume hospitals (bottom 20%)
 - HV: Matched patients at high-volume hospitals (top 20%)
- Controlling patient risk characteristics through matching provides a clearer comparison of predicted mortality rates between low- and high-volume hospitals.

Out-of-Sample Comparisons

	Low Volume	High Volume Matched	High Volume All
Observed Mortality	28.3	19.8	12.4
(C,C)	23.1	21.6	12.7
(SLI,L)	29.6	21.0	12.4

Table: Out-of-sample predicted mortality compared against observed mortality in the matched study of low and high volume hospitals.

At low-volume hospitals, (C,C) is poorly calibrated.
It predicted 23.1% when the actual observed was 28.3%.

Indirect and Direct Standardization

Standardizing Mortality Rates for Public Reporting

- Mortality rate P_h at hospital h influenced by patient case-mix.
- Standardize P_h to eliminate this effect of case-mix variation.
- Two approaches:
 - Indirect standardization (used by Hospital Compare)
 - Direct standardization
- Both approaches make use of the fact that the mortality rate for patient $\mathbf{x}_{hj}$ at *any* hospital h^* can be obtained via

$$p_{h^*}(\mathbf{x}_{hj}) = \text{logit}^{-1}(\alpha_{h^*} + \mathbf{x}'_{hj}\beta).$$

- Note that $p_{h^*}(\mathbf{x}_{hj})$ is a *counterfactual* unless $h^* = h$.

Indirect Standardization

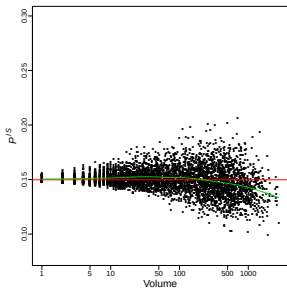
$$P_h^{IS} = (P_h/E_h) \times \bar{y},$$

where

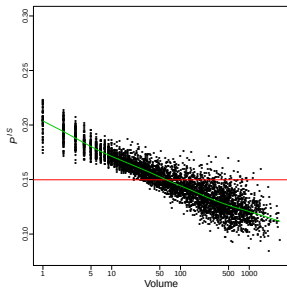
- $E_h = \frac{1}{n_h} \sum_{j=1}^{n_h} \left[\frac{1}{H} \sum_{h^*=1}^H p_{h^*}(\mathbf{x}_{hj}) \right]$
- E_h : Average mortality rate of hospital h patients had they been treated at all H hospitals.
- $\bar{y}$: national average mortality rate ($\approx 15\%$).

Some drawbacks:

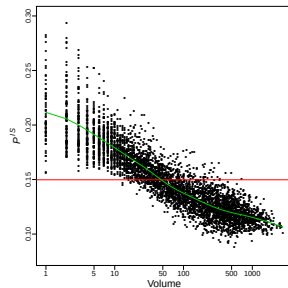
- lacks probabilistic justification.
- fails to eliminate case-mix variation effects, except for (C,C).
- systematically underestimates actual hospital mortality rates.



: (C,C)



: (L,C)



: (SLI,L)

Figure: Indirectly Standardized Mortality Rates P_h^{IS} vs. vol_h .

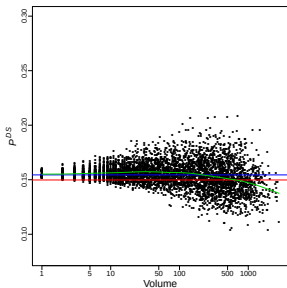
Direct Standardization

$$P_h^{DS} = \frac{1}{N} \sum_{h^*=1}^H \sum_{j=1}^{n_{h^*}} p_h(\mathbf{x}_{h^*j}), \quad N = \sum_{h^*=1}^H n_{h^*}$$

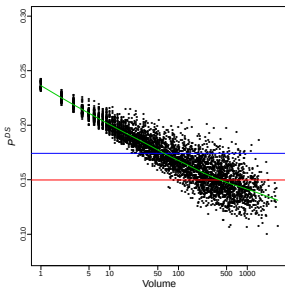
P_h^{DS} : Average mortality rate of all N patients had they been treated at hospital h .

Benefits of this approach

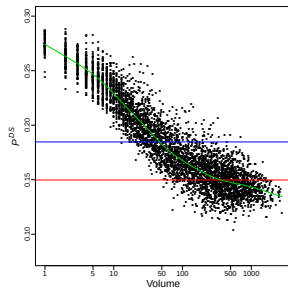
- + easier to understand.
- + an interpretable, almost linear scaling of α_h .
- + eliminates the effect of case-mix variation.
- + is correctly calibrated to actual mortality rates.



: (C,C)



: (L,C)



: (SLI,L)

Figure: Directly Standardized Mortality Rates P_h^{DS} vs. vol_h .

- red horizontal line - average mortality rate over patients
- blue horizontal line - average mortality rate over hospitals

Mortality Rate Uncertainty Quantification

- The variation of the P_h^{DS} posterior mean estimates is smaller at the low volume hospitals.
- However, the posterior mean variation should not be confused with posterior uncertainty of the estimates which is conveyed by the full posterior distribution of the P_h^{DS} values.

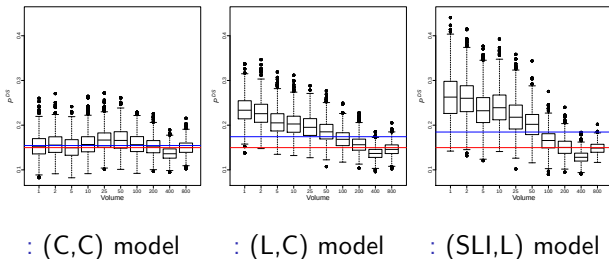


Figure: P_h^{DS} posterior uncertainty at 10 hospitals of varying volume.

Hospital Classification by Mortality Rates

- The credibility intervals for P_h^{DS} can be used to classify hospitals into Low, Average and High mortality according to whether its 95% interval is entirely below, intersects or is entirely above the overall average mortality rate of 15%.

Counts (%)	All Hospitals			Lower Volume Quartile			Upper Volume Quartile		
	Low	Average	High	Low	Average	High	Low	Average	High
(C,C)	33 (0.752)	4333 (98.57)	30 (0.68)	0 (0.00)	1116 (100.00)	0 (0.00)	32 (2.91)	1047 (95.27)	20 (1.82)
(SLI,L)	58 (1.32)	3310 (75.30)	1028 (23.38)	0 (0.00)	210 (18.82)	906 (81.18)	57 (5.19)	1038 (94.45)	4 (0.36)

Table: Hospital Classifications by Low, Average and High Mortality Rates.

Attribute Effects

- With P_h^{DS} values, meaningful insights into relationships between hospital mortality rates and PCI , $NTBR$ are readily obtained.

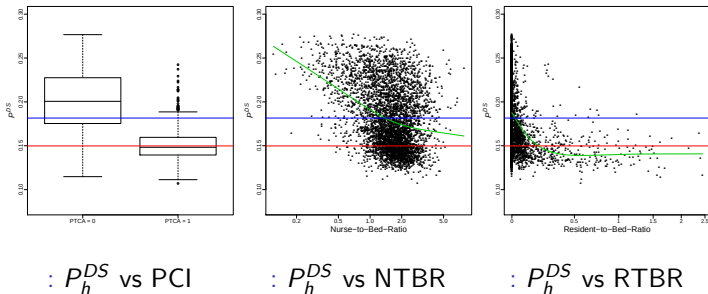


Figure: P_h^{DS} under the (SLI,L) model.

Conclusions

- Strong evidence that hospital characteristics and interactions should be included in the model.
- Indirect standardization fails to eliminate the effect of case-mix variation and underestimates actual mortality rates.
- Directly standardized rates should be the new gold standard for public reporting and for further analyses of what influences mortality.

Dilemma

- Should Medicare publicly report the alarmingly high mortality rates at the low volume hospitals?

Happy Birthday Aad!

Proost to you!

References:

- 1 *Mortality Rate Estimation and Standardization for Public Reporting: Medicare's Hospital Compare.*
E. George, V. Ročková, P. Rosenbaum, V. Satopää, J. Silber.
Journal of the American Statistical Association.
- 2 *Improving Medicare's Hospital Compare Mortality Model.*
J. Silber, V. Satopää, N. Mukherjee, V. Ročková, ..., P. Rosenbaum, E. George.
Health Services Research.