Module III: Profiling Health Care Providers – A Multi-level Model Application

Francesca Dominici
&
Michael Griswold

Outline
- What is profiling?
  - Definitions
  - Statistical challenges
  - Centrality of multi-level analysis
- Fitting Multilevel Models with Winbugs:
  - A toy example on institutional ranking
- Profiling medical care providers: a case-study
  - Hierarchical logistic regression model
  - Performance measures
  - Comparison with standard approaches

What is profiling?
- Profiling is the process of comparing quality of care, use of services, and cost with normative or community standards
- Profiling analysis is developing and implementing performance indices to evaluate physicians, hospitals, and care-providing networks

Objectives of profiling
- Estimate provider-specific performance measures:
  - measures of utilization
  - patients outcomes
  - satisfaction of care
- Compare these estimates to a community or a normative standard

Evaluating hospital performance
- Health Care Financing Administration (HCFA) evaluated hospital performance in 1987 by comparing observed and expected mortality rates for Medicare patients
- Expected Mortality rates within each hospital were obtained by:
  - Estimating a patient-level model of mortality
  - Averaging the model-based probabilities of mortality for all patients within each hospital
- Hospitals with higher-than-expected mortality rates were flagged as institutions with potential quality problems
Statistical Challenges
- Hospital profiling needs to take into account
  - Patients characteristics
  - Hospital characteristics
  - Correlation between outcomes of patients within the same hospital
  - Number of patients in the hospital
- These data characteristics motivate the centrality of multi-level data analysis

"Case-mix" bias
- Estimating hospital specific mortality rates without taking into account patient characteristics
  - Suppose that older and sicker patients with multiple diseases have different needs for health care services and different health outcomes independent of the quality of care they receive. In this case, physicians who see such patients may appear to provide lower quality of care than those who see younger and healthier patients
- Develop patient-level regression models to control for different case-mixes

Within cluster correlation
- Hospital practices may induce a strong correlation among patient outcomes within hospitals even after accounting for patients characteristics
- Extend standard regression models to multi-level models that take into account the clustered nature of the data

Health care quality data are multi-level!
- Data are clustered at multiple-levels
  - Patients clustered by providers, physicians, hospitals, HMOs
  - Providers clustered by health care systems, market areas, geographic areas
- Provider sizes may vary substantially
- Covariates at different levels of aggregation: patient-level, provider level
- Statistical uncertainty of performance estimates need to take into account:
  - Systematic and random variation
  - Provider-specific measures of utilization, costs

Sampling variability versus systematic variability
- "Sampling variability": statistical uncertainty of the hospital-specific performance measures
- "Systematic variability": variability between hospitals performances that can be possibly explained by hospital-specific characteristics (aka "natural variability")
- Develop multi-level models that incorporate both patient-level and hospital-level characteristics

Borrowing strength
- Reliability of hospital-specific estimates:
  - because of difference in hospital sample sizes, the precision of the hospital-specific estimates may vary greatly. Large differences between observed and expected mortality rates at hospitals with small sample sizes may be due primarily to sampling variability
- Implement shrinkage estimation methods: hospitals performances with small sample size will be shrunk toward the mean more heavily
Each point represents the amount of laboratory costs of patients who have diabetes deviates from the mean of all physicians (in US dollars per patient per year). The lines illustrate what happens to each physician's profile when adjusted for reliability (Hofer et al JAMA 1999).

Measures of Performance
- Patient outcomes (e.g., patient mortality, morbidity, satisfaction with care)
  - For example: 30-day mortality among heart attack patients (Normand et al JAMA 1996, JASA 1997)
- Process (e.g., specific medications given or tests done, costs for patients)
  - For example: laboratory costs of patients who have diabetes (Hofer et al JAMA, 1999)
- Number of physician visits (Hofer et al JAMA, 1999)

Fitting Multi-Level Models
- SAS / Stata
  - Maximum Likelihood Estimation (MLE)
  - Limitation: hard to estimate ranking probabilities and assess statistical uncertainty of hospital rankings
- BUGS and Bayesian Methods
  - Monte Carlo Markov Chains methods
  - Advantages: estimation of ranking probabilities and their confidence intervals is straightforward

Fitting Multilevel Models in Winbugs
A Toy example in institutional ranking
Toy example on using BUGS for hospital performance ranking

This example considers mortality rates in 12 hospitals performing cardiac surgery in babies. The data are shown below.

<table>
<thead>
<tr>
<th>Hospital</th>
<th>No of ops</th>
<th>No of deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>47</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>148</td>
<td>18</td>
</tr>
<tr>
<td>C</td>
<td>119</td>
<td>8</td>
</tr>
<tr>
<td>D</td>
<td>610</td>
<td>46</td>
</tr>
<tr>
<td>E</td>
<td>211</td>
<td>9</td>
</tr>
<tr>
<td>F</td>
<td>196</td>
<td>13</td>
</tr>
<tr>
<td>G</td>
<td>148</td>
<td>19</td>
</tr>
<tr>
<td>H</td>
<td>215</td>
<td>31</td>
</tr>
<tr>
<td>I</td>
<td>207</td>
<td>14</td>
</tr>
<tr>
<td>J</td>
<td>97</td>
<td>5</td>
</tr>
<tr>
<td>K</td>
<td>258</td>
<td>20</td>
</tr>
<tr>
<td>L</td>
<td>360</td>
<td>24</td>
</tr>
</tbody>
</table>

A Multi-level model for hospital ranking

- Let $r_i$ the number of deaths for hospital $i$
- Let $n_i$ the number of surgeries performed in hospital $i$

We assume
- $r_i \sim \text{Bin}(n_i, \pi_i)$
- $\log(\pi_i) = \mu + \lambda_i$
- $\lambda_i \sim N(0, \sigma^2)$

where
- $\pi_i$ is the hospital-specific probability of death
- $\bar{\pi} = \exp(\mu)/(1 + \exp(\mu))$ is the probability of death for $\lambda_i = 0$

Goal: identify the "aberrant" hospitals

Summary Statistics

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>p[1]</td>
<td>0.05357</td>
<td>0.01959</td>
</tr>
<tr>
<td>p[2]</td>
<td>0.1026</td>
<td>0.02203</td>
</tr>
<tr>
<td>p[3]</td>
<td>0.07102</td>
<td>0.01701</td>
</tr>
<tr>
<td>p[4]</td>
<td>0.05947</td>
<td>0.008078</td>
</tr>
<tr>
<td>p[5]</td>
<td>0.05252</td>
<td>0.01354</td>
</tr>
<tr>
<td>p[6]</td>
<td>0.06867</td>
<td>0.01401</td>
</tr>
<tr>
<td>p[7]</td>
<td>0.06796</td>
<td>0.01597</td>
</tr>
<tr>
<td>p[8]</td>
<td>0.1217</td>
<td>0.02196</td>
</tr>
<tr>
<td>p[9]</td>
<td>0.05943</td>
<td>0.01435</td>
</tr>
<tr>
<td>p[10]</td>
<td>0.07859</td>
<td>0.01938</td>
</tr>
<tr>
<td>p[11]</td>
<td>0.1019</td>
<td>0.01745</td>
</tr>
<tr>
<td>p[12]</td>
<td>0.06683</td>
<td>0.01165</td>
</tr>
<tr>
<td>pop.mean</td>
<td>0.07246</td>
<td>0.0105</td>
</tr>
</tbody>
</table>

BUGS Model specification

```b
model
{ 
  for (i in 1:N) {
    b[i] ~ dnorm(mu, tau)
    r[i] ~ dbin(p[i]*p[i], n[i])
    log(p[i]) ~ b[i]
  }
  pop.mean <- exp(mu) / (1 + exp(mu))
  mu ~ dnorm(0, 0.1, 0.1, 0.1)
  sigma <- 1 / sqrt(tau)
  tau ~ dgamma(0.001, 0.001)
}
```
Hospital Profiling of Mortality Rates for Acute Myocardial Infarction Patients (Normand et al JAMA 1996, JASA 1997)

- Data characteristics
- Scientific goals
- Multi-level logistic regression model
- Definition of performance measures
- Estimation
- Results
- Discussion

Data Characteristics

- The Cooperative Cardiovascular Project (CCP) involved abstracting medical records for patients discharged from hospitals located in Alabama, Connecticut, Iowa, and Wisconsin (June 1992- May 1993)
- 3,269 patients hospitalized in 122 hospitals in four US States for Acute Myocardial Infarction

Data characteristics

- Outcome: mortality within 30-days of hospital admission
- Patients characteristics:
  - Admission severity index constructed on the basis of 34 patient characteristics
- Hospital characteristics
  - Rural versus urban
  - Non academic versus academic
  - Number of beds

Scientific Goals:

- Identify “aberrant” hospitals in terms of several performance measures
- Report the statistical uncertainty associated with the ranking of the “worst hospitals”
- Investigate if hospital characteristics explain heterogeneity of hospital-specific mortality rates

Hierarchical logistic regression model

- I: patient level, within-provider model
  - Patient-level logistic regression model with random intercept and random slope
- II: between-providers model
  - Hospital-specific random effects are regressed on hospital-specific characteristics
Patient-level model

- $Y_{ij}$ is the binary indicator of death within 30 days of admission for patient $j$ at hospital $i$
- severity$_{ij}$ is the severity index for patient $j$ at hospital $i$
- severity is average severity index
  \[
  \logit(P(Y_{ij} = 1)) = \beta_0 + \beta_1 \text{severity}_{ij} - \text{severity}
  \]
- $\beta_0$ and $\beta_1$ are random intercept and slope
- $\beta_0$ denotes the log-odds ratio of death for hospital $i$ having patients with severity equal to the average
- $\beta_1$ denotes the hospital-specific association between severity and logit of probability of death

Hospital-level model

- Without hospital-specific covariates
  \[
  \beta_0 = \gamma_0 + N(0, \sigma_0^2)
  \]
  \[
  \beta_1 = \gamma_1 + N(0, \sigma_1^2)
  \]
- With hospital-specific covariates
  \[
  \beta_0 = \gamma_0 + \gamma_{0 \text{rural}} + \gamma_{0 \text{non-acad}} + \gamma_{0 \text{small}} + \gamma_{0 \text{medium}} + N(0, \sigma_0^2)
  \]
  \[
  \beta_1 = \gamma_1 + \gamma_{1 \text{rural}} + \gamma_{1 \text{non-acad}} + \gamma_{1 \text{small}} + \gamma_{1 \text{medium}} + N(0, \sigma_1^2)
  \]

where rural, non-acad, small, medium are indicators of a rural, non academic, small, and medium size beds hospitals, respectively.

The interpretation of the parameters are different under these two models.

Normand et al JASA 1997

Table 1. Patient and Hospital Characteristics in the Study Cohort

<table>
<thead>
<tr>
<th>Hospital characteristic</th>
<th>% of patients</th>
<th>% of Hospitals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural (yes, urban)</td>
<td>54</td>
<td>76</td>
</tr>
<tr>
<td>Nonacademic (yes, academic)</td>
<td>79</td>
<td>85</td>
</tr>
<tr>
<td>Number of beds</td>
<td>100-500</td>
<td>50</td>
</tr>
<tr>
<td>500-700 (medium)</td>
<td>27</td>
<td>21</td>
</tr>
<tr>
<td>&gt;700 (large)</td>
<td>44</td>
<td>52</td>
</tr>
</tbody>
</table>

Hospital-Performance Measures

- Let $\mu_i^A$ be the “adjusted” mortality rate for hospital $i$
- Let $\mu_i^S$ be the “standardized” mortality rate for a “reference” hospital
- We assume that a provider’s performance is poor if the probability that $\mu_i^A > \mu_i^S$ being bigger than some benchmark value is large. We estimate:
  \[
  P_{i}^{A < S} = P(\mu_i^A < \mu_i^S | \text{benchmark})
  \]
  where
  \[
  \alpha_i = \frac{1}{n} \sum_{j=1}^{n} \logit^{-1}(\beta_0 + \beta_1, \text{severity}_{ij} - \text{severity})
  = \frac{1}{n} \sum_{j=1}^{n} \logit^{-1}(\beta_0 + \beta_1, \text{severity}_{ij} - \text{severity})
  \]
  \[
  \beta_i = \gamma_0 + \gamma_{0 \text{rural}} + \gamma_{0 \text{non-acad}} + \gamma_{0 \text{small}} + \gamma_{0 \text{medium}} + N(0, \sigma_0^2)
  \]

Estimating Hospital Profiling Measures with MCMC

- Probability of a large difference between adjusted and standardized mortality rates:
  \[
  P_i^{A < S} = P(\mu_i^A < \mu_i^S)
  \]

- Where:
  \[
  \alpha_i = \frac{1}{n} \sum_{j=1}^{n} \logit^{-1}(\beta_0 + \beta_1, \text{severity}_{ij} - \text{severity})
  \]
  \[
  \beta_i = \gamma_0 + \gamma_{0 \text{rural}} + \gamma_{0 \text{non-acad}} + \gamma_{0 \text{small}} + \gamma_{0 \text{medium}} + N(0, \sigma_0^2)
  \]
- $H$ is a benchmark value that can be calculated based upon the distribution of $P_i^{A < S}$ across hospitals.
Comparing measures of hospital performance

- Three measures of hospital performance
  - Probability of a large difference between adjusted and standardized mortality rates
  - Probability of excess mortality for the average patient
  - Z-score

Results

- Estimates of regression coefficients under three models:
  - Random intercept only
  - Random intercept and random slope
  - Random intercept, random slope, and hospital covariates

- Hospital performance measures

Estimates of log-odds of 30-day mortality for a “average patient”

- Exchangeable model (without hospital covariates), random intercept and random slope:
  - We found that the 2.5 and 97.5 percentiles of the log-odds of 30-day mortality for a patient with average admission severity is equal to (-1.87, -1.56), corresponding to (0.13, 0.17) in the probability scale

- Non-Exchangeable model (with hospital covariates), random intercept and random slope:
  - We found that the 2.5 and 97.5 percentiles of the log-odds of 30-day mortality for a patient with average admission severity treated in a large, urban, and academic hospital is equal to (-2.15, -1.45), corresponding to (0.10, 0.19) in probability scale
Effect of hospital characteristics on baseline log-odds of mortality

- Rural hospitals have higher odds ratio of mortality than urban hospitals for an average patient
- This is an indication of inter-hospital differences in the baseline mortality rates

Estimates of II-stage regression coefficients (intercepts)

\[ \beta_0 = \gamma_0 + \gamma_{rural} + \gamma_{non-acad} + \gamma_{small} + \gamma_{medium} + \epsilon \]

<table>
<thead>
<tr>
<th>Hospital Type</th>
<th>Intercept ( \hat{\gamma} )</th>
<th>SE ( \hat{\gamma} )</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural</td>
<td>-1.79</td>
<td>0.17</td>
<td>(-2.18, -1.40)</td>
</tr>
<tr>
<td>Non-Academic</td>
<td>0.05</td>
<td>0.76</td>
<td>(-1.15, 1.25)</td>
</tr>
<tr>
<td>Small</td>
<td>-2.17</td>
<td>0.74</td>
<td>(-3.57, -0.77)</td>
</tr>
<tr>
<td>Medium</td>
<td>-3.28</td>
<td>0.74</td>
<td>(-4.70, -1.86)</td>
</tr>
</tbody>
</table>

Effects of hospital characteristics on associations between severity and mortality (slopes)

- The association between severity and mortality is "modified" by the size of the hospitals
- Medium-sized hospitals having smaller severity-mortality associations than large hospitals
- This indicates that the effect of clinical burden (patient severity) on mortality differs across hospitals

Estimates of II-stage regression coefficients (slopes)

\[ \beta_1 = \gamma_0 + \gamma_{rural} + \gamma_{non-acad} + \gamma_{small} + \gamma_{medium} + \epsilon \]

<table>
<thead>
<tr>
<th>Hospital Type</th>
<th>Severity ( \hat{\gamma} )</th>
<th>SE ( \hat{\gamma} )</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural</td>
<td>1.12</td>
<td>0.15</td>
<td>(0.82, 1.42)</td>
</tr>
<tr>
<td>Non-Academic</td>
<td>0.05</td>
<td>0.30</td>
<td>(-0.27, 0.36)</td>
</tr>
<tr>
<td>Small</td>
<td>-1.11</td>
<td>0.30</td>
<td>(-1.74, -0.48)</td>
</tr>
<tr>
<td>Medium</td>
<td>-0.08</td>
<td>0.20</td>
<td>(-0.47, 0.21)</td>
</tr>
</tbody>
</table>

Observed and risk-adjusted hospital mortality rates: Crossover plots

Display the observed mortality rate (upper horizontal axis) and corresponding risk-adjusted mortality rates (lower horizontal line). Histogram represents the difference = observed - adjusted

Substantial adjustment for severity!
What are these pictures telling us?

- Adjustment for severity on admission is substantial (mortality rate for an urban hospital moves from 29% to 37% when adjusted for severity)
- There appears to be less variability in changes between the observed and the adjusted mortality rates for urban hospitals than for rural hospitals

Ranking of hospitals

- There was moderate disagreement among the criteria for classifying hospitals as "aberrant"
- Despite this, hospital 1 is ranked as the worst. This hospital is rural, medium sized non-academic with an observed mortality rate of 35%, and adjusted rate of 28%

Discussion

- Profiling medical providers is a multi-faced and data intensive process with significant implications for health care practice, management, and policy
- Major issues include data quality and availability, choice of performance measures, formulation of statistical analyses, and development of approaches to reporting results of profiling analyses

Discussion

- Performance measures were estimated using a unifying statistical approach based on multi-level models
- Multi-level models:
  - take into account the hierarchical structure usually present in data for profiling analyses
  - provide a flexible framework for analyzing a variety of different types of response variables and for incorporating covariates at different levels of hierarchical structure

Discussion

- In addition, multi-level models can be used to address some key technical concerns in profiling analysis including:
  - permitting the impact of patient severity on outcome to vary by provider
  - adjusting for within-provider correlations
  - accounting for differential sample size across providers
- The multi-level regression framework permits risk adjustment using patient-level data and incorporation of provider characteristics into the analysis
Discussion

- The consideration of provider characteristics as possible covariates in the second level of the hierarchical model is dictated by the need to explain as large a fraction as possible of the variability in the observed data.
- In this case, more accurate estimates of hospital-specific adjusted outcomes will be obtained with the inclusion of hospital specific characteristics into the model.

Key words

- Profiling
- Case-mix adjustment
- Borrowing strength
- Hierarchical logistic regression model
- Bayesian estimation and Monte Carlo Markov Chain
- Ranking probabilities