Using Hierarchical Models to Attribute Sources of Variation in Consumer Assessments of Health Plans

Alan M. Zaslavsky
Harvard Medical School
Outline

• CAHPS overview
• Dimensions of quality and casemix
  – Modeling Level II variance structure
• Sources of variation in assessments
CAHPS® overview

• “Consumer Assessments of Health Plans Study”

• Methodology for surveys on quality of health care
  – Survey instruments and methods
  – Analysis, reporting formats
2002 Medicare Satisfaction Survey

CAHPS®
CAHPS Rating Items

• Respondent synthesizes experiences into summary rating

• 0 (worst possible) to 10 (best) scale

• 4 items (plan, doctor, care, specialists)
33. In the last 6 months, how often did doctors or other health providers listen carefully to you?
- Never
- Sometimes
- Usually
- Always
- I had no visits in the last 6 months.

34. In the last 6 months, how often did doctors or other health providers explain things in a way you could understand?
- Never
- Sometimes
- Usually
- Always
- I had no visits in the last 6 months.

35. In the last 6 months, how often did doctors or other health providers show respect for what you had to say?
- Never
- Sometimes
- Usually
- Always
- I had no visits in the last 6 months.

37. How would you rate all the health care you got in the last 6 months from all doctors and other health providers?
Using any number from 0 to 10 where 0 is the worst health care possible, and 10 is the best health care possible, what number would you use to rate all your health care?
- 0 WORST HEALTH CARE POSSIBLE
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 BEST HEALTH CARE POSSIBLE
- I had no visits in the last 6 months.
CAHPS Report Items

- Reports on specific aspects of care
- Many (30-40) items
- Need to synthesize these for reporting
CAHPS Report Items

• Various response formats:
  – Yes/No (0/1)
  – Frequency: Never/Sometimes/Usually/Always (1-4)
  – Big problem / small problem / Not a problem (1-3)
CAHPS Report Items

• Answered only by those with relevant experience
  – Screener items, skip patterns
  – Greatly varying item response rates (12% - 94%)
  – “Missing” values not conceptually meaningful
### OTHER HEALTH SERVICES

The next questions ask about your experience with other types of health services you may have had in the last 6 months.

38. In the last 6 months, did you have a health problem for which you needed special medical equipment, such as a cane, a wheelchair, or oxygen equipment?

- [ ] Yes
- [ ] No → If no, go to Question 40

39. In the last 6 months, how much of a problem, if any, was it to get the special medical equipment you needed through your Medicare health plan?

- [ ] A big problem
- [ ] A small problem
- [ ] Not a problem
- [ ] I didn’t need special therapy in the last 6 months.

41. In the last 6 months, how much of a problem, if any, was it to get the special therapy you needed through your Medicare health plan?

- [ ] A big problem
- [ ] A small problem
- [ ] Not a problem
- [ ] I didn’t need special therapy in the last 6 months.

42. Home health care or assistance means home nursing, help with bathing or dressing, and help with basic household tasks.

In the last 6 months, did you need someone to come into your home to give you home health care or assistance?

- [ ] Yes
- [ ] No → If no, go to Question 44 on Page 8

43. In the last 6 months, how much of a problem, if any, was it to get the home health care or assistance you needed through your Medicare health plan?
Standard CAHPS Analysis & Reporting

• Means for each item by health plan
• Case-mix adjustment (for age, health status, education) via linear regression
• For grouped items, means of item means
• Variance estimation: Taylor linearization of ratio (or mean of ratios)
Medicare Managed Care Implementation (CAHPS-MMC)

- About 5 million beneficiaries
- Part of quality efforts of country’s largest purchaser of health care (CMS)
- Largest single national study of consumer assessments (c. 900,000 surveyed over 7 years)
Structure of Data

• Content domains/items
• Patient subgroups
• Organizational:
  – Health plan
  – Delivery system, physician group, physician
• Geography:
  – Region, State, MSA, County, Zip code
• Time (multiple years)
(1) Dimensions of Consumer Assessments

• Question: how do items group (correlate) at the *plan* level?

• May be different from person-level (psychometric) analyses

• Relevant to reporting, incentives, case studies, research on determinants of quality
Plan-level covariance structure:

- Identify related items to summarize for reporting
- Understand relationships of plan attributes measured by survey
Data:
Years 2-3 MMC CAHPS

- 381 reporting units (risk contracts or parts)
- 290,739 respondents (duplicates removed)
- Items analyzed include:
  - 4 global ratings (plan, doctor, care, specialist)
  - 30 report items
Analytic approach

• Estimate plan-level covariance matrix
• Rotated principle factor analysis of report items
  – Guide to grouping items
• Regress rating items on report items (or composites) – validate items
Complication: Subgroups by health status

• General health status (among those who respond):
  – Excellent, very good, good = 70%
  – Fair, poor = 30%
“Sick” and “healthy” rate plans similarly, but not identically

• For each item, calculate correlation between plan means for “sick” and “healthy”

• Mean correlation (across items) of these means is 0.49 (corrected for sampling error)
Some items apply generally, some especially to sicker patients

<table>
<thead>
<tr>
<th>Item</th>
<th>Completion rate for: Hlthy</th>
<th>Sick</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating of plan</td>
<td>96%</td>
<td>96%</td>
</tr>
<tr>
<td>Rating of personal doctor</td>
<td>76%</td>
<td>79%</td>
</tr>
<tr>
<td>Doctor explains things</td>
<td>71%</td>
<td>82%</td>
</tr>
<tr>
<td>Doc understands health probs</td>
<td>11%</td>
<td>48%</td>
</tr>
<tr>
<td>Plan provided help</td>
<td>6%</td>
<td>37%</td>
</tr>
</tbody>
</table>
Hence different items represent different populations

• Rating of plan
  – Healthy = 96% X 70% = 67% of all
  – Sick = 96% X 30% = 29% of all

• Plan provided help
  – Healthy = 6% X 70% = 4.2% of all
  – Sick = 37% X 30% = 11.1% of all
**Sick patients report more problems**

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean rating for: Healthy</th>
<th>Sick</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating of plan (0-10)</td>
<td>8.81</td>
<td>8.31</td>
</tr>
<tr>
<td>Rating of care (0-10)</td>
<td>8.94</td>
<td>8.63</td>
</tr>
<tr>
<td>Doctor listens carefully (1-4)</td>
<td>3.73</td>
<td>3.58</td>
</tr>
<tr>
<td>Customer service helpful (1-4)</td>
<td>3.51</td>
<td>3.36</td>
</tr>
</tbody>
</table>
Analysis of subgroup-specific measures

• Form “sick”, “healthy” means for each item
• Apply factor analysis to 55 measures
• For which items do responses from sick and healthy group separately?
Factor analysis results (8-factor solution)

• Interactions with doctor / doctor’s office for healthy members (11 subitems)
• Interactions with doctor / doctor’s office for sick members (9 subitems)
• Finding satisfactory doctor, getting needed care/referrals for sick members (6 subitems) … etc.
Factor analysis results (continued)

• Customer service for all members (10 subitems)

• Getting plan-provided services for all members, getting doctor/referral for healthy members (9 subitems)
Factor analysis results (continued)

• Getting vaccinations, waiting over 15 minutes in doctor’s office for all members (6 subitems)

• Prescription drugs for all (4 subitems)

• Advice to quit smoking
Conclusions on report items

• Relatively few dimensions of quality
• Customer service interactions, preventive care, waits similar for healthy, sick.
• Different needs in direct care:
  – Plans/areas where healthy get good care not necessarily best for sick patients.
Conclusions on report items

• Getting care/referrals more distinct issue for sick patients.
• Reporting by subpopulation may be more informative than fine division of items
• Similar issue for other subgroups
  – Children vs. adults
Implications

• For measurement:
  – Analyze health information on CAHPS to define subgroups better (and possibly collect more).
  – Define skip patterns so information can be collected for all relevant subgroups.
Implications (continued)

• For reporting:
  – Report separately for sick, healthy on direct care composites
  – Consumers want to hear about experiences of “people like me”
Implications (continued)

• For analysis and research:
  – Distinguish health-status subgroups in analyses of determinants of quality
  – Research needed on how to define subgroups (conditions, health status, functional limitations, utilization)
(2) Risk selection among plans

- Do some plans get sicker/healthier members?
- Critical issue for Medicare: adversely selected plans might be driven out
- Similar analytic approach
Methodology

• Condition, health status measures from CAHPS survey
• Predicted costs from model fit to MCBS
• Analyses similar to assessment items:
  – Factor analysis to define dimensions
  – Variance components to attribute sources
Results: Selection dimensions (factors)

- Chronic: heart disease, stroke, COPD, diabetes, disability (2 items), self-assessed general health status
- Cancer
- Smoking
- [Substantial plan variation in each, within geographic areas]
Why hierarchical model?

- Plan-level measures are of varying reliability
  - Include component of individual-level variation
- We are interested only in plan-level relationships
Model

• Observation: \( y_p \sim [\theta_p, V_p] \)
  – \( y_p \) is plan-level observed mean
  – \( V_p \) is variance of this mean
  – \( \theta_p \) is underlying (population) plan mean

• Structural: \( \theta_p \sim [\mu, \Sigma] \)
  – Structure of \( \Sigma \) is what we are interested in!
Direct estimates of $\mathbf{V}$

- **Variances (item $i$):**

$$\hat{V}_i \approx \sum_k (y_{ik} - r_{ik} \bar{y}_i)^2 / \left( \sum_k r_{ik} \right)^2$$

- **Covariances (items $i, j$):**

$$\hat{V}_{ij} \approx \sum_k (y_{ik} - r_{ik} \bar{y}_i)(y_{jk} - r_{jk} \bar{y}_j) / \left( \sum_k r_{ik} \right) \left( \sum_k r_{jk} \right)$$
Naive estimation method

• Method of moments approach

$$\text{Var } y \approx \Sigma + \bar{V}$$

• Use survey sampler’s direct (approximately unbiased) estimate of \( \bar{V} \)

• Obtain \( \Sigma \) by subtraction
Naive estimation method: Problems

- Inefficient if sample sizes are unequal
- Small plans, low response items: sample sizes of 1 or 0 (no direct mean or variance)
  - Deletion or collapsing of plans or items throws away information
  - Sample sizes vary drastically across items
  - Ad hoc, arbitrary
Maximum likelihood?

• Full model for entire data structure might be very complex
  – Multinomial response categories
  – Jointly model nonresponse and responses

• Simplification: quasilikelihood modeling of means
Quasilikelihood model

- **Level I:** \( y_p \sim [\theta_p, V(\theta_p, R_p)] \)
  
  \( R_p = \) joint response pattern at plan \( p \)
  
  \( V \) is function of joint response pattern and plan population means

- **Level II:** \( \theta_p \sim [\mu, \Sigma] \)

- Approximate normality
Assumptions

• Mean is adequate summary of distribution
  – for modeling of variances

• Condition on response pattern
  – Treat as consequence of patient characteristics, uninformative about measures
  – Usual assumption in quality measurement: quality conditional on utilization
Variance functions

• Traditionally used to simplify presentation of variances for tables (see Wolter 1985)
• Used (implicitly) in hierarchical modeling with complex surveys, e.g.
  – Census SAIPE program - binomial (Fay, Bell)
  – No applications (?) with multivariate data
Variance-Covariance Functions

• Two part approach:
  – Variances (diagonal of $V$)
  – Correlations $C$
Variance functions

• Normalize by number of responses

\[ V_{ii} = r_i \tilde{V}_i = r_i \tilde{V}(\theta) \]

• Normalized variance is function of mean
  – Binomial-like function
  – Quadratic goes to 0 at extreme (high end)

\[ \tilde{V}(\theta) = b_1 (\theta_{\text{max}} - \theta) + b_2 (\theta_{\text{max}} - \theta)^2 \]
Variance functions

- Estimate coefficients by weighted least squares
  - Fit model to direct variance estimates
  - Separate parameters for each item
  - Estimate residual variance after subtracting sampling variability of direct estimates
  - $R^2$ typically 0.7 to 0.9
Correlation functions

• Componentwise prediction
• Decomposition into
  – correlation $C$ in overlap of item respondents
  – difference on item $j$ between respondents and nonrespondents to item $i$
• Model transformation of $C$ to $(-\infty, \infty)$
Correlation functions

• Less good fit (smaller $R^2$):
  – No obvious relationship to means

• For some plans, the componentwise predictions assemble to indefinite matrix!
  – Fewer problems with simpler models for $C$
Assembling $V$

- Combine model and direct variance estimates (precision-weighted componentwise)
- Combine model and direct correlation matrices (averaged weight)
- Combine variances and correlations
EM algorithm

• Given current estimate of \((\mu, \Sigma)\), estimate posterior mode and variance for each \(\theta_i\)

• Re-estimate \((\mu, \Sigma)\) given expected values of sufficient statistics \(\sum \theta_i\) and \(\sum \theta_i \theta_i^T\)

• Further adhockeries in estimation of \(V\) as algorithm converges
Future work: Bayesian hierarchical model

• Posit prior distributions for \((\mu, \Sigma)\)
  – Separate priors for variances, correlations
  – Priors that favor factor structure (some eigenvalues larger than others)

• Calculations more straightforward [?!] (no need to approximate integrals)
Interpretation of posterior

• Calculate factor analysis as descriptive summary supporting assignment of variables to groups

• Posterior distribution of $\Sigma$ induces posterior distribution of factor loadings, etc. and hence of groupings $\Sigma$
(3) Sources of variation

- Attribute variation to plan and geographical contributions
- Relevant to
  - validating measures
  - assessing discriminating power of measures
  - understanding quality variation
  - action to improve health services
Main analyses for “sources of variation”

• Geography by plan
• Health plan by time, within geography
Methodology

• 5 years of data, >700,000 responses
• Assign each to plan and MSA within state
• Analyze each rating and report item, summarize reports by groups
• Variance component models: estimate variation attributable to state, MSA, plan
### Plan × geography

#### % variance components

<table>
<thead>
<tr>
<th>Rating item:</th>
<th>Plan</th>
<th>Doctor</th>
<th>Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>1</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>State</td>
<td>22</td>
<td>28</td>
<td>23</td>
</tr>
<tr>
<td>MSA</td>
<td>7</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>Plan</td>
<td>53</td>
<td>33</td>
<td>35</td>
</tr>
<tr>
<td>Plan*MSA</td>
<td>17</td>
<td>21</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>(.171)</td>
<td>(.045)</td>
<td>(.045)</td>
</tr>
</tbody>
</table>
State & MSA effects

• Substantial state variation, some regional
• Biggest share of variance attributable to plan is for rating of plan
• Modest plan X MSA interaction, reasonable to report statewide for plans (usually)
### (State, MSA, Plan) × Time

#### % Variance Components

<table>
<thead>
<tr>
<th></th>
<th>Ratings</th>
<th>Report Groups (average)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Plan</td>
<td>Care</td>
</tr>
<tr>
<td>State</td>
<td>25</td>
<td>33</td>
</tr>
<tr>
<td>State* Year</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>MSA</td>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>MSA * Year</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Plan</td>
<td>48</td>
<td>40</td>
</tr>
<tr>
<td>Plan * Year</td>
<td>10</td>
<td>3</td>
</tr>
</tbody>
</table>
Time effects

• State effects extremely stable for every variable

• Plan effects also stable for all except rating of plan, customer service
  – Few significant changes from year to year (except for rating of plan)
Better longitudinal models

• Exchangeable structure across years doesn’t make sense
• AR-1 or similar structure would make more sense
• Software limitations
Plan Rating. Medicare Managed Care CAHPS, 1997–2001

Quartile
1 2 3 4

1 = Lowest, 4 = Highest Score
Shaded areas are non-MSA counties
Prescription Composite. Medicare Managed Care CAHPS, 1998–2000

Quartile

1 2

3 4

1 = Lowest, 4 = Highest Score
Shaded areas are non–MSA counties
Conclusions (variation)

• Consumer experiences affected by a variety of influences:
  – Interactions with the health plan are more specific to the plan
  – Direct care interactions affected by features of local care system impinging on all plans.
• May be less readily changed by plans
Levels of variation (below plan)

- Analysis of G-CAHPS (group) survey
- Designed to rate groups, doctors
- Focus on access and interactions at office
- Pilot study: regional units, groups and sites of care within integrated care delivery organization (Partners Healthcare in metro Boston)
Variation (below plan): Results

- Most variation attributable to lowest level (site), for most dimensions
- Some policies and procedures set at group or Regional Service Organization level
- Little effect of plan on any aspect of care received
- Possibly atypical, but very suggestive!
Implications

• Focus improvement efforts and incentives on the right level
• Report on subgroups and dimensions that differ, not repetition of the same information
• Methodologies have broad potential applicability
END