CESR

Center for Effectiveness & Safety Research

Using Item Response Theory to Summarize Health Care Quality Measures: An Application to HEDIS© Diabetes Measures in Medicare Advantage

John L. Adams

UCLA Center for Maximizing Outcomes and Research on Effectiveness, 16 September 2014

Kaiser Permanente Research

KAISER PERMANENTE

Collaborators

- Marc N. Elliott, Amelia M. Haviland, Carol A. Edwards

RAND

Sarah Gaillot

CMS

Funded by CMS: HHSM-500-2005-000281 to RAND

CESR

2

KAISER PERMANENTE

Overview of today's talk

- Building composite quality of care measures
- The potential of item response theory (IRT)
- IRT analysis of Medicare Advantage diabetes HEDIS© measures
- Conclusions and future directions

Composite quality of care measures (mostly at a more aggregate level)

- Opportunities scoring
  - #passed/#triggered
- Observed difficulty of delivery (ODD) adjustment
  - Sum(observed)/sum(expected) or Sum(observed)-sum(expected)
  - Not quite case mix adjustment, only population pass rates
- Standardization
  - (rate1-mean(rate1))/SDrate1+(rate2-mean(rate2))/SDrate2+…..
- Typically not as much case mix adjustment or fancy measurement methods as health status or outcomes

The item response theory (IRT) model

- A latent trait model
  - Assumes a person has a latent quality score
  - That quality score drives the passing of measures
  - Similar to factor analysis with a single factor
    - Correctly handles binary data
    - Can handle complicated missing data patterns
- The 2 parameter model for binary data looks like a logistic regression except the theta isn't known:

$$P_{jk}(+ \mid \theta_k) = \frac{e^{\alpha_j(\theta_k - \beta_j)}}{1 + e^{\alpha_j(\theta_k - \beta_j)}}$$

- Where k is the person and j is the item

CESR 5 KAISER PERMANENTE

Some useful references

- Hays, R. D., Morales, L. S., and Reise, S. P. (2000), "Item Response Theory and Health Outcomes Measurement in the Twenty-First Century," *Medical Care*, 38, Suppl. 9, 1128–1142.
- Reeve, B., Hays, R. D., Bjorner, J., Cook, K., Crane, P. K., Teresi, J. A., Thissen, D., Revicki, D. A., Weiss, D. J., Hambleton, R. K., Liu, H., Gershon, R., Reise, S. P., Lai, J. S., Cella, D., & on behalf of the PROMIS cooperative group. (2007). Psychometric evaluation and calibration of health–related quality of life item banks: Plans for the Patient–Reported Outcome Measurement Information System (PROMIS). *Medical Care*, 45(5), S22–31.
- Reeves D, Campbell S, Adams JL, Shekelle PG, Kontopantelis E, Roland MO. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. *Med Care*. 2007;45(6):489-496.
- Scholle SH, Roski J, Adams JL, Dunn DL, Kerr EA, Dugan DP, Jensen RE. Benchmarking Physician Performance: Reliability of Individual and Composite Measures. *Am J Manage Care*. 2008;14(12):829-838. PMCID: PMC2667340; PMID: 19067500; NIHMSID: NIHMS99203.

CESR 6 KAISER PERMANENTE

HEDIS Diabetes Measures in Medicare Advantage

- We have the HEDIS diabetes measures for reporting years 2008-2010
  - Today I will focus on the 2010 measurement year data
  - Approximately 450k unique individuals that triggered at least one diabetes measure
- Nine measures:
  - HbA1c Testing
  - HbA1c Poor Control >9% (reversed)
  - HbA1c Good Control
  - Retinal eye exam
  - LDL-C Screening
  - LDL-C Control <100mg/dl
  - Kidney disease / Nephropathy
  - Blood pressure control <130/80
  - Blood pressure control <140/90

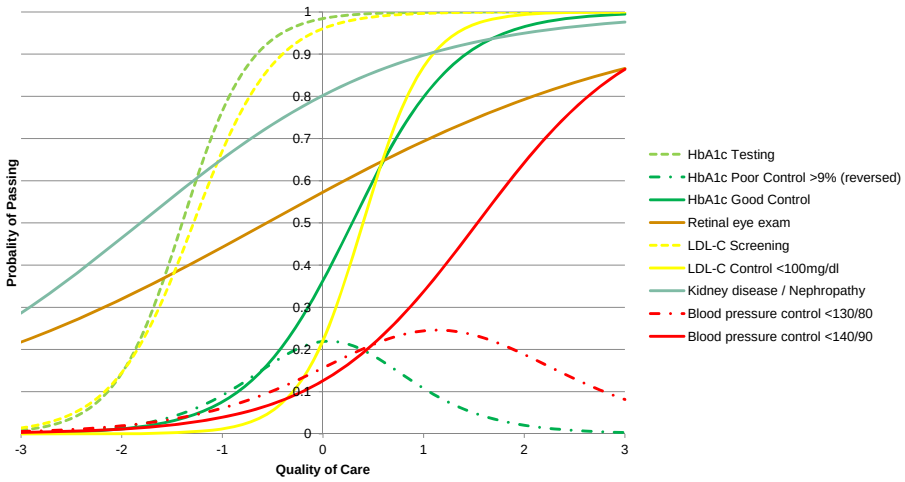
CESR 7 KAISER PERMANENTE

But some pairs of these measures have a natural ordinal structure

- Glycemic control
  - HbA1c Poor Control >9% (reversed)
  - HbA1c Good Control
- Blood pressure control
  - Blood pressure control <130/80
  - Blood pressure control <140/90
- We will treat these as ordered categorical data and combine them
  - Just like ordinal logistic regression

CESR 8 KAISER PERMANENTE

Item characteristic curves



CESR

9

KAISER PERMANENTE

The quality scores

- Each combination of triggering and pass/not has a quality estimate
  - Like a best linear unbiased predictor (BLUP)
  - On a z-score scale, roughly -2 to +2

CESR

10

KAISER PERMANENTE

Computation

- In the beginning
  - Used Proc Nlmixed in SAS 9.2*
  - There are only 3^5*4^2 ~4k possible patterns of triggering and scoring (but only about 500 occur in nature)
  - Weighted analysis makes the problem tractable
- And then a miracle occurred
  - Experimental Proc IRT in SAS 9.4
  - MIRT and many good things are now possible

*Sheu, C.-F., Chen, C.-T., Su, Y.-H., & Wang, W.-C. (2005). Using SAS PROC NLMIXED to fit item response theory models. *Behavior Research Methods,* 37(2), 202-218.

CESR 11 KAISER PERMANENTE

High scoring trigger/pass/fail combinations

Quality Score	Standard error	HbA1c Testing	HbA1c Poor Control >9% (reversed)	HbA1c Good Control	Retinal eye exam	LDL-C Screening	LDL-C Control <100mg/dl	Kidney disease / Nephropathy	Blood pressure control <130/80	Blood pressure control <140/90
1.47	0.63	1	1	-1	1	1	1	1	1	1
1.47	0.63	1	1	1	1	1	1	1	1	1
1.47	0.64	1	-1	-1	1	1	1	1	1	1
1.42	0.63	1	1	1	-1	1	1	1	1	1
1.29	0.61	1	1	-1	0	1	1	1	1	1
1.29	0.61	1	1	1	0	1	1	1	1	1
1.29	0.61	1	-1	-1	0	1	1	1	1	1
1.28	0.69	-1	-1	-1	-1	-1	-1	1	1	1
1.27	0.61	1	1	-1	1	1	1	0	1	1
1.27	0.61	1	1	1	1	1	1	0	1	1

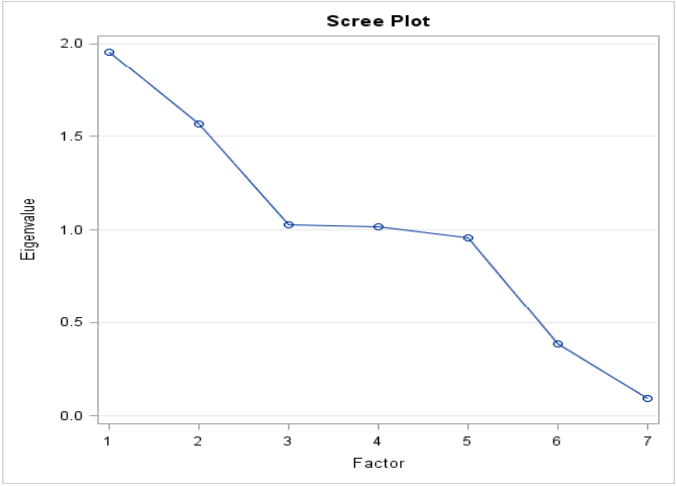
CESR 12 KAISER PERMANENTE

Low scoring trigger/pass/fail combinations

Quality score	Standard error	HbA1c Testing	HbA1c Poor Control >9% (reversed)	HbA1c Good Control	Retinal eye exam	LDL-C Screening	LDL-C Control <100mg/dl	Kidney disease / Nephropathy	Blood pressure control <130/80	Blood pressure control <140/90
-1.85	0.58	0	0	0	0	0	0	0	0	0
-1.85	0.58	0	0	-1	0	0	0	0	0	0
-1.85	0.58	0	-1	-1	0	0	0	0	0	0
-1.84	0.59	0	0	0	0	0	0	0	-1	-1
-1.83	0.59	0	-1	-1	0	0	-1	0	-1	-1
-1.79	0.58	0	0	0	-1	0	0	0	0	0
-1.72	0.58	0	-1	-1	0	0	-1	-1	-1	-1
-1.7	0.56	0	0	0	1	0	0	0	0	0
-1.7	0.56	0	0	-1	1	0	0	0	0	0
-1.7	0.56	0	-1	-1	1	0	0	0	0	0

CESR 13 KAISER PERMANENTE

What about the possibility of more than one latent dimension of quality?

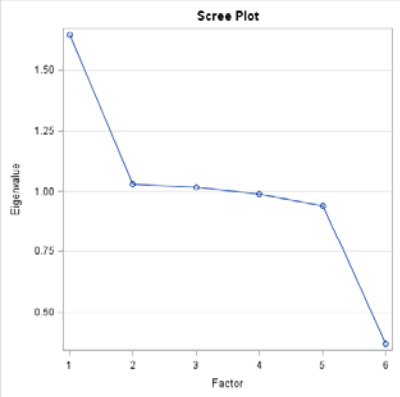


CESR 14 KAISER PERMANENTE

Exploring the second dimension

- Fitting the model parameters was very sensitive
- There is a conditional structure in the data not captured in the model
  - You can't be scored on a lab value if you didn't get the test
- Tried turning the LDL screening and control variables into an ordered categorical variable
  - Some conceptual and data issues with this

- New scree plot:



CESR

KAISER PERMANENTE

What I still need to work on

- I am not sure how important the local dependence assumption is for this problem

CESR

KAISER PERMANENTE

Conclusions and future directions

- The IRT model for HEDIS diabetes measures has good face validity
- IRT has the potential to sensibly:
  - Address “topped out” measures without dropping them
  - Fold in emergent measures while maintaining the meaning of the scale
- Future directions
  - Continue to explore multidimensionality (MIRT)
  - Explore differential item functioning (race, gender, etc.)
  - Expand models to include system level structure
  - Explore a wider range of measures

CESR 17 KAISER PERMANENTE

Questions?

CESR KAISER PERMANENTE