USING HIERARCHICAL MODELS TO ATTRIBUTE SOURCES OF VARIATION IN CONSUMER RATINGS OF HEALTH PLANS

Alan Zaslavsky, Harvard Medical School
Department of Health Care Policy, Harvard Medical School, Boston, MA 02115

Abstract

Surveys, notably the Consumer Assessments of Health Plans Survey (CAHPS), are widely used to elicit information from members about experiences with health plans and their affiliated care providers. Health care quality is affected by factors associated with the region, market area, health plan, and delivery system or provider group levels. To interpret quality data and determine where action can be taken for improvement of the various dimensions of quality, it is helpful to estimate the relative contribution of each level to the total variation in quality measures. We fit hierarchical models to data from several large CAHPS implementations and find that the decomposition of variance is different for each quality dimension and that these differences are plausible in view of the ways plans and providers affect different aspects of quality. These findings are useful in devising sampling and reporting plans for quality measurement. Similar approaches are useful in understanding processes involved in selection of health risks by managed care plans.

Keywords: Consumer assessments, CAHPS survey, health care.

1 Introduction

In an earlier JSM presentation (Zaslavsky 1998), I described the Consumer Assessments of Health Plans Survey (CAHPS®), a newly-developed survey instrument used to elicit from members of health plans their assessments of the quality of care provided by the plans. Since that time, the use of CAHPS has burgeoned, including large-scale implementations for surveys of Medicare Managed Care beneficiaries (currently over 300 health plans), Federal employees and dependents (almost 300 plans), and Medicare Fee-for-Service beneficiaries (going to the field in late 2000 with a population of about 20 million), as well as a multitude of smaller implementations.

Much of the earlier CAHPS research focused on instrument design and survey methodology, as well as issues involved in generalization of CAHPS to a variety of populations. (See the introduction by Crofton, Lubalin and Darby (1999) to a special issue of *Medical Care* devoted to CAHPS methodology.) Such analyses, based largely on pilot data, necessarily concentrated on effects measurable at the individual level. With the availability of data from uniform application of the survey across large collections of health plans, it is now possible to perform analyses in which the unit of analysis is the health plan, rather than the individual. This allows us to place the primary focus on factors that distinguish among plans, bringing us closer to the original motivations for the development of CAHPS.

In this paper I first briefly describe the CAHPS survey. (My summary refers to the CAHPS 1.0 survey which is the source of the data used here; since that data collection, a revised version, CAHPS 2.0, has replaced it.) I then summarize two investigations using data from the first year of the Medicare Managed Care CAHPS implementation to better understand variation in the CAHPS measure at the plan level.

2 The CAHPS survey

The CAHPS survey contains two types of evaluative items. Report items are designed to elicit descriptions of specific experiences with health care or the health plan, and are answered in a "never/sometimes/usually/always" format. A typical question is: "When you needed care right away for an illness or injury, how often did you get care as soon as you wanted?" Report items are typically grouped into reporting composites such as "getting needed care."

The four rating items elicit summary evaluations of experiences with the health plan and care providers, on a 0 to 10 scale. In addition, several items ask for sociodemographic characteristics (age, sex, race/ethnicity) and self-rated health status of the respondent. The Medicare version of the survey also asks about specific medical conditions and functional limitations.

The standard CAHPS analysis is described in the CAHPS implementation handbook (AHCPR 1997, 1999). Briefly, a linear model is fit to remove effects of individual characteristics such as age, health status, and education, of the form $y_{pj} = \mu_p + \beta x_{pj} + \epsilon_{pj}$, where y_{pj} is the response from person j at plan p, μ_p is a plan intercept, x_{pj} is that person's individual characteristics, and ϵ_{pj} is an error term. The adjusted mean is then $a_p = \mu_p + \beta x_{\mathrm{ALL}}$ where x_{ALL} is the mean of plan means on the adjuster variables. See Zaslavsky (1998) for further discussion. After adjusted plan means are calculated for every item, composites are calculated for the report items, using

equal weighting of plan means for items included in each composite. Variances are calculated by Taylor linearization, taking into account that cases for which there was no response on a particular item (usually because the item did not ap[ly and was skipped) have no influence on the plan mean for that item. A global F-test was calculated for each composite or rating, for the null hypothesis that the means for the plans included in a given comparison group (e.g. region, state, market area) are equal. Individual t-tests were also calculated for the comparison of each plan's mean score against the mean of all plans in its comparison group.

3 Dimensions of consumer assessments

A team of researchers at the Harvard Medical School analyzed data from the Medicare Managed Care survey. Our data included 89,414 survey responses from 212 reporting units (health plans or geographically defined sections of a health plan). As noted above, the availability of large datasets containing results from many plans provided new opportunities for research on plan-level patterns in the responses. The first question we addressed concerned the relationships of the items (Zaslavsky et al. 2000a). We applied factor-analytic methods to summarize these relationships at the plan level by using the mean values for each plan as input to the factor analysis.

A statistically interesting feature of the analysis arose because we wished to remove the effects of sampling variability within each plan from the factor analysis. Sampling variability could induce correlations that from our point of view are spurious, because they were due to correlated error. For example, the sampled individuals at a plan might happen to be particularly optimistic in their assessments across the board, yet this would not necessarily indicate that their favorable assessments of disparate elements reflected relationships at the plan level in the quality of these elements. Conversely, real relationships at the plan level might be attenuated in an uncorrected analysis of means, because two items that are related at the plan level might be answered by different subsets of the plan's sample and therefore their mean scores would include uncorrelated error components.

Our model for the calculation of corrected covariance matrices was of the form

$$\mathbf{a}_p = \theta_p + \delta_p$$

where Var $\delta_p = S_p$ and Var $\theta_p = \Sigma$. Here, S_p is a sampling covariance matrix estimated separately at each plan and Σ is the plan-level covariance matrix desired for the factor analysis. We used a simple method-of-moments estimate of Σ, first calculating the covariance matrix of the sample means $T = \text{Var } \mathbf{a}_p$ and then estimating $\hat{\Sigma} = T - \bar{S} = T - (1/n)\sum S_p$. The covariance matrix $\hat{\Sigma}$ estimated in this manner had several small negative eigenvalues. We did not regard as problematical, since for factor analysis only the large eigenvalues and corresponding eigenvectors are important. Furthermore, we found that in most cases subtracting sampling variability *increased* correlations among items, suggesting that the correlations were in fact due to plan-level systematic differences rather than individual reporting idiosyncracies.

Analyses of this form have been conducted in the past under the general rubric of "multilevel structural equations modeling" (Muthén 1989, 1994). Our case differs from those considered in those analyses due to the complex pattern of missing data, which prevents the use of standard structural equations modeling software that requires entry of complete individual-level data. We were able, however, to make a slight modification of the standard CAHPS analysis program to calculate the covariances as well as the variances of the influence statistics for each observation. This enabled us to obtain the required estimated sampling covariances S_p of the plan means.

Results of the factor analysis (with an oblique Promax rotation) are summarized in Table 1. We found that four factors explained most of the variance in the 31 report items that we analyzed, and most of the variables could readily be assigned to a group corresponding to one of these factors. The first group, comprising 16 variables, contained questions about health care provided at the doctor's office and interactions with healthcare providers. The second contained 5 items about customer service from the health plan. The third consisted of 8 items about services other than direct care, which (in a 6-factor alternative solution) split into groups for specialists and prescription drugs, services obtained through the plan, and ease of access at the doctor's office. The final factor included 2 items concerning health-promoting advice provided to patients, to quit smoking and to diet and exercise. An interesting feature of this last factor is that the two items measure experiences of very different populations. Almost all respondents (96%) answered the "diet/exercise" item, while only 11% answered the smoking advice item, primarily because only smokers were eligible. The fact that these items nonetheless were grouped by the factor analysis suggests

that the variations found were due to real differences among plans.

An interesting conclusion from these analyses is that although the original CAHPS reporting strategy differentiated among various aspect of care delivery, whatever the plans are doing to affect this seems to affect a large number of specific report items in very similar ways. On the other hand, we discovered an unexpected dimension of plan quality embodied in reports on advice. The remaining item groupings confirmed that customer service and provision of special health services are distinct from each other and from care provided at the doctor's office.

4 Components of variation among plans

A second analysis of the same data (Zaslavsky et al. 2000b) used the same data to examine variation among plans in a regional context. Each reporting unit was assigned to the metropolitan area (MSA) in which it had the largest-enrollment county, and also to one of 8 national regions. The first set of analyses tested (using F-tests in a 3-level nested ANOVA model) the significance of differences among plans. Columns (1)–(5) of Table 2 show the results of these tests. Plans differed significantly nationally (1) and within MSAs (5), and furthermore regions differed from each other, regardless of whether we treated the MSAs as fixed (2) so plans were the error term, or treated the MSAs as random (4). MSAs also differed significantly within regions.

Having established the existence of significant variation among plans at all three levels, we next fitted variance component models to estimate the magnitude of the component of variation contributed at each level (columns (6)–(7)). Concentrating on the ratings in the first four rows of the table, we found that evaluation of the health plan was largely driven by the plan itself. On the other hand, the majority of the variation in evaluations of care was driven by regional and MSA variation; the same was true, to a lesser extent, for variation in ratings of the doctor and specialists. Similar results were obtained for our newly-defined composites: the customer service composite resembled the plan rating, while the delivery composite resembled ratings of care. These findings are consistent with our expectations that direct care would be affected by regional factors because it is provided by networks of providers who are subject to local medical norms but tend to overlap across many plans. Conversely, customer service is a highly plan-specific function, run by personnel of the health plan itself, and may vary greatly across plans even within the same area. These findings may have important implications for quality improvement efforts, suggesting that the impact of any individual plan's efforts on care delivery may be limited. More useful information for measurement and improvement of the quality of direct patient care may be obtained from surveys at the provider group level, now being developed through another component of the CAHPS project.

5 Ongoing research

A number of ongoing projects using CAHPS data are now underway and will be reported in manuscripts now under preparation.

5.1 Extension of variance components methodology for assessments

We are extending our analyses to look at more fine-grained geographical variation in ratings of care, and interactions of geography with plan effects at local areas. We hope to find out whether a given health plan provides a consistent quality of care across its service area or whether conversely quality variation is driven by local characteristics of health care providers.

We now have three years of CAHPS Medicare data, and we have begun to analyze these longitudinally. By interacting plan and geographical effects with time, we can analyze the stability of each of these effects. Preliminary results suggest that geographic effects are very stable, while assessments of plans are more likely to change over time, possibly affected by changes in policies and benefits.

Because we found that so much of consumer assessments are related to geographical variation, we look forward to linking results for managed care with those from the CAHPS survey of the conventional (fee-for-service) component of Medicare. Results from that survey should begin to appear early in 2001.

5.2 Selection of health risks

A persistent issue in the economics of managed care is the extent to which health plans are able to "cherry-pick," that is to selectively recruit the healthiest potential members and therefore increase their profitability at the expense of alternative providers of health insurance. This has been a particular concern in the Medicare market.

We are using the health status, condition and functional limitation items on the CAHPS survey to study the extent to which such selection occurs among competing health plans within a market area. The statistical methodologies involved are similar to those described above, i.e. factor analy-

sis to determine what conditions tend to be selected for similarly, and variance components models to determine how much variation in health is driven by geographical distribution of sicker and healthier members and how much by selection among plans.

5.3 Subgroup and casemix effects

Casemix effects, i.e. variation between assessment results from beneficiaries with different characteristics, have generally been treated as a nuisance in most analysis of plan quality. We have performed analyses in smaller data sets of the extent to which such differences (for example, between reports from more and less healthy plan members) are the same at every plan (Zaslavsky, Zaborski and Cleary 2000). We are applying similar methodologies to the Medicare data in an effort to determine whether some health plans are better than others at closing the gap between experiences of more and less health members.

Acknowledgments

I acknowledge the major contributions of my collaborators, Nancy Dean Beaulieu, Paul D. Cleary, and Bruce E. Landon, and programming assistance by Matthew Cioffi and Lin Ding and document preparation by Virginia Wang. I thank Liz Goldstein, Thomas Reilly, and other collaborators at HCFA, Barents LLC, Westat, Inc., DRC, Inc., and the Picker Institute for their efforts in the CAHPS-MMC implementation that generated the data on which this paper is based. This research was supported by a grant from the Commonwealth Fund to Paul D. Cleary and by a contract with HCFA (500-95-007).

References

AHCPR (Agency for Health Care Policy and Research) (1997), *CAHPS 1.0 Survey and Reporting Kit.* Rockville, MD: AHCPR.

AHCPR (Agency for Health Care Policy and Research) (1999), *CAHPS(TM) 2.0 Survey and Reporting Kit,* Publication No. 99-039. Rockville, MD: AHCPR.

Crofton, Christine, James S. Lubalin, and Charles Darby (1999), "Forward". *Medical Care*, 37, MS1-9.

Muthén, Bengt O. (1989), "Latent variable modeling in heterogeneous populations." *Psychometrika*, 54:557–585.

Muthén, Bengt O. (1994), "Multilevel covariance structure analysis." *Sociological Methods and Research*, 22:376–398.

Zaslavsky, Alan M. (1998), "Issues in case-mix adjustment of measures of the quality of health plans". *Proceedings, Government and Social Statistics Sections, American Statistical Association*, 56-64.

Zaslavsky, Alan M., Nancy Dean Beaulieu, Bruce E. Landon, and Paul D. Cleary (2000a). "Dimensions of consumer-assessed quality of Medicare managed care health plans." *Medical Care*, 38:162-174.

Zaslavsky, Alan M., Bruce E. Landon, Nancy Dean Beaulieu, and Paul D. Cleary (2000b). "How consumer assessments of Medicare managed care vary within and among markets." *Inquiry*, 37:146-161, 2000.

Zaslavsky, Alan M., Lawrence Zaborski, and Paul D. Cleary (2000). "Does the effect of respondent characteristics on consumer assessments vary across plans?" *Medical Care Research and Review*, 57:379-394.