Anticipating market demand: tracking enrollee satisfaction and health over time

HARRIS ALLEN

Harris Allen Associates, Boston, MA, USA

Abstract

Objective. To assess guidelines, set by the National Committee for Quality Assurance, for the Health Plan Employer Data and Information Set (HEDIS) 1999 CAHPS 2.0H Survey (formerly the HEDIS 1999 Consumer Survey) in the light of user's needs to monitor health plan performance over time, monitor sick enrollees, and prioritize determinants (drivers) of enrollee experience.

Design. A two-wave, cross-sectional/longitudinal panel design, consisting of national surveys mailed to employees of three major USA corporations in 1993 and 1995.

Study participants. Samples included employees selected to represent 23 major managed care and indemnity plans in five regions of the USA. In 1993, 14,587 employees responded and in 1995 9,018 employees responded (response rates: 51 and 52%). The longitudinal panel sample included 5,729 employees who completed both surveys and stayed in the same plan for both years.

Study measures. The main 1993 and 1995 surveys consisted of 154 and 116 items, respectively. Panel survey content assessed care delivery, plan administration, functional status, well being, and chronic disease.

Results. CAHPS 2.0H's point-in-time, cross-sectional design was unable to detect selection bias and led to an inaccurate view of change in performance. CAHPS 2.0H's use of aggregate samples masked key differences between healthy and sick enrollees; e.g. the sick became less satisfied over time. The association-based, statistical techniques that many survey users will employ to prioritize the 'drivers' of enrollee experience in the absence of CAHPS 2.0H guidelines yielded a less efficient account of change than the multi-method/multi-trait approach developed for this project.

Conclusion. Consumer experience of plan performance is best understood when the separate contributions of longitudinal membership and movement in and out of plans are clarified, changes in health are identified, changes for sick and healthy enrollees are compared, and plan performance on satisfaction criteria is probed to give confirmation and detail. Changes to the CAHPS 2.0H approach in HEDIS 1999 will facilitate user application of these principles.

Key words: Consumer Assessment of Health Plans Survey (CAHPS), consumer experience, health plan performance, health care surveys, Health Plan Employer Data and Information Set (HEDIS)

Use of outcomes data to compare health plans, networks, and providers has reached critical mass in the USA marketplace. Few would argue that momentum for results-based comparisons – whether conducted on an external basis across competing entities or on an internal basis by entities seeking outside referents – has achieved a point of no return. Much debate probably lies ahead on how to define and conduct these comparisons, how to interpret and act on results, etc. Yet, the pivotal role that they play in the search for value will continue to grow, as users become more familiar with the uniquely relevant information they provide.

This paper focuses on one source for providing outcomes data, the consumer, and recent developments for eliciting data therefrom. The principal medium for this purpose, consumer surveys, is undergoing a fundamental shift in the USA. This shift towards a publicly sanctioned greater consistency across surveys and how they are implemented, is most evident at the level of health plans. Broad support
has emerged for two initiatives: Health Plan Employer Data Information Set (HEDIS) [1], sponsored by the National Committee for Quality Assurance (NCQA), and Consumer Assessment of Health Plans Survey (CAHPS) [2], sponsored by the Agency for Health Care Policy Research. HEDIS and CAHPS developers have recently endeavored to resolve differences between these initiatives and NCQA is now presenting a unified offering to users, CAHPS 2.0H [3].

Considerable experience has by now been gained with the use of consumer surveys to evaluate plan performance, under the aegis of prior versions of HEDIS and other report card initiatives [4–8]. This paper proceeds from the premise that lessons have emerged from this experience, about how a survey approach can and should be fashioned to maximize fit with user demand [9,10], which merit examination in light of CAHPS 2.0H. How does this offering’s survey instrument and design fare when held up against these lessons? This question is examined by posing the lessons as analytic issues against which CAHPS 2.0H guidelines are assessed, using results from the first reported consumer survey of plan performance over time [11].

**Giving shape to user demand**

Although each of the principal user groups – health plans, consumers, and purchasers – has unique interests for consumer survey data, all three share a growing demand for the best information that can be derived from results on each of the following issues. Each issue has corresponding prerequisites for survey approach, and can be illustrated by questions that one of the three user groups is raising out of the needs its constituents bring to survey data.

First, can consumer experience of plan performance change over time? If so, how? For example, health plans are increasingly developing empirical benchmarks for performance, both for self-appraisal and to represent their performance to current and potential customers. As they implement these benchmarks, they need to know if they are improving over time, relative to themselves and their competitors. The capacity of plans to make accurate links across successive survey administrations is paramount in this regard.

Second, what is the experience of sick enrollees, and how does this compare across plans? Hibbard et al. have suggested that for many individual consumers, an acculturation process is needed before they can make effective use of comparison data when shopping for plans [12]. As this occurs, many consumers will start to ask one or both of two questions: what happens when people in my plan become sick and how does the experience of sick enrollees compare in my plan versus my other plan options? Proper procedures for identifying and monitoring the sick will be key to any project’s attempts to address these questions.

Third, what drives (determines) consumer experience of plan performance? Which aspects of performance most affect what consumers say and do about their plan? Purchasers, too, are turning to benchmarks, not only to inform plan-related buying decisions but also to maximize the value of plan offerings for their covered populations by encouraging plans to improve. As they endeavor to characterize performance and to negotiate improvement expectations, purchasers need to be able to identify a short list of primary drivers. Which variables are likely to mean the most to the insured? Which will yield the best return on investment for quality improvement? Credible driver detection techniques for prioritizing among the many variables often available in collected data sets are an essential element of this exercise.

Examination of CAHPS 2.0H’s approach to these issues needs to take into account survey design and content. Each issue is mainly design-oriented, yet content concerns still apply in each case. Implications for content are therefore probed below to facilitate treatment of design issues.

**Applying CAHPS 2.0H**

Discussion of CAHPS 2.0H needs to be undertaken in light of guidelines that the NCQA and CAHPS investigators explicitly offer and how these guidelines are likely to be applied by users. To answer their questions, many users are likely to do what they can within limits set by the guidelines. What the guidelines do not say – what they default to in the absence of explicit prescription – may be as important as what they do say.

**Monitoring over time**

NCQA has for several years now called for periodic administration of the HEDIS survey. CAHPS 2.0H maintains this guideline. It asks plans to produce a new random sample for each administration and, in so far as analysis and reporting are concerned, contains no provision for linking data across administrations. For this point-in-time framework, monitoring performance over time defaults to a sequence of single administration/cross-sectional surveys. The accuracy of this design for over-time assessment is at issue.

**Monitoring the sick**

CAHPS 2.0H relies on random samples of entire memberships when representing health plans. No guidelines are offered for identifying or monitoring healthy and sick enrollees. Do random samples adequately represent the experience of the sick? Growing public concern about this group in health management organizations (HMOs), reflected in patient’s ‘bill of rights’ legislation now appearing across the USA [13], emphasizes the need to scrutinize this issue.

**Driver detection**

As with earlier versions, CAHPS 2.0H is silent on offering ways to prioritize the relative importance of variables in collected data sets. Left to their own devices, users to date (that is, those implementing empirical methods for this purpose) have often used correlation and multiple regression analyses where plan satisfaction is either associated with or regressed on competing candidates in single administration,
Anticipating demand for enrollee surveys

point-in-time data sets. Sole reliance on point-in-time, association-based statistical techniques for this purpose is at issue.

Methods

To test the CAHPS 2.0H approach, data are tapped from national domestic health care surveys that William Rogers and I conducted in 1993 and 1995 for three major US corporations. Further details are given in a technical appendix at this Internet address: http://ra.ne.mediaone.net/chpv.

Design

This project was designed to develop a longitudinal panel in a context that also represented plans in both years. In 1993, we administered the first wave instrument by mail to 14,587 employees. In 1995, we mailed the second wave instrument to a sample of 9294 of the 1993 respondents (responders) and to a new sample of employees from the three companies. A total of 9018 employees responded to the 1995 surveys, and they formed the 1995 sample for our cross-sectional study. Responders who replied to the survey in both years and who had remained in the same health plan between 1993 and 1995 (a total of 5729 employees) formed our longitudinal sample.

Survey content

The 1993 instrument was a 154-item questionnaire incorporating proven sets of items with a track record for efficiently delivering a comprehensive evaluation of plan satisfaction and consumer health that is relevant across plan types and across the range of clinical and demographic groups in general populations [14]. The 1995 instrument for two corporations was a 116-item survey patterned closely after the 1993 questionnaire. The third corporation's survey was a 126-item version of this questionnaire expanded to serve the company's desire to examine progress made on its 1993 quality improvement targets. Panel content across these instruments included over 80 items assessing care delivery and plan administration, management of care and coverage, as well as functional status and well being, chronic disease and health risk behaviors.

Sampling

Whereas our sampling strategy was national in scope, the over-time component focused on 23 leading managed care plans and company indemnity plans located in five areas: Eastern Massachusetts/New Hampshire, New York City, Tampa, San Francisco, and Los Angeles. The managed care plans spanned three plan types: independent practice association (IPA), prepaid group practice (PPGP), and point of service (POS). Our intent was to obtain representative probability samples of plan membership in 1993 and 1995, collect longitudinal panel data on 1993 respondents, and enhance information on sick employees.

Data collection

Data collection consisted of a well-documented, four-step mail procedure that yielded response rates of 51% in 1993 and 52% in 1995 [15].

Test: monitoring over time

The issue of over-time assessment is addressed by examining change in plan satisfaction in two ways: (i) comparing the 1995 cross-sectional sample with the 1993 cross-sectional sample (the analog to which CAHPS 2.0H defaults); and (ii) comparing the 1993 and 1995 scores of the longitudinal sample. See the technical appendix for details on how the samples were weighted for these comparisons.

Test: monitoring the sick

CAHPS 2.0H's use of aggregate samples to represent plan membership, including the sick, is scrutinized by comparing healthy versus unhealthy responders. The tests probe group differences on the following: (i) point-in-time comparisons in plan satisfaction in 1993 and 1995; and (ii) plan satisfaction by 2-year health transitions for four groups: 1993 healthy responders who stayed healthy in 1995; 1993 healthy responders who became sick in 1995; 1993 sick responders who stayed sick in 1995; 1993 sick responders who became healthy in 1995. We based health definitions on Physical Component Scores (PCS) and Mental Component Scores (MCS) from SF-12 Health Status Survey items [16] and on disease counts from a 17-item chronic condition checklist. Responders with PCS scores of 50 or more and MCS scores of 50 or more and three or fewer diseases were classified as healthy; responders with PCS scores of 42 or less, or MCS scores of 42 or less, or four or more diseases as sick.

Test: driver detection

The driver detection issue is explored by developing a short list of key determinants of enrollee experience that derive from two strategies. The first was the use of correlation and multiple regression analyses predicting satisfaction on the two point-in-time data sets (the 1993 and 1995 cross-sectional samples), which together comprise the method which CAHPS 2.0H users are most likely to employ in the absence of explicit guidelines. The second was a multi-method/multi-trait (MMMT) strategy developed for this project. This alternative added three components: a direct question ascertaining responder's preferences for different dimensions of care; prediction of actual disenrollment; and longitudinal analysis pitting the prediction of changes in satisfaction against changes in other domains. All four tests compared six domains: access, quality of care, physician choice, coverage, service, and employees' costs of care. Since our MMMT strategy used the point-in-time/association-based statistical approach, comparison of the two strategies amounts to the question, what is to be gained by adding the three components? This question is addressed by probing what drove change by plan type over the 2 years.
Table 1 describes sample demographic and health characteristics at the start of the evaluation period by giving 1993 scores for four groups: the overall 1993 sample; the 1995 target sample; 1995 responders; and 1995 non-responders. As shown, 1995 responders reported health and percent female scores in 1993 that were virtually identical to the corresponding 1993 scores for each of the other groups. The 1995 responders group was somewhat older and more likely to be white, yet these differences were correctable by weighting. These data rule out the possibility that initial demographic and health differences between the 1993 sample and those who were sampled, who responded, and who did not respond in 1995 were a factor in these tests of CAHPS 2.0H.

Figure 1 gives the cross-sectional view (i.e. the CAHPS 2.0H analog) of change in plan satisfaction. It shows that IPA and PPGP enrollee satisfaction was relatively high in 1993 and remained favorable in 1995. In both years, more than four out of five responders in these two plan types said they were satisfied. POS enrollees were somewhat less likely to be satisfied, with roughly three out of four responders reporting satisfaction. Cross-sectional comparisons, however, imply

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean age (years)</td>
<td>40.6</td>
<td>40.6</td>
<td>41.3</td>
<td>39.5</td>
</tr>
<tr>
<td>Female (%)</td>
<td>37.2</td>
<td>37.0</td>
<td>37.0</td>
<td>37.0</td>
</tr>
<tr>
<td>White (%)</td>
<td>85.9</td>
<td>85.7</td>
<td>87.4</td>
<td>82.9</td>
</tr>
<tr>
<td>Health</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall physical health</td>
<td>53.2</td>
<td>53.2</td>
<td>53.3</td>
<td>53.0</td>
</tr>
<tr>
<td>Overall mental health</td>
<td>50.4</td>
<td>50.5</td>
<td>50.5</td>
<td>50.4</td>
</tr>
<tr>
<td>Four or more chronic conditions</td>
<td>21.5</td>
<td>21.6</td>
<td>21.7</td>
<td>21.3</td>
</tr>
<tr>
<td>Total n</td>
<td>14 587</td>
<td>9294</td>
<td>5729</td>
<td>3565</td>
</tr>
</tbody>
</table>

1 Average score, normed to the 1990 USA general population using a T score distribution with a mean of 50 and standard deviation of 10.
2 Percent self-reporting four or more conditions from a checklist of 16 chronic conditions that are primarily physical (i.e. non-mental) in orientation.

**Key findings**

Monitoring over time

Figure 1 gives the cross-sectional view (i.e. the CAHPS 2.0H analog) of change in plan satisfaction. It shows that IPA and PPGP enrollee satisfaction was relatively high in 1993 and remained favorable in 1995. In both years, more than four out of five responders in these two plan types said they were satisfied. POS enrollees were somewhat less likely to be satisfied, with roughly three out of four responders reporting satisfaction. Cross-sectional comparisons, however, imply...

![Figure 1 Over-time changes in plan satisfaction: cross-sectional view.](image-url)
Anticipating demand for enrollee surveys

Table 2 Selection patterns for indemnity

<table>
<thead>
<tr>
<th>Group</th>
<th>1993</th>
<th>1995</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent satisfied¹</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>Stayed in sample</td>
<td>84.2</td>
<td>83.7</td>
</tr>
<tr>
<td>Switched</td>
<td></td>
<td></td>
</tr>
<tr>
<td>To another plan type</td>
<td>74.4</td>
<td>76.5</td>
</tr>
<tr>
<td>From another plan type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lost from sample</td>
<td>73.0</td>
<td>76.5</td>
</tr>
<tr>
<td>New to sample</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹Percent somewhat + very + completely satisfied.

downward shifts for all three managed care plan types during
the 2 years, with the highest rated plan type in 1993, IPA,
recording a significant drop. In contrast, indemnity enrollees
reported sharply lower ratings in 1993, but a major significant
gain in 1995. Fifteen percent more enrollees indicated satis-

The longitudinal view of change given in Figure 2 provides
a strikingly different picture. As with the cross-sectional
samples, the IPA and PPGP longitudinal samples recorded
relatively high 1993 scores and the POS sample recorded a
lower score. However, the longitudinal indemnity sample
scored 15% higher than its cross-sectional counterpart and,
contrary to the cross-sectional comparisons, all four plan
types in the longitudinal sample reported 2–5% declines. For
indemnity, this trend is in sharp contrast with the double-
digit gain implied by the difference between the 1993 and
1995 cross-sectional samples.

Further analysis identified substantial shifts in enrollment
that were notable for their size relative to the size of the
entire plan type sample and for their marked differences in
satisfaction. Those who stayed with indemnity were, on
average, 10% more satisfied than either those who voluntarily
switched to a managed care plan type or who were lost from
the 1993 sample. On the other hand, the 1995 satisfaction
level of new entrants to the indemnity sample in 1995 was
actually 2% higher than the level for those who had stayed
in the indemnity sample. Almost all new entrants came from
the pool of employees who had been with indemnity since
1993, sampled to make up for the loss due to the high rate
of disenrollment from this plan type during the 2 years.

The loss of less satisfied enrollees who switched out of
their plan was not unique to indemnity, as both IPA and
PPGP samples showed a similar trend (data not shown).
What was unique to indemnity was the proportionate size of
this loss. Each of the corporate sponsors had promoted a
managed care strategy that had succeeded in persuading many
employees to switch to a HMO. Less satisfied switchers
shifted principally to managed care plans, with a very small
percentage going to indemnity. The net result was selection
patterns underlying trends in satisfaction, in most cases having
no clear relationship to actual plan performance, that worked
very much to indemnity’s favor at the expense of the managed
care plans, particularly IPAs.

Comparison of the four groups shown in Table 1 revealed
another dynamic at work: non-response bias. The 1993
satisfaction levels of those who responded 2 years later
averaged 4% more than those who did not respond. Re-
sponders in 1995 recorded 84% satisfaction in 1993 whereas
1995 non-responders recorded 80% satisfaction in 1993. This
trend showed a positive relationship between satisfaction and
response, with more satisfied responders more likely to return
a completed survey. It suggests that follow-up non-responders
were more likely to have experienced negative transitions in
satisfaction from 1993 to 1995 than follow-up responders.

Figure 2 Over-time change in plan satisfaction: longitudinal view.
The net effect of this non-response bias on the relative accuracy of cross-sectional and longitudinal comparisons was not clear in our data. What was clear was that the CAHPS 2.0H approach alone detected neither selection bias or non-response bias and, in the former case, was sharply confounded by selection. Cross-sectional changes in plan satisfaction over time became understandable only in light of the separate contributions of longitudinal sample members and their selection in and out of plans. Furthermore, only by surveying employees longitudinally could the impact of non-response on satisfaction be analyzed in the cross-sectional results.

**Monitoring the sick**

Responders classified as sick in the longitudinal sample reported significantly less satisfaction than healthy responders in both years. In 1993, healthy responders were 88.7% satisfied whereas sick responders were 83.9% satisfied, while the corresponding numbers in 1995 were 85.7% and 81.8% respectively.

The longitudinal sample also, however, experienced declines in health that were beyond the level of chance. As shown in Figure 3, the average PCS score of longitudinal responders decreased by more than 1% whereas the average MCS score decreased by almost 1%. It is noteworthy that even in 1995 the health of this sample stayed favorable relative to the 1990 norm. Yet, while modest, the declines in health were not trivial. The decline in physical health, for example, was roughly five times the amount expected from the increase in age alone.

Elsewhere we have posited that health status determines satisfaction, not vice versa [11]. This position is supported by the 1993–1995 changes on PCS and MCS for the values of the self-rated outcomes measure on the 1995 overall sample: for the PCS: poor = −3.00, fair = −2.24, good = −1.34, very good = −0.81, excellent = −0.62; for the MCS: poor = −1.25; fair = −1.18, good = −1.11, very good = −0.67, excellent = −0.47. The strikingly linear changes on both measures attest to the validity of health status as the independent variable in this relationship.

Accordingly, the 2-year changes in health made an additional contribution to the change in overall plan satisfaction. Figure 4 follows overall satisfaction by health transitions during the 2-year period. As shown, the sharpest fall in satisfaction (almost 9%) was recorded by 1993 healthy responders who become sick in 1995. 1993 sick responders who stayed sick in 1995 also showed a decrease, but by less than half as much (∼4%). In contrast, those who stayed healthy recorded a comparatively negligible decrease (1%). Interestingly, the most satisfied group was those 1993 sick responders who became healthy, many of which may have known that their health was improving. This group's satisfaction score actually increased by almost 1%.

With its reliance on aggregate samples, the CAHPS 2.0H approach in effect assumes no meaningful differences between healthy and sick enrollees. Yet, in our combined data set, sick responders differed from healthy responders not only in point-in-time comparisons; they also showed different trends depending on their health transitions. Following CAHPS 2.0H, users would have missed not only that sick enrollees were consistently less satisfied throughout the study, but also that enrollees who became sick or who stayed sick played a profound role in the aggregate decrease in satisfaction.

**Driver detection**

Table 3 gives results of tests to determine drivers. Anticipating that most if not all domains would probably bear some noteworthy relationship to satisfaction, our analyses sought to prioritize among the domains under consideration. Table 3 characterizes the results of each test using this four-point
category scheme: (0) no statistical evidence ($P > 0.05$); (1) statistically significant evidence ($P < 0.05$); (2) compelling statistical evidence ($P < 0.0001$ and correlation of 0.2–0.4 or equivalent); (3) most compelling statistical evidence ($P < 0.0001$ and correlation of greater than 0.2 or equivalent and correlation larger than any other correlation).

The top panel gives results when we used correlation and multiple regression analyses conducted on the 1993 and 1995 cross-sectional data sets—the analog for the approaches that most users of CAHPS 2.0H will employ in the absence of guidance otherwise. As shown, correlations on 1993 data found support for coverage, but were not discriminating otherwise. Multiple regression techniques yielded mixed evidence concerning coverage and supporting the nomination of quality. Results in both cases assigned less importance to the access, choice of physician, customer service, and cost domains.

As shown in the bottom panel of Table 3, the additional approaches used in our study yielded a different picture. Strong, unambiguous support for assigning the quality domain highest priority came from three methods: the direct question asking responders to make choices; the longitudinal correlations predicting change in satisfaction; and longitudinal regressions predicting satisfaction changes. In addition, costs alone provided compelling evidence when predicting actual disenrollment. Combined, the evidence from our multiple approaches led us to assign high priority to the quality, coverage, and costs domains, and lower priority to access, physician choice, and customer service.

Table 3 Tests of major drivers of overall plan satisfaction

<table>
<thead>
<tr>
<th>Type of analysis</th>
<th>Access</th>
<th>Quality</th>
<th>Physician choice</th>
<th>Coverage</th>
<th>Service</th>
<th>Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approaches available to users of CAHPS 2.0H guidelines</td>
<td>Cross-sectional correlation: 1993</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Cross-sectional regression: 1993</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Cross-sectional regression: 1995</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Additional multi-method/multi-trait approaches</td>
<td>Respondents’ values</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Prediction of actual disenrollment</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Longitudinal correlation</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>(predicting change in satisfaction)</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Entries: 0 = no evidence ($P > 0.05$); 1 = statistically significant evidence ($P < 0.05$); 2 = compelling evidence ($P < 0.01$); 3 = most compelling evidence ($P < 0.001$).
Drivers of satisfaction by plan type: 1993 to 1995

<table>
<thead>
<tr>
<th>Driver variable</th>
<th>IPA</th>
<th>PPGP</th>
<th>POS</th>
<th>INDEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thoroughness of treatment</td>
<td>$-1.2$</td>
<td>$-2.8$</td>
<td>$-2.9$</td>
<td>$+2.8$</td>
</tr>
<tr>
<td>Outcomes of care</td>
<td>$-3.6$</td>
<td>$-2.9$</td>
<td>$-2.4$</td>
<td>$-5.7$</td>
</tr>
<tr>
<td>Coverage</td>
<td>$-2.4$</td>
<td>$-2.8$</td>
<td>$-2.4$</td>
<td>$-5.7$</td>
</tr>
<tr>
<td>Premium (employee share)</td>
<td>$-3.2$</td>
<td>$-2.4$</td>
<td>$-2.4$</td>
<td>$-5.7$</td>
</tr>
</tbody>
</table>

1All variables have a 5-point excellent/poor response scale and are scored positively on 0–100 point scale.
2IPA, Independent Practice Association.
3PPGP, Pre-Paid Group Practice.
4POS, Point-of-Service.
5INDEM, Indemnity.
6All – entries indicate non-significant results.

To select individual items within these domains, our analyses capitalized on the re-survey strategy, including use of cross-lagged correlations that removed bias attributable to individual rating differences. Our final short list of primary drivers included perceived thoroughness of treatment, self-rated outcomes of medical care, range of coverage, and premiums (the employee share). In this regard, other aspects of quality, for example interpersonal communication and time spent with physician, like access, choice, and customer service, showed positive trends during the 2-year period. Each driver making our final short list, in contrast, recorded decreases when dominating the trend in plan satisfaction. That some aspects of performance got better and others worse increased confidence that no unaccounted for method effect confounded this exercise.

The added value that resulted from the additional three components in our MMMT strategy becomes apparent when the 1993–1995 trend for each of these drivers is examined by plan type. The numerical entries in Table 4 report statistically significant estimates of change on the longitudinal sample; no entry indicates non-significant change.

The configuration that point-in-time, association-based statistical techniques suggested was most influential – coverage and to lesser extent quality – helped to explain change for POS. The quality element of this configuration was also useful for each of the other plan types, but coverage was not. In contrast, the configuration emerging from the MMMT strategy – quality, coverage, and costs, with the highest emphasis on quality – was useful in explaining change for all four plan types. For IPAs, the decrease in satisfaction was fueled mostly by decreases in outcomes, thoroughness of treatment, and employee premiums. For PPGPs, decreases in outcomes and, to a lesser extent, thoroughness were the major contributors. A modestly significant decrease in coverage was at work among POS plans, whereas sharp decreases in premiums for indemnity outweighed an increase in thoroughness. Applying the decision rules employed here, many CAHPS 2.0H users would probably have missed the key role that quality and especially costs played in these changes.

**Comment**

CAHPS 2.0H makes several noteworthy contributions. It brings the HEDIS and CAHPS initiatives into closer alignment. It includes new content that, the vetting process of the CAHPS focus groups suggests, taps concerns that are high on the list of consumers in terms that are salient and meaningful. These contributions represent progress that virtually all parties concerned would say is needed. They are made in addition to NCQA’s compelling track record for advancing routine survey comparisons of health plans as a standard for the field.

As outcomes measures gain currency in the marketplace, however, design issues – how data are collected, analyzed, and reported – are becoming critical [17]. Lessons supported by the results here suggest that consumer experience of plan performance is best understood when: (i) separate contributions of longitudinal members and selection in and out of plans are clarified; (ii) changes in health are identified; (iii) changes for sick and healthy enrollees are compared; and (iv) performance on specific satisfaction criteria is probed to give confirmation and detail.

Our tests of the CAHPS 2.0H approach have found it wanting. Its point-in-time, cross-sectional design will not enable users to account for selection bias. Their interpretation of change over time will be susceptible to serious errors. CAHPS 2.0H’s use of simple random samples will not sufficiently represent entire covered populations. Users may miss key differences between sick and healthy enrollees. With no guidance otherwise, many users will rely on point-in-time, association-based statistics to prioritize drivers. Important causal influences may be missed.

Given marketplace developments, several changes to CAHPS 2.0H merit review. Based on published work, these changes focus on survey design but also have implications for content.

**Monitoring over time**

Since migration across plans is likely to continue in USA health care, CAHPS 2.0H needs to work toward a design that blends longitudinal sampling into the current cross-sectional design. Users need to generalize to enrollee populations at various points in time; hence the cross-sectional design. But they also need to keep track of the stayers and switchers; hence the need for longitudinal sampling and tracking.

Routine implementation of blended cross-sectional/longitudinal designs in the marketplace will be no small feat. Our work offers one model, but others are needed because purchasers, not plans, supplied our sampling frames. A key issue is the capacity of users to re-survey disenrollees once they have left a plan. Momentum is building with several...
Anticipating demand for enrollee surveys

developments (e.g. the Health Care Financing Administration’s push toward universal identification numbers) that will make this task easier. Action is needed in the interim, however, as there is the still the issue of accuracy and the fact that users need to ‘link the dots’ is not going to lessen.

Earlier versions of HEDIS included a seven-point overall plan satisfaction item, which CAHPS 2.0H dropped for the CAHPS zero to 10 point item on overall plan evaluation. To support monitoring over time, re-instatement of the seven-point item in future versions of HEDIS warrants consideration. Its use of adjectival labels for each response option has amassed an excellent track record for both cross-sectional and longitudinal measurement. In contrast, the absence of labels for each of its 11 response options in the CAHPS item may become problematic, as users will be hard-pressed to say that a ‘3’ or an ‘8’ is itself not changing over time.

Monitoring the sick

Our comparisons of healthy versus sick responders suggest that there is a role for treating these two groups separately. CAHPS 2.0H’s use of aggregate samples needs to have folded within it provisions for identifying the sick from surveys and for over-sampling from this group in subsequent surveys. Appropriate steps for analyzing and reporting on healthy and sick enrollees cross-sectionally and longitudinally, with special attention to change in health status, are needed. The addition of health outcome measures that track functional status and well being over time would help CAHPS 2.0H to address the concerns of consumers directly. The Foundation for Accountability has recently proposed a strategy for this purpose [18].

To handle sample size issues associated with sick groups, NCQA might consider initially developing this paradigm on a community-wide (across-plan) basis. This would allow users to assess how managed care in general is caring for the sick, a feature that will not be lost on critics of HMOs. It could also stimulate interest in how well individual plans are doing and help to build momentum among plans and purchasers for greater investment in the NCQA survey undertaking.

To implement these changes, CAHPS 2.0H’s core survey specification for health needs to be reassessed. The core CAHPS 2.0H survey retains only one item (the excellent/poor measure of overall health) from the HEDIS 3.0 survey, which included the SF-12 Health Status Survey and a chronic condition checklist. This item is not a sufficient measure of the multiple dimensions of health, and, at a minimum, the SF-12, or a shorter version thereof, merits another look.

Driver detection

The three added components we brought to this task each ultimately yielded its own rank ordering of drivers. One interpretation of this work is that conclusions regarding the mix and importance of individual drivers depend on underlying study design. But, this emphasizes an even more fundamental interpretation: the assimilation and synthesis of results from multiple method/trait combinations will lead to the best driver determinations.

Given that plans and purchasers are now taking this natural next step and will increasingly do so in the immediate future, they are better served if future versions of HEDIS strive for standardization in this area. The objective should be credible, empirical determinations of the most important variables in each collected data set, with the recognition that this list could well vary across data sets and over time (even for the same plan). Full consensus would not need to be made a requirement, but it would be a worthwhile goal.

The four drivers identified in our work – ratings of thoroughness of care, outcomes of care, cover, and premiums – best met the multiple criteria imposed by our tests. They too were included in earlier versions of HEDIS but were omitted from CAHPS 2.0H. We know from previous work that these items can be measured with a high degree of reliability and validity [19]. Their re-instatement in an instrument that also keeps key CAHPS items would enable future versions of HEDIS to take advantage of the best of both bodies of work.

Acknowledgments

This paper expands on a previous submission made to the NCQA in response to its Call for Public Comment on a draft version of HEDIS 1999. The author thanks William H. Rogers, PhD, The Health Institute, co-author on this previous submission, whose collaboration made the work reported herein possible. Thanks also go to the following individuals, whose comments on an early draft of our response were incorporated in the final submission to NCQA: David Nerenz, PhD of Henry Ford Health Systems; Bruce Taylor of the GTE Corporation; Benjamin Druss, MD of Yale University; Pei Tang, BlueCross BlueShield Association, and David Langley of BlueCross BlueShield of Maine. Finally, a special note of thanks goes to the three employers that sponsored and funded the two national employee surveys used for this analysis.

References


*Accepted for publication 23 July 1998*