HEDIS Measures and Managed Care Enrollment

Dennis P. Scanlon
Pennsylvania State University

Michael Chernew
University of Michigan

This article examines the relationship between 1996 health plan enrollment and both HEDIS-based plan performance ratings and individual HEDIS measures. Data were obtained from a large firm that collected, aggregated, and disseminated plan performance ratings to its employees. Plan market share regressions are estimated controlling for out-of-pocket price and model type in addition to the plan ratings and HEDIS measures. The results suggest that employees did not respond strongly to the provided ratings. There are several potential explanations for the lack of response, including difficulty understanding the ratings and never having seen them. In addition, employees may base their plan choices on information that is obtained from their own past experience, friends, family, and colleagues. The pattern of results suggests that such information is important. Counterintuitive signs most likely reflect an inverse correlation between some HEDIS ratings (or measures) and attributes employees observe informally.

Many employers, particularly large employers, offer a choice of health plans. Rice et al. (1997) report that almost 50 percent of large employers offer several insurance options to their workers and retirees. A Robert Wood John-

son Foundation survey of establishments suggests that more than 25 percent of large establishments offer multiple managed care plans.

Managed care plans control many aspects of health care delivery including the incentives surrounding the provision of care and the set of physicians and hospitals from which enrollees can seek care. In an effort to facilitate choice and monitor quality, employers, the media, and accrediting organizations, such as the National Committee for Quality Assurance (NCQA) and the Joint Commission for Accreditation of Health Care Organizations (JCAHCO), have undertaken extensive, expensive initiatives to measure plan performance. For example, NCQA has developed, in conjunction with health plans and large employers, a set of plan performance measures (HEDIS) that many large purchasers require health plans to report (Scanlon and Hendrix 1998).[1] This effort was originally intended to facilitate quality improvement and evaluation of the plans by employers.

In recent years, many organizations have disseminated HEDIS-based plan performance ratings in the form of "report cards" to consumers. Periodicals such as *Newsweek* and *Consumer Reports* have published health plan report cards based, in part, on HEDIS measures, and several large employers have provided their employees and retirees with similar information (Scanlon et al. 1998). Two assumptions provide the foundation for these efforts.

- Employees are not adequately informed about important health plan attributes such as the measures contained in the HEDIS data set.
- Employees value the attributes measured by HEDIS and will use HEDIS-based ratings when choosing plans.

The first assumption is almost surely true in a literal sense. It is unlikely that individuals are aware of plan performance on specific HEDIS measures such as use rates for selected services (e.g., mammography, prenatal care, coronary artery bypass surgery), the percentage of physicians accepting new patients, physician board certification rates, and the percentage of enrollees satisfied with the plan. However, individuals may observe more general aspects of plan performance through sources other than the ratings. We label all of this additional information that might influence choice *informal information*, although it could reflect initiatives by others outside of the firm to evaluate plans. Much of this informal information may include employee experiences and contacts with friends, relatives, and colleagues. This general assessment of plan performance may capture a broader set of plan traits than HEDIS and influence plan choice even when plan ratings are provided.

The second assumption is more difficult to evaluate. Surveys and focus groups reveal that employees are interested in measures of plan performance

similar to those contained in HEDIS. However, strong conclusions from such studies are difficult to obtain because important traits such as the popularity of physicians and hospitals are not easily captured in hypothetical plan offerings. Hibbard and Jewett (1996) provide some evidence that employee behavior when faced with real choices may differ from that implied when asked about preferences for plan attributes. Few studies relate plan performance measures to actual plan enrollment (Scanlon, Chernew, and Lave 1997).

One study (Chernew and Scanlon 1998) examined the relationship between employer-reported plan performance ratings and health plan enrollment. The findings indicated that in several domains of performance, better plan ratings were not associated with a greater probability of plan enrollment. In some cases, better plan ratings were associated with a lower probability of enrollment. The pattern of results suggested several potential hypotheses. For example, an inverse relationship was found between enrollment and a rating of surgical care that rewarded health plans with low surgical use rates. The study hypothesized that the counterintuitive findings might reflect employee preference for plans with easier access to specialist services, a plan characteristic that may be correlated positively with high surgical use rates.

Similarly, the study postulated that an inverse relationship between a rating of satisfaction and plan enrollment might have reflected the inclusion of measures of physician availability, such as the percentage of physicians accepting new patients, in the satisfaction rating. Physician availability may be related inversely to physician popularity because, just as one does not see long lines at bad restaurants, one may not see substantial enrollment in plans with many physicians accepting new patients. However, because the underlying HEDIS data that comprised the ratings were not available in the study, Chernew and Scanlon (1998) were unable to explore these hypotheses.

This article builds on the previous work by modeling plan choice with more recent data (1996 enrollment) and including underlying HEDIS measures that comprise plan ratings in the analysis. Several specifications are estimated. The first specification relates plan choice to the ratings in a manner similar to that in Chernew and Scanlon (1998). It is useful to examine whether the previous findings still hold for several reasons. First, the presentation of the information was more clear and user friendly during 1996 enrollment such that the probability that employees would have noticed the information was likely greater during 1996 enrollment relative to 1995 enrollment. Second, the rating construction changed significantly between 1995 and 1996 enrollment. In 1995, only one rating ("superior") was assigned for each dimension of plan performance, but only 25 percent of plans received these ratings. Failure to achieve a superior rating was not noted on open enrollment materials. For 1996 enrollment, a rating of "above average," "average," "needs

improvement," or "missing data" was assigned to each plan on each dimension of performance. Third, the reliability of the data was improved due to better plan compliance with HEDIS. Comparison of the findings from 1995 and 1996 permits inferences regarding whether the change in presentation format altered earlier conclusions.

The second set of specifications included in the analysis models plan choice as a function of underlying HEDIS measures. This analysis allows exploration of the hypotheses raised in prior work. Specifically, the counterintuitive signs most likely reflect a correlation between the ratings and unobserved determinants of plan choice, which we have lumped into our broad definition of informal information. By disaggregating ratings into their components, we can assess the aspects of the ratings that are correlated with the informal information and gain insight into the nature of plan choice.

Although the underlying HEDIS data were not provided to employees, the coefficients will reflect the correlation between the HEDIS measures and information observed through informal sources as well as the firm-provided ratings. Studies of the hospital market commonly accept the idea that various aspects of performance (such as hospital mortality rates) may influence choice even if they are not directly observed by patients, physicians, or insurers (Luft et al. 1990; Burns and Wholey 1992), presumably because of their correlation with informal information.

The third set of specifications models plan choice as a function of both the ratings and individual HEDIS measures. Including both types of data in the model permits one to assess whether the underlying relationship between the ratings and plan choice changes when HEDIS measures are included. For example, does the surgical care rating matter holding surgery rates constant?

Throughout this analysis, we cannot directly measure attributes we believe are important determinants of plan choice, such as the popularity of physicians or the ease of referrals. Instead, we rely on the pattern of results regarding the HEDIS measures and ratings to explore these hypotheses.

DATA

Similar to 1995, health plan enrollment at this firm was conducted in a flexible-benefits environment in which employees were given a menu of plans from which to choose and flex dollars to spend across a variety of benefits, including health plans. The out-of-pocket price was determined by each business unit within the firm and did not equal the premium charged by the health plan. The benefit package for each health plan was standardized, and previous analysis suggests that exceptions were minor and noninfluential.

Chernew and Scanlon (1998) provide more detail regarding health plan selection at this firm.

SAMPLE

We separate individuals choosing single coverage from those choosing family coverage because these employees face different prices and because these groups of workers may differ in their valuation of plan performance measures. Workers choosing family coverage are older (41 years vs. 38), are more likely to be male (67 percent vs. 44 percent), and are more likely to be married and have dependents (2.95 covered dependents, on average, excluding themselves). Thus, employees choosing family coverage may be more interested in some of the underlying HEDIS measures used to determine plan ratings, such as childhood immunization rates or cesarean section rates. We estimate models for both the single and family samples; however, the reported results focus only on the sample of workers choosing family coverage because of sample-size issues and because the results do not differ significantly by sample.[2]

The restriction to active, nonunion employees that chose coverage for the 1996 calendar year yields a sample of 46,486 employees, of which 26,720 chose managed care plans.

HEALTH PLAN RATINGS

The firm that provided the data for this study required the plans with which it contracted to report HEDIS 2.0 or 2.5 data. In some cases, plans were unable to report all of the requested HEDIS measures. Since we use the underlying HEDIS data in our analysis, we imputed values for missing HEDIS measures using the EM algorithm if the health plan reported at least 50 percent of the requested HEDIS data. Plans that reported less than 50 percent of the HEDIS data were dropped from our analysis (nine plans). The firm used the data to rate plans on the same five domains as 1995 enrollment: physician quality, surgical care, medical treatment, participant satisfaction, and preventive care. Each domain was composed of several specific HEDIS measures; however, the composition of measures within each domain varied somewhat from the previous year (Table 1).[3] Unlike the previous year, all plans were assigned ratings for each domain.

The methodology used to determine the ratings for each plan for each domain also differed from the previous year. Specifically, an index score was assigned to each plan relative to the scores of other plans in the same geographic region. The index score for each domain was calculated by first

TABLE 1 Health Plan Quality Ratings and Their Measures

Rating	Measures
Surgical care	Cardiac catheterizations per 1,000 members (men 45-64)
	Coronary artery bypass graft surgeries per 1,000 members (men 45-64)
	Cholecystectomies per 1m000 members (women 30-64)
	Hysterectomies per 1,000 members (women 45-64)
	Laminectomies per 1,000 members (men 20-64)
	PTCAs per 1,000 members (men 45-64)
	Prostatectomies per 1,000 members (men 45-64)
	Cesarean section rate per 100 live births
Preventive care	Childhood immunization rate
	Percentage of patients receiving prenatal care in first trimester
	Cholesterol screening rate (ages 40-64)
	Mammography screening rate (women 52-64)
	Cervical cancer screening rate (women 21-64)
Participant satisfaction	Overall enrollee satisfaction
	Percentage visited PCP in last 3 years (ages 40-64)
	Telephone abandon rate
	Waiting time in days for nonurgent appointment
Physician quality	Percentage of PCPs that are board certified
	Percentage of specialists that are board certified
	Primary care turnover rate
	Percentage of PCPs accepting new patients
	Recredentialing score (verification of various aspects of physician licensing, etc.)
Medical care	Diabetic retinal exam rate
	Medical and surgical days/1,000 (ages 45-64)
	Mental health 90-day readmission rate
	Follow-up rate following major mental health disorder

Note: PCP = primary care physician.

computing a *z*-score for each measure in the domain based on the mean and variance of reported values for a national sample of plans, including many not offered by this firm. For example, the preventive care domain was made up of the following HEDIS measures: childhood immunization rate, first trimester prenatal care rate, cholesterol screening rate, mammography screening rate, and cervical cancer screening rate.[4] Within each domain, *z*-scores were averaged to compute a domain score, and then, within each of 14 regions, plans with the top-third scores were rated "superior quality," the middle third were rated "average quality," and the bottom third were rated "needs improvement."

As part of their open enrollment materials, each employee was given an information sheet for every plan in their choice set and a summary sheet comparing all plans in their choice set (Figures 1 and 2). These sheets conveyed the information used in our analysis including the out of pocket price, the relative size of the physician panel, the degree of integration (model type), and the HEDIS-based ratings for each of the five domains.[5] The plan sheets also contained three additional ratings, two of which were developed by the firm ("exceptional quality designation" and "benchmark HMO"), and a rating that indicated whether the plan was accredited by the NCQA. None of these ratings is included in our base analysis because they are not derived specifically from the HEDIS data set and because of concern regarding the number of explanatory variables relative to sample size and potential multicollinearity. These three ratings are generally not related to plan choice, and the inclusion of these ratings does not change the conclusions of our analysis.[6]fs20

As was the case with the ratings developed for 1995 enrollment, the ratings constructed by the firm for 1996 enrollment are imperfect measures of plan quality. For example, the scoring methodology and the process of aggregating the measures to form domain ratings may do a poor job of valuing plan attributes from the perspective of enrollees (Scanlon et al. 1998). In addition, just as for 1995 enrollment, the participant satisfaction rating for 1996 was based on surveys developed independently by each plan, which was the HEDIS requirement at the time. Because these surveys differed across plans, satisfaction rates are not necessarily comparable, although comparability improved between 1995 and 1996 because several plans adopted the standardized satisfaction survey that eventually became a requirement in HEDIS 3.0.

Tables 2 and 3 report descriptive statistics for plans that were available to active employees in our samples. Thirteen of the nonintegrated plans were preferred provider organization (PPO) or point-of-service (POS) plans. The unit of observation is the health plan. In the analysis, some plans are represented multiple times because they are offered in several markets. Note that the mean price, $592, is very high. This reflects a generous benefits package and a decision by the firm to charge high monthly prices and provide flex dollars to compensate employees. The prices used in the analysis reflect the prices charged to enrollees, and unspent flex dollars could be converted to pretax income.

METHOD

The underlying econometric model is based on the assumption that employees seek to maximize utility, and the utility derived by each individual *i* from health plan *j* can be expressed as a function of health plan attributes. The

Option 203:

Non-Medicare Provisions Only

MC Plan #1
1234 CABG Avenue
Utilization Review, PA 16802

HEALTH PLAN FACT SHEET

About the Rating Symbols

Depending on how this health plan has been rated, up to three different symbols may appear above. The absence of any symbols may reflect the unavailability or inadequacy of plan performance data.

Health plans that have met national standards of quality as determined by the **National Committee for Quality Assurance (NCQA)** receive the NCQA logo designation. By December 31, 1996, all firm health plans will be required to achieve NCQA accreditation.

Health plans that offer the best combination of quality and cost-effectiveness in each location are designated as "**Benchmark Plans.**"

Finally, the firm has identified health plans throughout the country that have earned an **Exceptional Quality Designation** based on Healthplan Employer Data Information SET (HEDIS) data and consumer surveys. The Exceptional Quality Designation logo indicates that the health plan has been rated as having the very best overall quality.

Quality Ratings

◈◈	**Participant Satisfaction**	This rating is taken from surveys of participants and includes questions about choice of physicians, waiting times for care, and overall satisfaction with care.
◈◈◈	**Preventive Care**	Disease prevention is the basis for this rating, which evaluates the amount of screening for cancers and high cholesterol, childhood immunizations, and smoking cessation and other wellness programs.
◈◈◈	**Medical Care**	This measure rates the health plan's success rate in caring for patients with serious medical conditions, such as asthma, diabetes, and mental illness.
◈◈◈	**Surgical Care**	Many medical problems can be best treated without surgery and health plans are rated on their ability to only perform surgical procedures that are medically necessary.
◈◈	**Physician Quality**	The quality of physicians within the health plan is rated by the number that are board certified in their specialty, by the turnover rate, and by the number of physicians available to new members.

Key : ◈ = Needs improvement, ◈◈ = average quality, ◈◈◈ = superior quality, NA = not applicable, ND = no data was provided by this plan

FIGURE 1   Health Plan Fact Sheet

Use this sheet to compare the options available to you. More specific information on each health plan is available on the individual Health Plan Fact Sheets for each option. Also, the back of this sheet lists the firm's standard HMO coverage for your reference.

Your 1996 Medical Coverage Options

Option	NCQA Accredited?	Benchmark HMO?	Exceptional Quality Designation?	Participant Satisfaction	Preventive Care	Medical Care	Surgical Care	Physician Quality	Monthly Cost of Coverage (You Only)
203. MC Plan #1	Yes	No	Yes	◈◈	◈◈◈	◈◈◈	◈◈◈	◈◈	$ 578
205. MC Plan #2	No	No	No	◈◈	◈◈◈	◈◈	◈◈	◈	$ 595
211. MC Plan #3	No	No	Yes	◈◈◈	◈◈	◈◈	◈	◈◈◈	$ 595
213. MC Plan #4	Yes	Yes	Yes	◈◈◈	◈◈◈	◈◈	◈◈	◈◈◈	$ 549
227. MC Plan #5	Yes	Yes	Yes	◈◈◈	◈◈◈	◈◈	◈◈	◈◈◈	$ 578
228. MC Plan #6	Yes	No	Yes	◈◈	◈◈◈	◈◈◈	◈◈◈	◈◈	$ 549

Key : ◈ = Needs improvement, ◈◈ = average quality, ◈◈◈ = superior quality, NA = not applicable, ND = no data was provided by this plan

FIGURE 2   Comparing Your Medical Options

TABLE 2   Descriptive Statistics for Plan Ratings (N = 96)

Domain	Rating	Mean
Preventive care	Superior quality	0.33
	Average quality	0.21
	Needs improvement	0.22
	Missing data	0.24
Participant satisfaction	Superior quality	0.32
	Average quality	0.32
	Needs improvement	0.11
	Missing data	0.24
Medical treatment	Superior quality	0.23
	Average quality	0.31
	Needs improvement	0.21
	Missing data	0.25
Physician quality	Superior quality	0.35
	Average quality	0.34
	Needs improvement	0.18
	Missing data	0.13
Surgical care	Superior quality	0.23
	Average quality	0.28
	Needs improvement	0.22
	Missing data	0.27

Note: Plans reporting less than half of the measures comprising a specific rating were labeled *missing data* for the respective rating.

methodology is outlined in detail elsewhere (Chernew and Scanlon 1998). Utility of each plan is a function of plan attributes, a plan-specific error term, and an error term that varies by chooser. This latter error is assumed to follow a type I extreme value distribution.

Because it is likely in this study that unobserved plan attributes are important and are systematically valued by potential enrollees, the plan-specific error component is likely not equal to zero, and standard error estimates from conditional logit models would be biased downward. A more appropriate estimation technique approximates the conditional logit model by using grouped data (Berry 1994; Chernew and Scanlon 1998). This approach models the log difference in market share between a given plan (j) in a specific market (m) and an arbitrarily chosen reference plan (j^*) in the same market as a function of differences between the plans in the explanatory variables. This model is outlined in Equation 1, in which υ_{jm} is a stochastic error and P_{jm} is the probability that plan j is chosen by each employee in market m.[7] As the number of

TABLE 3   Descriptive Statistics for Price and Nonprice Health Plan Attributes ($N = 96$)

	Mean	Standard Deviation	Percentage missing[a]
Price (dollars per month)[b]	592	32	0
Integration	0.260	0.441	0
Surgical care factor[a]	−0.250	0.855	0-5
Prevention factor[a]	0.121	0.559	1-15
Physician board certification factor[a]	0.086	0.542	7
CABG rate per 1,000	3.672	1.826	1
Cervical cancer screening rate (%)	0.746	0.071	1
Specialist certification rate (%)	0.836	0.074	7
Overall enrollee satisfaction	0.887	0.057	9
Percentage of physicians accepting new patients	0.880	0.093	5
Prenatal care first trimester rate (%)	0.875	0.081	9
Cesarean-section rate (%)	0.206	0.038	7
Physician turnover rate (%)	0.082	0.345	10
Follow-up after mental health disorder (%)	0.749	0.136	22
Readmission for mental health disorder (%)	0.085	0.243	43
Medical/surgical days per 1,000	325.800	87.529	23
Recredentialing score (range 1-4)[c]	3.825	0.724	4
Percentage visiting PCP in last 3 years (ages 40-64)	0.902	0.064	20
Diabetic retinal exam rate	0.398	0.134	9

Note: CABG = coronary artery bypass graft surgery; PCP = primary care physician.
a. Imputed data were used in computation of the factors. The range is the range of missing data for variables comprising the factor.
b. Because plans can be offered at different prices in different markets, the unit of observation for price is the (plan and market) combination.
c. Equals sum of four dummy variables indicating whether the plan checks specific physician credentials.

employees approaches infinity in each market, the last two terms approach zero, and the approximation becomes exact.

$$\log(S_{jm}) - \log(S_{jm}^*) \cong \beta'(X_j - X_j^*) + (\gamma_j - \gamma_j^*) + \upsilon_{jm}/P_{jm} - \upsilon_{jm}^*/P_{jm}^*. \qquad (1)$$

Each market contains $(n - 1)$ observations (in which n is equal to the total number of plans in a market), and the number of observations in the ordinary least squares regression is equal to the sum of the observations for all markets. The unit of analysis is the market, and we focus on workers in markets in

which at least 10 employees meet our sample criteria. The estimates are similar if we raise the threshold for inclusion to 20 employees.[8] For the family-coverage sample, there is sufficient data for 96 plans offered in 68 markets of adequate size. Because some plans are offered in multiple markets, there are a total of 222 plan/market observations. Because a reference plan is used for each market, the family-coverage sample has 154 observations (222 – 68), representing the choices of 9,719 employees. Because of fewer markets of adequate size, the analogous numbers for the single-coverage sample are 105 observations, representing 5,536 employees.

In all models, the X_j vector includes the employee's monthly out-of-pocket plan contribution and the extent of plan integration (model type). In the first specification, the X vector also includes dummy variables indicating the scores on each of the five HEDIS-based plan ratings (two diamonds, "average quality," being the omitted category).

Because our analysis is cross sectional, we cannot identify the impact the release of plan ratings had on choice. A positive relationship between ratings and plan choice could reflect a response to the information. It might also indicate that the ratings simply confirmed attributes employees knew already. Similarly, weak correlation between the ratings and enrollment is consistent with the hypothesis that employees did not respond to the ratings or that the ratings and informal information offset each other.

If the correlation between the ratings and enrollment is negative, we cannot definitively say the information release had no effect. Maybe without the release, the correlation between the ratings and enrollment would have been more negative. However, an inverse correlation does reveal that there are other attributes, not captured in the rating system, influencing choice. The inverse correlation reveals that these other attributes are inversely related to the values captured by the rating system.

In the second set of specifications, the plan ratings are omitted in favor of the HEDIS measures comprising those ratings.[9] Prior research revealed counterintuitive findings from the ratings analysis. Separate examination of the relationship between underlying HEDIS measures and plan choice can provide insight into which HEDIS measures accounted for the results from the ratings specifications.

Because the HEDIS measures are highly correlated with each other, we condensed the set of HEDIS measures into factors based on confirmatory factor analysis using principal axis factoring with varimax rotation, grouping variables that were conceptually related and empirically related in exploratory factor analysis. The factors used in our analysis and their component variables and Cronbach's alphas are reported in Table 4. Several HEDIS measures that did not load strongly on these factors are included individually in our share

TABLE 4   HEDIS-Based Factors

Factor	Measures
Surgical care	
α = .71	Prostatectomies per 1,000 members (men 45-64)
	PTCAs per 1,000 members (men 45-64)
	Coronary artery bypass graft surgeries per 1,000 members (men 45-64)
	Cardiac catheterizations per 1,000 members (men 45-64)
	Laminectomies per 1,000 members (men 20-64)
	Cholecystectomies per 1,000 members (women 30-64)
	Hysterectomies per 1,000 members (women 45-64)
Prevention	
α = .82	Percentage visited PCP in last 3 years (ages 40-64)
	Diabetic retinal exam rate
	Cholesterol screening rate (ages 40-64)
	Childhood immunization rate
	Cervical cancer screening rate (women 21-64)
	Mammography screening rate (women 52-64)
Physician board certification	
α = .72	Percentage of PCPs that are board certified
	Percentage of specialists that are board certified

Note: PTCA = percutaneous transluminal coronary angioplasty; PCP = primary care physician.

models. As a test of sensitivity, we also estimate models that include dummy variables indicating whether the analysis contained imputed HEDIS data for measures that were not reported by the health plan. Another model replaces the factor scores with the specific HEDIS measures that had the highest factor loading on each of the three latent variables included in our analysis.

Finally, the third specifications include both the ratings and underlying HEDIS measures. There are not a sufficient number of observations to estimate one model with all of the ratings variables and HEDIS measures. Hence, five specifications corresponding to each domain of performance were estimated. Each model included the ratings for the relevant domain and the relevant HEDIS factor and measures that comprised the rating.

If the ratings were dominant, one would expect the underlying HEDIS measures would not predict plan choice in these models. Such a test requires the assumption that important informal information is correlated with the HEDIS measures. If this assumption holds, the coefficients on the HEDIS measures should capture some of the impact of informal information. By

holding the underlying HEDIS measures constant, the coefficients on the ratings may more closely capture the true effect of the rating information on plan choice. For example, controlling for the value of the surgical factor score and the cesarean section rate, the coefficients on the surgery ratings may reflect the impact of the firm-provided ratings to a greater extent. Essentially, the coefficients on the ratings test for the effect of plan performance related to the discrete categorization of plans into "superior," "average," and "needs improvement."

RESULTS

HEDIS-BASED PLAN RATINGS

Parameter estimates for the model relating choice to the HEDIS-based performance ratings for workers choosing family coverage are presented in Table 5. For the performance ratings, the relative probabilities represent the probability of choosing a plan with the indicated rating relative to a plan that is identical except rated "average quality" on the dimension in question. If employees were responding strongly to the ratings, the relative probability for "superior quality" would be greater than 1, and the relative probability for "needs improvement" would be less than 1. The relative probability for price reflects a difference of $10 per month (the mean price is $592/month).

The hypothesis that none of the report card ratings enters the model can be rejected at the 0.05 level. Yet, the coefficients on the ratings are varied in their impact (Table 5).[10] For example, none of the "missing data" ratings is significantly different from zero at the 0.10 level, suggesting that employees were not systematically attracted or deterred from plans that did not report enough data to be assigned one of the three rating categories. For only one of the domains, satisfaction, do the estimates support the hypothesis that employees responded to the reported ratings in the manner hypothesized (the "superior quality" coefficient is positive, and the "needs improvement" coefficient is negative). Only the "needs improvement" result is statistically significant, however.

This finding contrasts with our previous work in which the satisfaction rating was consistently inversely related to plan choice. Several differences in the construction of the ratings may explain the difference in findings. For example, the 1995 satisfaction index included a measure of the percentage of physicians accepting new patients. This measure may be correlated inversely with enrollee preferences if acceptance of new patients is inversely related to physician popularity.

TABLE 5   Estimates of the Relationship between Plan Choice and Plan Ratings

		Coefficient (Standard Error)	Relative Probability (p Value)
	Price	−0.006	0.94
		(0.004)	(0.15)
	Integration	0.085	1.09
		(0.309)	(0.78)
Preventive care[a]	Superior	−0.029	0.97
		(0.340)	(0.93)
	Needs improvement	−0.079	0.92
		(0.395)	(0.84)
	Missing data	−0.147	0.86
		(0.875)	(0.87)
Participant satisfaction[a]	Superior	0.158	1.17
		(0.316)	(0.62)
	Needs improvement	−0.844	0.43
		(0.474)	(0.07)
	Missing data	−0.660	0.52
		(0.503)	(0.19)
Medical treatment[a]	Superior	0.534	1.71
		(0.297)	(0.07)
	Needs improvement	0.399	1.49
		(0.312)	(0.20)
	Missing data	−0.562	0.57
		(0.740)	(0.45)
Physician quality[a]	Superior	−0.276	0.76
		(0.291)	(0.34)
	Needs improvement	−0.070	0.93
		(0.274)	(0.80)
	Missing data	0.037	1.04
		(0.390)	(0.93)
Surgical care[a]	Superior	−1.254	0.29
		(0.289)	(<.001)
	Needs improvement	0.088	1.09
		(0.314)	(0.78)
	Missing data	0.031	1.03
		(0.399)	(0.94)

Note: $N = 154$; Adjusted $R^2 = 0.21$. p values and standard errors are adjusted for possible heteroscedasticity.
a. "Average quality" was the omitted category.

For three of the remaining domains—prevention, medical treatment, and physician quality—the "superior quality" and "needs improvement" coefficients have the same sign, suggesting a U or inverse-U shape between the rating and enrollment. For prevention, both coefficients are small and statistically insignificant, suggesting no relationship between the rating and enrollment. If outliers are removed, the coefficients are consistent with our earlier work and support the hypothesized relationship between ratings and enrollment. The estimates are not statistically significant however.[11]

For the medical treatment domain, the coefficient on "superior quality" is positive and statistically significant. The "needs improvement" coefficient is not statistically different from zero.[12] The point estimate suggests that enrollees are 49 percent more likely to enroll in a plan rated "needs improvement" than a plan rated "average performance" in this domain.

For the physician quality domain, the coefficients suggest that enrollees are most likely to enroll in plans rated "average quality." Although neither coefficient is statistically different from zero, the coefficient on "superior quality" is larger in absolute value than that on "needs improvement," which is close to zero. If outliers are omitted, the "needs improvement" coefficient becomes positive, although it remains small and statistically insignificant. This suggests that employees may prefer plans not rated "superior quality" on this dimension of plan performance. This may be due to the fact that for 1996 enrollment, the firm moved the "percentage of physicians accepting new patients" measure into the physician quality domain from the satisfaction domain.

The strongest result from the analysis is the inverse relationship between probability of plan enrollment and a rating of "superior quality" on the surgical care domain. This domain captures use of surgical procedures, with more use considered worse. Employees seem to prefer plans with higher rates of use, which may reflect easier access to specialists.

Price is correlated inversely with enrollment, although it is not statistically significant. The relative probability for the price coefficient indicates that active nonunion employees choosing family coverage are approximately 6 percent less likely to choose a plan with an out-of-pocket price of $602/month than an otherwise identical plan with a mean price of $592/month. The magnitude of the price coefficient is smaller than those estimated in other studies (Feldman et al. 1989). For any dollar change in price, we estimate a smaller response. Yet, our base prices are much higher than those observed in other studies. In Feldman et al. (1989), a $5 change in price approximated a doubling of price, a much larger percentage change than in our study. Because elasticities relate the percentage change in volume (enrollment) to a percentage

change in price, the elasticities implied by our work would be different than those implied by other studies.

The coefficient on integration is not statistically different from zero but suggests that employees prefer more integrated plans, controlling for price and the dimensions of performance captured by the ratings. The insignificance of the coefficient may reflect measurement error associated with blurring distinctions between plan types and heterogeneity within plan types.

On balance, the results examining the relationship between plan choice and the performance ratings support earlier conclusions that employees did not respond strongly to the firm-developed HEDIS-based ratings.

UNDERLYING HEDIS MEASURES

The results from the regressions that include the underlying HEDIS measures are generally consistent across models (Table 6). However, correlation among variables occasionally inflates or deflates the point estimates (and *p* values), although the sign of the coefficients does not change across specifications.

Consistent with the results from the surgical care ratings, employees prefer plans that perform more surgeries. The correlation among surgery rates makes it difficult to identify the impact of specific surgery measures. We could not identify any specific surgical category that might explain the inverse correlation between the surgery rating and enrollment. It appears as if there is a common factor related to all surgeries, such as ease of referral, that accounts for the finding. For example, the surgical care factor score, coronary artery bypass graft (CABG) surgery rate, and cesarean section rate are all positively related to the probability of plan enrollment. The findings do not necessarily suggest that employees prefer cesarean sections, for example, but they may prefer plans with management systems that do not discourage use of cesarean sections or other surgeries. Because the rating system would have encouraged employees into plans with fewer surgeries, we take this as evidence that employees were aware of differences in access to surgical treatment (or correlates of access, such as ease of specialist referrals) and chose plans with easy access. Because our analysis is cross sectional, we cannot determine if the ratings dampened this relationship, but they certainly did not overwhelm it.

The results regarding the prevention factor score and the cervical cancer-screening rate reveal a negative but statistically insignificant relationship between enrollment and use of these procedures. It may be that employees understand that these services will be available in each of the plans if desired.[13] However, the prenatal care rate is positively and significantly related to plan enrollment in all four models in Table 6, suggesting that

TABLE 6 Estimates of the Relationship between Plan Choice and HEDIS Measures

	Relative Probability (p *Value)*	Relative Probability (p *Value)*	Relative Probability (p *Value)*	Relative Probability (p *Value)*
Price	0.89	0.89	0.89	0.93
	(0.03)	(0.02)	(0.02)	(0.16)
Integration	1.07	1.06	1.61	1.03
	(0.83)	(0.85)	(0.08)	(0.94)
Surgical care factor	1.54	1.51	1.50	
	(0.01)	(0.01)	(< .005)	
Prevention factor	0.84	0.84	0.83	
	(0.46)	(0.46)	(0.35)	
Physician board certification factor	0.72	0.71	0.64	
	(0.04)	(0.05)	(0.01)	
CABG rate				1.14
				(0.35)
Cervical cancer screening rate				0.84
				(0.40)
Specialist certification rate				0.64
				(< .003)
Overall enrollee satisfaction	1.18	1.18	1.06	1.24
	(0.14)	(0.14)	(0.57)	(0.06)
Percentage of physicians accepting new patients	0.86	0.86	0.77	0.82
	(0.27)	(0.23)	(0.04)	(0.13)
Prenatal care first trimester rate	1.43	1.46	1.45	1.59
	(0.05)	(0.04)	(0.03)	(0.01)
Cesarean section rate	1.33	1.31	1.48	1.31
	(0.12)	(0.13)	(0.01)	(0.11)
Physician turnover rate	0.83	0.87	0.84	0.88
	(0.78)	(0.03)	(0.01)	(0.07)
Follow-up after mental health disorder	1.29	1.27	1.59	1.01
	(0.02)	(0.02)	(< .002)	(0.94)
Readmission for mental health disorder	1.05			
	(0.95)			
Medical/surgical days/1,000	0.49			
	(0.69)			
Recredentialing score	1.00			
	(0.97)			
Dummies for imputed values	No	No	Yes	No
N	154	154	154	154
Adjusted R^2	0.16	0.18	0.26	0.12

Note: CABG = coronary artery bypass graft surgery.

employees are between 43 and 59 percent more likely to enroll in plans that provide more prenatal care.

The certification factor is consistently related inversely to plan enrollment and statistically different from zero. We do not believe this reflects a preference of employees for non-board-certified physicians. Instead, this result could reflect a correlation between other plan attributes and certification. Although we cannot identify which type of informal information is generating this result, we take this finding as evidence that plan traits not observed by the researcher or contained in HEDIS are important.

The results regarding the percentage of physicians accepting new patients also support our earlier hypotheses (although they are statistically significant only when the imputation dummies are included). Specifically, employees are less likely to enroll in plans with many physicians accepting new patients. This may reflect a correlation between acceptance of new patients and physician popularity. We were unable to explore similar hypotheses with data on waiting times because these measures were dropped from the analysis because of too much missing data.

Survey-based measures of employee satisfaction (overall enrollee satisfaction) are positively related to enrollment, although the results are generally not statistically significant. This is consistent with the findings from the analysis of firm-reported ratings and confirms our hypothesis from previous work that could not be addressed with the earlier data. Specifically, the significant inverse relationship between the satisfaction rating and enrollment observed in prior work was not due to preferences for plans with low satisfaction scores but instead reflected other measures included in the index that were inversely related to enrollment (e.g., percentage of primary care physicians accepting new patients).

Finally, the coefficient on price reflects an inverse relationship between price and enrollment. The results from these models are typically stronger than the specification that included the performance ratings and are significantly different from zero in three of the four models. The coefficient on integration remains positive although generally not statistically significant.

RATINGS AND HEDIS MEASURES

These results also suggest that employees did not respond strongly to the plan ratings (Table 7). In a few cases, the coefficients on ratings move in the hypothesized direction (in the positive direction for the coefficient on "superior quality" and in the negative direction for the coefficient on "needs improvement"). Yet in other cases, the coefficients move in the opposite direction. In no case do the fundamentally counterintuitive signs on the ratings

variables reverse sign when the underlying HEDIS measures are controlled for in a linear fashion.

DISCUSSION

Considerable efforts are under way to convey plan performance information to consumers. These efforts are typically based on the notion that consumers are uninformed and will use plan performance ratings to aid in their enrollment decisions. Although longitudinal data would be necessary to assess the true impact of the release of plan performance ratings, this work suggests that the ratings did not have a major influence on plan enrollment at a large firm in 1996. However, because the analysis is cross sectional, the effects of information cannot be disentangled from the effects of plan attributes valued by the employees that would have been observed even without the release of formal ratings.

There are several potential explanations for why enrollment was not strongly related to the plan ratings. These explanations are not mutually exclusive. First, many employees may never have seen the ratings even though they were included as part of the open enrollment materials. It is possible that employees who were satisfied with their existing plan and had no intention of switching may have ignored the ratings. Prior work found no difference in the results for a sample of switchers.

Second, employees may have distrusted the source of the ratings—the employer. Third, employees may have found the ratings confusing or difficult to understand. Fourth, information obtained from informal channels may offset the impact of the ratings.

The counterintuitive signs indicate that the latter explanation may be important. The first three explanations would suggest no significant relationship between plan performance ratings (or measures) and enrollment. Yet, the finding of several statistically significant counterintuitive signs suggests that unobserved factors are important. Moreover, the general pattern of results is consistent with the hypothesis that informal information matters.

Substantial research indicates that consumers are interested in such measures (U.S. General Accounting Office 1995; Tumlinson et al. 1997; Robinson and Brodie 1997; Hibbard and Jewett 1996). The findings from this study suggest that additional research is needed to identify how (or if) consumers use plan performance ratings in an environment when other sources of information are available.

Finally, the results from this study should not be interpreted as evidence for abandoning health plan performance measurement or reporting. Instead, measurement and reporting efforts might be refined and directed toward the

TABLE 7   Estimates of the Relationship between Plan Choice, Plan Ratings, and HEDIS Measures

	Surgical Care Relative Probability (p Value)	Preventive Care Relative Probability (p Value)	Participant Satisfaction Relative Probability (p Value)	Physician Quality Relative Probability (p Value)	Medical Care Relative Probability (p Value)
Price	0.909 (0.05)	0.976 (0.62)	0.968 (0.47)	0.913 (0.06)	0.985 (0.73)
Integration	1.225 (0.59)	1.013 (0.97)	1.095 (0.77)	0.960 (0.91)	0.928 (0.85)
Superior	0.470 (< 0.01)	0.912 (0.77)	1.261 (0.31)	0.760 (0.27)	1.694 (0.16)
Needs improvement	0.881 (0.69)	1.266 (0.50)	0.384 (0.03)	1.247 (0.43)	1.149 (0.63)
Missing data	0.824 (0.62)	0.422 (0.18)	0.555 (0.14)	0.302 (0.03)	0.538 (0.28)
Surgical care factor	1.589 (< 0.01)				
Prevention factor		0.714 (0.37)			
Physician board certification factor				0.437 (< 0.01)	
Cesarean section rate	1.541 (0.21)				
Prenatal care rate		1.487 (0.02)			

Percentage visited primary care physician in last 3 years		1.187 (0.12)			
Overall satisfaction		1.007 (0.95)			
Physician turnover rate				0.820 (0.02)	
Percentage of physicians accepting new patients				0.844 (0.20)	
Recredentialing score				0.934 (0.47)	
Diabetic retinal exam rate					0.970 (0.85)
Medical and surgical days/1,000					1.184 (0.18)
Follow up following major MH disorder					0.922 (0.45)
Mental health 90-day readmission rate					0.965 (0.50)
N	154	154	154	154	154
Adjusted R^2	0.15	0.04	0.08	0.13	0.01

most appropriate audience.[14] For example, plan performance information may be most valuable to employers when selecting the set of plans to offer. Moreover, the possibility that the collection and reporting of performance measures may encourage health plans to improve the quality of care provided should not be forgotten.

NOTES

1. HEDIS is the Health Plan Employer Data and Information Set (National Committee for Quality Assurance 1995).
2. The results regarding performance measures (ratings and HEDIS variables) are similar for both the single-coverage and family-coverage samples. In the single sample, the price coefficient varies widely across specifications and is positive in specifications with the ratings, which we believe reflects overfitting of the model and correlation with other variables (notably integration).
3. For example, the percentage of physicians accepting new patients, a HEDIS measure, was included in the employee satisfaction domain in 1995 and the physician quality domain in 1996.
4. The measures of use are typically computed using an appropriate denominator (for example, cesarean section rates are computed as a percentage of live births). However, other than these adjustments, case-mix severity is not controlled for.
5. Information on physician panel size and model type (integration) was included on the back of the health plan summary sheet (Figure 1).
6. Only the "exceptional quality designation" rating is significantly related to plan enrollment (in the positive direction) when these three ratings are included in the base analysis.
7. The probability of choosing plan j is identical for all employees in a market since this probability is modeled only as a function of plan characteristics (which do not vary by employee). In our analysis, the probability differs for family- and single-coverage choosers by virtue of estimating separate regressions for each sample.
8. The model in Equation 1 is heteroscedastic because markets differ in size and plans differ in the probability they will be chosen. The base model, which included the HEDIS-based ratings, revealed heteroscedasticity. However, tests for heteroscedasticity could not reject the hypothesis that the errors were homoscedastic in the models that included the underlying HEDIS measures. This may indicate that the variance in the unobserved plan traits, the gs, dominates the variance in the aggregated individual-specific stochastic terms. An unweighted version of Equation 1 is reported using standard errors and p values robust to heteroscedasticity. Unadjusted standard errors and models weighted by market size yield much the same results.
9. Not all plans reported each HEDIS measure. We dropped nine plans reporting fewer than 50 percent of all measures. Similarly, we dropped HEDIS measures that were commonly not reported. The remaining missing data were imputed using the EM algorithm. Given the number of plans missing some data, this approach is pref-

erable to analysis of only complete cases (Little 1992). However, the imputation will affect our standard errors.

10. We can reject the hypothesis that the performance measures can be aggregated into one index representing the sum of superior ratings or that the four nonsatisfaction scores can be aggregated into a single index. Such aggregate indices would mask the heterogeneity of the estimated coefficients.

11. Outliers are defined as observations with a studentized residual greater than 2 (in absolute value) (Belsey, Kuh, and Welsch 1980).

12. In the weighted version of the model, the coefficient on "needs improvement" in the medical treatment domain is much larger and statistically significant and the coefficient on "superior quality" is about half as large as that reported (still positive) and not statistically significant.

13. This relationship becomes statistically significant in the weighted regression models.

14. Updated versions of HEDIS already exist that include expanded measures of outcomes and standardized measures of satisfaction. Other plan performance measurement systems, such as the Consumer Assessment of Health Plans Survey (CAHPS), have also been developed. These revised systems may provide more useful data. Virtually all of the measures in this analysis are included in the most recent version of HEDIS (HEDIS 3.0), although some of the detail surrounding construction of the measures has been refined. HEDIS 3.0 also contains additional measures of plan performance, generally related to the quality of care. For example, HEDIS 3.0 includes measures of the use of beta blocker treatment following heart attack, appropriate use of antibiotics for children's ear infections (otitis media), and the percentage of low-birth-weight babies delivered at facilities designed to handle them. HEDIS 3.0 also uses a standardized satisfaction survey. The refinement and expansion of HEDIS may strengthen the relationship between report card measures and plan choice.

REFERENCES

Belsey, D. A., E. Kuh, and R. E. Welsch. 1980. *Regression Diagnostics.* New York: John Wiley.

Berry, S. T. 1994. Estimating Discrete-Choice Models of Product Differentiation. *RAND Journal of Economics* 25:242-62.

Burns, L. R., and D. R. Wholey. 1992. The Impact of Physician Characteristics in Conditional Choice Models for Hospital. *Journal of Health Economics* 11 (1): 43-62.

Chernew, M. E., and D. P. Scanlon. 1998. Health Plan Report Cards and Insurance Choice. *Inquiry* 35 (Spring): 9-22.

Feldman, R., M. Finch, B. Dowd, and S. Cassou. 1989. The Demand for Employment-Based Health Insurance Plans. *The Journal of Human Resources* 24 (1): 115-42.

Hibbard, J. H., and J. J. Jewett. 1996. What Type of Quality Information Do Consumers Want in a Health Care Report Card? *Medical Care Research and Review* 53:28-47.

Little, R. J. 1992. Regression with Missing X's: A Review. *Journal of the American Statistical Association* 87 (420): 1227-37.

Luft, H. S., D. W. Garnick, D. H. Mark, D. J. Peltzman, C. S. Phibbs, E. Lichtenberg, and S. J. McPhee. 1990. Does Quality Influence Choice of Hospital? *Journal of the American Medical Association* 263 (21): 2899-906.

National Committee for Quality Assurance. 1995. *HEDIS 2.5: Updated Specifications for HEDIS 2.0.* Washington, DC: National Committee for Quality Assurance.

Rice, T., N. Puorat, R. Levan, L. Silbert, K. Hunt, and K. M. Hurst. 1997. Job Based Health Insurance and Choice: Plans Available, Trends, and Employee Take-up. Paper presented at the American Public Health Association Conference, November, Indianapolis, IN.

Robinson, S., and M. Brodie. 1997. Understanding the Quality Challenge for Health Consumers: The Kaiser/AHCPR Survey. *Journal on Quality Improvement* 23:239-44.

Scanlon, D. P., M. Chernew, and J. R. Lave. 1997. Consumer Health Plan Choice: Current Knowledge and Future Directions. *Annual Review of Public Health* 18:507-28.

Scanlon, D. P., M. Chernew, S. Sheffler, and A. M. Fendrick. 1998. Health Plan Report Cards: Exploring the Differences in Plan Ratings. *Joint Commission Journal on Quality Improvement* 24 (January): 5-20.

Scanlon, D. P., and T. J. Hendrix. 1998. Health Plan Accreditation: NCQA, JCAHO, or Both? *Managed Care Quarterly* 6 (4): 52-61.

Tumlinson, A., H. Bottigheimer, P. Mahoney, E. M. Stone, and A. Hendricks. 1997. Choosing a Health Plan: What Information Will Consumers Use? *Health Affairs* 16:229-39.

U.S. General Accounting Office. 1995. *Employers and Consumers Want Additional Information on Quality.* Report No. GAO/HEHS-95-201. Washington, DC: U.S. General Accounting Office.