# The Impact of Health Plan Report Cards on Managed Care Enrollment

July 12, 2000

Dennis P. Scanlon
Assistant Professor
Department of Health Policy & Administration
The Pennsylvania State University

Michael Chernew
Associate Professor
Department of Health Management & Policy
Department of Economics
Department of Internal Medicine
The University of Michigan

Catherine McLaughlin
Associate Professor
Department of Health Management & Policy
The University of Michigan

Gary Solon
Professor
Department of Economics
The University of Michigan

## Abstract

Belief in the importance of information in today's health insurance marketplace has led to the development and dissemination of health plan report cards by employers, the media, and consumer advocacy organizations. These report cards attempt to measure plan performance along various dimensions and are often based on the Health Plan Employer Data and Information Set (HEDIS).

We examine how the release of health plan performance ratings influences health plan enrollment. Our analysis is based on the health plan choices of employees of General Motors (GM) Corporation for two open enrollment periods. Specifically, for 1997 enrollment, GM produced a HEDIS-based report card for all HMOs they contract with nationally. The report card rated plans on a three-point scale in six domains of performance (e.g., preventive care, surgical care). The report card was disseminated for the first time during the 1997 open enrollment period to non-union employees.

In accordance with earlier research, our results confirm that employees are less likely to enroll in plans requiring relatively high out-of-pocket contributions. The point estimates suggest that a 10% increase in the relative price of a plan would generate about a 3% drop in its relative market share. The results with respect to the report card ratings are sensitive to the exact specification and sample, perhaps due to multicollinearity among report card rating variables. In most models the results are equivocal. The hypothesis that employees shifted towards the higher-rated plans is as likely to be rejected as accepted, though most of the estimated coefficients on ratings are not statistically significant. However, a specification that aggregates the ratings provides some evidence that employees avoided plans with many below-average ratings.

# I.    Background

For nearly two decades, many scholars and policymakers have advocated a competitive health insurance market in which plans compete for enrollees on the basis of price and quality.  The managed competition model requires that consumers choose from among several competing plans and pay the incremental cost associated with the plan they choose.  Because managed care plans integrate the financing and delivery of health care, the quality of health care plans should be broadly interpreted to include aspects of provider networks and clinical care as well as traditional measures of plan quality such as customer service. When choosing among competing plans, consumers must actively consider the trade-off between price and quality. Advocates of managed competition argue that the competitive model will promote efficient market outcomes.

A cornerstone of the managed competition model is the requirement that consumers be sufficiently informed about quality. Towards this end, employers and other organizations are increasingly compiling and releasing information about dimensions of plan performance thought to be related to plan quality.  The information on plan performance is based largely on standardized data systems, such as the Health Plan Employer Data and Information Set (HEDIS) and the Consumer Assessment of Health Plan Survey (CAHPS). Together these systems measure over 100 aspects of plan performance, which are typically aggregated into a more manageable number of 'domains' of performance, such as 'preventive care' and 'satisfaction.'

This paper examines the impact of the release of a health plan report card on the health plan choices of employees at General Motors (GM) Corporation, whose benefit program resembles the managed competition model. The firm provided a fixed contribution to employees to subsidize all benefit choices, including health insurance. Health insurance benefit packages were standardized across plans.  The firm created a HEDIS-based health plan 'report card' that was disseminated to non-union employees for the first time during the 1997 open enrollment period.  By using longitudinal data to examine the impact of the report card on plan enrollments while controlling for out-of-pocket price, this paper empirically assesses the impact of report card information on plan enrollment.

## II. Existing Literature

Although a substantial existing literature examines factors that influence health insurance enrollment, much of this literature pertains to indemnity plan enrollment and considers out-of-pocket price, deductibles, co-insurance rates, and benefit structure as the key determinants of plan choice (Scanlon, Chernew, and Lave, 1997). This literature consistently finds an inverse relationship between plan price and plan enrollment, congruent with economic theory.

Feldman et al. (1989) include a measure of the ability of enrollees to choose their physician in their model of plan choice, capturing an important aspect of how managed care plans differentiate themselves from one another. Other studies, such as Buchmueller and Feldstein (1996), also recognize the importance of the type of health plan in influencing plan choice.

Chernew and Scanlon (1998) examine the cross-sectional relationship between report card ratings of plan performance and health plan choice in a setting where employees were given report cards during open enrollment. The cross-sectional design prevents identification of the role of report card information per se, because individuals may have been aware of plan performance in the absence of the report cards. The authors are unable to detect the hypothesized relationship between the report card ratings and enrollment.  In several instances Chernew and Scanlon find counterintuitive relationships between ratings and enrollment that may be the result of correlation between the ratings and unmeasured plan attributes.  A follow-up study, using the detailed performance measures, yields results consistent with the hypotheses regarding correlation between the ratings and unobserved plan traits (Scanlon and Chernew, 1999).  The authors conclude that unobserved plan traits are probably important determinants of plan choice. The omission of important plan traits from the analysis could bias coefficient estimates due to the correlation between the unobserved attributes and the ratings.  Moreover, there is likely a correlation among the health plan choices of individuals in the same market. Thus analyses conducted at the individual level will underestimate the standard errors if independence of the individual observations is falsely assumed (Moulton, 1986).

Farley et al. (1999) study the impact of report card information on plan choice among Medicaid enrollees.  After randomly distributing a CAHPS-based Medicaid managed care report card, these authors examine the impact of the report card on

enrollment for new Medicaid beneficiaries in New Jersey. The authors find that the report card had little effect on enrollment patterns in aggregate. When survey results are examined to identify report card 'users,' they find a stronger link between enrollment and the ratings for this sub-sample. The commercially insured population may differ from the Medicaid population in their response to report cards for a variety of reasons. In general commercially insured individuals are better educated, which may increase their responsiveness to information, but they may also be more informed from other sources and thus less likely to respond to health plan report card data.

## III.    Experimental Setting and Data

### *Experimental Setting*

Our analysis is based on the health plan choices of employees at GM for two open enrollment periods, 1996 and 1997. GM employed a flexible benefits system in which employees and retirees received a fixed amount of 'flex dollars' that could be allocated across several benefit categories (e.g., health insurance, life insurance, disability insurance, and dental insurance). Within each benefit category there were various options, each with firm specified prices. If the cost of one's total benefit elections exceeded one's allotted flex benefit dollars, the difference was paid out of pocket. If the total cost was less than the allotted amount, the difference was received as taxable income. The firm determined the set of health insurance plans from which enrollees could choose as well as the prices they were charged for each plan. We define the price as the employee out-of-pocket premium.

During the 1997 open enrollment period, which occurred in the fall of 1996, the firm, for the first time, provided health plan performance ratings for all available HMOs to non-union employees as part of the open enrollment materials. The performance ratings were based on aggregated HEDIS data. No performance ratings were released for traditional fee-for-service (FFS) or preferred provider (PPO) plans. The release of 'report card' information provides the fundamental natural experiment that is the foundation of the analyses reported in this paper.

Our analysis identifies the impact of the report card on plan choice using the observed health plan choices both before and after the firm released plan ratings to non-union employees. Fortunately, the set of health plans offered was very stable during the

study period.  Because GM has employees in many markets, this is analogous to analyzing many different natural experiments. Because the union employees were not provided the report card for 1997 open enrollment, we also use changes in union enrollment patterns to control for unobserved, time-varying plan factors that may have affected health plan choice (e.g., changes in provider panels).

### *HMO Ratings Developed by GM*

HMOs were rated 'below expected performance,' 'average performance,' or 'superior performance' along six domains, labeled by GM as:

- preventive health care services
- medical and surgical care
- women's health issues
- access to care
- patient satisfaction
- operational performance

The ratings for one domain, operational performance, were based on plan site visits by GM staff.  The other five ratings were based on HEDIS results.  The HEDIS measures comprising the five domains are listed in Table 1.  The firm used individual HEDIS measures to create ratings via a Z-score methodology described in detail in Appendix A.  In addition to these HEDIS-based ratings, GM also reported whether the National Committee for Quality Assurance (NCQA) accredited each HMO, and whether GM considered the HMO a 'benchmark' plan.[1]

During the open enrollment period in 1996 (for 1997 enrollment), non-union employees were given an information sheet on each of the HMOs from which they could choose.  The information sheet designated each plan as 1, 2, or 3 diamonds for each domain to represent the plan ratings.  An example is provided in Figure 1, though no employee received the exact sheet represented in Figure 1 because no employees were offered all of the plans displayed.

### *Analytic Sample*

The firm covered over 1.6 million active employees, retirees, and dependents. The analysis is based on the health plan choices of the approximately 96,000 active

---

[1] GM considered a combination of factors in choosing which plans were labeled benchmark plans.  These factors included premiums, quality, and geographic location of the HMO.  Although the designation of

employees based in the U.S. that chose HMO coverage (Table 2). About 29,000 of these employees were salaried (non-union) and thus were given the report card information. Dependents were not analyzed separately because they almost always made the same choice as the employee with GM coverage eligibility. Retirees were excluded because they are frequently Medicare-eligible, making the nature of plan choice different than for the non-Medicare population. Employees who enrolled in FFS or preferred provider organizations (PPOs) were excluded because ratings were not provided for these plans.[2] Hence, the analysis pertains to the impact of report card information on HMO choice, conditional on HMO enrollment. We also excluded plans with zero enrollments in either year. In most cases, this would be due to offering of plans that were not realistic choices. In a few cases, plans that were not offered in one of the years were dropped from the analysis. Approximately 27,000 employees in HMOs included in our study. About 25% of the 1997 employees enrolled in HMOs were either in a different HMO, enrolled in another plan type, or not receiving benefits from GM in 1996.

For our analysis, employees are assigned to markets based on the set of health plans from which they could choose. All employees that share a common set of plan choices are grouped into the same geographic market. The firm determined the set of plans from which each employee could choose based on the employee's zip code of residence.

Markets are mutually exclusive, but plans may serve multiple markets. For example, plan A could be offered in San Francisco and south to Santa Cruz, and plan B could be offered in San Francisco and north through Marin County. This would result in three markets. Market 1 would represent Santa Cruz, with only plan A offered. Market 2 would represent Marin County, with only plan B offered. Market 3 would be San Francisco, with both plans offered. Other plans may serve only one market.

In addition to the market distinction, employees within a market could chose from four different coverage categories: single, employee and spouse, employee and children, and employee and family. Coverage category could affect plan preferences in a variety of ways. For example, employees with children may be more interested in the set of available pediatricians and plan performance in the area of pediatric care.

---

benchmark was based on these factors, the final decision was determined by a 'qualitative' judgment rather than a score resulting from a numerical algorithm.
[2] Plan ratings were not provided for FFS or PPO plans because the HEDIS data that were used to construct the ratings were collected only for HMOs.

For our analysis we define a 'cell' as a particular market/coverage category combination. After excluding cells with fewer than 5 employees and markets with only 1 plan, we have observations on 69 plans spread across 183 market/coverage category cells. On average, cells have 3.19 plans (minimum=2, maximum=6). In 1996, the mean number of choosers per cell was 178.78 (minimum = 5, maximum = 3,299). In 1997, the cells were slightly larger (mean=188.29, minimum=5, maximum=3,462). Descriptive statistics for the 69 plans are in Table 3. The descriptive statistics on annual price reported in this table reflect the difference between the out-of-pocket price and the allotted flex dollars to standardize across coverage categories.

## IV. Econometric Methods

Let $U_{ijct}$ represent person $i$'s utility from health plan $j$ in market/coverage cell $c$ in year $t$. This utility depends partly on observable characteristics of the health plan, such as its out-of-pocket premium and its report card ratings. It also depends on other characteristics that researchers do not observe, such as the popularity of physicians in the health plan's provider network, the convenience associated with using the plan providers, confidence that the plan will approve requests for care in unusual circumstances, and amenities of the hospitals affiliated with the plan. These unmeasured variables may differ by market (as in the case of provider location or quality) or coverage category (as in the case of quality of pediatric services). Individuals may observe these attributes through interactions with family, friends, co-workers, and health care professionals, regardless of whether the firm reports plan performance measures. Finally, different individuals have different valuations of the same plan because of idiosyncratic differences in individual preferences or circumstances.

We formalize these considerations with the random-utility model

$$(1) \qquad U_{ijct} = \boldsymbol{b}'X_{jct} + \boldsymbol{g}_{jc} + \boldsymbol{e}_{ijct}.$$

The vector $X_{jct}$ contains the measured health plan characteristics, and $\boldsymbol{b}$ is the associated coefficient vector. The $\boldsymbol{g}$ term represents the average individual's valuation of the unmeasured characteristics of the plan. The error term $\boldsymbol{e}_{ijct}$ reflects the individual's idiosyncratic deviation from that average valuation.

6

Each individual chooses the health plan that maximizes his or her utility. Therefore, given the $X$'s and the $\boldsymbol{g}$'s, the probability $\boldsymbol{p}_{jct}$ that an individual in cell $c$ chooses plan $j$ in year $t$ is

(2) $\quad \boldsymbol{p}_{jct} = \text{Prob}(U_{ijct} > U_{ikct})$ for all $k \neq j$

$\quad\quad\quad = \text{Prob}[\boldsymbol{e}_{ijct} - \boldsymbol{e}_{ikct} > \boldsymbol{b}'(X_{kct} - X_{jct}) + (\boldsymbol{g}_{kc} - \boldsymbol{g}_{jc})]$ for all $k \neq j$.

Specifying a functional form for the choice probability $\boldsymbol{p}_{jct}$ requires a distributional assumption for the $\boldsymbol{e}$'s. Following McFadden (1973, 1974), if the $\boldsymbol{e}$'s in year $t$ follow independent Type I extreme value distributions,[3] then $\boldsymbol{p}_{jct}$ takes the logit form

(3) $\quad \boldsymbol{p}_{jct} = \exp(\boldsymbol{b}'X_{jct} + \boldsymbol{g}_{jc}) / D_{ct}$

where $D_{ct} = \sum_{k=1}^{N_c} \exp(\boldsymbol{b}'X_{kct} + \boldsymbol{g}_{kc})$ and $N_c$ is the number of plans offered in cell $c$.

If all the $\boldsymbol{g}$'s were zero – that is, if unmeasured plan traits were of absolutely no consequence – then the model in equation (3) would simplify to the standard conditional logit model, which could be estimated by applying the conventional maximum likelihood estimator to individual-level data. This approach is common in the existing literature (Short and Taylor, 1989; Feldman et al., 1989; Garnick et al., 1989). But the assumption that consumers are indifferent to unmeasured factors, such as popularity of physicians and locational convenience, is quite implausible, and false imposition of this assumption creates two serious econometric problems. First, assuming the $\boldsymbol{g}$'s are zero when they really are not overlooks correlation across individual observations in the same cell. This oversight can lead to gross underestimation of standard errors (Moulton, 1986). Second, and more importantly, if some of the unobservables underlying $\boldsymbol{g}_{jc}$ are correlated with

---

[3] The assumption that the $\boldsymbol{e}$'s follow independent Type I extreme value distributions imposes independence of irrelevant alternatives (IIA). IIA implies that the probability of choosing one plan relative to a second plan is unaffected by changes in attributes of other plans in the individual's choice set. If some plans are closer substitutes to a given plan than others, IIA will be violated.

some of the observables in $X_{jct}$, failure to control for those unobservables generates an omitted-variables inconsistency in the estimation of $\boldsymbol{b}$.

To avoid these problems, we employ a methodology that does account for the $\boldsymbol{g}$'s.[4] Let $S_{jct}$ denote the market share of plan $j$ in market/coverage cell $c$ in year $t$, which is simply the fraction of sample individuals in the cell that choose that plan. The expected value of $S_{jct}$ is the choice probability $\boldsymbol{p}_{jct}$, but with finite samples the realized value of $S_{jct}$ deviates from $\boldsymbol{p}_{jct}$ because of random sampling error. Thus,

$$
(4) \qquad S_{jct} = \boldsymbol{p}_{jct} + v_{jct}
$$
$$
= \exp(\boldsymbol{b'}X_{jct} + \boldsymbol{g}_{jc})/D_{ct} + v_{jct}
$$

where $E(v_{jct}) = 0$ and $Var(v_{jct})$ varies inversely with the number of sample individuals in the cell. Taking natural logarithms yields

$$
(5) \qquad \ln(S_{jct}) = \ln[\exp(\boldsymbol{b'}X_{jct} + \boldsymbol{g}_{jc}) + D_{ct}v_{jct}] - \ln(D_{ct}),
$$

and a first-order Taylor series expansion of equation (5) around $v_{jct} = 0$ yields

$$
(6) \qquad \ln(S_{jct}) \cong \boldsymbol{b'}X_{jct} + \boldsymbol{g}_{jc} - \ln(D_{ct}) + v_{jct}/\boldsymbol{p}_{jct}.
$$

It follows that the difference between the log market share of plan $j$ and that of an arbitrarily selected reference plan $r$ in the same cell is

$$
(7) \qquad \ln(S_{jct}) - \ln(S_{rct}) \cong \boldsymbol{b'}(X_{jct} - X_{rct}) + (\boldsymbol{g}_{jc} - \boldsymbol{g}_{rc}) + (v_{jct}/\boldsymbol{p}_{jct} - v_{rct}/\boldsymbol{p}_{rct}).
$$

Inspection of equation (7) makes clear that, if one were to perform least squares estimation of the cross-sectional regression of the difference in log market shares on the differences in measured plan traits, the failure to control for the difference in the unobserved $\boldsymbol{g}$'s would generate an omitted-variables bias in the estimation of $\boldsymbol{b}$.

8

Fortunately, our access to longitudinal data from both 1996 and 1997 enables us to "difference out" the $\boldsymbol{g}$'s. Differencing equation (7) between the two years yields

$$(8) \qquad \Delta \ln( S_{jc}) - \Delta \ln( S_{rc}) \cong \boldsymbol{b}'(\Delta X_{jc} - \Delta X_{rc}) + [\Delta(v_{jc}/\boldsymbol{p}_{jc}) - \Delta(v_{rc}/\boldsymbol{p}_{rc})]$$

where the $\Delta$ notation denotes a change from 1996 to 1997.

Least squares estimation of equation (8) escapes the omitted-variables bias because the longitudinal differencing causes the $\boldsymbol{g}$'s to drop out of the equation. While the effects of observable, but time-invariant plan traits cannot be identified in this approach, we can identify the effects of two key types of traits – price and report card ratings. The out-of-pocket premia for different plans did change differentially between the two years, so we identify the price sensitivity of consumer choices off of the association between year-to-year changes in market shares and year-to-year changes in relative premia. GM did not provide report card ratings to its employees in 1996, but it introduced that information in 1997, so we identify the impact of report card ratings by relating the 1996-to-1997 changes in market shares to differences between plans in their newly introduced report card ratings.

With $N_c$ plans offered in cell $c$ and one plan used as the reference plan, cell $c$ contributes $N_c - 1$ observations to the regression sample. Ordinary least squares estimation of equation (8) would be inefficient because the error term is nonspherical. The error term is heteroskedastic because $v$ is inversely related to the number of individuals within the cell, so we weight each observation by the square root of the number of employees in that cell in 1997. We also perform a generalized-least-squares correction that accounts for five varieties of correlation across observations: (1) between observations for the same plan but different coverage categories in the same cell; (2) between observations for different plans in the same cell, which share the same reference plan; (3) between observations for the same plan in different cells; (4) between observations for different plans in different cells that share the same reference plan; and (5) between observations in different cells where the same plan is the reference plan in one cell and a plan observation in the other cell.[5]

---

[4] This approach is similar to that of Berry (1994).
[5] We estimated the model using the REML variation of the PROC MIXED procedure in SAS.

Our claim that our longitudinal estimation approach avoids omitted-variables bias depends on our assumption that the plan/cell unobservables represented by the $\boldsymbol{g}$'s are time-invariant. What if these "fixed effects" are not really fixed, i.e., what if there are important changes between 1996 and 1997 in factors like a plan's locational convenience and the popularity of its physicians? Then equation (8) changes to

$$(9) \qquad \Delta \ln( S_{jc} ) - \Delta \ln( S_{rc} ) \cong \boldsymbol{b}'(\Delta X_{jc} - \Delta X_{rc} ) + (\Delta \boldsymbol{g}_{jc} - \Delta \boldsymbol{g}_{rc} )$$
$$+ [\Delta(v_{jc} / \boldsymbol{p}_{jc} ) - \Delta(v_{rc} / \boldsymbol{p}_{rc} )] .$$

If the changes in the $\boldsymbol{g}$'s are correlated with the changes in the $X$'s, then our estimation may be subject to omitted-variables bias after all.

To treat this possibility, we will perform a supplementary analysis that exploits additional information on the plan choices of *union* employees. Report card ratings were not provided to union employees, and the out-of-pocket premia for union employees were zero in both 1996 and 1997. Therefore, if the choices of union employees obey the same model we have specified for non-union employees, the union version of equation (9) is

$$(10) \qquad \Delta \ln( S_{jc}^{u} ) - \Delta \ln( S_{rc}^{u} ) \cong (\Delta \boldsymbol{g}_{jc}^{u} - \Delta \boldsymbol{g}_{rc} ) + [\Delta(v_{jc}^{u} / \boldsymbol{p}_{jc}^{u} ) - \Delta(v_{rc}^{u} / \boldsymbol{p}_{rc}^{u} )]$$

where the $u$ superscript signifies a union variable. The union counterpart to the dependent variable in equation (9) therefore can be viewed as a proxy for the $(\Delta \boldsymbol{g}_{jc} - \Delta \boldsymbol{g}_{rc} )$ term we wish to control for in equation (9). It differs from that term only because of the sampling error involving the $v$'s. Unfortunately, the union information we have is not as finely detailed as our information on non-union employees. Our market-share information on union employees consists of state-level data aggregated over all coverage categories and markets. Nevertheless, as a step towards attempting to control for time-varying unobservables, we will try using that aggregated version of the left side of equation (10) as an additional control variable in our regression model for the changes in relative market shares among non-union employees.

## V.      Results

The first column of Table 4 reports the results for the base model.  Consistent with economic theory and the vast majority of existing literature, out-of-pocket price is inversely related to enrollment.  The magnitude of the estimated coefficient suggests that a 10% increase in relative price would generate approximately a 3% decrease in relative market share.  Accreditation status is estimated to be positively associated with enrollment changes. Plans that became NCQA-accredited gained relative market share.

The results for the rating variables are equivocal.  Of the twelve estimated coefficients on the superior or below-average ratings, only seven are of the hypothesized sign.  Of the six domains of performance, only two, women's health and access to care, have positive estimated coefficients on the superior rating and negative estimated coefficients on the below-average rating.  None of those estimated coefficients is statistically significant.  Two of the twelve relevant coefficient estimates are statistically significant, but neither has the hypothesized sign.  Despite the lack of significance of most rating coefficient estimates, the hypothesis that all rating coefficients equal zero can be rejected at $p < 0.01$.  The conclusions are robust to several specifications of the random effects.

To test whether time-varying factors might be influencing the results, we included the change in union share variable to capture unmeasured, time-varying plan effects (Table 4, column 2).  The estimated coefficient of this variable had the hypothesized sign, but was not statistically significant. Perhaps this was because unobserved plan traits were relatively stable over this two-year study period or perhaps the estimate reflects downward bias due to measurement error for the union market share. Regardless of the reason why the estimated coefficient on the change in union share is not statistically significant, its inclusion does not alter any of the conclusions regarding the ratings; they remain equivocal with frequently counterintuitive signs.

Two specifications explored the sensitivity of the results to outliers.  First, we raised the size of the cell required for inclusion in the model from 5 to 10 employees (Table 4, column 3).  Market shares in small markets may be relatively unstable so this restriction should eliminate some noise in the data.  The results from this exercise reveal no substantive difference from the base results. The estimated coefficients on price and accreditation are virtually unchanged and the estimated coefficients on the rating variables remain equivocal.

Second, we omitted outliers from the base model (Table 4, column 4). Outliers were defined as observations with studentized residuals greater than 2 (Belsley, Kuh & Welsch, 1980).[6] The estimated coefficient on price remains stable. However, the estimated coefficient on the accreditation variable, though still positive, drops about 70 percent and is no longer statistically significant. Although several estimated rating coefficients switch signs in this specification, and several lose statistical significance, the qualitative conclusions regarding the ratings remain unchanged. The estimated coefficients frequently have counterintuitive signs and only one domain, operational performance, has both coefficients estimated with the hypothesized sign.

Taken as a group, the results reported in Table 4 do not provide support for the notion that employees responded to the report card ratings. Most of the estimated coefficients on ratings were not statistically significant. Those that were significant often exhibited counterintuitive signs.

Table 5 reports results from specifications that aggregate the rating variables into summary measures. There are several reasons why aggregation might be useful. First, individuals may not be able to process information from all the domains. They may adopt simplifying decision rules such as focusing only on selected domains, or selecting plans with the most superior ratings or fewest below-average ratings (Hibbard, 1997).

Second, the number of parameters in the base model is relatively large compared to the sample size, and there is some collinearity among the explanatory variables in the base specification. The counterintuitive signs and general lack of statistical significance could reflect these issues.

The first two columns of Table 5 report specifications that include only the superior or below-average ratings. Both specifications are easily rejected relative to the base model (p<0.01). Nevertheless, the estimated price coefficient remains negative and statistically significant, though in the specification with only below-average ratings its magnitude increases by about a third. The estimated accreditation coefficient remains positive, but is much smaller than in the base model and not statistically significant.

The estimated coefficients on the rating variables when only superior ratings are included are of the hypothesized sign in five of six cases, though none is statistically significant. When only below-average ratings are included, the estimated coefficients

---

[6] Outliers were identified using the studentized residual in OLS estimation of the model.

have counterintuitive signs in three of the six cases, though the one statistically significant coefficient estimate (women's health) is of the hypothesized negative sign.

The third column of Table 5 imposes the restriction that employees respond to the sum of superior ratings and the sum of below-average ratings. This set of restrictions is rejected, relative to the base model, at the 0.10 level but not at the 0.05 level. The estimated price coefficient remains statistically significant and about -0.3, and the estimated accreditation coefficient remains positive, though not statistically significant. In this specification, both of the estimated rating coefficients have the hypothesized sign. The estimated coefficient on the sum of below-average ratings is negative and statistically significant, and the estimated coefficient on the sum of superior ratings is positive, though not statistically significant. Though this model is rejected at $p<0.10$ relative to the base model, it provides the strongest support for the impact of ratings on enrollment.

An alternative approach for limiting the parameters in the model is to selectively exclude various domains. Domains excluded were selected in a stepwise fashion. Exclusion of each domain was tested relative to the base model, and 'Prevention' was the domain whose exclusion had the least effect on the likelihood function. The hypothesis that all of the prevention coefficients were zero could not be rejected ($p=0.39$). With 'Prevention' excluded, we sequentially tested exclusion of each of the remaining five domains against the base model and against the model with prevention ratings excluded. The exclusion of 'Access to Care' had the smallest impact on the likelihood function, and the hypothesis that all the 'Prevention' and 'Access' coefficients equal zero could not be rejected ($p=0.197$ vs. base). An analogous process led to the further, sequential exclusion of the 'Satisfaction' domain ($p=0.182$ vs. base), 'Operations' ($p=0.106$ vs. base), and 'Women's Health' ($p=0.124$ vs. base). This left only 'Medical/Surgical Care' in the model.

The final column of Table 5 reports results excluding all domains except 'Medical/ Surgical Care.' Given the multicollinearity, these estimated coefficients are capturing any unmeasured effects from the omitted domains. As with all other models, the estimated price coefficient remains negative and statistically significant, with a magnitude of approximately -0.3. The accreditation coefficient estimate remains positive, but not statistically significant. The model provides some support for the hypothesis that ratings matter because the estimated coefficient on the below-average

rating is negative and statistically significant. However, the coefficient on the superior rating is also negative, though smaller in absolute value, and statistically significant, at least at the p<0.10 level.  Similar qualitative conclusions would be drawn from any of the estimates based on selectively omitted domains: some coefficient estimates support the hypothesis that ratings matter, but others have statistically significant counterintuitive signs.

## VI.      Discussion

Considerable resources have been devoted to collecting health plan performance measures.  Many organizations, including GM, have spent substantial additional resources to construct and disseminate health plan report cards based on that information. Evidence from focus groups and surveys regarding whether employees would use report card data in making their plan selections is mixed. Several studies report interest in measures of access, satisfaction, technical quality, and use of preventive services (Robinson and Brodie, 1997; Tumlinson et al., 1997; Hibbard and Jewitt, 1996).

Despite interest in such information, some focus-group evidence suggests that most employees do not use the information when it is provided (Robinson and Brodie, 1997; Meyer et al., 1998).  The work by Meyer et al. is particularly pertinent because the focus groups in that study, though small relative to our sample size, were conducted in 1997 using salaried employees at GM.  Evidence from focus groups should be supplemented with evidence based on actual plan choices because researchers have found a discrepancy between what individuals say and what they do (Hibbard and Jewitt, 1996).

The findings from this work confirm that price is inversely related to plan choice. The estimated price coefficient is very stable across specifications.  All specifications also indicate that NCQA accreditation is positively related to enrollment.  Other research suggests that employees are distrustful of employer-sponsored information.  The positive relationship between enrollment and accreditation could reflect greater trust in information generated by organizations other than the firm, such as NCQA.  However, the magnitude of the estimated accreditation effect varies dramatically by specification and is not statistically significant in models with less than the full set of rating variables.

The results regarding the impact of ratings on plan enrollment are equivocal. Estimates that relate plan enrollment to ratings for each domain (Table 4) or to subsets of domains (Table 5, column 4) reject the hypothesis that all of the ratings coefficients equal

zero. Yet most of the estimated coefficients are statistically insignificant and they often have counterintuitive signs. Given the multicollinearity in the rating data and our sample size, we are unable to achieve precise identification of the effects of ratings for specific domains.

The specification most favorable to rating effects includes only aggregated ratings variables, including the count of superior ratings and the count of below-average ratings. This specification suggests that individuals avoided plans with a large number of below-average ratings. This result is consistent with recent findings from focus groups that individuals are more likely to avoid bad plans than select good ones (Hibbard et al., 2000).

Given the instability of estimated coefficients across specifications, one should be wary of over-interpreting the findings. The restrictions on parameters that generate the most favorable specification for the ratings are rejected at $p<0.10$, suggesting that more information is contained in the ratings than we have identified.

Several limitations are worth noting. First, in our base specification, the ratio of observations to parameters is relatively small. This is because we aggregate the data to market/coverage category cells. The aggregation does not discard information because our model includes only plan attributes and the aggregation is important because of salient unobserved plan traits. Yet this process illustrates that the ability to investigate the questions we pose requires more than simply a large number of employees (we had 27,000 non-union employees in our sample). It also requires many plan/market combinations. The large number of employees helps, of course, by increasing the extent to which the observed market shares approximate what would be the true market shares with infinite cell sizes. In short, the sample size within each cell and the number of cells relative to parameters are both important.

As the number of cells decreases, time varying unobserved plan-specific factors could increasingly influence the findings. To some extent our inclusion of changes in the union shares will capture some of these unobserved, time-varying factors. Yet this control is imperfect both because our union data are at the state, not market level, and union workers could react differently than salaried workers. Hence, some of the counterintuitive signs likely reflect the low ratio of observations to parameters.

Second, it may be the case that certain subsets of employees are more influenced by the ratings than others. Our estimated coefficients could be dampened by substantial

inertia among employees. However, most report card efforts, including competitive model proposals, are aimed at similar audiences that are likely to have quite a bit of inertia. Moreover, this inertia does not prevent identification of hypothesized effects for the price variable. In addition, earlier work that examined the behavior of new hires in a cross-sectional setting found little difference in the conclusions relative to the full sample of workers (Chernew and Scanlon, 1998). Many other studies also focus on existing workers (Feldman et al., 1989; Buchmueller and Feldstein, 1996). Nevertheless, further work examining subgroups would be valuable.

Finally, the efforts to construct and disseminate report cards may have merit beyond influencing employee behavior. The plan performance information underlying report cards may be important for employers in selecting plans with which to contract, and report cards may push plans to improve their performance. Moreover, as employees become more familiar with report cards, the impact of such information may grow. Given our findings, continued study of how health plan ratings influence markets is important for assessing the responsiveness of consumers to commonly used plan performance measures.

**Figure 1**

**Example Information Sheet**

## COMPARING YOUR 1997 GM MEDICAL OPTIONS

The following table shows the rating of the HMO option(s) available in eight selected quality measures. The ratings are based on historical data and therefore may not necessarily represent the quality of care you will receive in the future. GM does not endorse or recommend any particular medical plan option. The medical plan you elect is your personal decision.

For a more complete description of the eight selected quality measures, see the *GM Medical Plan Guide*.

| | NCQA Accredited? | Benchmark HMO? | Operational Performance | Preventive Care | Medical/ Surgical Care | Women's Health | Access to Care | Patient Satisfaction |
|---|---|---|---|---|---|---|---|---|
| 0001 Basic Medical Plan | Information Currently Not Available | | | | | | | |
| 0002 Enhanced Medical Plan | Information Currently Not Available | | | | | | | |
| PPO 2190 Blue Preferred Plus | Information Currently Not Available | | | | | | | |
| HMO 2101 Care Choices HMO | Yes | No | ▲▲ | ▲▲▲ | ▲▲ | ▲▲▲ | ▲▲ | ▲▲▲ |
| HMO 2108 BCN G Lakes SW MI | Yes | No | ▲▲ | ▲ | ▲ | ▲ | ▲ | ▲▲ |
| HMO 2113 BCN Health Central | Yes | No | ▲ | ▲▲▲ | ▲▲▲ | ▲▲▲ | ▲ | ▲▲ |
| HMO 2116 Priority Health | Yes | No | ▲▲ | ▲ | ▲ | ▲▲ | ▲▲▲ | ▲▲▲ |
| HMO 2117 Care Choices HMO W MI | Yes | No | ▲▲ | ▲▲▲ | ▲▲ | ▲▲▲ | ▲▲ | ▲▲▲ |
| HMO 2118 BCN West Michigan | Yes | No | ▲▲ | ▲ | ▲ | ▲ | ▲ | ▲▲ |

Key: ▲ =below expected performance   ▲▲ =average performance   ▲▲▲=superior performance
ND=no data was available from this plan

Michigan - Lansing and West

**Table 1**

**Health Plan Performance Domains and Their Measures**

| Domain | Measures |
|---|---|
| **Prevention** | Childhood Immunization Rate |
| | Cholesterol Screening Rate |
| | Prenatal Visit Rate in the First Trimester* |
| | Cervical Cancer Screening Rate* |
| | Diabetic Retinal Examination Rate |
| | Mammography Rate* |
| | |
| **Medical/Surgical Care** | C-Section Rate |
| | Cardiac Catheterization Rate |
| | Coronary Artery Bypass Graft Rate |
| | Coronary Angioplasty (PTCA) Rate |
| | Laminectomy Rate |
| | Prostatectomy Rate |
| | Inpatient Admission Rate for Asthma |
| | |
| **Women's Care** | Prenatal Visit Rate in the First Trimester |
| | Cervical Cancer Screening Rate |
| | Hysterectomy Rate |
| | Mammography Rate |
| | |
| **Access** | Follow-up after a Major Mental Health Disorder |
| | Inpatient Readmission Rate |
| | Open Panel |
| | Primary Care Physician Turnover Rate |
| | Percent of Enrollees with a Primary Care Visit |
| | |
| **Satisfaction** | Enrollee Satisfaction Survey ** |

* Less weight was placed on the Prenatal Visit Rate, Cervical Cancer Screening Rate and Mammography Rate in the prevention domain because these measures were also used in the women's care domain.

** The satisfaction rating was modified to reflect GM preference for large sample sizes and phone vs. mail administration of the survey.

**Table 2**

**Active Employees by Plan Type (1997)**
**(rounded to nearest 1000)**

|  | HMO | PPO | FFS | Total |
|---|---|---|---|---|
| **Salary (non-union)** | 29,000 | 11,000 | 32,000 | 72,000 |
| **Union (hourly)** | 67,000 | 72,000 | 90,000 | 229,000 |
| **Total Active** | 96,000 | 83,000 | 122,000 | 301,000 |

**Table 3**

**Descriptive Statistics: Means and Frequencies**

|  | N | Mean | Std. | Min | Max |
|---|---|---|---|---|---|
| Annual Price* (1996) Single Coverage | 66[+] | 434.36 | 163.90 | 84 | 708 |
| Annual Price* (1996) Family Coverage | 69[+] | 1213.91 | 463.19 | 240 | 1956 |
| Annual Price* (1997) Single Coverage | 66[+] | 446.91 | 184.40 | 108 | 732 |
| Annual Price* (1997) Family Coverage | 69[+] | 1238.55 | 499.52 | 300 | 2004 |
| NCQA Accredited | 69 | .70 |  |  |  |
| Benchmark Status | 69 | .15 |  |  |  |
|  | N | Superior | Average | Below Average | No Data |
| Operational Performance | 69 | 29 | 22 | 18 | 0 |
| Preventive Care | 69 | 27 | 19 | 15 | 8 |
| Medical and Surgical Care | 69 | 18 | 26 | 18 | 7 |
| Access | 69 | 27 | 18 | 17 | 7 |
| Women's Health | 69 | 24 | 23 | 20 | 2 |
| Patient Satisfaction | 69 | 27 | 17 | 15 | 10 |

+ The sample sizes vary for price because markets with less than five employees were excluded from the regression analysis.

* The annual prices reported in Table 3 reflect the difference between the out-of-pocket price and the allotted flex dollars in order to standardize price across coverage categories.

**Table 4**
**Estimates of Model in Equation (8)**
**(t-stats in parentheses)**

|  |  | Base Model | Base with Union | Excluding Cells<10 | Omitting Outliers |
|---|---|---|---|---|---|
| Ln(Price) |  | -0.290 ** | -0.275 * | -0.291 ** | -0.309 *** |
|  |  | (-2.15) | (-1.90) | (-2.10) | (-2.89) |
| Union Share |  |  | 0.04519 |  |  |
|  |  |  | (0.57) |  |  |
| Accreditation |  | 0.761 *** | 0.7656 *** | 0.770 *** | 0.209 |
|  |  | (3.22) | (2.70) | (3.21) | (1.50) |
| Operational Performance | Superior | 0.307 | 0.2372 | 0.318 | 0.117 |
|  |  | (1.51) | (1.00) | (1.52) | (0.87) |
|  | Below Average | 0.707 *** | 0.6171 *** | 0.734 *** | -0.065 |
|  |  | (3.29) | (2.34) | (3.28) | -(0.48) |
| Preventive Care | Superior | 0.051 | 0.07794 | 0.030 | -0.013 |
|  |  | (0.22) | (0.28) | (0.13) | -(0.08) |
|  | Below Average | 0.164 | 0.08059 | 0.150 | -0.036 |
|  |  | (0.65) | (0.28) | (0.58) | -(0.25) |
| Medical/ Surgical Care | Superior | -0.472 ** | -0.4844 *** | -0.443 ** | -0.150 |
|  |  | (-2.51) | (-2.39) | (-2.27) | (-1.18) |
|  | Below Average | -0.282 | -0.3405 | -0.249 | -0.183 |
|  |  | (-1.19) | (-1.24) | (-1.02) | (-1.33) |
| Women's Health | Superior | 0.389 | 0.3912 | 0.405 | 0.293 |
|  |  | (1.31) | (1.12) | 1.34 | (1.42) |
|  | Below Average | -0.257 | -0.1802 | -0.253 | 0.074 |
|  |  | (-1.21) | (-0.69) | (-1.17) | (0.51) |
| Access to Care | Superior | 0.042 | -0.00562 | 0.038 | -0.123 |
|  |  | (0.24) | (-0.03) | (0.21) | -(1.38) |
|  | Below Average | -0.279 | -0.2591 | -0.289 | -0.205 |
|  |  | (-1.14) | (-0.96) | (-1.16) | (-1.50) |
| Patient Satisfaction | Superior | -0.342 | -0.3530 | -0.324 | -0.216 |
|  |  | (-1.60) | (-1.37) | (-1.50) | (-1.77) * |
|  | Below Average | 0.069 | 0.02748 | 0.108 | -0.178 |
|  |  | (0.29) | (0.10) | (0.44) | (-1.10) |
|  | "Missing data" dummy variables included | yes | yes | yes | yes |
|  | N | 274 | 274 | 242 | 264 |

*significant at p=0.10,   ** significant at p=0.05,   *** significant at p=0.01

21

**Table 5**

**Restricted Models**
**(t-stats in parentheses)**

| | | Only Superior | Only Below Average | Summing Ratings | Excluding Domains |
|---|---|---|---|---|---|
| Ln(Price) | | -0.282 ** | -0.396 *** | -0.299 ** | -0.338 |
| | | (-2.14) | (-3.37) | (-2.44) | (-3.13) *** |
| Accreditation | | 0.2144 | 0.241 | 0.062 | 0.236 |
| | | (1.16) | (1.13) | (0.38) | (1.51) |
| Operational Performance | Superior | 0.032 | | | |
| | | (0.19) | | | |
| | Below Average | | 0.161 | | |
| | | | (0.85) | | |
| Preventive Care | Superior | 0.153 | | | |
| | | (0.75) | | | |
| | Below Average | | -0.006 | | |
| | | | (-0.03) | | |
| Medical/Surgical Care | Superior | -0.030 | | | -0.280 |
| | | (-0.20) | | | (-1.93) * |
| | Below Average | | 0.012 | | -.4695 |
| | | | (0.07) | | (-2.86) *** |
| Women's Health | Superior | 0.412 | | | |
| | | (1.62) | | | |
| | Below Average | | -0.357 | | |
| | | | (-2.25) ** | | |
| Access to Care | Superior | 0.081 | | | |
| | | (0.58) | | | |
| | Below Average | | -0.216 | | |
| | | | (-1.29) | | |
| Patient Satisfaction | Superior | 0.037 | | | |
| | | (0.25) | | | |
| | Below Average | | 0.012 | | |
| | | | (0.06) | | |
| Number of Superior Ratings | | | | 0.013 | |
| | | | | (.21) | |
| Number of Below Average Ratings | | | | -0.126 | |
| | | | | (-2.29) ** | |
| Missing Data Dummies | | No | No | Summed | For Only Med/Surg |
| | N | 274 | 274 | 274 | 274 |
| | # of Restrictions | 11 | 11 | 14 | 14 |
| | p-value (vs. base) | 0.002 | 0.001 | 0.071 | 0.124 |

*significant at p=0.10,   ** significant at p=0.05,   *** significant at p=0.01

## Appendix A:  Construction of Plan Ratings

Ratings for all of the domains except operational performance were based on a subset of HEDIS, version 2.5, measures (Table 1).  For each measure, each HMO was assigned a Z-score, which was computed as

$$Z_{ij} = (X_{ij} - \overline{X}_i)/s_i$$

where

$X_{ij}$ = score for plan $j$ on HEDIS measure $i$,

$\overline{X}_i$ = average score on HEDIS measure $i$ for a sample of managed care health plans, including all of those offered by the firm and about 80 other plans offered by other firms that used the same methodology (about 200 plans in total), and

$s_i$ = standard deviation of HEDIS measure $i$ for the sample of managed care health plans, including all of those offered by the firm and about 80 other plans offered by other firms that used the same methodology.

Within each domain, the Z-scores for each measure were averaged for each plan to compute a domain-level score.  If fewer than half of the data elements were missing in a given domain for a given plan, the average was computed over the measures provided. If more than half of the data elements were missing for a specific domain, the plan was given a rating of 'No Data' for that domain.  For each domain of performance, the top third of the HMOs (including those not offered by this firm) were rated 'superior performance,' the middle third were rated 'average performance,' and the bottom third were rated 'below expected performance.'

The 'operational performance' rating, which captured non-clinical aspects of plan management such as claims payment and customer service, was based on evaluations from site visits to the plans by GM staff.

## References

Belsey, DA, Kuh E , and R.E. Welch. *Regression diagnostics: Identifying Influential Data and Sources of Collinearity,* Wiley, New York. 1980.

Berry ST. Estimating discrete-choice models of product differentiation. *RAND Journal of Economics.* 1994; 25:242-262.

Buchmueller TC, Feldstein PJ. The effect of price on switching among health plans. *Journal of Health Economics 16.* 1997;231-247.

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

Farley, D.O., Short, P.F., Elliot, M., Kanouse, D., Brown, J., and R.D. Hays. 1999 Use of CAHPS Information in Health Plan Choices By New Jersey Medicaid Beneficiaries, Unpublished Manuscript (draft). Santa Monica, CA. RAND Corporation

Feldman R, Finch M, Dowd B, Cassou S. The Demand for Employment-Based Health Insurance Plans. *The Journal of Human Resources.* 1989;XXIV:115-142.

Garnick DW, Lichtenberg E, Phibbs CS, Luft HS, Peltzman DJ, McPhee SJ. The Sensitivity of Conditional Choice Models for Hospital Care to Estimation Technique. *Journal of Health Economics.* 1989;8:377-397.

Hibbard, JH. 2000. Presentation at the *Quality and the Consumer Perspective Research Meeting*, March 10, Columbia, MD.

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

Hibbard, JH, Slovic, P, and JJ Jewett. 1997. "Informing Consumer Decisions in Health Care: Implications from Decision-Making Research". *Milbank Quarterly*, Vo. 75(3), 395-414.

McFadden DL. Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers In Econometrics*. New York: Academic; 1973.

McFadden DL. The Measurement of Urban Travel Demand. *Journal of Public Economics.* 1974;3:303-328.

Meyer, J.A., E.K. Wicks, L.S. Rybowski, and M.J. Perry. March 1998. Report on Report Cards. Economic and Social Research Institute, Washington, D.C.

Moulton, B.R. 1986. Random Group Effects and the Precision of Regression Estimates. *Journal of Econometrics*. 32(3), 385-97

Robinson S, Brodie M. Understanding the Quality Challenge for Health Consumers: The Kaiser/AHCPR Survey. *Journal on Quality Improvement.* 1997;23:239-244.

Scanlon, D.P. and M.E. Chernew. 1999. HEDIS Measures and Managed Care Enrollment. *Medical Care Research and Review*.

Scanlon, DP, Chernew, ME, and J Lave. "Consumer Health Plan Choice: Current Knowledge and Future Directions". *Annual Review of Public Health. 1997, 18:507-28.*

Short, PF. "Early Lessons from the CAHPS Demonstrations and Evaluations." CAHPS User's Meeting Sponsored by the Agency for Health Care Policy and Research, Baltimore, MD, October 15-16, 1998.

Short PF, Taylor AK. Premiums, Benefits, and Employee Choice of Health Insurance Options. *Journal of Health Economics.* 1989;8:293-311.

Tumlinson A, Bottigheimer H, Mahoney P, Stone EM, Hendricks A. Choosing a Health Plan: What Information Will Consumers Use? *Health Aff (Millwood).* 1997;16:229-238.