HEDIS® Performance Trends in Medicare Managed Care

Terry R. Lied, Ph.D., and Steven Sheingold, Ph.D.

The authors analyzed performance trends between 1996 and 1998 for health plans in the Medicare managed care program. Four measures from the Health Employer Data and Information Set (HEDIS®) were used to track performance changes: adult access to preventive/ambulatory health services, beta blocker treatment following heart attacks, breast cancer screening, and eye exams for people with diabetes. Using a cohort analysis at the health plan level, statistically significant improvements in performance rates were observed for all measures. Health plans exhibiting relatively poor performance in 1996 accounted for the largest share of overall improvement in these measures across years.

INTRODUCTION

While the debate concerning the future structure of the Medicare program continues, CMS has already made considerable progress in its transformation from a payer to a value-based purchaser of health care. Critical to these efforts has been the extensive development of performance measures in Medicare managed care, which includes the Medicare+Choice (M+C) program.[1] Internal steps have also been taken to develop similar measures for the fee-for-service program.

CMS arguably has the most comprehensive health-related performance measurement database in the world to support its managed care program. This system can:
- Support policy development.
- Monitor and enforce contract standards.
- Inform beneficiaries about their choices.
- Guide targeted quality improvement efforts.

There is still substantial work to be done, however, in making performance data optimally useful for these purposes. Although significant advances have been made in this area, we still face a substantial learning curve in our application of performance data to improve health care.

Several studies have demonstrated or suggested that performance measurement, itself, contributes to improved patient outcomes. For example, Kazandjian and Lied (1998), in a retrospective cohort study of cesarean section rates, found that continuous participation in a performance measurement project was associated with performance improvement. In a study on coronary artery bypass grafting (CABG) outcomes, Hannan and others (1994) found that risk-adjusted CABG mortality in New York dropped substantially between 1989 and 1992 during the first 4 years of New York's Cardiac Surgery Reporting System, a decline that has sometimes been attributed to reporting itself. The National Committee for Quality Assurance (NCQA) (2000), in its report on the state of managed care quality in 1999, found that managed care plans that consistently monitor and report on quality show significant improvements in quality. NCQA also found that health plans that score high on clinical quality also have the most satisfied members. However, Jencks (2000) indicates that performance data do not automatically

---

[1] M+C is a term used to describe the various health plan options available to Medicare beneficiaries. Technically, the M+C program did not begin until January 1999, however, Medicare managed care programs have been reporting performance data since 1996.

The authors are with the Centers for Medicare & Medicaid Services (CMS). The views expressed in this article are those of the authors and do not necessarily reflect the views of CMS.

improve clinical performance. He suggests that health plans, with their substantial purchasing power, can create a market force for better performance in health plans.

Faulty, unrefined, or inefficient measures have sometimes made it difficult to measure performance changes accurately. For example, according to Palmer (1996), "many measure sets in current use do not provide meaningful comparisons of clinical performance or, at worst, are actually misleading because they are limited in scope, insufficiently detailed, methodologically flawed, or not standardized across providers." Roper and Cutler (1998) identify several obstacles to measuring performance. These include difficulty constructing valid measures of health care outcomes, slow development of accessible data systems, lack of ready systems to obtain clinical information beyond the reach of the individual health plan, and lack of consensus about appropriate performance measures. Eddy (1998) views efforts to measure performance as currently limited because of the probabilistic nature, rarity, and confounding of many health outcomes; the inadequacies of information systems; the multiplicity of measurers and measures; the complexity of health plans; and the availability of funding. An overriding problem is that there are few generally accepted or explicit models of health care performance other than the structure-process-outcome model of Donabedian (1988). Lied and Kazandjian (1999) have recently proposed a multidisciplinary and conceptual model of health care performance with a mathematical basis that incorporates quality, access, cost, and patient satisfaction. This approach allows for a theoretical calculation of performance levels (as opposed to quality), but the model has not been empirically tested.

In recognition of these issues, CMS has taken a very careful and systematic approach to performance measurement. Over the past few years, several performance measurement sets, such as HEDIS®, the Consumer Assessment of Health Plans Study (CAHPS®), and the Health Outcomes Survey (HOS), have been developed or adapted to the M+C program. We have now begun to analyze the initial years of these data in order to determine how useful they will be for the program's objectives. In this article, we provide an analysis of the first 3 years of HEDIS® data.

METHODS

HEDIS® was initially developed by NCQA in concert with a cooperative group of health plans and large employers to help employers in the United States understand what they are getting for their health care dollars. In 1996, HCFA contracted with NCQA to develop a database of HEDIS® measures to be used in assessing performance of the Medicare program. (There is also a Medicaid version of HEDIS®.) In addition, HCFA required a number of HEDIS® measures, including the four in this study, to be audited according to the methodology developed by NCQA. HCFA contracted with a peer review organization to conduct the audits for 1996 and 1997. NCQA-licensed audit firms under contract with the health plans conducted the audit of the 1998 data.

As specified by contract, NCQA supplied HCFA with the raw data files in this study. The file formats were different for 1996 than for 1997 and 1998. The 1996 file was the public use file contained on HCFA's Web site (http://www.hcfa.gov/stats/pufiles.htm). This file contained the rates and some plan contract identifying information.

The 1997 and 1998 data files contained many data elements, including: contract number, market area code, eligible member population, minimum required sample size, sampling method, numerator and denominator values for the performance rates, the rates themselves, and 95-percent confidence intervals for the rates.

The number of separately reporting plan contracts participating in Medicare managed care was 289, 371, and 320 for service years 1996, 1997, and 1998, respectively. Many of these plans did not report on the four audited HEDIS® measures for all 3 years. Health plans were required to use a statistically valid sample of beneficiaries for compiling aggregate data for their organization based on technical specifications provided by NCQA.

The current study reports on performance changes in Medicare managed care for the first 3 years for which Medicare HEDIS® data were available. The study analyzes trends in plan performance for a cohort of plans using the four HEDIS® measures that were fully audited (100 percent) for both measurement years 1997 and 1998 and partially audited for measurement year 1996 services. (A fifth HEDIS® measure, "frequency of selected procedures," was also audited for these years.) A total of 27.4 percent of HEDIS® 3.0 submissions were fully validated by on-site review for the 1996 measurement year (Centers for Medicare & Medicaid Services, 2001).

Only Medicare reporting entities that were fully or substantially compliant with HEDIS® 3.0 specifications were included in this study. For measurement year 1997 (calendar year 1997 services), compliance rates were greater than 80 percent for all four measures: adult access to preventive/ambulatory health services (AAP) (90.7 percent); beta blocker treatment after heart attack (BB) (80.6 percent); breast cancer screening (BCS) (87.6 percent); and eye exams for people with diabetes (EE) (84.8 percent) (Centers for Medicare & Medicaid Services, 2001). Three of these HEDIS® measures involve effectiveness of care, and the fourth pertains to access to care. A brief description of these measures follows:

AAP—Measures the percentage of enrollees age 20 or over who were continuously enrolled during the measurement year and who had an ambulatory or preventive care visit during the measurement year. Although this measure reports rates for three age groupings (20-44, 45-64, 65 or over), this study used the 65-or-over rate. Because this study is limited to Medicare, the vast majority of Medicare enrollees were, in fact, age 65 or over.

BB—Measures the percentage of enrollees age 35 or over during the measurement year who were hospitalized and discharged alive between January 1 and December 24 of the measurement year with a diagnosis of acute myocardial infarction (AMI) and who received an ambulatory prescription for beta blockers upon discharge.

BCS—Measures the percentage of women age 52-69 years who were continuously enrolled during the measurement year and the preceding year and who had a mammogram during the measurement year or the preceding year. Enrollees may have no more than one gap in enrollment of up to 45 days during each year of continuous enrollment.

EE—Measures the percentage of Medicare beneficiaries with Type 1 or Type 2 diabetes age 18-75 years who were continuously enrolled during the measurement year who had an eye screening for diabetic retinal disease.

We use a less cumbersome term, "plan," in this study in referring to the "Medicare reporting entity," this study's unit of analy-

sis. Medicare reporting entities, i.e., plans, prepare a separate Medicare HEDIS® report for each contract. Separate reporting is done within Medicare managed care contracts for market areas that are not geographically contiguous if the contract covers more than one major community in which there were at least 5,000 Medicare enrollees for that organization. This study used a cohort analysis in which only the Medicare plans that reported data for 1996, 1997, and 1998 service periods were included. As an additional requirement for inclusion, the denominator value, reflecting the target population, had to be at least 30 for each plan for 1996, 1997, and 1998. (Prior to 1999, NCQA suppressed HEDIS® rates for effectiveness-of-care measures if there were fewer than 30 eligible members for a given measure within a reporting entity.)

NCQA allows plans to use one of two methods for data collection: administrative or hybrid. The administrative method requires that the plan identify a target population and search for evidence of the intervention using administrative data. The hybrid method requires the plan to choose a sample of patients identified through administrative records. For most measures, a sample of 411 patients is required. The plan then searches for evidence of the intervention by examining administrative records. If no evidence is found, the plan reviews patient charts. The calculated rate is based on a numerator reflecting administrative and patient chart data. Plan performance was measured using the aggregate data for each Medicare reporting entity. Mean performance rates were equal to the sum of aggregate performance rates (percentages) across all plans (Medicare reporting entities) divided by the number of plans.

Changes in performance in the cohort were tested for statistical significance using the paired samples t-test procedure from SPSS® 10.0. Changes in performance were judged to be significant if their probability of occurrence by chance was less than 5 percent (two-tailed t-test). The number of plans included in the cohort for each measure is reported in Table 1. Summary data for all Medicare risk plans reporting in each of the 3 years were also analyzed for comparison.

A second analysis examined if performance gains could be attributed to improved performance by plan contracts with relatively poor performance in the first year of collection (1996). This analysis involved an examination of changes in rates corresponding to given percentiles in 1996, 1997, and 1998. If the lower percentile rates (5, 10, 25) greatly improved between 1996 and 1998, and there was minimal improvement in the upper percentile rates (75, 90, 95), most of the overall performance gains between 1996 and 1998 could be attributed to improved performance by initially lower performing plan contracts.

The final analysis examined the stability and predictability of plan performance. This analysis examined the stability of plan performance on given measures over time as well as the relationships among different performance measures within the same plans. The main questions to be answered were:
- Does a plan's relative performance on a given measure tend toward stability over time (i.e., do plans that perform at a given level relative to other plans in one year tend to perform at that same level relative to other plans in subsequent years)?
- Does a plan's performance on one measure tend to predict its performance on other measures (i.e., are there generally low, mid-level, and high-performing plans on all measures)?

To address these questions, Spearman rank-order correlations were computed among all the measures over the 3-year period, and a correlation matrix was constructed.

Table 1

Descriptive Statistics of HEDIS® Measures for a Cohort of Medicare Managed Care Risk Plans: 1996-1998

Measure	Statistic	Year 1996	Year 1997	Year 1998	t-Test Comparisons 1996 Versus 1997	t-Test Comparisons 1997 Versus 1998	t-Test Comparisons 1996 Versus 1998
Adult Access to Preventive/	Mean	84.90	87.43	88.55	*2.00	1.77	*2.90
Ambulatory Care	SD	16.63	10.70	9.83			
(n = 167)	Range	90.10	67.73	59.37			
	Minimum	9.90	32.27	40.63			
	Maximum	100.00	98.74	99.64			
Beta Blocker Administration	Mean	60.38	78.52	85.14	*7.76	*4.33	*11.16
After Heart Attack	SD	18.84	15.13	11.57			
(n = 55)	Range	77.20	66.30	52.98			
	Minimum	19.30	32.39	47.02			
	Maximum	96.50	98.69	100.00			
Breast Cancer Screening	Mean	72.08	72.73	74.48	1.02	*4.24	*4.14
(n = 151)	SD	9.10	9.67	8.16			
	Range	46.50	51.70	45.63			
	Minimum	42.90	38.87	43.26			
	Maximum	89.40	90.57	88.89			
Eye Exams for People	Mean	52.86	52.55	55.72	-0.27	*3.52	*2.37
with Diabetes	SD	17.91	15.66	14.92			
(n = 156)	Range	91.00	82.76	79.42			
	Minimum	6.70	1.69	6.69			
	Maximum	97.70	84.44	86.11			

*$p < 0.05$.

NOTES: HEDIS® is Health Employer Data Information Set. SD is standard deviation.

SOURCE: Authors' tabulations from the Centers for Medicare & Medicaid Services HEDIS® files, 1998.
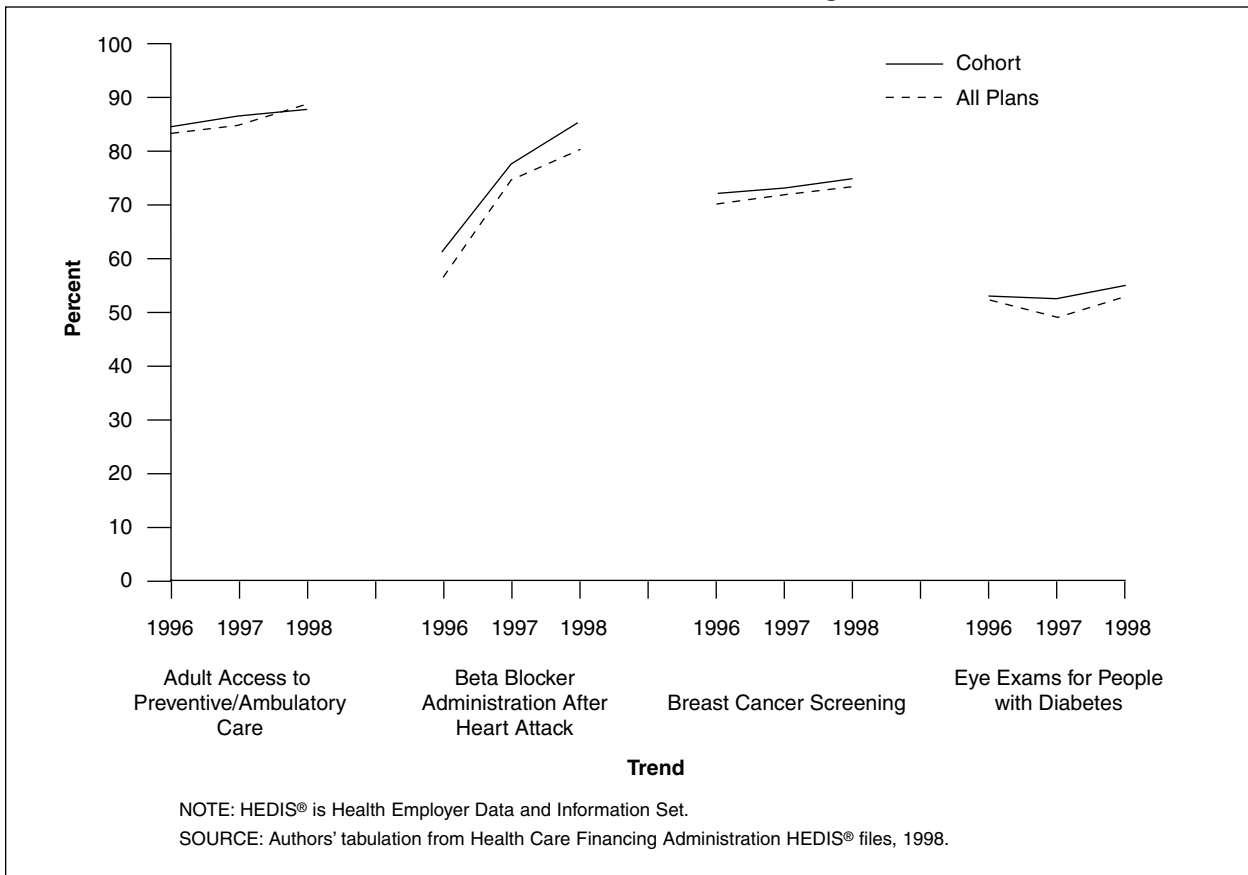
RESULTS

Table 1 presents the descriptive statistics (means and standard deviations [SDs]) for the cohort of Medicare managed care HEDIS® reporting entities for 1996-1998, along with the paired samples t-test results for years 1996 versus 1997, 1997 versus 1998, and 1996 versus 1998 for the four HEDIS® measures in this study. Figure 1 displays the trends in HEDIS® rates for the cohort in this study compared with all Medicare managed care plans between 1996 and 1998. As Figure 1 illustrates, the mean rates for all plans for each of the 3 years did not differ substantially from the means when just the cohort group was analyzed.

AAP Measure

This analysis was limited to the age group of 65 years or over. A total of 167 health plan contracts met the criteria for inclusion in the cohort group. In calendar year 1996, the mean rate for the cohort was 84.90 percent (SD=16.63; range=90.10). In 1997, the rate increased to 87.43 percent (SD=10.70; range=67.73). By 1998, the rate increased to 88.55 percent (SD=9.83; range=59.37). Using two-tailed t-tests, the paired samples statistical results (cohort group) indicated that the difference in the means between 1996 and 1997 was statistically significant ($t=2.00$, $p < 0.05$), as was the difference in the mean rates between 1996 and 1998 ($t=2.90$, $p < 0.01$). However,

Figure 1

Trends in HEDIS® Rates for a Cohort Versus All Medicare Managed Care Risk Plans: 1996-1998

NOTE: HEDIS® is Health Employer Data and Information Set.

SOURCE: Authors' tabulation from Health Care Financing Administration HEDIS® files, 1998.

there was not a significant change (increase) in mean plan rates between 1997 and 1998 (t=1.77, p > 0.05). Over the 3-year period, there was a steady increase in the mean rate accompanied by declines in the SD and range. Between 1996 and 1998, the mean performance rate increased by 4.12 percent. Performance rate means for the AAP measure for all Medicare managed care risk plans for 1996, 1997, and 1998 were 83.34 percent (n=220), 85.97 percent (n=286), and 89.45 percent (n=299), respectively.

BB Measure

A total of 55 health plan contracts met the criteria for inclusion in the cohort group for this analysis. As with all the measures, a denominator of at least 30 was required for inclusion in the analysis. Many health plans—in fact most of them—did not have denominator values for this measure of 30 or more for each of the 3 years, accounting for the relatively low sample. The mean performance rate increased substantially over 3 years. It was 60.38 percent (SD=18.84; range=77.20) in 1996 but increased to 78.52 percent (SD=15.13; range=66.30) by 1997. By 1998, the mean rate was 85.14 percent (SD=11.57; range=52.98). All differences among the pairs (1996 versus 1997, 1997 versus 1998, 1996 versus 1998) were statistically significant. Over the 3-year period, both the SD and range declined. Between 1996 and 1998, the mean performance rate increased by 29.08 percent. Performance rate means for the BB measure for all Medicare managed care risk plans for 1996, 1997, and

Table 2

Rates Corresponding to Selected Percentiles for Cohorts of Medicare Managed Care Risk Plans: 1996-1998

Measure and Year	Percentile						
	5th	10th	25th	50th	75th	90th	95th
Adult Access to Preventive/Ambulatory Care							
1996	42.16	59.58	83.80	91.00	94.40	95.50	97.22
1997	64.97	74.74	83.87	91.14	94.26	95.73	96.69
1998	73.38	77.50	86.55	91.82	94.41	95.49	96.50
Beta Blocker Administration After Heart Attack							
1996	30.00	34.92	43.80	61.30	74.70	86.36	89.50
1997	43.22	55.67	71.74	82.93	88.77	96.11	97.01
1998	58.74	69.33	79.71	87.76	93.32	96.30	97.47
Breast Cancer Screening							
1996	55.60	60.60	66.40	71.90	79.10	83.78	86.64
1997	55.78	59.22	68.23	73.28	80.47	83.96	86.65
1998	60.10	63.22	71.26	75.00	80.37	84.33	85.37
Eye Exams for People with Diabetes							
1996	19.42	27.85	40.40	53.00	67.95	75.12	80.72
1997	22.37	28.99	44.06	54.20	64.00	70.18	74.85
1998	26.76	36.06	47.89	57.37	66.15	73.39	75.94

SOURCE: Authors' tabulations from the Centers for Medicare & Medicaid Services HEDIS® files, 1998.

1998 were 57.33 percent (n=86), 74.83 percent (n=129), and 80.76 percent (n=169), respectively.

BCS Measure

A total of 151 health plan contracts met the criteria for inclusion in the cohort group in this analysis. In calendar year 1996, the mean breast cancer screening rate was 72.08 percent (SD=9.10; range=46.50). In 1997, the rate had increased to 72.73 percent (SD=9.67; range=51.70). By 1998, the mean breast cancer screening rate of this cohort was 74.48 percent (SD=8.16; range=45.63). The differences among the pairs of means for 1997 versus 1998 and 1996 versus 1998 were significant, indicating an improvement in mean performance rates on this measure between 1997 and 1998 and 1996 and 1998. The change in the screening rates between 1996 and 1997 was not statistically significant. The SD declined steadily over the 3-year period with little change in the range. Between 1996 and 1998, the mean performance rate increased by 3.22 percent. Performance

rate means for the breast cancer screening measure for all Medicare managed care health plans for 1996, 1997, and 1998 were 70.56 percent (n=188), 71.66 percent (n=247), and 73.07 percent (n=266), respectively.

EE Measure

A total of 156 health plan contracts met the criteria for inclusion in the cohort group of plans in this analysis. In 1996, the mean rate of people with diabetes who received eye exams was 52.86 percent. The SD across plan contracts was 17.91 and the range was 97.70. In 1997, the mean rate was 52.55 percent, the SD was 15.66, and the range was 82.76. Finally, for 1998, the mean rate was 55.72 percent, the SD was 14.92; and the range was 84.83. The differences among the pairs of means for 1997 versus 1998 and 1996 versus 1998 were significant, indicating an improvement in mean performance rates on this measure between 1997 and 1998 and 1996 and 1998. The decrease in the eye exam rates between 1996 and 1997 was not sta-

Table 3

Spearman Rank-Order Correlations Among HEDIS® Measures for Cohorts of Medicare Managed Care Risk Plans: 1996-1998

Measure and Year	AAP			BB			BCS			EE		
	1996	1997	1998	1996	1997	1998	1996	1997	1998	1996	1997	1998
AAP												
1996		*0.56	*0.55	*0.64	*0.76	*0.54	*0.44	*0.50	*0.46	*0.56	*0.38	*0.44
1997			*0.84	*0.52	*0.56	*0.51	*0.48	*0.52	*0.53	*0.45	*0.45	*0.53
1998				*0.42	*0.50	*0.52	*0.48	*0.44	*0.48	*0.41	*0.41	*0.53
BB												
1996					*0.61	*0.66	*0.54	*0.65	*0.58	*0.55	*0.41	*0.49
1997						*0.68	*0.57	*0.62	*0.60	*0.37	*0.50	*0.52
1998							*0.41	*0.52	*0.56	*0.30	*0.33	*0.55
BCS												
1996								*0.71	*0.65	*0.55	*0.55	*0.59
1997									*0.86	*0.62	*0.57	*0.64
1998										*0.58	*0.53	*0.67
EE												
1996											*0.67	*0.67
1997												*0.77
1998												

*$p>0.05$.

NOTES: AAP is Adult Access to Preventive/Ambulatory Services. BB is Beta Blocker Administration After Heart Attack. BCS is Breast Cancer Screening. EE is Eye Exam for People with Diabetes.

SOURCE: Authors' tabulations from the Centers for Medicare & Medicaid Services HEDIS® files, 1998.

tistically significant. Over the 3 years, there was a decrease in the SDs. The range dropped considerably between 1996 and 1997, then rose slightly in 1998. Between 1996 and 1998, the mean performance rate increased by 5.13 percent. Performance rate means for the EE measure for all Medicare managed care plans for 1996, 1997, and 1998 were 52.06 percent (n=195), 49.75 percent (n=275), and 52.33 percent (n=304), respectively.

Changes in Rates Corresponding to Percentiles

This aspect of the study compared the rates for given percentiles in 1996, 1997, and 1998 in order to determine whether the positive shift in performance was largely attributable to low performers doing better in the subsequent year. Table 2 displays the rates corresponding to the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles

in 1996, 1997, and 1998. The analysis generally supports the proposition that performance rate gains were largely attributable to performance improvement in initially low-performing plan contracts. For example, 5th percentile rates for AAP for the 3 years were 42.16, 64.97, and 73.38, respectively, a rate gain of more than 30 points between 1996 and 1998. However, the 90th and 95th percentile rates were nearly the same for these years. The largest gains for BB treatment occurred at the 5th, 10th, and 25th percentiles. For BCS, the greatest gains occurred below the 50th percentile. Similar results were found for the EE measure, i.e., slight gains in rates occurred at the 50th percentiles and below.

Stability of Plan Performance

Table 3 displays the Spearman rank-order correlations between each of the measures across the 3-year period. There was a tendency

for plans to perform at the same level relative to other plans on each of the four measures across the 3-year period. For example, the correlations between BCS rates between 1996 and 1997 and 1997 and 1998 were 0.71 and 0.65, respectively ($p < 0.05$ in both cases). Similarly, the correlations between EE rates between 1996 and 1997 and 1997 and 1998 were 0.67 and 0.77, respectively (again, $p < 0.05$ in both cases). The two remaining measures also displayed significant correlations between years. For the BB measure, the correlations were 0.61 and 0.68 for years 1996 and 1997 and 1997 and 1998, respectively. For the AAP measure, the correlations for years 1996 and 1997 and 1997 and 1998 were 0.56 and 0.84, respectively.

Predictability of Plan Performance

Plan performance was predictable in the sense that if a plan did well or poorly on one measure, there was a tendency for that plan to perform similarly on other measures. For example, using 1998 data, the rank-order correlations were as follows: AAP and BB, 0.52; AAP and BCS, 0.48; AAP and EE, 0.53; BB and BCS, 0.56; BB and EE, 0.55; and BCS and EE, 0.67. All of these correlations were statistically significant at the 0.05 level.

SUMMARY AND DISCUSSION

The objective of this analysis was to begin the significant task of analyzing the Medicare managed care performance database. We analyzed four HEDIS® measures to examine trends in plan performance over a 3-year period. A secondary objective of the analysis was to examine the reliability of these measures in preparation for similar studies in the future—studies that will include other performance measures such as CAHPS®, the Medicare HOS, and administrative data.

Based on the analysis of the four HEDIS® measures—measures selected because of their clinical importance and the fact that they were 100 percent audited for 1997 and 1998—this study is encouraging with regard to performance improvement in Medicare managed care. The unit of analysis in this study was the Medicare reporting entity. We conducted a cohort analysis of health plans that participated in managed care contracts for 3 years (1996-1998). We found that there were statistically significant improvements for three of the four selected HEDIS® measures between 1997 and 1998 (BB, BCS, EE). Only the access measure (AAP) did not show a statistically significant improvement between 1997 and 1998; however, the gain between 1996 and 1998 in the mean rate for this measure was statistically significant.

The BB measure showed statistically significant gains in all possible yearly combinations (1996 versus 1997, 1997 versus 1998, and 1996 versus 1998). The improvement in performance appears to be practically significant as well as statistically significant for at least some of the measures (Figure 1). For example, mean BB rates increased from 60.38 percent to 78.52 percent to 85.14 percent across the 3 years of reporting. Mean EE rates actually decreased from 52.86 percent in 1996 to 52.55 percent in 1997; however, this rate increased to 55.72 percent in 1998 ($p < 0.01$). The remaining two measures, BCS and AAP, also showed important gains in mean rates between 1997 and 1998—from 72.73 percent to 74.48 percent and from 87.43 percent to 88.55 percent, respectively, although these gains were not large.

NCQA (2001) reported HEDIS® results for commercial enrollees in all health plans reporting HEDIS® data on effectiveness-of-care measures between 1996 and 1998. Their study did not use a cohort analysis;

still, a comparison of their results with our results is instructive. The HEDIS® performance rates for Medicare beneficiaries in managed care plans as reported in our study generally exceeded the overall averages for commercial enrollees in managed care health plans as reported by NCQA (2001). NCQA reports that beta blocker administration rates increased from 62.2 percent in 1996 to 74.5 percent in 1997 and then to 79.9 percent in 1998 for enrollees in commercial health plans reporting HEDIS® data. With the exception of 1996, rates for Medicare enrollees in the cohort were higher than the commercial rates reported by NCQA. BCS rates for commercial health plans were 70.3 percent in 1996, 71.3 percent in 1997, and 72.2 percent in 1998; the rates reported for the Medicare cohort in our study were slightly higher for all 3 years. EE rates were substantially lower for commercial enrollees in health plans reporting HEDIS® data to NCQA than for Medicare enrollees in this study. NCQA reports these rates as varying from 38.0 percent in 1996 to 38.8 percent in 1997 and then to 41.4 percent in 1998. The rates for Medicare beneficiaries were all greater than 50 percent across this time period. Three-year AAP rates were not reported in the NCQA study.

The gains in plan performance, as evidenced by improvements in the mean rates of these HEDIS® measures, suggest that reporting requirements may have provided an incentive for performance improvement. Moreover, over the 3-year period, there were consistent decreases in the variability of performance rates (as measured by the SD) for all four measures, further evidence of an impact on performance. Along with declines in the SD, there was a general tendency for the range in performance rates to decrease over the 3-year period.

There are a number of factors that could have affected the results of this study. Improved documentation across the 3-year period could have been a factor contributing to improved scores on the HEDIS® measures. In addition, plan differences in the sophistication of their information systems could have contributed to some of the variability among plans in HEDIS® scores. Finally, the method of data collection could have affected the results, because the hybrid method tends to yield slightly higher rates (Himmelstein et al., 1999). Although all of these factors could have influenced the results of this study, we believe that improved plan performance was the single highest contributor to the improved HEDIS® scores. The basis for this interpretation is the consistency of results across all four measures over the 3-year period and the fact that these measures were fully audited for 1997 and 1998.

Despite the encouraging results in this study, there is considerably more opportunity for improvement in health plan performance and quality. Arguably, the rates of the HEDIS® measures in this study should approach 100 percent, because these measures require removal of contraindications from the denominators. The populations are virtually all at risk for the various effectiveness-of-care or access-to-care measures in this study. One of the most gratifying findings was that the positive shift in performance from one year to the next was largely attributable to improved performance in plan contracts that did poorly initially. It is difficult to argue that this result is entirely a function of the "regression to the mean" phenomenon or to a "ceiling effect." The percentiles corresponding to particular rates changed primarily at the lower end of the distribution. Under pure statistical regression and no shifting of the mean, only the ordering would change—

lower plan contracts would gain, higher ones would lose—not the rates corresponding to given percentiles. The positive shifting of the mean was in part accounted for by the improvement of relatively poor performers over the 3-year period. The yearly mean rates were well below 100 percent, arguing against a ceiling effect, although this could occur in the future with continued improvement on the measures. Most likely candidates for a ceiling effect are the BB and AAP measures.

A rank-order correlation analysis demonstrated that a plan's relative performance on a particular measure with respect to other plans tended to be fairly stable across 1996, 1997, and 1998 reporting years. In addition, most of the rank-order correlation relationships between the four measures for a given year were significant, and all were in the positive direction. Together, these findings provided considerable support for the notion of plan consistency of performance—consistency across different measures for a given reporting period and consistency across different reporting periods for the same measure. This is important because, within the context of shrinking variation in performance, there is an implied stability in measure reporting. That is, if plans were to vary considerably on any measure from year to year, we would be concerned about the reliability and, consequently, the validity of measuring health plan performance.

Despite a study that supplied plenty of reason for optimism with respect to plan performance, there was evidence that plan contracts that do relatively poorly initially continue to perform below average in subsequent years. In a subsidiary analysis not previously reported here, we found that, of 30 plan contracts that had BCS rates in the lowest quintile in 1996, 19 (63.3 percent) continued to be in the lowest quintile in the subsequent year. Eighteen of these plan contracts (60.0 percent) remained in the lowest quintile 2 years later. It appears that, although low performers improved the most, their improvement was generally not enough to place them among the mid-level or high-performing plans in the subsequent years in this study. The implication for purchasers is that initial performance results are predictive of future performance for specific measures, and the identification of relatively low performers at an early stage may lead through intervention to better-than-predicted future performance.

This study did not address the relationship between HEDIS® measures and other measures of performance. We still do not know if improved performance on HEDIS® measures contributes to beneficiary satisfaction, improved functional status, or higher enrollee retention. In addition, we still do not know the effect of improved data collection on performance data. Research in these areas will greatly assist in the development of approaches to quality improvement.

REFERENCES

Center for Medicare & Medicaid Services: Medicare HEDIS® 3.0 1996 Data Audit Report. 2001. Internet address: http://www.hcfa.gov/quality/3I1.htm.

Donabedian, A.: The Quality of Care: How Can It Be Assessed? *Journal of the American Medical Association* 260(12):1743-1748, 1988.

Eddy, D.M.: Performance Measurement: Problems and Solutions. *Health Affairs* 17(4):7-25, 1998.

Hannan, E.L., Kumar, D., Racz, M., et al.: New York State's Cardiac Surgery Reporting System: Four Years Later. *Annals of Thoracic Surgery* 58(6):1852-1857, 1994.

Himmelstein, D.U., Woolhandler, S., Hellander, I., and Wolfe, S.M.: Quality of Care in Investor-Owned vs. Not-for-Profit HMOs. *Journal of the American Medical Association* 282(2):159-163, 1999.

Jencks, S.F.: Clinical Performance Measurement— A Hard Sell. *Journal of the American Medical Association* 283(15):2015-2016, 2000.

Kazandjian, V.A., and Lied, T.R.: Cesarean Section Rates: Effects of Participation in a Performance Measurement Project. *The Joint Commission Journal on Quality Improvement* 24(4):187-196, 1998.

Lied, T.R., and Kazandjian, V.A.: Performance: A Multidisciplinary and Conceptual Model. *Journal of Evaluation in Clinical Practice* 5(4):393-400, 1999.

National Committee for Quality Assurance: *The State of Managed Care Quality—1999*. Washington, DC. 2000.

National Committee for Quality Assurance: *The State of Managed Care Quality—2000*. Washington, DC. 2001.

Palmer, H.: Measuring Clinical Performance to Provide Information for Quality Improvement. *Quality Management in Health Care* 4(2):1-2, 1996.

Roper, W.L., and Cutler, C.M.: Health Plan Accountability and Reporting: Issues and Challenges. *Health Affairs* 17(2):152-155, 1998.

Reprint Requests: Terry R. Lied, Ph.D., Centers for Medicare & Medicaid Services, C4-13-01, 7500 Security Boulevard, Baltimore, MD 21244-1850. E-mail: tlied@cms.hhs.gov