The contribution of health plans and provider organizations to variations in measured plan quality

LAURENCE C. BAKER [1,2] AND DAVID S.P. HOPKINS [3]

[1]Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, USA, [2]National Bureau of Economic Research, Cambridge, MA, USA, and [3]Pacific Business Group on Health, San Francisco, CA, USA

Address reprint requests to: Laurence C. Baker, Department of Health Research and Policy, Stanford University School of Medicine, HRP Redwood Building, Room no 110, 259 Campus Drive, Stanford, CA 94305-5405, USA. Tel: +1-650-723-4098; E-mail: laurence.baker@stanford.edu

Abstract

Objective. Some argue that health plans have minimal impacts on quality of care and that quality data collection should focus only on physician organizations. We investigate the relative impact of physician organizations and health plans on quality measures.

Design. Statistical analysis of data on 9 Healthcare Effectiveness Data and Information Set (HEDIS) measures from 6 health plans and 159 provider organizations. We use regression analyses to examine the amount of variation in HEDIS measures accounted for by variation across provider organizations, and whether accounting for health plans explains additional variation. We also examine whether accounting for provider organizations explains away variation in HEDIS scores across health plans.

Setting. Six health plans and 159 contracted provider groups in California.

Main Outcome Measures. Nine HEDIS scores.

Results. For all nine measures studied, variation across provider organizations explains much of the HEDIS score variation. But, after accounting for variation across providers, variation across plans statistically significantly explains additional variation. We also find statistically significant differences across health plans in HEDIS rates that are not substantially affected when we control for the provider organization that cared for the patient.

Conclusions. On their face, these results suggest that plans can influence quality independent of the selection of physician organizations with which they contract, in contrast to hypotheses that plans are 'too far' from patients to have an influence. Continued attention to collecting plan-level data is warranted. Further work should address other possible sources of variations in HEDIS scores, such as variability in plan administrative databases.

Keywords: quality measurement, setting of care, HEDIS scores, health plans, provider groups

Introduction

Providing measures of quality that will enable purchasers and consumers of health care to select health plans and providers based on evidence about their performance is a key part of strategies to improve health care quality [1–3]. Measuring and reporting quality data at the health plan level has become routine through the National Committee on Quality Assurance's (NCQA) national accreditation program, which uses standardized measures from the Healthcare Effectiveness Data and Information Set (HEDIS). Nearly all major health plans in the USA receive NCQA accreditation, and their HEDIS results are published annually on the NCQA website. However, despite the prominence of health plan measures, some have argued that collecting data on the performance of health-care providers would be more valuable [4–9].

In recent years, federal government agencies have increasingly focused on measuring and rewarding provider performance through initiatives such as the Department of Health and Human Services' Four Cornerstones initiative and the Center for Medicare and Medicaid Services' Physician Quality Reporting Initiative. Likewise, private purchasers have been encouraging their contracted health plans to use data to measure and reward provider performance and offer their subscribers information and incentives for making good

choices [8]. The public and private sectors have been working together to define appropriate measures and data collection strategies for carrying measurement down to the individual provider level.

Arguments in favor of measurement at the provider level emphasize the proximity of health-care providers to patients, suggesting that since health-care providers interact most directly with patients, they are in the best position to influence care and should be the ones held accountable for health-care delivery. One might go so far as to question whether measurement of health plan quality contributes any new information beyond what could be gathered by provider-level measurement. If health plans are simply contracting entities with little or no direct impact on quality of care, then any differences across plans observed in studies of HEDIS rates or other related plan-level quality measures simply reflect differences in the provider networks with which they contract. If this is the case, data on health plans merely reflects their degree of success in contracting with high performing providers and data collection efforts should be focused only on measuring providers. Efforts to get consumers to select better health plans might also then be targeted away from health plans *per se*, and instead aimed at getting consumers to pay attention to the providers with which plans have contracted. It would also suggest that there would be convergence in the measured quality of plans with overlapping provider networks. Some work suggests that significant overlap in the provider networks of plans is common [10].

Whether or not this strong hypothesis is true depends on the extent to which plans influence quality over and above the effect of their contracted providers. In fact, health plan sometimes argue against this view, claiming that they are able to influence quality of care independently from their providers through things such as disease management programs, the use of reminder systems to encourage the use of preventive services and utilization monitoring with feedback to physicians to encourage better processes of care.

Despite the importance of understanding the locus of control over quality, evidence on these issues remains sparse. The one previous study of which we are aware examined the relationships between health plan and provider influence on HEDIS scores for six preventive care measures, using data from 21 health plans and 22 physician groups [11]. That paper reported that health plans appeared to have some influence on HEDIS scores independent of providers—in this case physician groups—although it was based on a small number of physician groups and used data on only a few basic preventive care measures.

In this paper, we use a unique database to study the relative effects of plans and organized physician groups on measured quality of care. The database contains data on quality measures from seven health plans and more than 200 of the physician organizations with which they contract. The provider networks offered by these plans have substantial overlap, and all of the plans provided quality data specific to each of the many participating physician groups with which they contracted. We can thus use statistical analyses to study the relative contribution of plans and physician organizations

to explain variations in the quality measures. We compute the share of the variance in quality measures that can be explained by just the physician organizations, and then the incremental amount that is explained by adding information about the plan. This essentially asks whether knowing the health plan from which a quality measure came contributes any explanatory power over and above what could be explained from knowledge of just the physician organization from which it came. Similarly, we examine whether controlling for physician groups substantially affects variations across plans in quality measures.

Methods

Data

Our data come from the 2006 Integrated Healthcare Association (IHA) pay-for-performance program, which is a unique collaboration among major purchasers, health plans and physician organizations in the California market, including seven network model health plans covering more than 90% of the commercial HMO enrollees outside of Kaiser, and 206 of the organized physician groups that contract with these plans under capitated payment arrangements. Each plan reported rates for several HEDIS 2006 measures specific to each physician organization with which it contracted, producing a number of what we term 'plan-group' observations. The data were collected in 2006, for use in the 2006 pay-for-performance program, but reflect care delivered in 2005. In all, the baseline data contained 911 plan-group observations for each HEDIS measure. The 911 observations represent all 7 plans, and each of the 206 separate physician organizations appears in at least 1 plan-group observation. Each plan appears many times, since each plan provided data for many physician organizations with which it contracts. Most physician organizations also appear multiple times since they contract with more than one plan.

The HEDIS measures studied include both preventive care measures and measures related to the care of patients with illness. They were all specified for collection at the physician group level by NCQA. Data for 15 HEDIS measures were originally collected: the use of appropriate medications for people with asthma (reported separately for patients age 5−9, 10−17 and 18−56), breast cancer screening, cervical cancer screening, chlamydia screening in women (reported separately for women of age 16−20 and 21−25), childhood immunization status, cholesterol management for patients with cardiovascular conditions, comprehensive diabetes care (composed of 5 separate measures) and appropriate treatment for children with upper respiratory infection.

Some of these measures are normally collected at the health plan level using NCQA's so-called 'hybrid' methodology, in which data from medical charts are used to fill in relevant information when it is missing from administrative records for sampled members (cervical cancer screening, childhood immunizations, cholesterol management for patients with cardiovascular conditions, and diabetes care).

The others are normally collected using administrative data only. We could only use administrative data for this study. This should not affect measures that normally use administrative data only. We elected to also consider the measures that normally use the hybrid methodology, but which we can measure here using only administrative data. In many of these measures, administrative data consistently collected across plans could still provide a useful opportunity to address the study questions. As we discuss below, we did end up having to exclude some of the hybrid measures that incorporate test results, for which chart data is important and the administrative data appear quite incomplete.

Data for each measure for each physician organization were compiled by each plan, using consistent specifications. All plans computed rates relying on their administrative data systems to identify patients who would be candidates for each of the reported HEDIS services, and then used their administrative data systems again to identify patients who had received the indicated services. According to HEDIS specifications, patients for whom receipt of the service could not be verified for any reason, including possibly missing data, are counted as not having received the service. The data were submitted by the plans directly to the NCQA, which performed validity checks to ensure consistency in the way the standard specifications were applied.

For each plan-group combination for each measure, we observe the number of patients eligible for the measure (the denominator) and the number of times the HEDIS specification was met, which can be used to compute the 'HEDIS completion rate.' We also observe the geographic region in which the physician organization is located. We did not observe other information about the organization or its patients. Note that HEDIS measure specifications do not call for adjustments based on patient characteristics.

After examining the data, we made several adaptations to the measures for our study purposes. The standard comprehensive diabetes care measure aggregates several measures including screening and control of low-density lipoprotein (LDL) levels, screening and control of HbA1c levels and screening for nephropathy. In the data we use, the reported information about LDL and HbA1c control appeared to be incomplete in many cases, perhaps because we rely on administrative data and these measures frequently require chart data. In addition, the information about nephropathy monitoring was experimental in the IHA project in the year we studied. As a result, we studied LDL screening rates and HbA1c screening rates in the diabetic population as individual measures, but did not examine the aggregated HEDIS standard comprehensive diabetes care measure. We also removed the cholesterol screening for cardiovascular patients from consideration since we were advised by the collaborative that reports health plan HEDIS rates in California that there may have been flaws in the technical specification for this measure in 2006. Since the number of patients on which the measures of appropriate medication use for asthma patients in the 5–9 and 10–17 age groups was frequently small, we combined these measures to create a single asthma medication measure for patients aged 5–17. After making these

adaptations, the measure set we examined contained nine measures as shown in Table 1.

From the full set of plan-group observations, we extracted our final analysis sample by excluding plan-group observations that did not meet IHA standards for completeness of the underlying encounter data. On a measure-by-measure basis, we also dropped any plan-group observations for which the measure denominator contained less than 10 patients. In sensitivity analyses, dropping observations with less than 5 or less than 15 in the denominator produced substantially similar results. On a measure-by-measure basis we excluded data from any plan-group observations from a physician organization that contracted with only a single plan, ensuring that we had data with overlap between physician organizations and health plans so we could separately identify the influence of plans and providers. Finally, after making the above exclusions, we observed that the study-eligible plan-group combinations for one of the plans came from only a handful of physician organizations that were highly geographically concentrated. We thus excluded this plan from further analyses. (We verified that key results are not affected by including this plan in the analysis.)

The final analysis data set contains information representing 6 plans and 159 physician organizations. (Further information about sample definition, exclusions and plan-group overlap is provided in a supplementary material, technical appendix to this paper.)

Analysis

Our analytic approach is built around the results of three regression analyses per measure. In all three, the dependent variable is the observed HEDIS rate. In the first regression, the independent variables are a set of dummy variables for physician organizations. In the second, the independent variables are a set of dummy variables for health plans. In the third regression, the independent variable set includes both dummies for physician organizations and dummies for health plans.

Our first analysis focuses on measuring the proportion of the variation in HEDIS rates explained by variation across physician organizations, and then computing the incremental share explained by variation across health plans. We implement this method by comparing the R-squared measure from the first regression, including just controls for physician organizations, with the R-squared from the third regression that controls for physician organization and plan. Since the R-squared measure from each regression provides a measure of the share of the variance in the HEDIS scores that is explained by the independent variables, the difference between the two R-squared measures is the incremental amount of variance explained by health plans, once the variance explained by physician organizations has been accounted for. We compute this difference, and use F-statistics to test the hypothesis that the difference is equal to zero. If plans have no impact independent of the providers groups they contract with, the first regression will show some amount of variance explained by provider organizations, but the second will show no additional variance accounted for when we add health plans to the set of

Table I Study measure descriptions

Measure name	Description
Appropriate medications for asthma patients, age 5–17	Percentage of members (5–17 years of age) who were identified as having persistent asthma and who were appropriately prescribed medication (≥ 1 prescription for inhaled corticosteroids, nedocromil, cromolyn sodium, leukotriene modifiers, or methylxanthines) during the measurement year
Appropriate medications for asthma patients, age 18–56	Percentage of members (18–56 years of age) who were identified as having persistent asthma and who were appropriately prescribed medication (≥ 1 prescription for inhaled corticosteroids, nedocromil, cromolyn sodium, leukotriene modifiers, or methylxanthines) during the measurement year
Breast cancer screening	Percenatge of women (40–69 years of age) who had a mammogram to screen for breast cancer during measurement or 1 prior year
Cervical cancer screening	Percenatge of women (18–64 years of age) who received one or more Pap tests to screen for cervical cancer during measurement or 2 prior years
Chlamydia screening in women	Percenatge of women (16–25 years of age) who were identified as sexually active and who had at least one test for chlamydia during measurement year
Childhood immunizations	Percenatge of children (2 years of age) who had at least three diphtheria/tetanus, hepatitis B, H influenza type B and polio vaccinations, and at least one MMR and one varicella vaccination by their second birthday
Diabetes care: LDL screening	Percenatge of members (18–75 years of age) with type 1 or type 2 diabetes who received LDL-C screening in the measurement year or 1 prior year
Diabetes care: HbA1c screening	Percenatge of members (18–75 years of age) with type 1 or type 2 diabetes who had HbA1c testing during measurement year
Appropriate treatment for children with upper respiratory infection	Percenatge of children (3 months to 18 years of age) with a diagnosis of URI and were not dispensed an antibiotic prescription on or within 3 days after the first encounter

independent variables. That is, we will observe a difference of zero between the R-squared measures from the two regressions. If, on the other hand, health plans exert some influence on HEDIS scores independent of the physician organizations with which they contract, there will be a difference between the two R-squared measures. One could reframe this question more generally, asking whether or not the addition of health plan controls improves the model fit. One could then use alternate statistics, such as the adjusted R-squared, to test the specification. We prefer the standard measure of R-squared since it is easily interpretable as the percent of variance explained by the independent variables. In any case, results obtained comparing the adjusted R-squared and the standard R-squared produce nearly identical conclusions.

Our second method asks the same fundamental question in a different way, this time by comparing results from the second regression, containing only plan controls and the third regression that contains controls for provider organizations and plans. We identify variations across plans in their observed HEDIS rates from the second regression, and then ask whether these differences can be explained away by controlling for the physician organizations with which the plans contract. The coefficients on the plan dummy variables in the second regression indicate differences across plans in the HEDIS measures. The coefficients on the plan dummy variables in the third regression indicate differences across plans, after the influence of physician organizations has been

statistically removed. By comparing the two sets of plan coefficients, we can observe whether or not accounting for the effects of physician organizations influences the observed differences across plans. If plans have no impact independent of the physician organizations with which they contract, any variations across plans observed in the first regression will be eliminated in the second regression. If plans do have an independent impact, variation across plans will persist even after the physician organization controls are added. We use t-statistics to test the equality of the plan coefficients estimated in the second and third regressions.

We conducted large numbers of statistical tests in this analysis. We investigated the use of Bonferroni adjustments, but they did not alter our conclusions, so we report traditional hypothesis test results.

Both of these analytic measures rely on ordinary least squares (linear) regression. Since we are analyzing rates of HEDIS service provision, which are derived from dichotomous trials at the level of individual patients, and that are thus constrained to vary between 0 and 100%, it is possible that logistic regression could provide a more precise fit to the data. We nonetheless use linear regression. Linear regression will produce statistically consistent (unbiased) estimates in this setting. Moreover, it has other advantages. In the case of the first analytic approach, an examination of the share of variation attributable to physician organizations and health plans cannot be conducted using logistic regression because logistic regression has no measure of variance explained comparable

to the *R*-squared of linear regression. We did some analyses using a logistic regression framework and comparing models with only physician organization controls and models that added plan controls. We computed differences in the 'pseudo *R*-squared' measure, which is a measure of goodness of fit, though it cannot be interpreted as a measure of the share of variance explained. We found the conclusions nearly identical to those we report. In the case of the second analytic approach, we continue to use linear regression since in our view it provides more easily interpretable information about the magnitude of variations across plans. We verified that our conclusions in the second analysis are not affected by the choice of analytic technique, specifically, we conducted comparable analyses using logistic regressions and reporting odds ratios across plans, and found results that were qualitatively nearly identical to those shown.

This analysis treats each measure separately. Since in many cases the same plan-group combination produces multiple measures, it may be that a modeling approach that explicitly incorporated this feature of the data, such as a hierarchical model, would perform better. In our data, however, the fact that a substantial number of the plan-group observations are missing at least one of the measures substantially complicates using hierarchical analysis. In our view, using as much of the sample as we can, but treating measures independently, provides the stronger results. We did perform exploratory analyses using the much smaller of plan-group observations with data on all nine measures using hierarchical analyses and found consistent results.

Results

Summary statistics

Table 2 reports the overall average HEDIS rates by measure, which range from 37% for Chlamydia screening in women

to 93% for appropriate medications for asthma in the pediatric population. These rates are generally comparable to rates reported by others for health plans in California during this time period (see supplementary material, technical appendix for a comparison). Table 2 also reports information about the extent of variation in the measures across the plan-group observations in our sample.

Analytic approach 1

Our first analytic approach examines the share of variation in HEDIS scores that can be explained by variation across physician organizations and, incrementally, by variations across health plans. Table 3 reports the shares of the variation in the measures explained by physician organization alone and by physician organization and health plan together. In every case, the incremental share explained by health plans is statistically significantly greater than zero, ranging across measures from 1 to 12% points. In all of the cases, the incremental amount of variation explained by health plans is smaller than the amount explained by physician organizations. That is, of variation in HEDIS scores across our plan-group combinations that can be explained by physician organizations and health plans, physician organizations explain the majority, but health plans always make at least some independent contribution.

Analytic approach 2

The second analytic approach we use examines variations across plans, and the extent to which observed variations across plans can be explained away by statistically accounting for the physician organizations with which plans contract. Figure 1 illustrates results of this approach for two example measures, appropriate medications for asthma in children and HbA1c screening for diabetic patients. The solid line represents the observed differences across the six study

Table 2 Observed HEDIS rates and variations across plan-group observations in HEDIS rates

Measure	*n* plan-group observations	*n* patients	Percentage completed (HEDIS rate)	Variation in HEDIS rate across plan-group observations		
				Standard deviation	10th percentile	90th percentile
Appropriate medications for asthma, age 5–17	226	6884	93	5	85	100
Appropriate medications for asthma, age 18–56	383	18 165	89	5	79	96
Breast cancer screening	717	279 616	71	7	55	78
Cervical cancer screening	719	736 544	73	8	56	80
Chlamydia screening in women	657	102 368	37	10	19	50
Childhood immunizations	659	161,844	69	12	37	83
Diabetes care: LDL screening	699	154 766	84	10	69	93
Diabetes care: HbA1c screening	699	154 766	74	11	55	85
Appropriate treatment for children with URI	673	115 800	81	10	60	94

Table 3 Percent of measure variance explained by physician organization and health plan

| Measure	Percentage variance explained by physician organization alone	Percentage variance explained by physician organization and health plan	Incremental share of variance explained by health plan (percentage points)	P-value for H0: incremental variance explained = 0
Appropriate medications for asthma, age 5–17	39	51	12	<0.001
Appropriate medications for asthma, age 18–56	27	31	4	0.009
Breast cancer screening	61	62	1	<0.001
Cervical cancer screening	70	73	3	<0.001
Chlamydia screening in women	66	67	1	0.002
Childhood immunizations	52	57	5	<0.001
Diabetes care: LDL screening	47	59	12	<0.001
Diabetes care: HbA1c screening	51	59	8	<0.001
Appropriate treatment for children with URI	78	83	5	<0.001
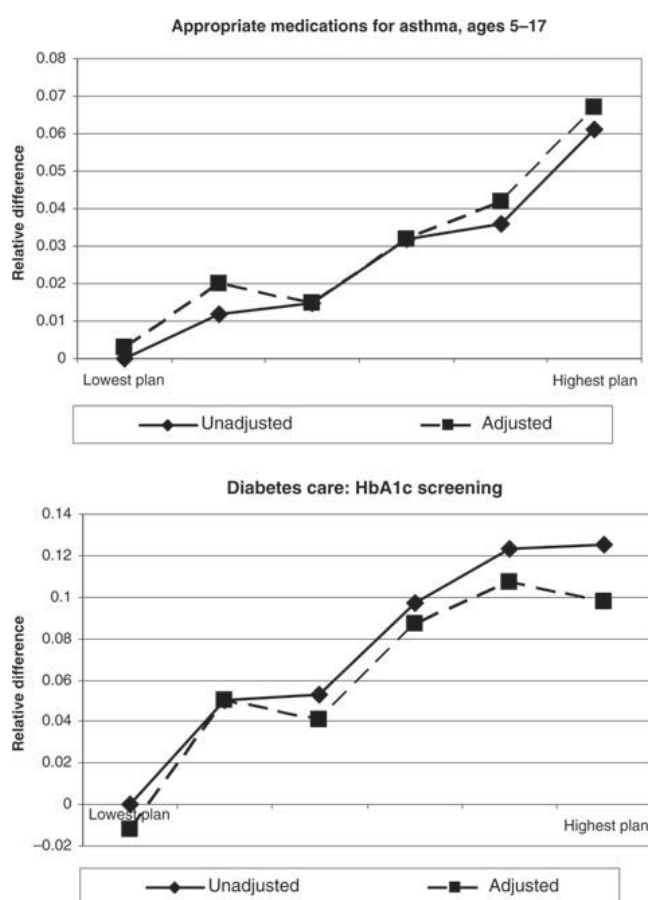
Figure 1 Variations in two example HEDIS scores across plans, before and after adjusting for provider groups.

health plans in HEDIS rates for these measures, before any adjustment is made for physician organizations. For each measure, we have organized the data so that the plans are ordered from lowest to highest unadjusted HEDIS rate. In the case of asthma medications for children, the plan with the highest rate has a rate about 6% points higher than the plan with the lowest rate. For HbA1c screening, the plan with the highest rate has a rate about 12% points higher than the plan with the lowest rate.

The dashed lines show the differences across the same plans after we adjust for the physician organizations with which the plans contract. If variation across plans is entirely a function of the physician organizations with which plans contract, the observed differences shown by the solid lines would be eliminated, and the dashed lines would be flat at 0. This does not happen. In both cases, the dashed lines are quite similar to the solid lines—there are still noticeable differences across plans even after controlling for physician organizations. In fact, the observed differences across plans, and the ordering of plans from lowest to highest rates, is not strongly affected by the addition of controls for physician organizations. This is not to say that it is exactly the same—there are a couple of instances for these two measures where differences between any two plans change by a percentage point or two, and a couple cases where this would affect the exact ordering from lowest to highest rate.

Table 4 reports results from this exercise for all measures. For each measure, the rate for the plan with the lowest rate in the unadjusted model is set as the baseline, and the other values are shown relative to that level. We tested the hypothesis that the unadjusted and adjusted differences are equal. Across 9 measures and across 6 plans per measure, there is no case in which adjusting for the physician organizations with which plans contract has a statistically significant impact on the measured differences across plans. All of the P-values are above 0.05, and all but one are above 0.10. As in Fig. 1, it is true that there are some cases where the variations across plans move somewhat with adjustment for physician

Table 4 Differences in HEDIS scores across health plans, before and after adjusting for physician organizations.[a]

Measure	Lowest plan					Highest plan
Appropriate medications for asthma, age 5–17						
Unadjusted	Baseline	+1.2	+1.5	+3.2	+3.6	+6.1
Adjusted	+0.3	+2.0	+1.5	+3.2	+4.2	+6.7
Appropriate medications for asthma, age 18–56						
Unadjusted	Baseline	+1.7	+2.0	+2.4	+2.6	+4.5
Adjusted	+0.5	+1.1	+2.0	+2.7	+3.1	+4.8
Breast cancer screening						
Unadjusted	Baseline	+0.6	+2.3	+3.2	+3.5	+4.0
Adjusted	+0.0	+1.4	+2.9	+2.7	+2.7	+4.0
Cervical cancer screening						
Unadjusted	Baseline	+0.6	+2.4	+3.2	+5.3	+5.5
Adjusted	+0.0	+0.4	+3.1	+2.9	+3.8	+4.6
Chlamydia screening in women						
Unadjusted	Baseline	+0.2	+0.5	+1.4	+2.7	+3.1
Adjusted	−0.1	+1.2	+0.5	+1.3	+3.2	+3.8
Childhood immunizations						
Unadjusted	Baseline	+2.4	+3.9	+7.2	+10.5	+11.8
Adjusted	+1.9	+2.4	+4.5	7.4	+12.4	+12.1
Diabetes care: LDL screening						
Unadjusted	Baseline	+5.2	+9.1	+11.2	+12.1	+12.3
Adjusted	−0.9	+5.2	+8.0	+10.5	+10.5	+11.2
Diabetes care: HbA1c screening						
Unadjusted	Baseline	+5.0	+5.3	+9.7	+12.3	+12.5
Adjusted	−1.2	+5.0	+4.1	+8.7	+10.7	+9.8
Appropriate treatment for children with URI						
Unadjusted	Baseline	+1.1	+5.3	+6.2	+7.5	+10.4
Adjusted	−0.3	+1.1	+4.6	+5.8	+6.7	+9.1

[a]Values shown are absolute percentage point differences from the unadjusted level of the measure in the baseline plan, defined to be the plan with the lowest unadjusted score on the given measure, e.g. +1.0 would indicate that a plan had a HEDIS rate 1.0% points higher than the HEDIS rate in the baseline case. Note: For each measure, for each plan, the hypothesis that the unadjusted difference is equal to the adjusted difference was tested. In no case a statistically significant difference was found. Test statistics had $P > 0.10$ for all cases except HbA1c testing in the highest plan, where the P-value was 0.06.

organizations, but the overall patterns across plans are not strongly affected by controlling for physician organizations.

Conclusions

We find significant differences among plans in reported HEDIS rates for nine services, even after controlling for the physician organizations with which the plans had contracted to provide care for their enrollees. Perhaps the most straightforward interpretation of the fact that plans have an influence on HEDIS scores that is statistically independent of the influence of providers is that there are activities that some plans engage in that result in higher HEDIS scores, independent of the physician organizations with which they contract. The persistent influence of plans contrasts with the hypothesis that health plans are 'too far' from patients to impact their care and thus would have minimal influence on HEDIS rates once provider choice was taken into account. This finding is particularly significant given that the physician

organizations were being rewarded for their performance on the measures we used through the IHA pay-for-performance program. Thus, one would expect that many of them were taking their own actions to maximize performance on these measures independent of the health plans.

The fact that the results of this analysis concur with one earlier analysis that used different data, albeit with a much smaller group of generally very large provider organizations, lend credence to the view that there are plan-specific components of HEDIS quality scores.

When we study the share of variation explained by different sources, we find evidence that the amount of variation attributable to variation across provider groups is much higher than the amount of variation across health plans. In some of the measures, most notably breast cancer screening, cervical cancer screening, and Chlamydia screening in women, the amount of variation explained by health plan after accounting for physician organization was quite small—physician organizations were much more important than health plans in explaining variation. In other cases, though, health plans

played a larger role. In the case of appropriate medications for asthma in children, variation across physician organizations explained 39% of the total variation, while variation across health plans contributed an explanation for an additional 12% of the variation. Health plans were also relatively larger contributors to the explanation of variation in the cases of the two diabetes care measures. Even in these cases, however, physician organizations explained the majority of the variation.

One interesting feature of the results is the fact that plans appear to explain a somewhat higher share of variance in HEDIS hybrid measures than in measures that are specified as administrative data only. We find this intriguing. It is unclear what would explain this pattern. Our analytic data are based on only administrative data, even for the hybrid measures, so that variable implementation of hybrid procedures by health plans could not be the explanation for this finding. Since variations in administrative data capture proficiency across plans should affect all measures, it seems unlikely that this would explain the pattern.

On their face, these findings suggest that plans can contribute to health-care quality. It is, in fact, quite plausible that plans could have an impact on HEDIS scores by influencing processes of care. For preventive services, there are a number of activities plans can include, which could improve scores, including efforts to educate patients, the development of reminder systems, and the use of financial incentives. Plans may also work to educate physicians, although for provider education efforts to play a role in our findings, their impact would have to be specific to an individual plan's patients, whereas it may be more likely that physicians would respond to education campaigns by changing their practices for all of their patients, not just those in the plan conducting the campaign. For chronic care, all the plans in our study employed formal disease management programs of various kinds to engage their members in better self-care as well as in obtaining periodic screening tests on the recommended schedule. Some of these programs even provide personal 'health coaches' free of charge to their members. Hence, we expected to find a plan effect on the diabetes measures; we also expected to find that some plans were more proficient than others. Regardless of the specific source, if some plans are better than others at improving the delivery of services measured by HEDIS rates, this would produce variation in plan scores that are independent of the providers with which plans contract.

There are also other potential explanations for the persistent variations across plans. One derives from the fact that plans collected and reported the data themselves, relying on administrative data systems. There may be differences in the ability of plans to collect and manage data. Plans that are not adept at collecting or maintaining data could end up with more cases with missing data and thus, because HEDIS specifications count cases with missing data as negative, with worse rates. This could lead us to observe variations in scores due to variations in data collection ability, rather than the ability of plans to actually influence processes of care. In other analyses (described in the technical appendix) we were able to conduct a limited investigation of this possibility by studying three measures using data on administrative data capture reliability derived from another reporting initiative. Results from this analysis suggested that there remained plan-related variation in HEDIS scores, even after we adjusted for observable variations in administrative data capture across plans.

There are two other factors that could play roles in our finding of persistent variations in HEDIS scores across plans after controlling for providers. We control for physician organizations, not individual physicians. All of the physician organizations represented in our sample are multi-physician entities, some of which are quite large. If it is really individual physicians or small groups of physicians that influence care, we might not observe this because our data on providers are not sufficiently detailed.

Finally, we may observe variations across plans because of differences in patient characteristics across plans. If, for example, some health plans attract the most prevention-oriented patients, this could drive better scores on some measures. We do not expect this to play a large role in our findings. The use of the HEDIS measure specifications should limit the potential for bias from patient selection to some extent. The official HEDIS specifications do not provide for risk adjustment, and HEDIS measures in general are defined in ways that aim to minimize bias across plans from variation in patient characteristics. Moreover, since all of the health plans we include are large—all with enrollments of more than 250 000—and have enrollment spanning multiple areas of the state, the possibility that some plans have identified and selected enrollees with specific, difficult-to-observe characteristics seems small. Finally, we note that our sample is derived from the same baseline data that the health plans and provider groups themselves have agreed to rely upon for payment purposes. Nonetheless, we cannot rule out the existence of some such variations in our data. Previous work by Zaslavsky and Epstein [12] examined the relationship between patient sociodemographic characteristics and HEDIS scores for California health plans. They found that patient characteristics were frequently not associated with HEDIS scores, but in some cases they did find associations. Of the measures they examined and that we also use, the breast cancer screening and cervical cancer screening measures had the strongest associations with patient demographics.

A related point might apply to provider groups. While our focus has been on whether or not there is variation across plans after accounting for provider groups, it should be noted that variations across groups could be generated by differential patient selection. Since groups are smaller, more geographically localized, and sometimes more reputationally differentiated, the potential for patient selection across groups seems stronger. Without data on specific patient characteristics, we are unable to specifically investigate these issues, but it does seem possible that there are variations across groups in demographic, socio-economic and cultural factors that would account for some of the variation in performance across groups.

Even in the presence of alternate explanations, we find the fact that there are persistent variations in plan scores after controlling for providers intriguing. This finding leaves distinctly open the possibility that plans do have an impact on

quality of care independent of providers, in sharp contrast with hypotheses that plans can have little to do with quality of care. Our findings suggest that continued attention to plans in quality data collection efforts is warranted. Although there are sometimes calls to concentrate data collection efforts on providers rather than plans, the existence of persistent variation across plans, even with caveats, suggests that moving away from plan-level data collection is premature. We would also note that an independent contribution of health plans to quality is not the only thing that makes collecting data at the plan level useful. It makes sense to collect data at the plan level in order to hold plans accountable to employers or others who purchase their services, even if some or all of the impact of plans were due to the set of providers that they contracted with rather than plan actions *per se*.

Our findings should not deter continued efforts to collect data on physician organizations. In fact, one can find in our results indications that provider groups have impacts on measured quality, independent of plans. Efforts to identify the best provider groups could continue to help patients select providers and help plans identify the best groups with which to contract.

Another question raised by the findings is whether it would still benefit plans to focus on careful selection of the providers with which they contract. Our analysis does not suggest that plans should rely on their own efforts to improve quality to the exclusion of efforts to identify and contract with the best provider groups. In fact, given the independent impacts of plans and providers, our results would suggest that improvements in HEDIS scores could be obtained if the best plans and providers were to systematically work together.

Finally, we stress that the measures we studied are for a specific set of services. It may be easier for plans to have an impact on some services than on others. For example, there may be more mechanisms by which plans can affect the provision of preventive care services than services associated with acute events or with other measures of quality. Plans may be able to influence primary care providers in different ways than specialists. More generally it is not clear that the presence of persistent plan effects here necessarily implies that they would also be observed if other services were studied.

In all, we believe that our finding of significant associations between plans and HEDIS rates, independent of providers, should lead to continued attention to collecting data on health plan quality and further discussion and analysis of the role of plans and providers in influencing care.

Supplementary material

Supplementary material is available at *INTQHC Journal* online.

References

1. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, D.C: Institute of Medicine, 2001.

2. Institute of Medicine, Committee on the National Quality Report on Health Care Delivery. *Envisioning the National Health Care Quality Report.* Washington, D.C: Institute of Medicine, 2001.

3. Galvin RS, McGlynn EA. Using performance measurement to drive improvement: a road map for change. *Med Care* 2003;**41**:I48–I60.

4. Milstein A, Lee TH. Comparing physicians on efficiency. *N Engl J Med* 2007;**357**:2649–51.

5. Galvin RS. Are performance measures relevant? *Health Aff* 1998;**17**:29–31.

6. McGlynn EA. Six challenges in measuring the quality of health care. *Health Aff* 1997;**16**:7–21.

7. Rosenthal MB, Landon BE, Normand SL *et al.* Pay for performance in commercial HMOs. *N Engl J Med* 2006;**355**:1895–902.

8. Galvin R, Milstein AM. Large employers' new strategies in health care. *N Engl J Med* 2002;**347**:939–42.

9. Mukamel DB, Mushlin AI. The impact of quality report cards on choice of physicians, hospitals, and HMOs: a midcourse evaluation. *Jt Comm J Qual Improv* 2001;**27**:20–7.

10. Chernew ME, Wochis W, Scanlon DP *et al.* Overlap in HMO Physician Networks. *Health Aff* 2004;**23**:91–101.

11. Baker LC, Hopkins D, Dixon R *et al.* Do health plans influence quality of care? *Int J Qual Assess Health Care* 2004;**16**:19–30.

12. Zaslavsky AM, Epstein AM. How patients' sociodemographic chacteristics affect comparisons of competing health plans in California on HEDIS quality measures. *Int J Qual Health Care* 2005;**17**:67–74.