Is Risk-Adjustor Selection More Important Than Statistical Approach for Provider Profiling? Asthma as an Example

I-Chan Huang, PhD, Francesca Dominici, PhD, Constantine Frangakis, PhD, Gregory B. Diette, MD, MHS, Cheryl L. Damberg, PhD, Albert W. Wu, MD, MPH

Objectives. *To examine how the selections of different risk adjustors and statistical approaches affect the profiles of physician groups on patient satisfaction.* ***Data sources.*** *Mailed patient surveys. Patients with asthma were selected randomly from each of 20 California physician groups between July 1998 and February 1999. A total of 2515 patients responded.* ***Research design.*** *A cross-sectional study. Patient satisfaction with asthma care was the performance indicator for physician group profiling. Candidate variables for risk-adjustment model development included sociodemographic, clinical characteristics, and self-reported health status. Statistical strategies were the ratio of observed-to-expected rate (OE), fixed effects (FE), and the random effects (RE) approaches. Model performance was evaluated using indicators of discrimination (C-statistic) and calibration (Hosmer-Lemeshow χ^2). Ranking impact of using different risk adjustors and statistical approaches was based on the changes in absolute ranking (AR) and quintile ranking (QR) of physician group performance and the weighted kappa for quintile ranking.* ***Results.*** *Variables that added significantly to the discriminative power of risk-adjustment models included sociodemographic (age, sex, prescription drug coverage), clinical (asthma severity), and health status (SF-36 PCS and MCS). Based on an acceptable goodness-of-fit (P > 0.1) and higher C-statistics, models adjusting for sociodemographic, clinical, and health status variables (Model S-C-H) using either the FE or RE approach were more favorable. However, the C-statistic (=0.68) was only fair for both models. The influence of risk-adjustor selection on change of performance ranking was more salient than choice of statistical strategy (AR: 50%–80% v. 20%–55%; QR: 10%–30% v. 0%–10%). Compared to the model adjusting for sociodemographic and clinical variables only and using OE approach, the Model S-C-H using RE approach resulted in 70% of groups changing in AR and 25% changing in QR (weighted kappa: 0.88). Compared to the Consumer Assessment of Health Plans model, the Model S-C-H using RE approach resulted in 65% of groups changing in AR and 20% changing in QR (weighted kappa: 0.88).* ***Conclusions.*** *In comparing the performance of physician groups on patient satisfaction with asthma care, the use of sociodemographic, clinical, and health status variables maximized risk-adjustment model performance. Selection of risk adjustors had more influence on ranking profiles than choice of statistical strategies. Stakeholders employing provider profiling should pay careful attention to the selection of both variables and statistical approach used in risk-adjustment.* ***Key words:*** *fixed effects model; physician group; random effects model; report cards; risk adjustment.* ***(Med Decis Making 2005;25:20–34)***

With the growth of managed care and health care costs, quality of care has become a major concern to payers and patients in the United States and worldwide. Performance measurement has the potential to increase provider accountability to patients, encourage health care managers to monitor and improve quality of care, and help consumers to choose providers or health plans. An increasing amount of informa-

DOI: 10.1177/0272989X04273138

Received 13 December 2003 from the Department of Health Policy and Management, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, Maryland (I-CH, AWW); Department of Biostatistics, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, Maryland (FD, CF); Department of Epidemiology, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, Maryland (GBD, AWW); Department of Medicine, School of Medicine, The Johns Hopkins University, Baltimore, Maryland (GBD, AWW); and Pacific Business Group on Health, San Francisco, California (CLD). Conference presentation: 1) 19th Annual Research Meeting of Academy of Health Services Research & Health Policy, 23–25 June

tion about the provider performance is being released to the public, often in the form of "provider profiles" or "report cards." However, accurate performance reporting depends on appropriate risk adjustment.[1,2]

Risk adjustment is intended to allow fair comparison in situations where it is difficult to randomly assign cases to different treatments or exposures.[3] Conventionally, risk adjustment emphasizes the concept of proper selection of risk adjustors. Most studies of risk-adjustor selection for profiling have been from clinical settings using clinical and administrative variables (e.g., Acute Physiology and Chronic Health Evaluation [APACHE], All Patient Refined Diagnosis-Related Group [APD-DRGs], Charlson Comorbidity Index, Computerized Severity Index [CSI], and Diagnostic Cost Groups-Hierarchical Coexisting Conditions [DCGs-HCCs]).[3] Only a few studies have focused on the impact of sociodemographic factors,[4–6] and empirical evidence is limited on the impact of different risk adjustors for profiling health plans or physician groups.[7,8]

There have been relatively few comparisons of different statistical approaches for risk adjustment.[9–16] For profiling of physicians or institutions, there are at least 4 approaches available. Perhaps the most frequently used approach is comparing the ratios of observed-to-expected (OE) performance rates (e.g., AMI mortality) across providers.[17] The 2nd approach is use of regression-based models with dummy variables for providers or institutions, using a multiple linear regression when the performance measure is continuous or a logistic regression when the performance is binary.[12] This method is called a "health plan fixed effects" model in the Consumer Assessment of Health Plans Study (CAHPS), a standard patient survey for assessing health plans performance in the United States.[18,19] One limitation of both of these approaches is that they ignore the effect of small numbers of cases within individual providers, thus increasing the variance in provider performance (regression-to-the-mean bias).[9] The random effects model (RE, or multilevel model) is an approach to deal with the situation in which some providers have smaller numbers of cases. This model can adjust for the regression-to-the-mean bias using shrinkage techniques. Furthermore, the RE model can be more appropriate than a fixed effects (FE) model because it takes into account the natural heterogeneity across providers, a key source of uncertainty of these analyses.[9,20,21] The most sophisticated but less used approach is the full Bayesian hierarchical model, which allows one to combine prior profiling information with the currently observed data to obtain the posterior probability distribution of profiling results.[9,11,22]

Previous studies have generally examined separately the impact of risk-adjustor selection and statistical approach.[3] We are not aware of any studies that have systematically evaluated the joint effect of different risk-adjustor selection schemes together with different statistical approaches. The goal of this study was to evaluate how the selection of risk adjustors and statistical model affects the profiles for physician groups. We used satisfaction with asthma care as the profiling indicator. Asthma is a useful example for profiling because it is one of the most common chronic conditions in the United States, and much of the mortality and morbidity associated with asthma are avoidable when adequately managed by providers.[23]

Specifically, we examined whether the performance ranking of physician groups was affected by 1) the selection of different risk adjustors, including sociodemographic, clinical, and health status variables, and 2) the use of a ratio of OE rate approach versus a FE versus a RE model. We also compared the favorable risk-adjustment models developed in this study with the standard risk-adjustment models used in profiling physicians or institutions (i.e., OE approach adjusting for sociodemographic and clinical variables) and health plans (i.e., CAHPS model—using FE model adjusting for age and health status). We hypothesized that the use of more sophisticated risk-adjustor selection schemes and statistical approaches would have a significant impact on ranking profiles.

METHODS

Sample and Data Collection

This study was conducted in 20 California-based physician groups participating in the 1998 Asthma Outcomes Survey (AOS). The AOS was initiated by the Pacific Business Group on Health (PBGH), a health care purchasing coalition in California, in conjunction with the HealthNet, a California-based health plan, for the purpose of evaluating, improving, and reporting on the quality of asthma care at the level of physician group.

2002, Washington, DC. 2) 19th International Conference of the International Society for Quality in Health Care (ISQua), 5–8 November 2002, Paris, France. Revision accepted for publication 13 October 2004.

Address correspondence and reprint requests to Albert W. Wu, MD, MPH, Department of Health Policy and Management, Bloomberg School of Public Health, The Johns Hopkins University, 624 North Broadway, Baltimore, MD 21205-1901; e-mail: awu@jhsph.edu.

Revised transcript of an invited presentation at the annual meetings of the Society for Medical Decision Making, Chicago, October 2003.

The cross-sectional component is described in this study.[24]

Experts have suggested that there are benefits of performance reporting at the level of the physician group or medical group.[2,25] On the west coast of the United States, large physician groups with full-risk contracts with HMOs are the main providers of medical care. Although health plans may set quality of care policy, most clinical decisions are made by the physicians within the physician groups. Health plans may be less able to affect the outcome of patients who receive care from physician groups. Evidence suggests that the use of a health plan as the unit of reporting can make performance differences appear very small, particularly within a given region or market.[26,27]

In this study, 20 participating physician groups were instructed to use administrative materials to identify all managed care patients with at least 1 asthma-related visit or admission in the outpatient, emergency, or inpatient setting (identified by ICD-9 code of 493.xx) between 1 January 1997 and 31 December 1997. Patients had to be continuously enrolled with the physician group for that calendar year. Patients were dropped if their addresses were unavailable (either through the administrative records or US Postal Service's National Change of Address process). From these eligible patients, the study randomly selected a sample of 650 patients in each physician group. If a physician group had fewer than 650 eligible patients, then all eligible patients were included in the survey sample.

Patient data were collected by mailed patient survey. The survey was fielded by the PBGH and the HealthNet using identical methodologies. The survey period began 14 July 1998 and ended 28 February 1999. The survey was administered by mail using a prenotification postcard, a mailed survey, a reminder postcard, 2 remailings of the survey, and a follow-up reminder phone call. A total of 2515 responses were obtained for a response rate of 32.2%.

Study Instrument

The survey was largely based on the "Health Survey for Asthma Patients" developed at the Johns Hopkins Health Services Research & Development Center for the Outcomes Management System Consortium Asthma Project of the Managed Health Care Association.[28–30] The survey included questions relating to patient characteristics, general health, asthma symptoms, effect of asthma on functioning, asthma medications and treatment, self-management knowledge and activities, access to care, and patient satisfaction. In this study, patients' satisfaction with asthma care was used as the performance indicator. In the survey instrument, patients were asked, "Overall, how would you rate the quality of care you received for your asthma during the past 12 months?" The satisfaction indicator was rated on a 5-point Likert-type scale (poor/fair/good/very good/excellent), which was dichotomized to greater satisfaction (very good/excellent) versus lesser satisfaction (poor/fair/good).

Risk-Adjustment Model Building

In theory, the characteristics of patients and physician groups are potential confounders that may influence physician group performance. For profiling, we would like to adjust for the effect of exogenous factors (mainly patient characteristics, i.e., those in which the providers have no influence, such as patient's age, sex, education, and baseline severity) rather than endogenous factors (mainly physician group characteristics, i.e., those characteristics that providers can influence, such as physician group specialty and number of ancillary staff).[31] Because the latter reflect quality of care, risk adjustment that accounts for characteristics of the physician groups may mask the true performance of physician groups. Adjusting for exogenous factors reflects that, for example, younger people tend to give lower responses to satisfaction questions rather than differences in care delivered to these groups.

In this study, we adjusted for age, sex, education level, type of health insurance, severity, number of comorbidities, and health status. All of the variables used in risk-adjustment models were collected from the patient survey. The study measured asthma severity using questions to approximate the National Heart, Lung, and Blood Institute severity strata (mild-intermittent, mild-persistent, moderate-persistent, and severe-persistent).[32] Classification of severity was based on patient reports of the frequency of symptoms (cough, sputum, wheezing, chest tightness, and shortness of breath), the frequency of nocturnal symptoms, and the chronicity of symptoms between attacks. Severity was determined by the greatest severity in the responses to any of these questions.[24] Comorbid conditions included rhinitis, sinusitis, chronic bronchitis, heartburn (gastroesophageal reflux), emphysema, and congestive heart failure. We also adjusted for prescription drug coverage because drug coverage is determined at the level of health plan rather than the physician group and therefore is an exogenous variable. In addition, lack of drug coverage could reduce access to health services and thus affect satisfaction with

care.[33,34] We did not adjust for patient race because evidence suggests that African American patients may receive poorer quality of care than white patients.[35]

These risk adjustors were grouped into 3 categories: 1) sociodemographic; that is, age, sex, education level, type of health insurance, and prescription drug coverage; 2) clinical; that is, severity and number of comorbidities; and 3) health status; that is, the SF-36 physical component score (PCS) and mental component score (MCS). We developed 3 risk-adjustment models: 1) Model S: including sociodemographic variables; 2) Model S-C: including sociodemographic and clinical variables; and 3) Model S-C-H: including sociodemographic, clinical, and health status variables. Comparisons among these models allowed us to evaluate the importance of these dimensions of risk adjustors on the physician group performance.

Statistical Modeling

We compared the risk-adjusted performance (i.e., patient satisfaction) across physician groups using 3 statistical methods: a ratio of OE rate approach, an FE model, and an RE model. The ratio of OE rate approach calculated the actual satisfaction rate with asthma care within each physician group divided by expected satisfaction rate of each physician group. The expected satisfaction rate for each participant was estimated using a logistic regression model, adjusting for patient characteristics based on all patients across 20 physician groups. The expected satisfaction rate for each physician group was estimated by summarizing the expected rates for subjects within a physician group and then divided by total subjects in that group.

The FE model using 19 dummy variables for 20 physician groups was as follows:

$$\text{logit } P(Y_{ij}=1)=\beta_0+\sum_{k=1}^{p}\beta_k(X_{kij}-\overline{X}_k)+\sum_{h=1}^{q}\beta_h X^*_{hij}+\sum_{i=1}^{19}\lambda_i Z_{ij}$$

where $P(Y_{ij}=1)$: probability of satisfaction with asthma care for subject j in physician group i; β_0: average log-odds of satisfaction across the 20 physician groups for a subject whose characteristics that are continuous equal to cohort average, and who is female, with high school & below education, public health insurance, and no prescription drug coverage; β_k: overall mean slope for continuous characteristic k; β_h: overall mean slope for binary characteristic h; X_{kij}: continuous characteristic k of subject j in physician group i; $\overline{X}_k$: overall mean of subjects' characteristic k, $k=1,\ldots p$; X^*_{hij}: binary characteristic h of subject j in physician group i; Z_{ij}: binary indicator of the physician group i for the subject j.

Applying the FE model, to estimate the risk-adjusted odds ratio of satisfaction (greater satisfaction v. lesser satisfaction) attributable to the ith physician group relative to the average group, we assigned a "1" into the binary indicator (Z_{ij}) for patients belonging to the 1 of the first 19 corresponding physician groups, and assigned a "–1" into all 19 binary indicators for patients belonging to the 20th physician group.[10] The performance of the physician groups 1~19 can be calculated by exponentiating the coefficient of the binary indicator (λ_i) of a specific physician group i. The performance for the 20th physician group can be calculated by exponentiating the sum of the negative coefficients of binary indicator ($\sum_{i=1}^{19}(-\lambda_i)$) for physician group 1~19.

For the RE model, conditional on the (p+1) – dimensional vectors of coefficients $\beta_i=(\beta_{i0},\beta_{i1},\ldots\beta_{ip})$ for the group ($i=1,\ldots 20$), logit $(Y_{ij}=1)=\beta_{ki}+\sum_{k=1}^{p}\beta_{ki},X_{kij}$. The β_{i0}, $i=1,\ldots 20$ are independently (and normally) distributed $N(\beta_0,\sigma_0^2)$. Usually the assumption on β_i is that it follows a multivariate normal distribution (MVN) with mean vector $\alpha=(\alpha_0,\alpha_1,\ldots\alpha_p)$ and covariance matrix T.[36] In this study, we used a random intercept model (i.e., a random parameter for β_{i0}). We assumed that the potential confounding effects of the patient-level risk factors on performance estimation are the same across 20 groups, and therefore we modeled the regression coefficients ($\beta_1,\beta_2,\ldots\beta_p$) as FE. The RE model was modeled as follows:

$$\text{logit } P(Y_{ij}=1\,|\,\beta_{0i})=\beta_{0i}+\sum_{k=1}^{p}\beta_k(X_{kij}-\overline{X}_k)+\sum_{h=1}^{q}\beta_h X^*_{hij}$$

$$\beta_{0i}\sim N(\beta_0,\sigma_0^2)$$

where $P(Y_{ij}=1\,|\,\beta_{0i})$: probability of satisfaction with asthma care for subject j in physician group i conditional to the random effect; β_{0i}: log-odds of satisfaction with asthma care specific to the physician group i; β_0: average log-odds of satisfaction across the 20 physician groups for a subject whose characteristics that are continuous and equal to cohort average, and who is female, with high school and below education, public health insurance, and no prescription drug coverage; β_k: overall mean slope for continuous characteristic k; β_h: overall mean slope for binary characteristic h; X_{kij}: continu-

ous characteristic k of subject j in physician group i; $\overline{X}_k$: overall mean of subject's characteristic k; X^*_{hij}: binary characteristic h of subject j in physician group i. The parameter $\exp(\beta_{0i})$ denotes the group-specific risk-adjusted odds ratio of satisfaction (greater satisfaction v. less satisfaction) with respect to all covariates that are continuous for the average patient, who is female, has high school and below education, public health insurance, and no prescription drug.[10]

An RE model is a more appropriate statistical approach than an FE model for estimating the performance of each physician group because it takes into account 1) the statistical uncertainty of each group-specific performance (within-group variance) and 2) the natural heterogeneity of the true group-specific performances (between-group variance). Therefore, under a random effect model, the group-specific performance takes into account both of these key sources of uncertainty.[9,11] In summary, we developed a 2-level model to better address the clustering effect of patients nested within a specific physician group. At level 1 (patient level), for the ith physician group, patient covariates are related to the probability of the dichotomous outcome by a multiple logistic regression. At level 2 (physician group level), the intercept term of logistic regression at level 1 is assumed to vary randomly across physician groups (i.e., RE), which allows for the odds of the outcome (for an average patient) to vary across physician groups. The random effects of physician groups were assumed to follow a normal distribution.[20,21] We did not adjust for physician group characteristics at level 2 because these factors are elements of quality of care of physician groups.[31]

Comparisons of Risk-Adjustment Models

We used discrimination and calibration to compare the performance of risk-adjustment models.[3] Discrimination measures the model's ability to distinguish between patients who have an outcome and those who do not (i.e., greater satisfaction v. lesser satisfaction). A model's discrimination can be measured by calculating the area under a receiver operator characteristic (ROC) curve (equivalent to the C-statistic). A model's C-statistic can range from 0.5 (no discriminative power) to 1.0 (perfect discriminative power). Separate C-statistics were compared for statistical differences across different risk-adjustment schemes using a univariate Z-test with the adjustment for correlations of areas under ROC curves in the standard errors, suggested by Hanley and McNeil.[37,38] Calibration measures the extent to which the model's predicted probability rate matches the observed rate for various risk groups of patients, which can be tested by using the Hosmer-Lemeshow goodness-of-fit test.[39] Models with smaller χ^2 values and larger P values have better goodness-of-fit.

For the RE model, there has been little research into use of the C-statistic and Hosmer-Lemeshow goodness-of-fit tests. In this study, for the Hosmer-Lemeshow goodness-of-fit test, we computed the estimated probability of satisfaction rate for each patient based on an RE model, rank ordered them to create deciles, and then statistically tested the expected and observed number of outcomes in each decile. For the C-statistics, we plotted the true-positive satisfaction rate (i.e., sensitivity) against the false-positive rate (i.e., 1-specificity) at cutoff points on the continuum of expected satisfaction rates and then calculated the area under the ROC curve.

Comparison of Ranking Impact

Rankings of physician groups were compared based on the odds ratio (OR) of performance for specific physician groups. Although using ranking profiling to compare provider performance has some methodological limitations (e.g., resulting in unreliable profiling results because there is a substantial overlap in terms of confidence interval for the estimated performance),[40–42] rank-based measures are very popular in the practice of comparing provider profiling.[5–7,43–49] In this study, 2 methods were used, including percentage changes in absolute ranking (AR) and percentage changes in quintile ranking (QR). Percentage changes in AR represented the portion of physician groups that changed in ranking, which was tested by the Spearman rank test.[5,6,43–47] A Spearman rank test $P < 0.05$ suggests that there is evidence to reject the null hypothesis of no correlation between ranking changes.

The percentage changes in QR are more useful for consumer choice or rewarding performance than the percentage changes in AR.[6] The QR represented the portion of physicians groups that moved into a different quintile of ranking, which was evaluated using a weighted-kappa statistic. The purpose of using the kappa statistic was to adjust for the effect of ranking changes due to chance. We used quadratic-weighted kappa rather than standard kappa (no weight) to reflect the ordinal nature (quintile) of the ranking scale.[50]

The 3 risk-adjustment selection schemes using RE model were compared to 4 reference models: 1) the ratio of OE rate without risk adjustment, 2) the FE approach without risk adjustment, 3) the ratio of OE rate adjusting for sociodemographic and clinical variables,

and 4) the FE model adjusting for age and health status. Approach 3) is the most commonly used model for profiling individual providers and institutions,[17] and approach 4) is recommended by AHRQ for profiling of health plans (i.e., the CAHPS model).[18,19] These comparisons would theoretically reflect the joint effect of different risk-adjustor selection schemes combined with different statistical approaches on provider profiling and provide readers with a full complement of perspectives regarding impact on profiling.

The statistical packages used in this study were SAS 8.1 with the Glimmix Macro for analyzing the RE model and STATA 7.0 for other analyses. The RE model was created using a linearization of the Lindstrom and Bates methodology, which produces estimated best linear unbiased predictors (EBLUPs) of the random effects.[51]

RESULTS

Characteristics of Physician Groups and Respondents

Of the 20 participating physician groups, 8 were located in Northern California, and 12 were in Southern California. The case number in each physician group ranged from 31 to 218, with a mean of 125.8 (*s* 56.0). Table 1 shows the characteristics of the 2515 patients who participated in this study. Patients ranged in age from 18 to 56 y, with a mean age of 39.9 y (*s* 9.5); 71.2% were female. In terms of clinical characteristics, 14.4% had mild intermittent asthma, 19.2% had mild persistent asthma, 49.3% had moderate persistent asthma, and 17.1% had severe persistent asthma.

Table 1 Characteristics of Patients with Asthma ($N = 2515$)

Characteristics	Percentage or Mean (*s*)
Sociodemographic	
Age, %	
Overall, $\bar{x}$ (*s*)	39.91 (9.45)
18–24	7.20
25–34	21.95
35–44	34.59
45–54	33.16
55 and above	3.10
Sex, %	
Males	28.83
Females	71.17
Education, %	
High school or below	18.41
College	65.29
Graduate	16.30
Health insurance status, %	
Private—through employer	69.07
Private—through self-purchase	24.77
Public—Medicare, Medicaid	1.35
Others	4.87
Drug insurance coverage, %	96.50
Clinical	
Asthma severity, %	
Mild intermittent	14.39
Mild persistent	19.24
Moderate persistent	49.30
Severe persistent	17.06
Number of comorbidity, $\bar{x}$ (*s*)	2.08 (1.43)
Health status-SF36 2 component scores	
Physical component score, $\bar{x}$ (*s*)	45.73 (10.31)
Mental component score, $\bar{x}$ (*s*)	47.43 (10.67)
Satisfaction with asthma care	
Greater satisfied with asthma care	55.35
Lesser satisfied with asthma care	44.65

Model Comparison Based on Different Risk Adjustors and Statistical Approaches

We first compared the importance of risk adjustors on patient satisfaction for physician group profiling. Table 2 shows that regardless of which risk-adjustor selection scheme and statistical approach were used, important risk adjustors included age, sex, asthma severity, and SF-36 PCS and SF-36 MCS ($P < 0.05$). Drug coverage was of a borderline significance ($P < 0.1$). Insignificant risk adjustors included education level, type of health insurance, and number of comorbid conditions ($P > 0.1$).

Comparison of risk-adjustor selection schemes based on discrimination showed that the Model S-C-H (sociodemographic, clinical, and health status dimensions) using either FE or RE approach had greater discriminative power (larger C-statistic) than Model S-C (sociodemographic and clinical dimensions) or Model S (sociodemographic only) (Table 2). The pairwise comparisons of discriminative power using a univariate *Z*-test across different risk-adjustment schemes given the same statistical approach were statistically significant ($P < 0.05$). In terms of calibration, 3 risk-adjustor selection schemes had a Hosmer-Lemeshow χ^2 value of $P > 0.1$, indicating acceptable calibration (Table 2).

Based on the above comparison, Model S-C-H using either the FE or RE approach appeared to be the favorable model for adjusting for physician group performance in terms of satisfaction with asthma care.

Table 2 Adjusted Odds Ratios of Satisfaction with Asthma Care Using Different Risk-Adjustor Selection Schemes and Statistical Approaches

Risk-Adjustor Selection Scheme and Statistical Approach	Model S			Model S-C			Model S-C-H		
	OE[b]	FE[b]	RE[b]	OE	FE	RE	OE	FE	RE
Sociodemographic dimension									
Age	1.03***	1.03***	1.03***	1.03***	1.03***	1.03***	1.03***	1.03***	1.03***
Sex (reference: males)	1.16	1.15	1.15	1.20*	1.19*	1.19	1.25**	1.25**	1.25**
Education (reference: high school & below)									
College	1.10	1.07	1.08	1.06	1.05	1.05	1.04	1.02	1.23
Graduate	1.33*	1.19	1.22	1.20	1.09	1.12	1.19	1.07	1.10
Health insurance (reference: public insurance)									
Private—through employer	1.30	1.16	1.19	1.14	1.02	1.05	0.99	0.88	0.91
Private—through self-purchase	1.60	1.34	1.40	1.43	1.20	1.25	1.26	1.07	1.11
Others	1.24	1.11	1.14	1.10	1.00	1.02	0.93	0.82	0.85
Drug insurance coverage (reference: no)	1.58*	1.45	1.48*	1.58*	1.46	1.49*	1.60*	1.48*	1.52*
Clinical dimension									
Asthma severity				0.80*	0.80***	0.80***	0.84***	0.85***	0.85***
Number of comorbidities				0.99	0.97	0.97	1.02	1.00	1.00
Health status dimension									
SF36 physical component score (PCS)							1.01*	1.01**	1.01*
SF36 mental component score (MCS)							1.02***	1.02***	1.02***
Model performance comparison									
C-statistic	0.60	0.64	0.64	0.62	0.65	0.65	0.66	0.68	0.68
Hosmer-Lemeshow χ^2 value	8.13	6.67	5.10	9.97	6.77	5.94	5.23	7.31	9.19
(P value)	(>0.1)	(>0.1)	(>0.1)	(>0.1)	(>0.1)	(>0.1)	(>0.1)	(>0.1)	(>0.1)

Note: Model S = adjusts for sociodemographic dimension; Model S-C = adjusts for sociodemographic and clinical dimensions; Model S-C-H = adjusts for sociodemographic, clinical dimensions, and SF-36 PCS and MCS; OE = observed/expected model; FE = fixed effects model; RE = random effects model.

*$P < 0.1$. **$P < 0.05$. ***$P < 0.01$.

Ranking Impact Comparison

Tables 3 through 5 show the ranking changes associated with different risk-adjustor selection schemes and statistical approaches. In general, within each statistical approach used, the more risk adjustors included in the model, the more changes there were in ranking profiles. Comparing risk-adjustor selection schemes (Models S, S-C, and S-C-H) to the null model, the AR changed 60% to 80% using the ratio of OE rate approach, 50% to 65% using the FE model, and 50% to 55% using the RE model (Table 3). The QR changed 15% to 25% (kappa: 0.8–0.88) using the ratio of OE rate approach, and 10% to 30% (kappa: 0.81–0.94) and 20% to 30% (kappa: 0.81–0.88) using the FE model and the RE model, respectively. It should be noted that although kappa is good for detecting change in performance ranking, we cannot say that the influence of risk adjustment is ignorable. For example, about 10% to 30% of provider performance in terms of QR change will be misconstrued if we use a less favorable risk-adjustment model.

The impact of statistical approach on ranking changes is shown in Table 4. Given the same risk-adjustment selection scheme, the OE approach versus FE model had slightly smaller ranking impacts (20%–45% for AR ranking changes and 5%–10% for QR ranking changes) than the OE approach versus RE model and FE versus RE models. By contrast, the OE approach versus RE model had slightly larger ranking impacts (35%–55% for AR ranking changes and 5%–10% for QR ranking changes) than the OE ratio approach versus FE model and FE versus RE models. The low percent change in QRs, together with the very high kappa observed, reinforces that the risk-factor selection was more important than choice of regression strategy.

Examining the joint impact of different risk-adjustor schemes and statistical approaches, Table 5 shows the comparisons of different risk-adjustor selection using the RE model to 4 reference models, including 1) the ratio of the OE rate approach without risk adjustment, 2) the FE approach without risk adjustment, 3) the ratio of the OE rate approach adjusting for sociodemographic and clinical variables, and 4) the FE model adjusting for age and health status (i.e., CAHPS model). In general, the results suggested that the changes in AR and QR ranged from 40% to 70% and 10% to 30% (kappa statistics: 0.81–0.94), respectively. Specifically, compared to the commonly used model for individual providers and institutions profiling (i.e., the OE ratio approach adjusting for sociodemographic and clinical variables), model S-C-H using the RE model had a 70% change in AR and 25% in QR. When compared to the model recommended by AHRQ to profile health plans (i.e., the CAHPS model), model S-C-H using the RE model had a 65% change in AR and 20% in QR. The agreement levels between the RE model using full-risk adjustment versus reference models 1), 2), 3), and 4) are shown in Table 6.

DISCUSSION

Inadequate risk adjustment has the potential to cause erroneous profiling and can mislead purchasers and patients, who think they are obtaining care from better providers than they actually are. Incorrect classification on profiling results can also unfairly penalize or reward providers. Using data from an asthma survey conducted by the PBGH and the HealthNet, we demonstrated the importance of risk-adjustment variable selection and statistical approach for physician group profiles on patient satisfaction. Different approaches had an important impact on the ranking of physician groups.

Our results confirm previous studies in which age, sex, and health status were significant risk adjustors for patient satisfaction.[52,53] Our results also suggest that asthma severity is an important risk adjustor for physician group profiling. Severity of illness has been widely accounted for in clinical studies[3]; however, it had been seldom emphasized for adjustment in health plan or physician group profiling. In addition, we found that drug coverage was of borderline significance as a risk adjustor. Drug coverage is important to physician group profiling because it can affect patients' access to health care, and its absence can reduce the satisfaction with health care. The risk adjustors that were not significantly associated with performance indicator are education level, type of health insurance, and number of comorbidities. However, because previous studies suggested that these variables can influence provider performance,[4,54–56] we included them in risk-adjustment models to assuage potential concerns about their effects.

Only a few studies have examined the impact of using different statistical approaches on profiling. The comparison between OE rate ratio approach or an FE model versus an RE model reflects the usual variance-bias tradeoff inherent in most of the statistical procedures.[57] By using an RE model, we estimated group-specific performance for groups with small case numbers by borrowing strength across physician groups (i.e., using shrinkage techniques to shrink their estimates toward the grand mean), obtaining biased but more efficient estimates of the group-specific performance.[9,20,21] In this study, because of varying distribu-

Table 3 Effect of Different Risk-Adjustor Selection Schemes on Percentage Change in Absolute Ranking, Quintile Ranking, and Agreement in Quintile Ranking[a]

	Statistical Approach					
	OE Model		FE Model		RE Model	
Risk-Adjustor Selection Scheme	% Change in Absolute Ranking[b]	% Change in Quintile Ranking (K_w)	% Change in Absolute Ranking	% Change in Quintile Ranking (K_w)	% Change in Absolute Ranking	% Change in Quintile Ranking (K_w)
Null model	Reference		Reference		Reference	
Model S	60%*	15% (0.88)	50%*	10% (0.94)	55%*	20% (0.88)
Model S-H	70%*	25% (0.80)	65%*	15% (0.88)	55%*	20% (0.88)
Model S-H-C	80%*	25% (0.80)	65%*	30% (0.81)	50%*	30% (0.81)

Note: OE = observed/expected model; FE = fixed effects model; RE = random effects model; Null model = no risk adjustment; Model S = adjusts for sociodemographic dimension; Model S-C = adjusts for sociodemographic and clinical dimensions; Model S-C-H = adjusts for sociodemographic, clinical dimensions, and SF-36 PCS and MCS; K_w = weighted-kappa statistic.

a. Comparing different statistical approaches using the same risk-adjustor selection scheme.

b. Spearman rank test for % change in absolute ranking.

*$P < 0.05$.

Table 4 Effect of Different Statistical Approaches on Percentage Change in Absolute Ranking, Quintile Ranking, and Agreement in Quintile Ranking[a]

	Statistical Approach					
	OE v. FE Model		FE v. RE Model		OE v. RE Model	
Risk-Adjustor Selection Scheme	% Change in Absolute Ranking[b]	% Change in Quintile Ranking (K_w)	% Change in Absolute Ranking	% Change in Quintile Ranking (K_w)	% Change in Absolute Ranking	% Change in Quintile Ranking (K_w)
Null model	20%*	5% (0.94)	50%*	10% (0.94)	35%*	5% (1.00)
Model S	25%*	10% (0.94)	35%*	0% (1.00)	45%*	10% (0.94)
Model S-C	20%*	10% (0.94)	35%*	10% (0.94)	45%*	10% (0.94)
Model S-C-H	45%*	10% (0.94)	40%*	0% (1.00)	55%*	10% (0.94)

Note: OE= observed/expected model; FE = fixed effects model; RE = random effects model; Null model = no risk adjustment; Model S = adjusts for sociodemographic dimension; Model S-C = adjusts for sociodemographic and clinical dimensions; Model S-C-H = adjusts for sociodemographic, clinical dimensions, and SF-36 PCS and MCS; K_w = weighted-kappa statistic.

a. Comparing different risk-adjustor selection schemes to null model using the same statistical approach.

b. Spearman rank test for % change in absolute ranking; *$P < 0.05$.

Table 5 Joint Effect of Different Risk-Adjustor Selection Schemes Combined with Statistical Approaches on Percentage Change, in Absolute Ranking, Quintile Ranking, and Agreement in Quintile Ranking

		RE Model	
Reference Model	**Risk-Adjustment Selection Scheme**	**% Change in Absolute Ranking[a]**	**% Change in Quintile Ranking (K_w)**
OE model without risk adjustment	Model S	60%*	20% (0.88)
	Model S-C	65%*	20% (0.88)
	Model S-C-H	65%*	30% (0.81)
FE model without risk adjustment	Model S	55%*	10% (0.94)
	Model S-C	60%*	20% (0.88)
	Model S-C-H	70%*	30% (0.88)
OE model adjusting for sociodemographic and clinical dimensions	Model S	60%*	20% (0.88)
	Model S-C	40%*	10% (0.94)
	Model S-C-H	70%*	25% (0.88)
FE model adjusting for age and health status (CAHPS model)	Model S	65%*	10% (0.94)
	Model S-C	65%*	10% (0.94)
	Model S-C-H	65%*	20% (0.88)

Note: OE = observed/expected model; FE = fixed effects model; RE = random effects model; Model S = adjusts for sociodemographic dimension; Model S-C = adjusts for sociodemographic and clinical dimensions; Model S-C-H = adjusts for sociodemographic, clinical dimensions, and SF-36 PCS and MCS; CAHPS model = adjusts for age and health status; K_w = weighted-kappa statistic.

a. Spearman rank test for % change in absolute ranking. *$P < 0.05$.

tions of case numbers (31–218 [*s* 56.0]) within physician groups, the standard profiling modeling would result in inaccurate estimates for small groups. In addition, evidence suggests that the application of shrinkage techniques could identify fewer statistical outliers of profiles.[10,12,40,45] However, caution should be taken, because the effect of shrinkage is also to shrink poor providers' outcomes toward the null value, implying that we may not identify those who provide poor quality of care.

In comparing the impact of different risk adjustors versus statistical approaches, our results suggest that the selection of risk adjustors played a more important role than the use of statistical strategies on profile ranking. It seems that the success of provider profiling appears to depend heavily on the conventional issues of selecting appropriate risk adjustors.[3,58] However, we could not conclude definitively that the choice of risk factors matters more than the selection of statistical models because, in theory, the relative importance of risk adjustment and statistical approach depends on the heterogeneity of physician groups and group size. If physician groups are more heterogeneous in terms of patient characteristics, risk adjustors will be more important than the statistical approach. If physician groups are less heterogeneous, particularly when the group variability is small relative to patient variability, then the statistical approach becomes more important. If the group size shrinks, then the statistical approach will play a more important role than risk adjustment. In addition, the study's conclusions could be changed radically if other covariates are introduced, if a different endpoint is considered, or if different practices are studied. Accordingly, more studies are needed to elucidate these issues.

Our findings have policy implications for profiling of health plans or physician groups. First, standard profiling systems that adjust only for age, sex, and health status[59–61] and apply standard statistical modeling[7,8,17] might be limited. When we compared our fully risk-adjusted RE model to the standard OE approach, adjusting for sociodemographic and clinical variables or the CAHPS model, profiling results suggested that the use of the standard risk-adjustment model may result in substantial ranking changes. Although in this study we cannot evaluate exactly how many providers are jeopardized and how many consumers are misled based on the standard model, these findings remind performance oversight agencies, such as the National Committee for Quality Assurance or the Center for Medicare and Medicaid Services, that the use of inappropriate risk variables and analytic methods will lead to improper ranking profiles.

Second, performance oversight agencies might consider using a clinical or administrative dataset in conjunction with survey data to collect import risk adjus-

Table 6 Agreement in Quintile Rankings between the Random Effects Model versus 4 Reference Models

	Random Effect Model Adjusting for Full Risk Adjustors[a]					
Reference Models	**Quintile 1**	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5**	**Agreement Level (%)**
OE model without risk adjustment						14/20 (70)
Quintile 1	2	2	0	0	0	2/4 (50)
Quintile 2	2	1	1	0	0	1/4 (25)
Quintile 3	0	1	3	0	0	3/4 (75)
Quintile 4	0	0	0	4	0	4/4 (100)
Quintile 5	0	0	0	0	4	4/4 (100)
FE model without risk adjustment						14/20 (70)
Quintile 1	2	2	0	0	0	2/4 (50)
Quintile 2	2	1	1	0	0	1/4 (25)
Quintile 3	0	1	3	0	0	3/4 (75)
Quintile 4	0	0	0	4	0	4/4 (100)
Quintile 5	0	0	0	0	4	4/4 (100)
OE model adjusting for sociodemographic and clinical dimensions						15/20 (75)
Quintile 1	2	1	0	0	0	2/4 (50)
Quintile 2	2	2	1	0	0	2/4 (50)
Quintile 3	0	1	3	0	0	3/4 (75)
Quintile 4	0	0	0	4	0	4/4 (100)
Quintile 5	0	0	0	0	4	4/4 (100)
FE model adjusting for age and health status (CAHPS model)						16/20 (80)
Quintile 1	2	2	0	0	0	2/4 (50)
Quintile 2	2	2	0	0	0	2/4 (50)
Quintile 3	0	0	4	0	0	4/4 (100)
Quintile 4	0	0	0	4	0	4/4 (100)
Quintile 5	0	0	0	0	4	4/4 (100)

Note: OE = observed/expected model; FE = fixed effects model; RE = random effects model.
a. Adjusts for sociodemographic, clinical dimensions, and SF-36 PCS and MCS.

tors, such as age, gender, prescription drug coverage, and health status. It is generally agreed that disease severity is desirable for all risk adjustment. By contrast, sociodemographic and health status variables are important risk adjustors for the non-clinically oriented outcomes like patient satisfaction. Some of the latter variables are less universally available than the clinical data available from the medical record. However, we should acknowledge the costs of conducting a patient survey. According to the CAHPS reports, the costs for the CAHPS survey are $20 to $40 per completed interview, which is more expensive than collecting data using patient discharge ($17 per record).[62,63] Therefore, we would recommend that data from patient surveys be used where patient reported outcomes or patient satisfaction is the outcome of interest and if the budgets permit.

Our study has some limitations. First, there was a low response rate to the patient surveys, and the response rates differed somewhat across physician groups despite not being statistically different. A lower response rate (35% to 50%) is a common phenomenon in satisfaction surveys, especially using a mailed survey.[64] The impact of lower response rate on performance comparison among providers depends on whether the satisfaction score between respondents and nonrespondents is similar. We would like to be able to compare respondents with nonrespondents on other characteristics. Unfortunately, we do not have these data. We believe that the low response rate would be more likely to affect the estimates of satisfaction for different groups and perhaps their ranking. However, it seems unlikely to affect the comparison of relative merits of methodology in provider profiling, including the impact of different risk adjustors and statistical approaches.

Second, in this study, the models were developed using a homogeneous population with a high educational level and a high percentage privately insured. It is likely that risk-adjustment models using a more heterogeneous population will perform much better to discriminate satisfied versus dissatisfied participants (i.e., higher C-statistics) than a model using a homogeneous population. However, evidence did not suggest that the use of homogeneous populations will affect the changes of ranking profiles.

Third, the differences between the models were small. Notably, the discriminative power (C-statistic) of all risk-adjustment models was less than the agreed acceptable level of 0.70 to 0.80. The lower discriminative power probably reflects the fact that adjustment using basic sociodemographic (age, sex, education level, type of insurance, and prescription status), clinical (severity and comorbidity), and health status variables may not be enough. However, we believe that this study should instigate further work in this area, including the possible need to adjust for additional risk adjustors, such as income, family size, or context and market characteristics (e.g., health plan or physician group penetration rate).[47,65]

Fourth, we only used patient satisfaction as the performance indicator. Although patient satisfaction has been widely used as an indicator to compare performance of health care delivery systems,[66,67] it represents only one aspect of performance or quality of care. Further studies need to examine the impact of other indicators such as process or outcome to reflect the impact on provider profiling.[1]

Fifth, evidence regarding the impact of risk adjustor and statistical approach on ranking profiles of physician groups was based on a single disease (i.e., asthma) and used data collected from 20 physician groups from a single state (California). Therefore, we cannot be certain that our results will be generalizable to other conditions or states.

Finally, the data collected in this study were cross-sectional. Therefore, the results only provide a point-in-time report card and cannot be used to assess quality improvement over time. If the data were longitudinal, then within-group changes over time would be more important to examine than the absolute rank.

In conclusion, the evaluation of risk-adjustment techniques to support provider profiling is complicated but critical. We found that selection of both risk adjustors and statistical approaches causes ranking changes in physician group profiling. The large shifts in rankings that we observed suggest that current risk-adjustment methods for profiling health plans are imperfect. Administrators, researchers, and policy makers engaging in provider profiling should take care to adjust for proper risk adjustors and apply appropriate statistical approaches.

REFERENCES

1. Brook RH, McGlynn EA, Cleary PD. Quality of health care. Part 2: measuring quality of care. N Engl J Med. 1996;335:966–70.

2. Epstein AM. Rolling down the runway: the challenges ahead for quality report cards. JAMA. 1998;279:1691–6.

3. Iezzoni LI. Risk Adjustment for Measuring Healthcare Outcomes. Chicago (IL): Health Administration Press; 1997.

4. Fiscella K, Franks P. Influence of patient education on profiles of physician practices. Ann Intern Med. 1999;131:745–51.

5. Hargraves JL, Wilson IB, Zaslavsky A, et al. Adjusting for patient characteristics when analyzing reports from patients about hospital care. Med Care. 2001;39:635–41.

6. Alter DA, Austin PC, Naylor CD, Tu JV. Factoring socioeconomic status into cardiac performance profiling for hospitals: does it matter? Med Care. 2002;40:60–7.
7. Zaslavsky AM, Hochheimer JN, Schneider EC, et al. Impact of sociodemographic case mix on the HEDIS measures of health plan quality. Med Care. 2000;38:981–92.
8. Elliott MN, Swartz R, Adams J, Spritzer KL, Hays RD. Case-mix adjustment of the national CAHPS benchmarking data 1.0: a violation of model assumptions? Health Serv Res. 2001;36:555–73.
9. Christiansen CL, Morris CN. Improving the statistical approach to health care provider profiling. Ann Intern Med. 1997;127:764–8.
10. DeLong ER, Peterson ED, DeLong DM, Muhlbaier LH, Hackett S, Mark DB. Comparing risk-adjustment methods for provider profiling. Stat Med. 1997;16:2645–64.
11. Normand SLT, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: issues and applications. J Am Stat Assoc. 1997;92:803–14.
12. Leyland AH, Boddy FA. League tables and acute myocardial infarction. Lancet. 1998;351:555–8.
13. Sullivan LM, Dukes KA, Losina E. Tutorial in biostatistics. An introduction to hierarchical linear modelling. Stat Med. 1999;18:855–88.
14. Austin PC, Naylor CD, Tu JV. A comparison of a Bayesian vs. a frequentist method for profiling hospital performance. J Eval Clin Pract. 2001;7:35–45.
15. Bronskill SE, Normand SL, Landrum MB, Rosenheck RA. Longitudinal profiles of health care providers. Stat Med. 2002;21:1067–88.
16. Austin PC, Alter DA, Tu JV. The use of fixed- and random-effects models for classifying hospitals as mortality outliers: a Monte Carlo assessment. Med Decis Making. 2003;23:526–39.
17. Salem-Schatz S, Moore G, Rucker M, Pearson SD. The case for case-mix adjustment in practice profiling. When good apples look bad. JAMA. 1994;272:871–4.
18. Agency for Health Care Policy and Research. CAHPS Implementation Handbook. Rockville (MD): Agency for Health Care Policy and Research; 1997.
19. Agency for Health Care Policy and Research. CAHPS TM 2.0 Survey and Reporting Kit. Rockville (MD): Agency for Health Care Policy and Research; 1999.
20. Leyland AH, Goldstein H. Multilevel Modelling of Health Statistics. Chichester (UK): Wiley; 2001.
21. Raudenbush SW, Bryk AS. Hierarchical Linear Models: Applications and Data Analysis Methods. Thousand Oaks (CA): Sage; 2002.
22. Austin PC. A comparison of Bayesian methods for profiling hospital performance. Med Decis Making. 2002;22:163–72.
23. U.S. Centers for Disease Control and Prevention. Forecasted state-specific estimates of self-reported asthma prevalence—United States, 1998. MMWR Morb Mortal Wkly Rep. 1998;47:1022–5.
24. Masland M, Wu AW, Diette GB, Dominici F, Skinner EA. The 1998 Asthma Outcomes Survey. San Francisco (CA): Pacific Business Group on Health; 2000.
25. Goldfield N. Physician Profiling and Risk Adjustment. Gaithersburg (MD): Aspen; 1999.
26. Meng YY, Leung KM, Berkbigler D, Halbert RJ, Legorreta AP. Compliance with US asthma management guidelines and specialty care: a regional variation or national concern? J Eval Clin Pract. 1999;5:213–21.
27. Safran DG. Measuring, monitoring and reporting functional health outcomes: opportunities and challenges in a bold national initiative. Int J Qual Health Care. 2001;13:7–8.
28. Steinwachs DM, Wu AW, Skinner EA. How will outcomes management work? Health Aff (Millwood). 1994;13:153–62.
29. Diette GB, Wu AW, Skinner EA, et al. Treatment patterns among adult patients with asthma: factors associated with overuse of inhaled beta-agonists and underuse of inhaled corticosteroids. Arch Intern Med. 1999;159:2697–704.
30. Wu AW, Young Y, Skinner EA, et al. Quality of care and outcomes of adults with asthma treated by specialists and generalists in managed care. Arch Intern Med. 2001;161:2554–60.
31. Welch HG, Black WC, Fisher ES. Case-mix adjustment: making bad apples look good. JAMA. 1995;273:772–3.
32. National Heart, Lung, and Blood Institute. Guidelines for the Diagnosis and Management of Asthma: Expert Panel Report No. 2. Bethesda (MD): US Department of Health and Human Services; 1997.
33. Katz PP, Yelin EH, Leong A, Sonneborn DR. Health insurance characteristics and health care evaluations among persons with rheumatic diseases in California. Arthritis Rheum. 1999;42:2710–8.
34. Eisenberg JM, Power EJ. Transforming insurance coverage into quality health care: voltage drops from potential to delivered quality. JAMA. 2000;284:2100–7.
35. Kahn KL, Pearson ML, Harrison ER, et al. Health care for black and poor hospitalized Medicare patients. JAMA. 1994;271:1169–74.
36. Gatsonis CA, Epstein AM, Newhouse JP, Normand SL, McNeil BJ. Variations in the utilization of coronary angiography for elderly patients with an acute myocardial infarction. An analysis using hierarchical logistic regression. Med Care. 1995;33:625–42.
37. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29–36.
38. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983;148:839–43.
39. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: John Wiley; 2000.
40. Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. J R Stat Soc A. 1996;159:385–443.
41. Marshall EC, Spiegelhalter DJ. Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates. BMJ. 1998;316:1701–4.
42. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. JAMA. 1999;281:2098–105.
43. Aron DC, Harper DL, Shepardson LB, Rosenthal GE. Impact of risk-adjusting cesarean delivery rates when reporting hospital performance. JAMA. 1998;279:1968–72.
44. Ivanov J, Tu JV, Naylor CD. Ready-made, recalibrated, or remodeled? Issues in the use of risk indexes for assessing mortality after coronary artery bypass graft surgery. Circulation. 1999;99:2098–104.
45. Peterson ED, DeLong ER, Muhlbaier LH, et al. Challenges in comparing risk-adjusted bypass surgery mortality results: results from the Cooperative Cardiovascular Project. J Am Coll Cardiol. 2000;36:2174–84.
46. DiGiuseppe DL, Aron DC, Payne SM, Snow RJ, Dierker L, Rosenthal GE. Risk adjusting cesarean delivery rates: a comparison of hospital profiles based on medical record and birth certificate data. Health Serv Res. 2001;36:959–77.
47. Fiscella K, Franks P. Impact of patient socioeconomic status on physician profiles: a comparison of census-derived and individual measures. Med Care. 2001;39:8–14.
48. Johnson ML, Gordon HS, Petersen NJ, et al. Effect of definition of mortality on hospital profiles. Med Care. 2002;40:7–16.
49. Selim AJ, Berlowitz DR, Fincke G, et al. Risk-adjusted mortality rates as a potential outcome indicator for outpatient quality assessments. Med Care. 2002;40:237–45.
50. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford (UK): Oxford University Press; 1995.

51. Witte JS, Greenland S, Kim LL, Arab L. Multilevel modeling in epidemiology with GLIMMIX. Epidemiology. 2000;11:684–8.
52. Cohen G. Age and health status in a patient satisfaction survey. Soc Sci Med. 1996;42:1085–93.
53. Carlson MJ, Blustein J, Fiorentino N, Prestianni F. Socioeconomic status and dissatisfaction among HMO enrollees. Med Care. 2000;38:508–16.
54. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40:373–83.
55. Agency for Healthcare Research and Quality. Annual Report of the National CAHPS Benchmarking Database 2000. Rockville (MD): Agency for Healthcare Research and Quality; 2000.
56. Shen JJ, Wan TT, Perlin JB. An exploration of the complex relationship of socioecologic factors in the treatment and outcomes of acute myocardial infarction in disadvantaged populations. Health Serv Res. 2001;36:711–32.
57. Greene WH. Econometric Analysis. Upper Saddle River (NJ): Prentice Hall; 2003.
58. Greenfield S, Sullivan L, Silliman RA, Dukes K, Kaplan SH. Principles and practice of case mix adjustment: applications to end-stage renal disease. Am J Kidney Dis. 1994;24:298–307.
59. Mainous AG III, Talbert J. Assessing quality of care via HEDIS 3.0. Is there a better way? Arch Fam Med. 1998;7:410–3.
60. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? A review of the evidence. JAMA. 2000;283:1866–74.
61. Simon LP, Monroe AF. California provider group report cards: what do they tell us? Am J Med Qual. 2001;16:61–70.
62. Iezzoni LI. How much are we willing to pay for information about quality of care? Ann Intern Med. 1997;126:391–3.
63. Zaslavsky AM, Beaulieu ND, Landon BE, Cleary PD. Dimensions of consumer-assessed quality of Medicare managed-care health plans. Med Care. 2000;38:162–74.
64. Fowler FJ Jr, Gallagher PM, Stringfellow VL, Zaslavsky AM, Thompson JW, Cleary PD. Using telephone interviews to reduce nonresponse bias to mail surveys of health plan members. Med Care. 2002;40:190–200.
65. Jarman B, Gault S, Alves B, et al. Explaining differences in English hospital death rates using routinely collected data. BMJ. 1999;318:1515–20.
66. Davies AR, Ware JE Jr. Involving consumers in quality of care assessment. Health Aff (Millwood). 1988;7:33–48.
67. Rosenthal GE, Shannon SE. The use of patient perceptions in the evaluation of health-care delivery systems. Med Care. 1997;35: NS58–68.