Title

The impact of risk-adjustor selection and statistical approach on physician group profiles for satisfaction with asthma care

Running title

Physycian profiling for asthma

Authors

I-Chan Huang [*], PhD, Francesca Dominici [†], PhD,
Constantine Frangakis [†], PhD, Gregory B. Diette [‡, §], MD, MHS,
Cheryl L. Damberg [¶], PhD, Albert W. Wu [*, ‡, §], MD, MPH

Affiliation

[*] Department of Health Policy and Management, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, Maryland

[†] Department of Biostatistics, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, Maryland

[‡] Department of Epidemiology, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, Maryland

[§] Department of Medicine, School of Medicine, The Johns Hopkins University, Baltimore, Maryland

[¶] Pacific Business Group on Health, San Francisco, California

Correspondent author

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, USA
Phone: (410) 955-6567, E-mail: awu@jhsph.edu

Conference presentation

(1) 19th Annual Research Meeting of Academy of Health Services Research & Health Policy (Formerly AHSR Annual Meeting). June 23-25, 2002. Washington DC, USA.

(2) 19th International Conference of the International Society for Quality in Health Care (ISQua). November 5-8, 2002. Paris, France.

Title

The impact of risk-adjustor selection and statistical approach on physician group profiles for satisfaction with asthma care

Abstract

Objectives

To examine how the selection 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 2,515 patients responded.

Study design

A cross-sectional study. Patient satisfaction with asthma care was the performance indicator for physician group profiling. Candidate variables for risk-adjustment model building included sociodemographic, clinical characteristics, and self-reported health status. Statistical approaches were the fixed effects vs. the random effects models. Model performance was evaluated using indicators of discrimination (C-index) and calibration (Hosmer-Lemeshow χ^2). Ranking impact of using different risk adjustors and statistical approaches was based on the changes in absolute and quintile ranking of physician group performance, and the weighted kappa for quintile ranking. Models were also compared to the Consumer Assessment of Health Plans (CAHPS) risk-adjustment model.

Principle findings

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). A model that included sociodemographic, clinical, and health status variables had greater discrimination (C-index = 0.68) than models including sociodemographic and clinical variables (C = 0.65) or sociodemographic variables only (C = 0.64). Both variable selection scheme and statistical analytic approach resulted in changes in ranking of physician groups. Compared to the CAHPS model, the best model resulted in 45-60% of groups changing in absolute ranking, and 20-30% changing in quintile ranking (weighted kappa: 0.81-0.88).

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. Statistical approach had less impact on ranking profiles than selection of risk adjustors. Provider profiling stakeholders should pay careful attention to both the selection of variables and of statistical approach in risk-adjustment.

Key words

Fixed effects model; physician group; random effects model; report cards; risk adjustment

Introduction

With the growth of managed care and health care costs, quality of care has become a major concern to payers and patients in the US and around the world. 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 information 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. APACHE, APD-DRGs, CSI, DCGs-HCCs, DS, MedisGroups, CI, Charlson Comorbidity Index) [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 [3]. For provider profiling, the common approaches are the regression-based models that use a multiple linear regression, and a logistic regression with dummy variables for providers. 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 US [9,10]. One limitation of this approach is that it ignores the effect of small numbers of cases within individual providers, thus increasing the variance in provider performance (regression-to-the-mean bias) [11]. The random effects model (or multilevel model) is a technique that may be applied to data in which some providers have smaller numbers of cases. This model can adjust for the regression-to-the-mean bias using the shrinkage techniques [11-13].

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 US [14], and much of the death and morbidity associated with asthma are avoidable when adequately managed by providers [15].

Specifically, we examined whether the performance ranking of physician groups was affected by (1) the selection of different risk adjustors, (2) the use of a fixed effects vs. a random effects model, and (3) the use of different risk adjustors combined with the two statistical approaches. We also compared the risk-adjustment models developed in this study with the standard risk-adjustment model used in the Consumer Assessment of Health Plans Study (CAHPS). We expected 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 (PGBH), 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. The initial cross-sectional component is described in this study [16].

Experts have suggested that there are benefits of performance reporting at the level of the physician group or medical group [2,17]. On the West Coast of the US, 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 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 health plan as the unit of reporting can make performance differences very small, particular within a given region or market [18,19].

In this study, 20 participating physician groups were instructed to use administrative materials to identify all managed care patients with at least one asthma-related visit or admission in the outpatient, emergency or inpatient setting (identified by ICD-9 code of 493.xx) between January 1, 1997 and December 31, 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

U.S. 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 in July 14, 1998 and ended in February 28, 1999. The survey was administered by mail using a pre-notification postcard, a mailed survey, a reminder postcard, two re-mailings of the survey, and a follow-up reminder phone call. A total of 2,515 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 (OMS) Consortium Asthma Project of the Managed Health Care Association (MHCA) [20-22]. 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, patient's 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)" vs. "lesser satisfaction (Poor/Fair/Good)."

Risk-adjustment model building

When data are collected for quality improvement, we do not risk adjust because we are interested in detecting all differences that could be used to direct improvement efforts. However, when data are collected for performance comparisons, we want to err on the side of making fair comparisons. 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) [23]. 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 these 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 comorbidity, 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 (NHLBI) severity strata (mild-intermittent, mild-persistent, moderate-persistent, and severe-persistent) [15]. 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 [16]. Comorbid conditions included rhinitis, sinusitis, chronic bronchitis, heartburn (gastroesophageal reflux), emphysema, and congestive heart failure. We also adjusted for prescription drug coverage since 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 [24,25]. We did not adjust for patient race because evidence suggests that African American patients may receive poor quality of care than white patients [26].

These risk adjustors were grouped into three categories: (1) sociodemographic: age, sex, education level, type of health insurance, and prescription drug coverage, (2) clinical: severity and number of comorbidity, and (3) health status: the SF-36 physical component score (PCS) and mental component score (MCS). We developed three 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 compare the importance of these dimensions of risk adjustors on the physician group performance.

Statistical modeling

We compared the risk-adjusted performance (patient satisfaction) of physician groups by using two statistical methods: a fixed effects model and a random effects model. The fixed effects model used 19 dummy variables for 20 physician groups (physician group 1 as the reference group) was as follows:

$$\textbf{Logit P(}\mathrm{Y}_{ij}\textbf{=1) =}\ \beta_0 + \sum_{h=1}^{p} \beta_h \chi_{hij} + \sum_{j=1}^{19} \lambda_j \mathrm{Z}_{ij} + \varepsilon_{ij}$$

where, i: index of subject; j: index of physician group; χ_{hij} : characteristic h of subject i in the physician group j; Z_{ij} : binary indicator of the physician group j for the subject i; ε_{ij} : error term.

Applying the fixed effects model, the performance of physician groups can be calculated by exponentiating the coefficient of the dummy variable (λ_j) of a specific physician group, which represents the risk-adjusted odds ratio of satisfaction (greater satisfaction vs. lesser satisfaction) attributable to the jth physician group relative to the reference group [27].

For the random effects model, a two-level model was developed to better address the clustering effect of patients nested within a specific physician group. At level 1 (patient level), for the jth 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. random effects), 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 [12,13]. We did not adjust for physician group characteristics at level 2 because these factors are elements of quality of care of physician groups [23]. The random effects model was modeled as follows:

$$\textbf{Logit P (}\mathrm{Y}_{ij}\textbf{=1) = } \beta_0 + \sum_{h=1}^{p} \beta_h \mathrm{X}_{hij} + \mu_{0j}$$

where, i: index of subject; j: index of physician group; χ_{hij}: characteristic h of subject i in the physician group j; β_0: overall mean intercept adjusting for all physician groups; β_h: overall mean slope for subject characteristic h; μ_{oj}: random effect of physician group j.

Applying the random effects model, the performance of physician groups can be calculated by exponentiating the difference in the random effects of the j^{th} physician group and reference group, which represents the risk-adjusted odds ratio of satisfaction (greater satisfaction vs. lesser satisfaction) attributable to the j^{th} physician group relative to the reference group [28].

Comparisons of risk-adjustment models

We used the 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 vs. 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 using a univariate Z-test described by Hanley and McNeil [29,30]. 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 [27]. Models with smaller χ^2-values and larger p-values have better goodness-of-fit. We used the method developed by Horton et al. to test Hosmer-Lemeshow goodness-of-fit for the random effects model [31].

We also tested whether adding a specific risk-adjustment dimension to a previous risk-adjustment model (e.g. adding clinical dimensions to Model S that had only sociodemographic dimension) makes a significant contribution to model performance.

Comparison of ranking impact

Rankings of physician groups were compared based on the odds ratio (OR) of performance for specific physician groups. To date, although there is no consensus on how to quantify ranking impact on provider performance [32-34], rank-based measures are very popular in the practice of comparing provider profiling [5-7,35-41]. In this study, two 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 Spearman rank test [5,6,35-39]. A Spearman rank test $p<0.05$ suggests that there is evidence to reject the null hypothesis of no correlation between ranking changes.

In practice, 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 we 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 [42].

The three risk-adjustment models using the random effects model were compared to (1) the null model with no risk adjustment using the fixed effects model, and (2) the CAHPS model, which only adjusts for age and health status using the fixed effects model [9,10]. These comparisons would reflect the joint effect of different risk-adjustor selection schemes combined with different statistical approaches on provider profiling.

The statistical packages used in this study were SAS 8.1 with the Glimmix Macro for analyzing the random effects model [43], and STATA 7.0 for other analyses.

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 [SD: 56.0]. Table 1 shows the characteristics of the 2,515 patients who participated in this study. Patients ranged in age from 18-56 years with a mean age of 39.9 years [SD: 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.

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, the 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 Model S-C-H (sociodemographic, clinical, and health status dimensions) had greater discriminative power (larger C-statistic) than Model S-C (sociodemographic and clinical dimensions) (Table 2). Model S (sociodemographic only) had the lowest discriminative power. All pairwise comparisons of discriminative power were statistically significant ($p<0.05$). In terms of calibration, three risk-adjustor selection schemes had a Hosmer-Lemeshow χ^2–value of $p>0.05$, indicating acceptable calibration. Significance testing indicated that all dimensions (sociodemographic, clinical, and health status) were statistically significant ($p<0.05$), except Model 1 using the random effect model ($p=0.18$) (Table 2).

Based on the above comparison using different approaches, Model S-C-H appeared to be the best model for adjusting for physician group performance in terms of satisfaction with asthma care.

Ranking impact comparison

Tables 3-5 show the ranking changes associated with different risk-adjustor selection schemes and statistical approaches. Comparing risk-adjustor selection schemes (Model S, S-C, and S-C-H) to the null model, the absolute ranking (AR) changed 60%-65% using the fixed effects model, and 50%-55% using the random effects model (Table 3). The quintile ranking (QR) changed 15%-30% and 20%-30% using the fixed effects model and the random effects model, respectively. Specifically, Model S caused a larger change in absolute ranking (AR) for both fixed effects and random effects models. Model S-C-H caused a larger change in quintile ranking (QR) for both fixed effects and random effects models.

The impact of statistical approach on ranking changes is shown in Table 4. Comparing the random effects model to the fixed effects model, Model S showed a larger change in absolute ranking (AR) (45%; $p<0.05$) than the other models. In terms of quintile ranking (QR), the null model, Model S, and Model S-C-H had no changes in quintile ranking (0%). Consequently, the effect of statistical approach seems to have less impact than variable selection on risk-adjusted performance.

Examining the joint impact of different risk-adjustor selection schemes and statistical approaches, Table 5 shows that when compared different risk-adjustor selection schemes using the random effects model to the null risk adjustment using the fixed effects model (Reference 1), Model S-C-H using the random effects model had a larger change in quintile ranking (QR) (30% ($p<0.05$); K_w: 0.81). Table 5 also shows that comparing different risk-adjustor selection schemes using the random effects model to the CAHPS model using the fixed effects model (Reference 2), Model S-C-H using the random effects model caused a larger change in quintile ranking (QR) (30% ($p<0.05$); K_w: 0.81).

Discussion

Inadequate risk adjustment has the potential to cause erroneous profiling that can mislead consumers and unfairly penalize 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 [44,45]. 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 a borderline significant risk adjustor. Drug coverage is important to physician group profiling because it can affect patient's access to health care and its absence can reduce the satisfaction with health care. The risk adjustors that do not significantly predict performance indicator are education level, type of health insurance, and number of comorbidities. However, previous studies suggest that these variables can importantly influence provider performance [4,46-48]. We therefore included these variables in risk adjustment models to assuage the concerns about their effects on provider performance.

Only a few studies have examined the impact of using different statistical approaches on profiling [28,37,49,50]. Because we analyzed clustered data within physician groups, it is natural to consider among competing statistical approaches a logistic regression model with a random intercept for the physician group. A random effects model is in general a more appropriate approach than a fixed effects model because it takes into account the natural heterogeneity across physician groups, a key source of uncertainty of these analyses. Under a random effects model we estimate group-specific performance by borrowing strengths across physician groups, obtaining biased but more efficient estimates of the group-specific performance. The comparison between a fixed vs. a randoms effect model reflects the usual variance-bias tradeoff inherent in most of the statistical procedures [51]. Another preferred way of dealing with the problem when small samples are suspected is to use the Chamberlain conditional fixed effects logit instead of just putting dummy variables into a usual logit [52].

In this study, because of varying distributions of case numbers (31-218 [S.D. 56.0]) within physician groups, the standard regression-based profiling modeling would result in inaccurate estimates for small groups. Use of the random effects model with shrinkage techniques allows those groups with smaller case numbers to shrink their estimates toward the grand mean, thus leading more robust estimates [11-13]. In addition, the application of shrinkage techniques could identify fewer statistical outliers of profiles, and thus avoids penalizing providers who do not actually have poor performance [28,32,37,49].

In comparing the impact of different risk adjustors vs. statistical approaches, in our study the selection of risk adjustors played a more important role than the use of statistical approaches on profile ranking. It seems that the success of provider profiling appears to depend heavily on the conventional issues of selecting appropriate risk adjustors [3,53]. However, we could not definitely conclude that the choice of risk factors matters more than the fixed effects vs. random effects models because the relative importance of risk adjustment and statistical approach depend on the heterogeneity of physician groups and group size. If physician groups are more heterogeneous in terms of patient characteristics, the 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 different practices are studied.

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 [54-56], and apply standard regression modeling [9,10] might be limited. When we compared our fully risk-adjusted random effects model to the standard CAHPS model, profiling results suggested that the use of 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 NCQA or CMS (formerly HCFA) that the use of inappropriate risk variables and analytic methods will lead to improper ranking profiles.

Second, performance oversight agencies might consider using survey data in conjunction with clinical or administrative dataset to collect import risk adjustors such as age, gender, prescription drug coverage and health status. It is generally agreeable that disease severity is a university quantity for 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-$40 per completed interview, which is more expensive than collecting data using patient discharge ($17 per record) [57,58]. 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. A lower response rate (35-50%) is a common phenomenon on satisfaction survey, especially using a mailed survey [59]. The impact of lower response rate on performance comparison among providers depends on whether the satisfaction score between respondents and non-respondents is similar. We would like to be able to compare respondents with non-respondents 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, the differences between the models were small. Notable, the discriminative power (C-statistic) of all risk-adjustment models was less than the generally agreed acceptable level of 0.70-0.80. The lower discriminative power probably reflects the fact that adjustment using sociodemographic (age, sex, education level, type of insurance, and prescription status), clinical (severity and comorbidity), and health status variables that were demonstrated in our study may not be enough. To achieve better risk adjustment, further studies might require additional data, including income, family size, or context and market characteristics (such as health plan or physician group penetration rate) [39,60].

Third, in this study, 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 system [61,62], 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 [1,63].

Fourth, 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 using 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 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 provider profiling is complicated, but critical. We found that both selection of risk adjustors and statistical approaches cause 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. For health plan or physician group profiling, we recommend the use of sociodemographic, clinical, and health status variables to maximize the risk-adjustment model performance.

Reference List

1. Brook RH, McGlynn EA, Cleary PD. Quality of health care. Part 2: measuring quality of care. *N Engl J Med* 1996; **335**:966-970.

2. Epstein AM. Rolling down the runway: the challenges ahead for quality report cards. *J Am Med Assoc* 1998; **279**:1691-1696.

3. Iezzoni LI. *Risk adjustment for measuring healthcare outcomes*. Illinois, Chicago: Health Administration Press, 1997.

4. Fiscella K, Franks P. Influence of patient education on profiles of physician practices. *Ann Intern Med* 1999; **131**:745-751.

5. Hargraves JL, Wilson IB, Zaslavsky A, James C, Walker JD, Rogers G et al. Adjusting for patient characteristics when analyzing reports from patients about hospital care. *Med Care* 2001; **39**:635-641.

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**:60-67.

7. Zaslavsky AM, Hochheimer JN, Schneider EC, Cleary PD, Seidman JJ, McGlynn EA et al. Impact of sociodemographic case mix on the HEDIS measures of health plan quality. *Med Care* 2000; **38**:981-992.

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**:555-573.

9. Agency for Health Care Policy and Research (AHCPR). *CAHPS implementation handbook*. Rockville, Maryland: AHCPR, 1997.

10. Agency for Health Care Policy and Research (AHCPR). *CAHPS TM 2.0 survey and reporting kit*. Rockville, Maryland: AHCPR, 1999.

11. Christiansen CL, Morris CN. Improving the statistical approach to health care provider profiling. *Ann Intern Med* 1997; **127**:764-768.

12. Leyland AH, Goldstein H. *Multilevel modelling of health statistics*. Chichester, UK: Wiley, 2001.

13. Raudenbush SW, Bryk AS. *Hierarchical linear models applications and data analysis methods*. Thousand Oaks, California: Sage Publications, 2002.

14. US CDC. Forecasted state-specific estimates of self-reported asthma prevalence--United States, 1998. *MMWR Morb Mortal Wkly Rep* 1998; **47**:1022-1025.

15. National Heart, Lung, and Blood Institute. *Guidelines for the Diagnosis and Management of Asthma: Expert Panel Report No. 2*. Bethesda, Maryland: US DHHS, 1997.

16. Masland M, Wu AW, Diette GB, Dominici F, Skinner EA. *The 1998 asthma outcomes survey*. San Francisco, California: Pacific Business Group on Health, 2000.

17. Goldfield N. *Physician profiling and risk adjustment*. Gaithersburg, Maryland: Aspen, 1999.

18. 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**:213-221.

19. 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.

20. Steinwachs DM, Wu AW, Skinner EA. How will outcomes management work? *Health Aff (Millwood)* 1994; **13**:153-162.

21. Diette GB, Wu AW, Skinner EA, Markson L, Clark RD, McDonald RC 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**:2697-2704.

22. Wu AW, Young Y, Skinner EA, Diette GB, Huber M, Peres A et al. Quality of care and outcomes of adults with asthma treated by specialists and generalists in managed care. *Arch Intern Med* 2001; **161**:2554-2560.

23. Welch HG, Black WC, Fisher ES. Case-mix adjustment: making bad apples look good. *J Am Med Assoc* 1995; **273**:772-773.

24. 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**:2710-2718.

25. Eisenberg JM, Power EJ. Transforming insurance coverage into quality health care: voltage drops from potential to delivered quality. *J Am Med Assoc* 2000; **284**:2100-2107.

26. Kahn KL, Pearson ML, Harrison ER, Desmond KA, Rogers WH, Rubenstein LV et al. Health care for black and poor hospitalized Medicare patients. *J Am Med Assoc* 1994; **271**:1169-1174.

27. Hosmer DW, Lemeshow S. *Applied logistic regression*. New York, New York: Wiley, 2000.

28. DeLong ER, Peterson ED, DeLong DM, Muhlbaier LH, Hackett S, Mark DB. Comparing risk-adjustment methods for provider profiling. *Stat Med* 1997; **16**:2645-2664.

29. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**:29-36.

30. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; **148**:839-843.

31. Horton NJ, Bebchuk JD, Jones CL, Lipsitz SR, Catalano PJ, Zahner GE et al. Goodness-of-fit for GEE: an example with mental health service utilization. *Stat Med* 1999; **18**:213-222.

32. Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. *J R Statist Soc A* 1996; **159**:385-443.

33. Marshall EC, Spiegelhalter DJ. Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates. *BMJ* 1998; **316**:1701-1704.

34. 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. *J Am Med Assoc* 1999; **281**:2098-2105.

35. Aron DC, Harper DL, Shepardson LB, Rosenthal GE. Impact of risk-adjusting cesarean delivery rates when reporting hospital performance. *J Am Med Assoc* 1998; **279:**1968-1972.

36. 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**:2098-2104.

37. Peterson ED, DeLong ER, Muhlbaier LH, Rosen AB, Buell HE, Kiefe CI et al. Challenges in comparing risk-adjusted bypass surgery mortality results: results from the Cooperative Cardiovascular Project. *J Am Coll Cardiol* 2000; **36**:2174-2184.

38. 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**:959-977.

39. 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.

40. Johnson ML, Gordon HS, Petersen NJ, Wray NP, Shroyer AL, Grover FL et al. Effect of definition of mortality on hospital profiles. *Med Care* 2002; **40**:7-16.

41. Selim AJ, Berlowitz DR, Fincke G, Rosen AK, Ren XS, Christiansen CL et al. Risk-adjusted mortality rates as a potential outcome indicator for outpatient quality assessments. *Med Care* 2002; **40**:237-245.

42. Streiner DL, Norman GR. *Health measurement scales a practical guide to their development and use*. Oxford, UK: Oxford University Press, 1995.

43. Witte JS, Greenland S, Kim LL, Arab L. Multilevel modeling in epidemiology with GLIMMIX. *Epidemiology* 2000; **11**:684-688.

44. Cohen G. Age and health status in a patient satisfaction survey. *Soc Sci Med* 1996; **42**:1085-1093.

45. Carlson MJ, Blustein J, Fiorentino N, Prestianni F. Socioeconomic status and dissatisfaction among HMO enrollees. *Med Care* 2000; **38**:508-516.

46. 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**:373-383.

47. Agency for Healthcare Research and Quality (AHRQ). *Annual report of the national CAHPS benchmarking database 2000*. Rockville, Maryland: AHRQ, 2000.

48. 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**:711-732.

49. Leyland AH, Boddy FA. League tables and acute myocardial infarction. *Lancet* 1998; **351**:555-558.

50. Sullivan LM, Dukes KA, Losina E. Tutorial in biostatistics. An introduction to hierarchical linear modelling. *Stat Med* 1999; **18**:855-888.

51. Greene WH. *Econometric analysis*. Upper Saddle River, New Jersey: Prentice Hall, 2003.

52. Chamberlain G. Analysis of covariance with qualitative data. *Rev Econ Stud* 1980; **47**:225-238.

53. 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.

54. Mainous AG, III, Talbert J. Assessing quality of care via HEDIS 3.0. Is there a better way? *Arch Fam Med* 1998; **7**:410-413.

55. 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. *J Am Med Assoc* 2000; **283**:1866-1874.

56. Simon LP, Monroe AF. California provider group report cards: what do they tell us? *Am J Med Qual* 2001; **16**:61-70.

57. Iezzoni LI. How much are we willing to pay for information about quality of care? *Ann Intern Med* 1997; **126**:391-393.

58. Zaslavsky AM, Beaulieu ND, Landon BE, Cleary PD. Dimensions of consumer-assessed quality of Medicare managed-care health plans. *Med Care* 2000; **38**:162-174.

59. 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.

60. Jarman B, Gault S, Alves B, Hider A, Dolan S, Cook A et al. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999; **318**:1515-1520.

61. Davies AR, Ware JE, Jr. Involving consumers in quality of care assessment. *Health Aff (Millwood)* 1988; **7**:33-48.

62. Rosenthal GE, Shannon SE. The use of patient perceptions in the evaluation of health-care delivery systems. *Med Care* 1997; **35**:NS58-NS68.

63. Rubin HR, Pronovost P, Diette GB. The advantages and disadvantages of process-based measures of health care quality. *Int J Qual Health Care* 2001; **13**:469-474.

Table 1: Characteristics of patients with asthma (n=2,515)

Characteristics	Percentage or mean (SD)
Sociodemographic	
Age, %	
Overall, mean (SD)	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, mean (SD)	2.08 (1.43)
Health status–SF36 two component scores	
Physical component score, mean (SD)	45.73 (10.31)
Mental component score, mean (SD)	47.43 (10.67)
Satisfaction with asthma care	
Greater satisfied with asthma care	55.35
Lesser satisfied with asthma care	44.65

Table 2: Adjusted odds ratio of satisfaction with asthma care using different risk-adjustor selection schemes and statistical approaches

Risk-adjustor selection scheme and statistical approach	Model S [1]		Model S-C [1]		Model S-C-H [1]	
	FE [2]	**RE** [2]	**FE**	**RE**	**FE**	**RE**
Sociodemographic dimension						
Age	1.03***	1.03***	1.03***	1.03***	1.03***	1.03***
Sex (reference: males)	1.15	1.15	1.19*	1.19	1.25**	1.25**
Education (reference: high school & below)						
College	1.07	1.08	1.05	1.05	1.02	1.23
Graduate	1.19	1.22	1.09	1.12	1.07	1.10
Health insurance (reference: public insurance)						
Private---through employer	1.16	1.19	1.02	1.05	0.88	0.91
Private---through self-purchase	1.34	1.40	1.20	1.25	1.07	1.11
Others	1.11	1.14	1.00	1.02	0.82	0.85
Drug insurance coverage (reference: no)	1.45	1.48*	1.46	1.49*	1.48*	1.52*
Clinical dimension						
Asthma severity			0.80***	0.80***	0.85***	0.85***
Number of comorbidity			0.97	0.97	1.00	1.00
Health status dimension						
SF36 Physical component score (PCS)					1.01**	1.01*
SF36 Mental component score (MCS)					1.02***	1.02***
Model performance comparison						
C-statistic	0.64	0.64	0.65	0.65	0.68	0.68
Hosmer-Lemeshow X^2-value	6.67	6.63	6.77	2.03	7.31	5.94
(p-value)	(0.57)	(0.58)	(0.56)	(0.98)	(0.50)	(0.65)
Dimension significance test, p-value	<0.0001	0.18	<0.0001	<0.0001	<0.0001	<0.0001

[1] Model S: adjust for sociodemographic dimension; Model S-C: adjust for sociodemographic and clinical dimensions; Model S-C-H: adjust for sociodemographic, clinical dimensions, and SF-36 PCS and MCS

[2] FE: the fixed effects model; RE: the random effects model

[3] * p<0.1; ** p<0.05; *** p<0.01

Table 3: Effect of different risk-adjustor selection schemes on percentage change in absolute ranking, quintile ranking, and agreement in quintile ranking [1]

	Statistical approach					
	Fixed effects model			Random effects model		
Risk-adjustor selection scheme [2]	% Change in absolute ranking	% Change in quintile ranking	Agreement in quintile ranking (K_w) [3]	% Change in absolute ranking	% Change in quintile ranking	Agreement in quintile ranking (K_w) [3]
Null model		Reference			Reference	
Model S	65%	20%**	0.88	55%	20%**	0.88
Model S-H	60%	15%**	0.88	50%**	20%**	0.88
Model S-H-C	60%	30%**	0.81	50%	30%**	0.81

[1] Comparing different risk-adjustor selection schemes to null model using the same statistical approach

[2] Null model: no risk adjustment; Model S: adjust for sociodemographic dimension; Model S-C: adjust for sociodemographic and clinical dimensions; Model S-C-H: adjust for sociodemographic, clinical dimensions, and SF-36 PCS and MCS

[3] K_w: weighted-kappa statistic

[4] * Spearman rank test: $p<0.05$; ** Spearman rank test: $p<0.01$

Table 4: Effect of different statistical approaches on percentage change in absolute ranking, quintile ranking, and agreement in quintile ranking [1]

	Random effects model			
	Fixed effects model	Random effects model		
Risk-adjustor selection scheme [2]		% Change in absolute ranking	% Change in quintile ranking	Agreement in quintile ranking (K_w) [3]
Null model	Reference	35%	0%**	1.00
Model S	Reference	45%**	0%**	1.00
Model S-C	Reference	35%**	10%**	0.94
Model S-C-H	Reference	40%**	0%**	1.00

[1] Comparing the random effects model to the fixed effects model using the same risk-adjustor selection scheme

[2] Null model: no risk adjustment; Model S: adjust for sociodemographic dimension; Model S-C: adjust for sociodemographic and clinical dimensions; Model S-C-H: adjust for sociodemographic, clinical dimensions, and SF-36 PCS and MCS.

[3] K_w: weighted-kappa statistic

[4] * Spearman rank test: $p<0.05$; ** Spearman rank test: $p<0.01$

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

	Statistical approach			
	Fixed effects model	Random effects model		
Risk-adjustor selection scheme [1]		% Change in absolute ranking	% Change in quintile ranking	Agreement in Quintile ranking (K_w) [2]
Null model	Reference 1	---	---	---
Model S	---	70%	20%**	0.88
Model S-C	---	65%	20%**	0.88
Model S-C-H	---	65%	30%**	0.81
CAHPS model	Reference 2	---	---	---
Model S	---	60%	20%**	0.88
Model S-C	---	45%*	20%**	0.88
Model S-C-H	---	50%	30%**	0.81

[1] Null model: no risk adjustment; Model S: adjust for sociodemographic dimension; Model S-C: adjust for sociodemographic and clinical dimensions; Model S-C-H: adjust for sociodemographic, clinical dimensions, and SF-36 PCS and MCS; CAHPS model: adjust for age and health status

[2] K_w: weighted-kappa statistic

[3] * Spearman rank test: $p<0.05$; ** Spearman rank test: $p<0.01$