
Methods

The Performance of Administrative and Self-Reported Measures for Risk Adjustment of Veterans Affairs Expenditures

Matthew L. Maciejewski, Chuan-Fen Liu, Ann Derleth, Mary McDonnell, Steve Anderson, and Stephan D. Fihn

Objective. To evaluate the performance of different prospective risk adjustment models of outpatient, inpatient, and total expenditures of veterans who regularly use Veterans Affairs (VA) primary care.

Data Sources. We utilized administrative, survey and expenditure data on 14,449 VA patients enrolled in a randomized trial that gave providers regular patient health assessments.

Study Design. This cohort study compared five administrative data-based, two self-report risk adjusters, and base year expenditures in prospective models.

Data Extraction Methods. VA outpatient care and nonacute inpatient care expenditures were based on unit expenditures and utilization, while VA expenditures for acute inpatient care were calculated from a Medicare-based inpatient cost function. Risk adjusters for this sample were constructed from diagnosis, medication and self-report data collected during a clinical trial. Model performance was compared using adjusted R^2 and predictive ratios.

Principal Findings. In all expenditure models, administrative-based measures performed better than self-reported measures, which performed better than age and gender. The Diagnosis Cost Groups (DCG) model explained total expenditure variation ($R^2 = 7.2$ percent) better than other models. Prior outpatient expenditures predicted outpatient expenditures best by far ($R^2 = 42$ percent). Models with multiple measures improved overall prediction, reduced over-prediction of low expenditure quintiles, and reduced under-prediction in the highest quintile of expenditures.

Conclusions. Prediction of VA total expenditures was poor because expenditure variation reflected utilization variation, but not patient severity. Base year expenditures were the best predictor of outpatient expenditures and nearly the best for total expenditures. Models that combined two or more risk adjusters predicted expenditures better than single-measure models, but are more difficult and expensive to apply.

Key Words. Risk adjustment, health status, expenditures, VA, risk segmentation

Various approaches have been taken in observational and experimental studies to reduce differences among patient samples that may inherently be at

differential risk. Risk adjustment has also been used in an attempt to reduce biases in prospective payments because of systematic differences in patient risk (Pope et al. 1998). Risk adjustment measures have been developed from various data sources including: diagnoses from medical records or health insurance claims (e.g., Adjusted Clinical Group [ACG] and Diagnostic Cost Group [DCG]), self-reported demographic and health status (e.g., SF-36) data from patient surveys, and pharmaceutical indicators of clinical conditions (e.g., RxRisk).

Medicare uses the DCG diagnosis-based measure in setting the Average Adjusted Per Capita Cost capitation paid to Medicare+Choice health plans. The Chronic Illness and Disability Payment System (CDPS) diagnosis-based measure is currently being used in eight states to adjust payments for Medicaid patients (personal communication with Todd Gilmer). Heretofore, the nation's largest integrated health care system, the Veterans Health Administration of the Department of Veterans Affairs (VA) has not used any of these risk adjusters in allocating its Congressionally determined global budget to each of its 21 component networks, even though there are known to be substantial regional differences in health and utilization (Kazis et al. 1998; Au et al. 2001). VA provides an ideal setting to assess the performance of differing strategies to adjust for patient risk differences in observational or experimental studies because of the availability of extensive demographic, clinical, pharmacy, and economic data on several million veterans who use VA services. In addition, many studies conducted within VA collect patient-reported survey data that can be coupled with administrative data.

Studies of risk adjustment in VA have examined the correlation between self-reported health status and claims-based risk adjusters (Wang et al. 2001); risk adjustment of ambulatory encounters and the combination of inpatient length of stay and outpatient visit days (Rosen et al. 2001; Rosen et al. 2002); and risk adjustment of total costs (Sales et al. 2003). No study to date has compared the full array of administrative-, pharmacy-based, and self-reported risk adjusters on the same patient data set, partly because it is rarely possible to

Address correspondence to Matthew L. Maciejewski, Ph.D., Health Services Research and Development Northwest Center of Excellence, Veterans Affairs Puget Sound Health Care System, Seattle, WA. Matthew L. Maciejewski, Chuan-Fen Liu, Ph.D., Ann Derleth, M.S.P.H., Mary McDonnell, M.S., Steve Anderson, M.S., and Stephen D. Fihn, M.D., M.P.H., are with the Health Services Research and Development Northwest Center of Excellence, Veterans Affairs Puget Sound Health Care System, Seattle, WA. Matthew L. Maciejewski, Chuan-Fen Liu and Stephan D. Fihn are also with the Department of Health Services and Stephan D. Fihn is also with the Department of Medicine, University of Washington, Seattle, WA.

compile self-report, pharmacy and diagnosis data on a large sample of patients from multiple sites.

This study employs administrative, clinical, and survey data on 14,449 veterans that were collected as part of a randomized trial of a quality improvement intervention (Fihn et al. 2004). The purpose of this analysis was to compare the predictive accuracy of eight risk adjustment measures derived from routinely collected administrative data or from self-reported health status questionnaires in prospective models of outpatient, inpatient, and total expenditures of VA care. In addition, this study examined the performance of models incorporating more than one measure of risk.

This comparison of risk adjustment measures may assist researchers in choosing which measures are most suitable for inclusion in future prospective experimental and observational studies. In addition, use of the most predictive risk adjuster in experimental studies could generate more precise cost predictions for cost-effectiveness analysis. Finally, VA medical center directors could use risk adjustment to determine which types of patients to attract, based on prospective risk adjustment models that accurately predict expenditures of low cost patients.

METHODS

Sample

Data for our analysis came from the Ambulatory Care Quality Improvement Project (ACQUIP), which was a multicenter, randomized trial conducted in the General Internal Medicine (primary care) Clinics at seven VA medical centers (VAMCs) in six states during 1997–2000. The objective of the trial was to determine whether providing regular reports to providers about their patients' self-reported health status, combined with routine clinical data and information about clinical guidelines would enhance patient care outcomes. Eligible patients included those who were assigned a primary care provider and had had a least one primary care visit in the year prior to the study intervention. The study examined whether quality of care and guideline-concordant care would change as a result of these provider-based assessments of patient health.

Patients in the intervention group were sent an initial screening questionnaire 6 months prior to the study start about their chronic medical conditions. All patients were sent the same questionnaire every 6 months thereafter until the end of the study. Those who completed the initial health-

screening questionnaire were then sent SF-36 and other condition specific surveys at 6 and 3 months prior to baseline, at baseline, and at 6-month intervals after baseline for 2 years. Control group patients who responded to the initial screening questionnaire were sent the same surveys at baseline, 1 and 2 years (Fihn et al. 2004).

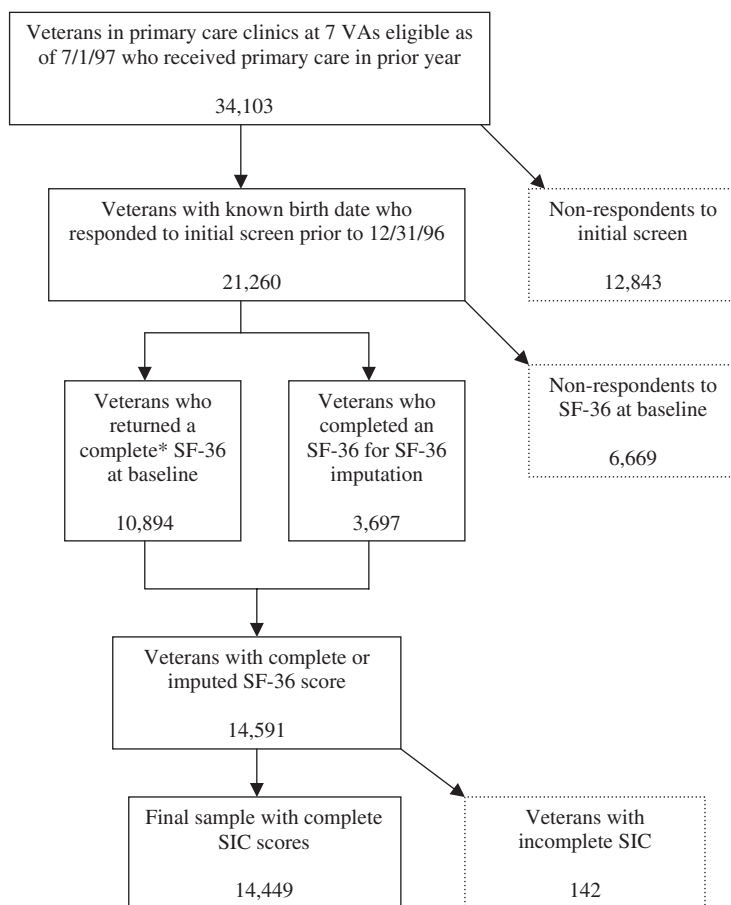
There were 34,103 subjects eligible for the ACQUIP study in this cohort, of whom 21,260 (62 percent) responded to the initial health condition screen before December 31, 1996 and for whom we had a recorded date of birth (Figure 1). Of these 21,260 respondents, 10,894 completed the SF-36 at baseline, permitting calculation of Physical Component Scores (PCS) and Mental Component Scores (MCS). Another 3,697 veterans from the intervention group had an incomplete or missing baseline SF-36 survey for which SF-36 values were imputed because these veterans had completed up to two SF-36 surveys prior to the baseline survey. If the baseline survey was missing and two surveys prior to the baseline were available, an average of these two was imputed for the missing baseline survey. Otherwise, we used a carry-forward of the SF-36 values from the closest survey prior to baseline. The imputed values did not significantly alter the mean PCS and MCS values in our sample, so those imputed observations were retained. Finally, 142 veterans were dropped from the sample because they were missing data necessary to compute the Seattle Index of Comorbidity (SIC) (Fan et al. 2002), yielding a final sample of 14,449. Patients who died in the baseline year were excluded from the final sample because they did not have prediction year expenditures for use in the prospective cost model.

The final step was to determine an approximate date for the end of the baseline year for risk adjustment and the beginning of the follow-up year for expenditures, so prospective expenditures in the subsequent year could be regressed on patient risk in the base year. The index date was calculated as the initial screening questionnaire return date plus a lag of 70 days to ensure that the baseline SF-36 survey represented patient health status in the baseline period.

Inpatient, Outpatient, and Total Expenditures

The dependent variable in this analysis was the total expenditures for VA care during a 1-year period from the index date based on the sum inpatient and outpatient expenditures. VA expenditures were calculated for each inpatient hospital admission and each outpatient visit, because patient-level expenditure data were not yet widely available for the study period (Barnett and

Figure 1: Flow Diagram of VA Sample Construction



* Enough responses to allow scoring of PCS, MCS

Rodgers 1999). As a result, expenditures had to be generated from utilization data that has been routinely collected by VA for many years in the Outpatient Care File (Kashner 1998). Outpatient encounter expenditures were calculated using an algorithm that assigns unit costs to specific clinics based on data from VA cost accounting system (the Cost Distribution Report), which in the past had been used to generate clinic-level expenditure estimates.

Inpatient expenditures per admission for acute care were calculated using a cost function developed by the VA Health Economics Resource

Center that generates patient-specific inpatient expenditures. These expenditures are based on age, sex, discharge disposition, bedsection(s), length of stay, and Medicare DRG weights, based on the Patient Treatment File that tracks inpatient utilization (Barnett and Rodgers 1999). Length of study was multiplied by the average cost per day for nine categories of nonacute days, based on the Cost Distribution Report to generate inpatient expenditures for nonacute care. Expenditures were measured in 1998 dollars.

We assumed that all outpatient visits in a specific clinic (e.g., primary care) had the same unit cost regardless of patient's severity or intensity of the visit. A similar assumption is applied to nonacute inpatient days in a specific care category (e.g., rehabilitation). The VA costing approach may be different from a pricing-based expenditure approach most commonly used in other health care systems. The VA costing approach may affect the predictive power of risk adjustment measures. However, VA expenditure distribution issues generally apply to all cost and expenditure estimation situations.

VA pharmacy expenditures were excluded because the listed medication price and quantity data in administrative databases were highly variable and difficult to reliably standardize without manual inspection of 47,000 unique medications listed for patients in this sample. Non-VA expenditures were excluded because they were not collected during the trial. For patients who died during the prediction year, mortality weights equal to the percentage of the year they were alive were created to annualize their partial-year VA expenditures. This weight was then used to down-weight these patients' expenditures in the regression models. The mortality weight for people who survived the entire time period was equal to 1.0 because these patients had a full year of expenditure data.

Risk Adjustment Measures

We evaluated eight different techniques for risk adjustment—two measures based on self-reported conditions or health status, five based on administrative data such as diagnoses or prescribed medications, and one based on prior year expenditures.

SF-36. The two SF-36 summary scores—Physical Summary Score (PCS) and Mental Summary Score (MCS)—were chosen because of their use in previous risk adjustment studies (Newhouse et al. 1989; Pope et al. 1998) and explanatory power in mortality, morbidity, and cost models. The PCS and MCS are standardized to a general population with a mean of 50

(standard deviation = 10) and higher scores indicate better health (Ware and Sherbourne 1992).

SIC. The SIC is an index based on chronic condition indicators, age, and smoking status listed in the initial questionnaire used to screen patients for various conditions. The SIC was developed to predict clinical events and was validated against 2-year mortality and hospital admission (Fan et al. 2002).

Charlson Index. The Charlson index was developed as a means of classifying the number and seriousness of co-morbid conditions to predict risk of death at 1-year based on diagnoses from medical charts (Charlson et al. 1987). We used the Deyo modification of the Charlson index that permits use of inpatient diagnoses from administrative data in place of chart abstraction (Deyo, Cherkin, and Ciol 1992). In some settings, the Charlson index has provided R^2 values of 8.7 percent in predicting one-year mortality and 16 percent in ten-year mortality but few studies have applied the Charlson score to costs (Beddhu et al. 2002).

Diagnosis Cost Groups—Hierarchical Cost Categories (DCG-HCC version 6.0). DCG-HCC model was originally developed to predict Medicare payments using inpatient diagnoses (Ellis et al. 1996), which was later expanded to include all outpatient diagnoses (Ash et al. 2000). The DCG-HCC model has outperformed other diagnosis-based risk adjusters and has been used adjust capitated payments to Medicare HMOs since 2000. The DCG-HCC model predicted VA outpatient visits (Rosen et al. 2001) and total VA expenditures (Sales et al. 2003) better than other risk adjusters, and has been used in Medicaid (Kronick et al. 2000) and Medicare cost analyses as well (Pope et al. 1998). The DCG-HCC model has also been shown to predict costs reasonably accurately in high cost cases (Ash et al. 2000), which is an important issue for veterans who seek care regularly.

ACG (Version 4.2). ACGs were originally developed using only outpatient diagnoses and were later expanded to include inpatient diagnoses (Weiner et al. 1995). ACGs have accurately predicted costs incurred by low-cost users and have been applied in to patient groups sponsored by VA (Rosen et al. 2001; Sales et al. 2003), Medicare (Weiner et al. 1995), and Medicaid (Kronick et al. 2000).

CDPS. The CDPS was developed to adjust Medicaid capitation payments using both inpatient and outpatient diagnoses (Kronick et al. 1996; Kronick et al. 2000). The CDPS extends the original Disability Payment System (DPS) classification to include additional diagnoses and uses a much larger database of Medicaid beneficiaries for definitions and validation (Kronick et al. 2000).

RxRisk-V. RxRisk (formerly the Chronic Disease Score) was originally developed in a staff-model HMO, using outpatient pharmacy data on drugs used to treat chronic diseases to classify patients into 29 overlapping disease conditions (Von Korff, Wagner, and Saunders 1992). RxRisk has been used to predict primary care visits, outpatient costs, hospitalizations, and total costs in different populations (Von Korff et al. 1992). RxRisk-V was tailored to the veteran population because of pharmacy data availability and completeness and was shown to explain 12 percent of VA total expenditures (Sales et al. 2003).

Prior Year Expenditures. Expenditures at baseline were also modelled for completeness and because they are highly correlated with expenditures incurred during the next year (Pope et al. 1998). Prior expenditures were routinely used to control for unobserved health status differences in expenditure analyses until conventional risk adjustment measures became widely available.

In addition to the eight single-measure models, we consider four models that combine multiple measures with the goal of improving explanatory power in terms of R^2 and predictive ratios in the expenditure quintiles.

Estimation

We estimated prediction year expenditures for DCG-HCC, ACG, CDPS, and RxRisk-V for each individual using the weights developed for the veteran population based on previous studies (Sales et al. 2003). Standardized risk scores were then calculated for the risk adjusters for each person by dividing the predicted expenditure measure by the average total expenditures of the study sample (\$5,739).

Prospective expenditure models, which regress expenditures in prediction year $t+1$ against each risk adjuster and age-sex categories at year t , were estimated. The 14 age-sex categories that were constructed and used in each model included men between 18-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+, and women in 18-44, 45-64, and 65+ age ranges.

Women make up a very small percentage of all veterans and were only 1.5 percent of the current sample, so their age categories were necessarily broad to ensure sufficient cell sizes. Age and sex categories were included in all prediction models.

We estimated the prospective expenditure models using weighted ordinary least squares and without adjusting for covariates that are known to explain variation in VA costs, such as whether the veteran received free care (Ash et al. 2000; Sales et al. 2003). Of the numerous statistics for comparing predictive accuracy that include adjusted R^2 , grouped R^2 , predictive ratios by expenditure groups, and predictive ratios by important demographic or clinical categories (Ash and Byrne-Logan 1998), we selected adjusted R^2 values. To further examine predictions might have been more accurate among patients with the lowest or highest expenditures, we calculated predictive ratios by quintiles of actual expenditures. Inpatient and outpatient expenditures were also analyzed to evaluate the relative predictive power of these risk adjusters for the two components of total expenditures.

RESULTS

The average age of the sample was 64.5 years and 98.5 percent were men (Table 1), making them somewhat older and more likely to be male than the average veteran seeking VA care (Rosen et al. 2001). The patients in our sample reported an average of two chronic conditions, with 63 percent having hypertension, 44 percent angina, 38 percent depression, 28 percent lung disease, and 26 percent diabetes (not shown). Total expenditures in the prediction year averaged \$5,739 for our sample with a maximum value of \$589,019. Inpatient expenditures averaged \$3,312 and were highly skewed because 83 percent of veterans had no inpatient expenditures while average expenditures for 17 percent who were hospitalized were nearly \$20,000. The mean PCS and MCS from the SF-36 were 33.3 and 45.7. The average SIC score was 6.4, reflecting a high level of comorbidity. Similarly, three of the four measures based on administrative data—ACG, CDPS, and Charlson score—also indicated a substantial burden of chronic illness. In every case, the mean was 50–400 percent greater than the standard deviation.

Three diagnosis-based measures—DCG, ACG, and CDPS—are highly correlated despite having different coding algorithms (Table 2). Surprisingly, DCG and RxRisk are also highly correlated in this sample even though RxRisk is based on pharmacy data. These four measures appear to have some

Table 1: Descriptive Statistics of Sample of Veterans Regularly Using Primary Care

	Mean (SD)	Minimum	Maximum
Prediction year total cost	5,739 (17,876)	0	589,018.60
Prediction year outpatient cost	2,428 (3015)	0	68,116.39
Prediction year inpatient cost	3,312 (17,135)	0	589,018.60
Age	64.46 (11.34)	21	97
Male (%)	98.48	–	–
PCS	33.32 (11.61)	6	73.54
MCS	45.65 (13.17)	2.35	75.80
SIC	6.40 (2.98)	0	23
Charlson score	0.24 (0.92)	0	14
ACG	1.07 (0.79)	0.05	3.30
DCG	1.28 (1.04)	0	10.58
CDPS	1.93 (1.37)	0.06	13.86
RxRisk	1.27 (0.83)	0	11.46
Base year total cost	4,383 (8825)	0	180,490.4
Base year outpatient cost	2,308 (2772)	0	72,319.38
Base year inpatient cost	2,076 (7750)	0	176,306.4
Sample size		14,449	

Note: Standard deviation in parentheses.

PCS, Physical Component Scores; MCS, Mental Component Score; SIC, Seattle Index of Comorbidity; ACG, Adjusted Clinical Group; DCG, Diagnostic Cost Group; CDPS, Chronic Illness and Disability Payment System.

underlying classification similarities, even though the data source or classification of specific diseases may vary across measures. Most other correlations ranged between 0.2 and 0.6 in absolute value.

In all instances, measures based on administrative data performed better than measures based on self-reported health, while age and gender explained the least variation (Table 3). Adjusted R^2 values were generally higher for the outpatient models and lowest for the inpatient models. In predicting total expenditures, the model incorporating PCS/MCS was the self-report model that exhibited the highest adjusted R^2 (1.8 percent), while of the models derived from diagnostic information, DCGs had the highest adjusted R^2 (7.2 percent). The pharmacy-based measure (RxRisk) had an adjusted R^2 of 4.7 percent and prior total expenditures had an adjusted R^2 of 4.8 percent. In predicting inpatient expenditures, the PCS/MCS model explained only 1.0 percent, the SIC 0.9 percent, and DCG 3.4 percent. The Charlson score was marginally more accurate ($R^2 = 2.4$ percent) than models based on CDPS, ACG, RxRisk, and prior inpatient expenditures.

The performance of outpatient cost models incorporating administrative data generally outperformed those based on self-report, while the DCG model

Table 2: Correlation Matrix of Eight Risk Adjustment Measures

	<i>MCS</i>	<i>PCS</i>	<i>RxRisk</i>	<i>SIC</i>	<i>Charlson</i>	<i>ACG</i>	<i>CDPS</i>	<i>DCG</i>	<i>Prior Total Cost</i>
PCS	0.18								
RxRisk	-0.21	-0.36							
SIC	-0.03	-0.29	0.34						
Charlson	-0.06	-0.12	0.35	0.20					
ACG	-0.17	-0.25	0.50	0.22	0.36				
CDPS	-0.12	-0.21	0.48	0.25	0.41	0.60			
DCG	-0.16	-0.26	0.58	0.29	0.55	0.73	0.72		
Prior total cost	-0.11	-0.14	0.39	0.12	0.52	0.50	0.42	0.58	
Prior outpatient cost	-0.15	-0.15	0.39	0.09	0.26	0.54	0.44	0.55	0.52

PCS, Physical Component Scores; MCS, Mental Component Score; SIC, Seattle Index of Comorbidity; ACG, Adjusted Clinical Group; DCG, Diagnostic Cost Group; CDPS, Chronic Illness and Disability Payment System.

was the best conventional risk adjuster ($R^2 = 20.6$ percent) followed closely by the ACG model ($R^2 = 18.0$ percent). Prior year outpatient expenditures, however, were more predictive than any other model ($R^2 = 41.9$ percent).

Table 3: Alternative Risk Adjustment Measures in Prospective Cost Models

	<i>Adjusted R²</i>		
	<i>Total Cost</i>	<i>Inpatient Cost</i>	<i>Outpatient Cost</i>
Age and gender alone (%)	0.22	0.24	0.12
Prior year cost	4.79	1.95	41.93
Self-report measures			
PCS+MCS (SF-36 summaries)	1.79	1.04	3.96
SIC	1.36	0.94	1.86
Administrative data-based measures			
Charlson (%)	3.56	2.40	4.63
ACG	5.14	2.22	18.00
DCG-HCC	7.16	3.43	20.63
CDPS	3.78	1.59	14.40
RxRisk	4.73	2.31	13.51
Combined models			
DCG, Prior cost	7.74	3.64	43.20
DCG, RxRisk	7.74	3.69	22.38
DCG, RxRisk, prior cost	8.23	3.88	43.82
DCG, RxRisk, prior cost, SF	8.33	3.95	43.90
Sample size		14,449	

Note: All self-report and diagnosis-based measures include age and gender.

PCS, Physical Component Scores; MCS, Mental Component Score; SIC, Seattle Index of Comorbidity; HCC, Hierarchical Cost Categories; ACG, Adjusted Clinical Group; DCG, Diagnostic Cost Group; CDPS, Chronic Illness and Disability Payment System.

Based on the correlation matrix and the variance explained by single-measure models in Table 3, four sets of measures were combined into multi-measure models because the correlations between measures ranged widely, measures had different data sources, and each model had at least one variable that predicted expenditure variation well. The four models were: (1) DCG and prior expenditure, (2) DCG and RxRisk, (3) DCG, RxRisk, and prior expenditure, and (4) DCG, RxRisk, prior expenditure, and PCS/MCS from the SF-36. These models are compared with the eight single-measure models to determine if combined models had greater predictive power. The combined models had marginally higher adjusted R^2 s than the single-measure total and inpatient expenditure models discussed above, with the greatest prediction coming from the model that incorporated DCG, DCG, RxRisk, prior expenditure, and SF-36 measures. This model was also the most predictive of outpatient expenditures.

When total costs were segmented by quintiles, all models significantly over-estimated expenditures in the lowest quintile and under-estimated those in the highest quintile, although this was less apparent for models incorporating total costs, ACGs and DCGs that exhibited the highest adjusted R^2 values (Table 4). The combined models improved on the single measure models by further reducing over-prediction in the two lowest quintiles and modestly improving under-prediction in the highest expenditure quintile. The adjusted R^2 was identical (7.74 percent) for the combined model of DCG and prior expenditure and that of DCG and RxRisk, but the predictive ratio for the highest quintile was most improved in the combined model of DCG and RxRisk. The predictive ratios for the remaining quintiles were similar between the two models.

DISCUSSION

The purpose of this analysis was to compare the performance of different risk adjustment measures and examine predictive ability for low and high expenditure veterans receiving primary care from VA. Accurate prospective risk adjustment is desirable because failure to allocate resources properly can generate biased treatment effects in observational studies and over- or under-payment for certain types of patients. The VA budget allocation system to distribute congressional approved funds to VA networks is based on a capitation methodology rather than risk-adjusted payments. As a result, over-payment for low expenditure veterans creates incentives for VAMCs to attract healthy veterans to enroll. VAMCs have the incentive to minimize

Table 4: Predictive Ratios in Prospective Total Cost Models

	<i>Lowest Cost Quintile</i>	<i>Second Lowest Cost Quintile</i>	<i>Middle Cost Quintile</i>	<i>Second Highest Cost Quintile</i>	<i>Highest Cost Quintile</i>
Age and gender alone	26.33	5.49	2.87	1.45	0.23
Prior cost	22.05	4.64	2.67	1.55	0.31
Self-report measures					
PCS+MCS (SF-36 summaries)	23.87	4.96	2.84	1.56	0.26
SIC	24.96	5.21	2.84	1.50	0.25
Administrative data-based measures					
Charlson	24.38	4.99	2.73	1.48	0.29
ACG	17.16	4.10	2.76	1.75	0.33
DCG-HCC	17.13	3.90	2.66	1.72	0.36
CDPS	20.05	4.65	2.82	1.64	0.29
RxRisk	19.08	4.44	2.80	1.68	0.31
Combined models					
DCG, prior cost	17.13	3.85	2.61	1.71	0.37
DCG, RxRisk	15.80	3.76	2.67	1.77	0.37
DCG, RxRisk, prior cost	15.90	3.73	2.62	1.75	0.38
DCG, RxRisk, prior cost, SF	15.73	3.65	2.62	1.77	0.38
Sample size			14,449		

PCS, Physical Component Scores; MCS, Mental Component Score; SIC, Seattle Index of Comorbidity; HCC, Hierarchical Cost Categories; ACG, Adjusted Clinical Group; DCG, Diagnostic Cost Group; CDPS, Chronic Illness and Disability Payment System.

expenditures if they are under-paid for treating high risk, high expenditure veterans. In addition, VAMCs have an incentive to find ways to identify and attract low risk, low expenditure patients for which they are over-paid.

We found that adjustment strategies utilizing diagnostic and pharmacy information predicted total, inpatient and outpatient expenditures better than self-reported measures. No model predicted inpatient expenditures well, largely because only 17 percent of the sample was hospitalized. Estimates for outpatient expenditures were more stable and were predicted more accurately. Baseline outpatient expenditures predicted prospective outpatient expenditures vastly better than any other measure ($R^2 = 41.9$ percent). Prior year expenditures are not a practical risk adjuster for payment setting because of the implicit incentive to obtain higher payments in the next year by generating higher expenditures in the current year. Models that combined risk adjustment measures performed slightly better than single-measure models for all expenditures.

Prediction of VA total expenditures was two to three times lower than the best models used in general populations (Fowles et al. 1996; Pope et al. 1998; Ash et al. 2000; Rosen et al. 2001; Shen and Ellis 2002). Given a wide array of available risk adjustment measures, it is surprising that prospective total expenditure models did not approach 20 percent, the likely asymptote for R^2 in risk adjustment models (Newhouse et al. 1989). The poor prediction in this sample of VA primary care users compared with previous general population studies may be because of two factors. First, the unit cost approach in this analysis to estimate VA expenditures for outpatient and nonacute inpatient utilization does not vary by visit length. Accordingly, the attributed expenditures were not directly related to patient severity of illness, as they are in other health care systems that use pricing based expenditure systems. VA generates most of its revenue through Congressional allocation, so expenditures are calculated to track how this revenue allocation is used within and across VAMCs. A new VA cost-accounting system generates expenditures that do vary by visit duration, but these data were not available at the time of the clinical trial. If visit-level expenditures varied to indirectly reflect case complexity, the correlation between expenditures and risk (the numerator of R^2) might have been higher.

The second factor was that total expenditures are comprised of inpatient and outpatient expenditures that have very different distributions. For example, inpatient expenditures are much more skewed (skewness = 13.6) than outpatient expenditures (skewness = 5.9). In this sample of veterans regularly using VA primary care, outpatient expenditures are partly based on routine care that is highly predictable and partly on care for acute events, while inpatient expenditures most likely represent acute exacerbations of chronic illness and acute events. The predictability and relatively higher R^2 of outpatient expenditures was mitigated by the unpredictability of inpatient expenditures, resulting in total expenditure prediction that tended toward R^2 of inpatient expenditures.

This study had several limitations, including the application of unit costs to outpatient and nonacute inpatient utilization in this study although patient-level variation did exist in acute inpatient care, which was based on a Medicare cost function (Barnett and Rodgers 1999). The lack of data on pharmacy expenditures was also a limitation, but does not invalidate our findings because the goal of the paper is to compare several measures on the same metric. The metric itself may be imperfect but it is consistent across measures. Another limitation is that coding of diagnoses may have been incomplete and variable across medical centers and comparability of data capture between

diagnosis and pharmacy data may have contributed to poor prediction, although this variation should not affect the relative performance of different models. Coding variation in diagnosis data is possible, since VA does not submit claims for patients who lack private insurance, making accuracy less important for such patients. In addition, the capture of pharmacy data is more comparable across facilities than that of diagnosis data, which are more likely to be affected by coding variation (Liu et al. 2003).

Moreover, the data were derived from a clinical trial for which the primary purpose was not risk adjustment and is generalizable only to the primary care patients at the seven VAMCs participating in this clinical trial (Fihn et al. 2004). Last, the omission of expenditures incurred outside of VA makes our results useful only from the perspective of those responsible for VA budgets. If non-VA use differs by clinical condition, race, or geographic region, then under-estimating the inpatient and total expenditures of patients who seek care outside of the VA may bias risk assessments based only on VA data.

Researchers and payors interested in improved risk adjustment for stratified groups of homogeneous risk may want to develop new measures to increase predictive power: (1) combined models of risk that include diagnosis-, pharmacy-based and self-report measures or (2) generic severity measures. Development of risk-class specific risk adjustment may improve the predictive power of these conventional measures, but will be costly to develop.

Creation of combined models of risk adjustment incorporating three types of measures can improve predictive power at the time and financial expense of collecting survey data and pharmacy data that is not routinely available in many health care systems. The combined models improved inpatient and total expenditure prediction in absolute terms (15–16 percent R^2 increase) and reduced over- and under-prediction in lowest and highest expenditure quintiles, even though the models included measures that were (relatively) highly correlated.

The combined models improved expenditure variation over single-measure models partly because the measures predicted different segments of the distribution, as can be seen in the over-/under-prediction within expenditure quintiles as seen in Table 4. Models with highly correlated risk adjusters explained more variation than combined models with less correlated measures, because these models did not have any one measure that was a strong predictor of expenditures. It appears that the predictive ability of multimeasure models is linked heavily to the predictive ability of the “strongest” single measure. Payors may value the marginal improvement in risk adjustment via combined models if the cost of collecting pharmacy, diagnosis, and self-report

data, along with the cost of proprietary risk adjustment software, is not burdensome. In VA, pharmacy and diagnosis data are systematically collected, but self-report data is not yet available system-wide for all VA users.

Generic severity measures may improve predictive power because most risk adjustment measures assess the presence of major illness and related comorbidities, but not the severity of them. Development of generic severity measures presents an interesting challenge, because such a measure would have to be applicable across a general population of healthy people, healthy people with acute conditions, and chronically ill people (Shen and Ellis 2002). Risk adjustment continues to be an empirical challenge in a decade of renewed cost inflation and declining insurance coverage. As health care systems have contained expenditures via risk selection and reducing hospitalisations and length of stay, experiments with new models of disease management and prevention are being considered. Accurate risk adjustment in this environment will be even more important.

ACKNOWLEDGMENTS

This research was supported by Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development (HSR&D) Service, LIP 61-105. Drs. Maciejewski, Liu, and Fihn are presently investigators at the VA Puget Sound Health Care System's HSR&D Center of Excellence. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or the University of Washington. The authors are grateful for comments from the editors, two anonymous reviewers, and seminar participants at the University of Washington, the 2003 International Health Economics Association, and 2003 AcademyHealth research meetings. We are indebted to Mark Perkins for capable research assistance.

REFERENCES

- Ash, A. S., and S. Byrne-Logan. 1998. "How Well Do Models Work? Predicting Health Care Costs." *Proceedings of the Section on Statistics in Epidemiology*, American Statistical Association, pp. 42-9.
- Ash, A. S., R. P. Ellis, G. C. Pope, J. Z. Ayanian, D. W. Bates, H. Burstin, L. I. Iezzoni, E. MacKay, and W. Yu. 2000. "Using Diagnoses to Describe Populations and Predict Costs." *Health Care Financing Review* 21: 7-28.

- Au, D. H., M. B. McDonell, D. C. Martin, and S. D. Fihn. 2001. "Regional Variations in Health Status." *Medical Care* 39 (8): 879–88.
- Barnett, P. G., and J. H. Rodgers. 1999. "Use of the Decision Support System for VA Cost-Effectiveness Research." *Medical Care* 37 (4, Suppl): AS63–70.
- Beddhu, S., M. L. Zeidel, M. Saul, P. Seddon, M. H. Samore, G. J. Stoddard, and F. J. Bruns. 2002. "The Effects of Comorbid Conditions on the Outcomes of Patients Undergoing Peritoneal Dialysis." *American Journal of Medicine* 112: 696–701.
- Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie. 1987. "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation." *Journal of Chronic Diseases* 40: 373–83.
- Deyo, R. A., D. C. Cherkin, and M. A. Ciol. 1992. "Adopting a Clinical Comorbidity Index for Use with ICD-9 Administrative Databases." *Journal of Clinical Epidemiology* 45: 613–9.
- Ellis, R. P., G. Pope, L. I. Iezzoni, J. Z. Ayanian, D. W. Bates, H. Burstin, and A. Ash. 1996. "Diagnosis-Based Risk Adjustment for Medicare Capitation Payments." *Health Care Financing Review* 17: 101–28.
- Fan, V. S., D. Au, P. P. Heagerty, R. A. Deyo, M. B. McDonell, and S. D. Fihn. 2002. "Validation of Case-Mix Measures Derived from Self-Reports of Diagnoses and Health." *Journal of Clinical Epidemiology* 55: 371–80.
- Fihn, S. D., M. B. McDonell, P. Diehr, S. M. Anderson, K. A. Bradley, D. H. Au, J. A. Spertus, M. Burman, G. E. Reiber, C. I. Kiefe, M. Cody, K. M. Sanders, M. A. Whooley, K. Rosenfeld, L. A. Baczek, and A. Sauvigne. 2004. "Effects of Sustained Audit/Feedback on Self-Reported Health Status of Primary Care Patients." *American Journal of Medicine* 116: 241–8.
- Fowles, J. B., J. P. Weiner, D. Knutson, E. Fowler, A. M. Tucker, and M. Ireland. 1996. "Taking Health Status into Account When Setting Capitation Rates: A Comparison of Risk-Adjustment Methods." *Journal of the American Medical Association* 276: 1316–21.
- Kashner, T. M. 1998. "Agreement between Administrative Files and Written Medical Records: A Case of the Department of Veterans Affairs." *Medical Care* 36 (9): 1324–36.
- Kazis, L. E., D. R. Miller, J. Clark, K. Skinner, A. Lee, W. Rogers, A. Spiro, S. Payne, G. Fincke, A. Selim, and M. Linzer. 1998. "Health-Related Quality of Life in Patients Served by the Department of Veterans Affairs: Results from the Veterans Health Study." *Arch Intern Med* 158 (6): 626–32.
- Kronick, R., T. Dreyfuss, L. Lee, and Z. Zhou. 1996. "Diagnostic Risk Adjustment for Medicaid: The Disability Payment System." *Health Care Financing Review* 17: 7–33.
- Kronick, R., T. Gilmer, T. Dreyfus, and L. Lee. 2000. "Improving Health-Based Payment for Medicaid Beneficiaries: CDPS." *Health Care Financing Review* 21: 29–64.
- Liu, C. F., A. E. Sales, N. D. Sharp, P. Fishman, K. L. Sloan, J. Todd-Stenberg, W. P. Nichol, A. K. Rosen, and S. Loveland. 2003. "Case-Mix Adjusting Performance Measures in a Veteran Population: Pharmacy- and Diagnosis-Based Approaches." *Health Services Research* 38 (5): 1319–37.

- Newhouse, J. P., W. G. Manning, E. B. Keeler, and E. M. Sloss. 1989. "Adjusting Capitation Rates Using Objective Health Measures and Prior Utilization." *Health Care Financing Review* 10: 41-54.
- Pope, G. C., K. W. Adamache, E. G. Walsh, and R. K. Khandker. 1998. "Evaluating Alternative Risk Adjusters for Medicare." *Health Care Financing Review* 20: 109-29.
- Rosen, A. K., S. Loveland, J. J. Anderson, J. A. Rothendler, C. S. Hankin, C. C. Rakovski, M. A. Moskowitz, and D. R. Berlowitz. 2001. "Evaluating Diagnosis-Based Case-Mix Measures: How Well Do They Apply to the VA Population?" *Medical Care* 39: 692-704.
- Rosen, A. K., S. A. Loveland, J. J. Anderson, C. S. Hankin, J. N. Breckenridge, and D. R. Berlowitz. 2002. "Diagnostic Cost Groups (DCGs) and Concurrent Utilization among Patients with Substance Abuse Disorders." *Health Services Research* 37: 1079-103.
- Sales, A. E., C. F. Liu, K. L. Sloan, J. Malkin, P. Fishman, A. Rosen, S. Loveland, P. Nichol, N. Suzuki, E. Perrin, N. D. Sharp, and J. Todd-Stenberg. 2003. "Predicting Costs of Care Using a Pharmacy Based Measure: Risk Adjustment in a Veteran Population." *Medical Care* 41 (6): 753-60.
- Shen, Y., and R. P. Ellis. 2002. "How Profitable Is Risk Selection? A Comparison of Four Risk Adjustment Models." *Health Economics* 11: 165-74.
- Von Korff, M., E. H. Wagner, and E. Saunders. 1992. "A Chronic Disease Score from Automated Pharmacy Data." *Journal of Clinical Epidemiology* 45: 197-203.
- Wang, M. C., A. K. Rosen, L. Kazis, S. Loveland, J. Anderson, and D. Berlowitz. 2001. "Correlation of Risk Adjustment Measures Based on Diagnoses and Patient Self-Reported Health Status." *Health Services and Outcomes Research Methodology* 1: 251-65.
- Ware, J. E., and C. D. Sherbourne. 1992. "The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual Framework and Item Selection." *Medical Care* 30: 473-83.
- Weiner, J. P., S. T. Parente, D. W. Garnick, J. Fowles, A. G. Lawthers, and R. H. Palmer. 1995. "Variation in Office-Based Quality." *Journal of the American Medical Association* 273: 1503-8.