Case-Mix Adjusting Performance Measures in a Veteran Population: Pharmacy- and Diagnosis-Based Approaches

Chuan-Fen Liu, Anne E. Sales, Nancy D. Sharp, Paul Fishman, Kevin L. Sloan, Jeff Todd-Stenberg, W. Paul Nichol, Amy K. Rosen, and Susan Loveland

Objective. To compare the rankings for health care utilization performance measures at the facility level in a Veterans Health Administration (VHA) health care delivery network using pharmacy- and diagnosis-based case-mix adjustment measures.

Data Sources/Study Setting. The study included veterans who used inpatient or outpatient services in Veterans Integrated Service Network (VISN) 20 during fiscal year 1998 (October 1997 to September 1998; $N = 126,076$). Utilization and pharmacy data were extracted from VHA national databases and the VISN 20 data warehouse.

Study Design. We estimated concurrent regression models using pharmacy or diagnosis information in the base year (FY1998) to predict health service utilization in the same year. Utilization measures included bed days of care for inpatient care and provider visits for outpatient care.

Principal Findings. Rankings of predicted utilization measures across facilities vary by case-mix adjustment measure. There is greater consistency within the diagnosis-based models than between the diagnosis- and pharmacy-based models. The eight facilities were ranked differently by the diagnosis- and pharmacy-based models.

Conclusions. Choice of case-mix adjustment measure affects rankings of facilities on performance measures, raising concerns about the validity of profiling practices. Differences in rankings may reflect differences in comparability of data capture across facilities between pharmacy and diagnosis data sources, and unstable estimates due to small numbers of patients in a facility.

Key Words. Case-mix adjustment, pharmacy data, profiling

The Veterans Health Administration (VHA) is the largest public-sector health care system in the United States, and one of the largest nationwide health care systems in the country. Over the last seven years, VHA has developed regional network structures and enhanced integration among tertiary and non-tertiary facilities in regionally based delivery systems (Kizer 1996; Ciesco and Greenblatt 1999). In addition, VHA initiated performance measures at the

regional network level, holding Network Directors accountable for their network's performance (Kizer, Demakis, and Feussner 2000; Veterans Health Administration 2000). Networks are ranked by their performance on these measures, and directors' compensation packages depend to some degree on how their network performs. The Performance Measurement Plan has been evolving since 1996. During the period covered by the analysis presented here, the plan included four domains: Quality, Access, Satisfaction, and Cost.

Because of the national performance measures, VHA provides an important model for health care systems examining the performance of profiling and case-mix measures across several facilities within an integrated delivery system. However, VHA performance measures are not adjusted for case-mix differences. Studies of outcomes and utilization measures in the VHA, consistent with studies in other health care systems, have found that case-mix adjustment affects rankings of facilities and outlier status in clinical outcomes and utilization measures (Phibbs, Swindle, and Recine 1997; Berlowitz et al. 1998; Selim et al. 2002).

The veterans who use the VHA system have been shown to be sicker than those who do not (Wilson and Kizer 1997; Kazis et al. 1999). The VHA is similar to other highly segmented health care markets, experiencing significant adverse selection, and providing specialized services to a specialized population. The factors driving market segmentation are not well understood, but in general, the more alternatives in a market area (that is, the more competitive health care is in an area), the lower the VHA market share.

This research was supported by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service Project IIR 99001-1. The views expressed in this report are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or the Health Services Research and Development Service.

Address correspondence to Chuan-Fen Liu, Ph.D., VA Puget Sound Health Care System, 1660 S. Columbian Way (HSR&D 152), Seattle, WA, 98108. Chuan-Fen Liu, Ph.D., Anne E. Sales, M.S.N., Ph.D., Nancy D. Sharp, Ph.D., are with the VA Puget Sound Health Care System Seattle, WA, and the Department of Health Services, University of Washington, Seattle, WA. Paul Fishman, Ph.D., is with the Department of Health Services, University of Washington, Seattle, WA, and the Center for Health Studies, Group Health Cooperative, Seattle, WA. Kevin L. Sloan, M.D., is with the VA Puget Sound Health Care System Seattle, WA, and the Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA. Jeff Todd-Stenberg is with the VA Puget Sound Health Care System Seattle, WA. W. Paul Nichol, M.D., is with the VA Puget Sound Health Care System Seattle, WA, and the Department of Medicine, University of Washington, Seattle, WA. Amy K. Rosen, Ph.D., and Susan Loveland, M.A.T., are with the Center for Health Quality, Outcomes and Economic Research, Bedford VAMC, Bedford, MA, and the Boston University School of Public Health, Department of Health Services, Boston, MA.

This suggests that veterans' selection of VHA as their health care system is largely market-driven. This study and other analyses suggest that patient populations across VHA facilities vary by age, gender, health status, and proportion of patients suffering from major illness conditions (Au et al. 2001; Sloan et al. 2003). Further, the rate of using of VHA services varies across facilities (Ashton et al. 1998). Therefore, the risk across VHA facilities may be different.

In this paper, we assess the ability of the RxRisk-V model, a risk assessment model based on automated pharmacy data refined for use in the VHA (Sloan et al. 2003; Sales et al. 2003), to predict concurrent health service utilization measured as bed days of care and number of ambulatory provider visits. The bed days of care measure has been included in the Network Directors' Performance Measurement set since 1996 (<http://vaww.opq.med.va.gov/>). The ambulatory provider visits measure is not currently included in the Performance Measurement set, but the provider visits measure is often used by other health care systems as a performance or profiling measure and has been discussed as a potential measure within VHA, particularly in the context of provider profiling. We also compare the power of the RxRisk-V to predict health services utilization with three commonly used diagnosis-based case-mix adjustment measures, Diagnostic Cost Groups (DCG), Adjusted Clinical Groups (ACG), and Chronic Illness and Disability Payment System (CDPS). Finally, we compare the rankings for specific measures of health utilization across facilities using different case-mix adjustment measures.

CASE-MIX ADJUSTMENT MEASURES

Case-mix adjustment measures developed to predict health service utilization can be used to allocate resources within health care systems and to case-mix adjust patient populations in clinical evaluation, monitoring outcomes, and research. In addition, case-mix adjustment measures are commonly used to adjust for performance comparisons among providers, facilities, or health care systems in dimensions such as access to care, provider efficiency, and quality of care. Studies comparing unadjusted and adjusted utilization measures have consistently found that adjusting for case mix affects results (Weiner et al. 1991; Newman, White, and Burman 1996; Chang and McCracken 1996; Salem-Schatz et al. 1994; Franks et al. 2000).

Currently, the most commonly used case-mix adjustment measures are diagnosis-based measures, which classify patients into different risk or disease categories based on inpatient and outpatient diagnoses (Weiner et al. 1991; Ash et al. 2000; Kronick et al. 2000). The ACG and DCG are proprietary, diagnosis-based measures; ACG classifies patients into broad clusters of conditions and DCG classifies patients into more refined hierarchical condition categories. The CDPS is a nonproprietary diagnosis-based measure with a focus on Medicaid populations.

Adjusted Clinical Groups (ACGs). The ACG is one of the most widely used diagnosis-based case-mix instruments in HMO (health maintenance organization), private, and VHA settings (Salem-Schatz et al. 1994; Parente et al. 1996; Rosen et al. 1999; Pietz, Byrne, and Petersen 2000). The ACGs measure a population's illness burden by grouping ICD-9-CM diagnosis codes into broad clusters of diagnoses and conditions based on health services resource consumption. The ACGs were initially developed using outpatient diagnosis codes in an HMO population to predict number of ambulatory visits (Starfield and Mumford 1991; Weiner et al. 1991). The current version of ACG classifies patients using both inpatient and outpatient diagnoses codes into 32 Ambulatory Diagnostic Codes (ADGs) based on the specific clinical dimensions such as duration and severity. Finally, each person is then classified into one of 93 mutually exclusive ACGs based on age, gender, and total number of ADGs.

Diagnostic Cost Group/Hierarchical Condition Categories (DCG/HCC). The DCG was developed to predict Medicare payments (Ellis and Ash 1995; Ellis et al. 1996; Pope et al. 2000, Ash et al. 2000), and has been used to adjust Medicare capitation payments (Pope et al. 2000). The DCG model assigns both inpatient and outpatient ICD-9-CM codes to "DxGroups" that are clinically related and similar with respect to levels of resource use. DxGroups are aggregated into 118 Condition Categories (CCs) that include DxGroups belonging to a major body system or disease type, grouped by cost and clinical relation. One patient can have multiple CCs. To avoid double counting within related CCs, the algorithm imposes hierarchies on the CCs based on disease severity, choosing only the highest ranked CC among sets of related conditions (Ellis et al. 1996). The DCG/HCC system handles comorbidity by allowing a person to be classified in multiple HCCs.

Chronic Illness and Disability Payment System (CDPS). The CDPS is an open-source diagnosis-based case-mix approach developed for case-mix adjusting Medicaid capitation payments (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 category definition and validation. Currently, two states have implemented a DPS payment system, and four additional states are planning DPS or CDPS implementation (Kronick et al. 2000). The CDPS algorithm classifies ICD-9-CM codes from both inpatient and outpatient claims into 20 major categories corresponding either to body systems or to specific types of illness or disability based on predicted expenditure. Major categories of diagnoses are divided into subcategories based on predicted expenditures. All of the major categories apply the hierarchical counting rule in which only the single most severe diagnosis within the major category is counted. A person can be classified into multiple major categories.

There has been growing interest in pharmacy-based case-mix adjustment models (Von Korff, Wagner, and Saunders 1992; Johnson, Hornbrook, and Nichols 1994; Roblin 1994; Clark et al. 1995; Roblin 1998; Lamers 1999), because individuals with chronic illnesses such as diabetes or hypertension are frequently prescribed a set of specific, identifiable drugs. One concern regarding diagnosis-based risk adjustment measures is the variation among health care systems, facilities, and providers in the degree of completeness and reliability of coding measures (Von Korff, Wagner, and Saunders 1992; Johnson, Hornbrook, and Nichols 1994; Roblin 1994; Kashner 1998; Hughes and Ash 1997; Szeto and Goldstein 1999). There are several advantages to using a pharmacy-based case-mix adjustment measure. First, it is a direct measure of treated morbidity and a direct link to clinical treatment. Second, pharmacy data may be more complete in closed-model and capitated systems of care than diagnostic data. Third, pharmacy data may be available timelier than diagnostic information. Finally, pharmacy data are possibly less subject to “gaming” by providers than diagnosis data because of the greater consequences of manipulating drug choices (Hornbrook and Goodman 1991; Johnson, Hornbrook, and Nichols 1994).

Pharmacy-based case-mix adjustment measures have disadvantages as well. These measures can only include conditions with established pharmacotherapy, can be reasonably applied only to populations with drug benefits, and will require more frequent updating than diagnosis-based measures because new drugs come on the market much more quickly than new diagnostic codes or coding systems.

RxRisk-V. RxRisk, formerly the Chronic Disease Score (CDS), is the most extensively described pharmacy-based case-mix adjustment measure in the literature. It was originally developed in Group Health Cooperative (GHC) of Puget Sound, a staff-model HMO, using outpatient pharmacy data

on drugs used to treat chronic diseases to classify patients into 29 nonmutually exclusive disease conditions (Von Korff, Wagner, and Saunders 1992; Johnson, Hornbrook, and Nichols 1994; Clark et al. 1995; Fishman et al. 2003). The RxRisk has been used to predict primary care visits, outpatient costs, hospitalizations, and total costs (Von Korff, Wagner, and Saunders 1992; Johnson, Hornbrook, and Nichols 1994; Clark et al. 1995; Fishman and Shay 1999). Since the RxRisk is nonproprietary, it has been tailored and revised for different populations (Fishman and Shay 1999; Lamers 1999; Gilmer, Kronick et al. 2001).

Currently, RxRisk is used in research as a case-mix or disease severity measure. In addition, GHC uses it to adjust capitated payments for a point-of-service insurance product, and to adjust for comorbidity in assessing panel size among primary care physicians. The national health care system in the Netherlands currently uses a revised version of the RxRisk, Pharmacy Cost Groups, to adjust for capitation payments (Lamers 1999).

The RxRisk-V is a newly created refinement of the private-sector RxRisk instrument, tailored for use in the VHA (Sloan et al. 2003). It was constructed using the VHA formulary in an attempt to better predict resource use in VHA. In developing the RxRisk-V, an expert panel conducted a clinical review of the RxRisk classification at the drug level, the drug class level, the multiple drug class level, or a combination of strategies (Sloan et al. 2003). Two principles guided this review. The first was to add previously unused, mutually exclusive groups of drugs corresponding to clinically significant disease entities. The second was to add disease categories important to VHA, such as alcohol dependence, which may not be regarded as equally important in other health care systems. The final RxRisk-V model includes 45 categories.

The VHA is a good environment for using a pharmacy-based risk adjuster because of availability and completeness of pharmacy data. The co-payment for drugs in VHA pharmacies is low. The typical co-payment increased in February 2002 to \$7 per prescription filled, lower than the \$10 or more charged by many private health insurance plans. During the study period, the VHA pharmacy co-payment was only \$2 per prescription filled. Therefore, VHA enrollees had strong financial incentives to obtain all their medications from VHA pharmacies. In two previous VHA studies, 98–100 percent of patients reported obtaining all outpatient medications at VHA facilities (Steiner et al. 1988; Elixhauser et al. 1990). Computerized drug dispense records and databases are widely available from individual VHA facilities and in a single national database.

METHODS

Study Population

The study population consisted of all veteran users ($N = 126,076$) of eight VHA medical facilities in the Veterans Integrated Service Network (VISN) 20 during Fiscal Year (FY) 1998 (October 1, 1997, to September 30, 1998). Veteran users were defined as veterans who visited any outpatient treatment clinic or were hospitalized in any of these facilities during FY1998. The term “outpatient treatment clinics” includes all regular medical, surgical, mental health, and emergency clinics but specifically excludes administrative, Compensation and Pension (typically a single-visit health examination for purposes of determining eligibility for benefits), laboratory and radiology clinics. The VHA is divided into 22 VISNs. These are regional groups of hospitals and other health care facilities that provide services to veterans living within their catchment area. The VISN 20 is composed of eight VHA facilities in Oregon, Washington, Alaska, and Idaho. The eight facilities in VISN 20 represent both urban and rural areas.

The VHA differs from many other health care systems in the United States in that veterans are encouraged to enroll for care through VHA facilities if they meet eligibility criteria, but are not required to do so. Most veterans in the United States are eligible for care, but only about 10 percent of veterans actually use VHA facilities in a given year (Kazis et al. 1999).

Health Care Utilization Measures

Two health care utilization outcome measures were used in this study—bed days of care for inpatient admissions and number of provider visits in FY1998. The bed-days-of-care measure was defined as the total number of acute inpatient days in medicine or surgery beds, including acute psychiatric and substance abuse stays. Long-term care bed days were excluded. The number of provider visits was derived by examining specific Evaluation and Management (E&M) CPT-4 codes indicating a clinical encounter (Rosen et al. 2001) associated with each individual outpatient clinic indicator to identify ambulatory provider-related “face-to-face” encounters. Ambulatory encounters that were not provider-related included laboratory, x-ray, admission/screening, and other miscellaneous clinics.

Data Sources

The primary data sources included VHA national administrative databases and the VISN 20 Data Warehouse. Two VHA national administrative

databases—the Patient Treatment File (PTF) and the Outpatient Clinic File (OPC)—were used to construct the two utilization outcome measures. The PTF file contains records of acute inpatient stays in VHA facilities. The OPC file contains all outpatient care services provided in VHA facilities.

Data extracted from the VISN 20 Data Warehouse, a relational database containing data from the clinical information systems of each of the VISN 20 medical facilities, included patient demographics (age and gender), inpatient and outpatient ICD-9 CM diagnoses, and outpatient pharmacy records. Outpatient pharmacy dispenses, which were used to construct the RxRisk-V classification, were identified using VHA National Product Name, Drug Class, and site-specific drug name in the electronic record. Inpatient and outpatient diagnoses were obtained to construct the three diagnosis-based risk-adjustment measures.

Data Analysis

We estimated concurrent regression models in which pharmacy or diagnosis classifications from the four different case-mix systems in FY1998 were used to predict health care resource use in the same year. Parameters were estimated by linear regression. Age and sex categories were included in all prediction models. We estimated regression models using ordinary least squares (OLS). Several researchers have proposed methods other than single-equation OLS regression to estimate health services utilization (Duan et al. 1983, 1984; Cameron and Trivedi 1998). However, an OLS regression model performs well as a prediction model relative to more sophisticated functional forms as it always yields unbiased estimates of parameter means (Judge et al. 1985).

To compare utilization measures across facilities and risk-adjustment measures, we calculated predicted utilization measures after separately adjusting the RxRisk-V and the three diagnosis-based models for each facility. Then we calculated standardized utilization ratios by dividing the actual or predicted utilization measure by the grand means of the study sample (0.97 bed days of care and 6.37 provider visits). We ranked the facilities by actual and predicted utilization ratios, and examined the changes in rankings between the actual and predicted values. To examine the rankings across case-mix adjustment models, we conducted the rank correlation test for agreement for multiple judgments (Kanji 1993, p. 115).

RESULTS

Population Characteristics

Table 1 presents characteristics of the study sample and utilization outcomes in FY1998. The study population was 94 percent male, with a mean age of 57; 36 percent were age 65 or older. About 3 percent of patients died during the year. Seventy percent of patients were classified into at least one RxRisk-V category, with a mean of 2.5 categories per patient. Almost all patients were classified into at least one medical or disease category by ACG and DCG/HCC, while only 73 percent of patients were classified into at least one disease category by CDPS. About 10 percent of the study population had at least one hospital admission, with an average of 0.97 bed days (SD = 4.0) per patient among the study population. More than three quarters of the study sample (77 percent) had at least one provider visit, with an average of 6.4 provider visits (SD = 12.6) per patient among the study population. Table 2 presents characteristics of case-mix measures by site. The variation in mean number of unique diagnoses among facilities is greater than variation in mean number of unique drugs. The mean number of unique drugs per patient ranged from 4.6 to 3.1 across facilities, while the mean number of unique diagnoses per

Table 1: Characteristics of Study Population

<i>Patient Characteristics</i>	<i>(N = 126,075)</i>
Mean age (SD)	56.85 (15.22)
Proportion of male	93.9%
Deaths in 1998	2.6%
Case-mix measure classification in 1998	
RxRisk-V	
Proportion with one or more	70.1%
Mean number of categories (SD)	2.46 (2.50)
Proportion in an ACG category*	100%
Proportion with one DCG/HCC category or more	100%
Proportion with one CDPS category or more	73.3%
Health service utilization in 1998	
Bed days of care	
Proportion with any hospital admissions	10.0%
Total bed days of care, days (SD)	0.97 (4.00)
Provider visits	
Proportion with at least one provider visit	76.7%
Number of provider visits (SD)	6.37 (12.61)

*Includes all ACG categories, except the categories of unclassified diagnoses and nonusers.

Table 2: Characteristics of Case-Mix Measures by Facility

	<i>Overall</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>
Percent overall population		32.33	21.38	8.75	9.85	11.05	6.80	5.88	3.97
Mean unique diagnoses per patient in 1998(SD)	8.40 (8.58)	8.65 (9.26)	9.40 (9.41)	10.76 (8.79)	7.99 (9.64)	7.24 (8.03)	7.49 (6.75)	5.83 (5.34)	5.34 (4.08)
Mean unique drugs per patient in 1998* (SD)	3.74 (4.27)	3.45 (4.23)	3.87 (4.35)	4.56 (4.60)	4.42 (4.40)	3.12 (4.08)	3.08 (3.91)	4.13 (3.91)	4.18 (3.96)
Case-mix measure classification									
Mean number of RxRisk-V categories (SD)	2.46 (2.50)	2.30 (2.51)	2.53 (2.54)	3.05 (2.67)	2.77 (2.42)	2.01 (2.37)	2.02 (2.26)	2.85 (2.40)	2.72 (2.28)
Mean number of DCG/HCC categories (SD)	4.76 (3.81)	4.55 (3.83)	5.34 (4.24)	5.86 (3.94)	4.96 (3.65)	4.33 (3.76)	4.32 (3.14)	4.01 (2.80)	3.60 (2.53)
Mean number of CDPS categories (SD)	1.44 (1.82)	1.29 (1.79)	1.59 (1.95)	1.94 (2.05)	1.71 (1.86)	1.28 (1.75)	1.11 (1.55)	1.46 (1.52)	1.23 (0.38)

*Includes drugs used in the RxRisk-V classification system.

patient ranged from 10.8 to 5.3. In addition, the range in mean numbers of disease categories per patient across sites was 3.1 to 2.0 for the RxRisk-V, 5.9 to 3.6 for the DCG/HCC, and 1.9 to 1.1 for the CDPS.

Case-Mix Adjustment Model Comparisons

In Table 3 we use R^2 to compare concurrent prediction of health care utilization across case-mix risk adjustment models. For bed days of care, the HCC performed best ($R^2 = 0.45$), followed by CDPS ($R^2 = 0.33$), ACG ($R^2 = 0.24$), and RxRisk-V ($R^2 = 0.12$). For provider visits, the HCC ($R^2 = 0.26$) and CDPS ($R^2 = 0.25$) had the best predictive ability, followed by the RxRisk-V ($R^2 = 0.21$) and the ACG ($R^2 = 0.21$).

Table 3: R-Square Comparisons

<i>Models</i>	<i>Number of Parameters</i>	<i>Bed Days of Care</i>	<i>Provider Visits</i>
Age/Sex	12	0.012	0.008
DCG/HCC	119	0.447	0.263
ACG	71	0.241	0.207
CDPS	59	0.325	0.251
RxRisk-V	51	0.118	0.208
DCG/HCC + RxRisk-V	120	0.449	0.299
CDPS + RxRisk-V	60	0.323	0.281
ACG + RxRisk-V	72	0.263	0.266

We also examined the improvement in concurrent explanatory power obtained by adding pharmacy information to the CDPS, ACG, and DCG/HCC models, respectively. We added the summary score of the RxRisk-V in these models, rather than including 45 RxRisk-V condition categories. The summary score of the RxRisk-V was calculated as the predicted value estimated from the RxRisk-V model for each individual. The improvement in predictive ability varies by outcome measure. The RxRisk-V summary score did not add much explanatory power to the DCG/HCC or CDPS in predicting bed days of care, but it improved the R^2 for DCG/HCC by 3.7 percentage points and for CDPS by 3 percentage points for provider visits. The R^2 of the combined ACG and RxRisk-V model increased by 2.2 percentage points in predicting bed days of care and by 4.9 percentage points for provider visits. For bed days of care, the R^2 values of the combined models (ACG + RxRisk-V and CDPS + RxRisk-V) remain smaller than in the DCG/HCC-only model.

Comparisons of Health Care Utilization among Facilities

Table 4 compares the actual and concurrently predicted utilization measures adjusted by RxRisk-V, DCG/HCC, CDPS, and ACG. The ranking of predicted utilization measures across facilities varies by case-mix adjustment measure. The rankings appear more consistent within diagnosis-based models than between pharmacy- and diagnosis-based models.

Differences in facility rankings between the RxRisk-V and DCG/HCC were observed in the concurrent prediction of bed days of care. The RxRisk-V, CDPS, and ACG models ranked Facility C as highest (ratio = 1.26 – 1.38), while DCG/HCC ranked Facility A as highest (ratio = 1.56). Facilities G and H were ranked differently by the RxRisk-V compared to the three diagnosis-based models. In addition, for these two facilities, predictive use ratios calculated for the RxRisk-V were very different from those generated by the three diagnosis models. Using the RxRisk-V, Facility G was ranked as the second highest number of bed days of care (ratio = 1.15), while the three diagnosis-based models ranked it as the seventh highest (ratio = 0.38 to 0.61). Also using the RxRisk-V, Facility H ranked as the fifth highest number of bed days of care (ratio = 0.98), while the three diagnosis-based models ranked it as the lowest number of bed days of care (ratio = 0.24 to 0.58). The rank correlation test indicates that the rankings across models were significantly different ($F_{7, 24} = 8.455$, $p < 0.01$). The mixed models, combining both the diagnosis model and the RxRisk-V summary score, show similar rankings to

Table 4: Comparisons of Health Care Utilization among Facilities

Bed Days of Care ¹									
Facility	Predicted								
	Actual			RxRisk-V			DCG/HCC		
	Use	Ratio	Rank	Use	Ratio	Rank	Use	Ratio	Rank
A	1.51	1	1.01	4	1.56	1	0.89	4	1.35
B	1.16	2	0.95	6	0.94	3	1.35	2	0.96
C	1.12	3	1.34	1	1.26	2	1.38	1	1.36
D	0.81	4	1.09	3	0.88	4	1.08	3	1.01
E	0.80	5	0.88	7	0.84	5	0.81	5	0.83
F	0.07	8	0.73	8	0.63	6	0.71	6	0.57
G	0.32	6	1.15	2	0.38	7	0.61	7	0.53
H	0.31	7	0.98	5	0.24	8	0.53	8	0.49
Number of Provider Visits ²									
Facility	Predicted								
	Actual			RxRisk-V			DCG/HCC		
	Use	Ratio	Rank	Use	Ratio	Rank	Use	Ratio	Rank
A	1.12	1	0.99	6	1.12	2	0.99	4	1.14
B	1.12	1	1.01	4	1.01	3	1.06	3	0.97
C	1.07	3	1.18	1	1.23	1	1.20	1	1.18
D	0.97	4	1.03	3	0.97	4	1.07	2	1.01
E	0.96	5	0.81	8	0.86	6	0.81	8	0.86
F	0.75	6	0.87	7	0.91	5	0.91	6	0.96
G	0.57	7	1.00	5	0.68	8	0.83	7	0.76
H	0.52	8	1.15	2	0.68	7	0.94	5	0.83

¹The rank correlation test indicates that the rankings across models in bed days of care were significantly different ($F_{7, 24} = 8.455, p < 0.01$).

²The rank correlation test indicates that the rankings across models in provider visits were significantly different ($F_{7, 24} = 9.822, p < 0.01$).

those of the diagnosis-only models. This result may reflect the fact that the diagnosis-based models have better predictive power than the RxRisk-V model. Therefore, adding the summary score of RxRisk-V has limited impact on rankings.

For provider visits, all four models ranked Facility C as having the highest number of provider visits. Two models (RxRisk-V and CDPS) ranked Facility E with the lowest number of provider visits; however, the use ratios of the four models were in a small range (0.81 – 0.86). The other two models (DCG/HCC and ACG) ranked Facility G with the lowest number of provider visits with a greater range of ratios across the four models (0.68 – 1.00). Facilities A and H had very different rankings under the RxRisk-V compared with the three diagnosis-based models. Facility A was ranked sixth highest in number of provider visits using the RxRisk-V (ratio = 0.99), in second place using the HCC (ratio = 1.12) and ACG (ratio = 1.14), and in fourth place using CDPS (ratio = 0.99). In contrast, facility H was ranked as the second highest in number of provider visits using the RxRisk-V (ratio = 1.15), while it was ranked in seventh place by the DCG/HCC and ACG, and in fifth place by the CDPS. The rank correlation test indicates that the rankings across models were significantly different ($F_{7, 24} = 9.822, p < 0.01$).

DISCUSSION

This study shows that the choice of case-mix adjustment measure affects rankings of facilities on two performance measures. There is greater consistency within the diagnosis-based models than between the diagnosis- and pharmacy-based models. For example, Facility H was ranked as the second highest in number of provider visits after adjusting for the RxRisk-V, but seventh highest after adjusting for any one of the three diagnosis models. There are several reasons for these differences, relating to characteristics of Facility H. First, Facility H is the smallest facility and may be unstable with respect to the case-mix adjustment measures. Because of the small numbers of patients, the rank may be susceptible to fluctuations based, for example, on the number of diagnoses recorded for each patient. This may be due to lack of resources to train coders or to ensure consistency across coders. This facility also has no inpatient acute care beds, which means that their incentives to code to insurance industry standards is lower than the other facilities in the study, all of which have inpatient acute care services. While VHA bills third party

payers (except Medicare and Medicaid) to recover costs of care, the primary emphasis is on third party billing for inpatient services.

Second, the difference in rankings may be related to the difference in comparability of data capture across facilities between two data sources. Facility H has a mean number of unique drugs per patient above the population mean, while its mean number of unique diagnoses per patient is below the population average. The number of unique diagnoses is more likely to be affected by coding practices, which vary across facilities, than the number of unique drugs obtained from outpatient pharmacy prescription fills. For example, providers in Facility H may code fewer diagnoses than those in other facilities. In other words, the capture of pharmacy data is more comparable across facilities than that of diagnosis data. Differences between these two measures in comparability of data capture across facilities could affect facility rankings.

Differences in facility rankings across risk-adjustment measures raise a concern about the validity of case-mix adjusted profiling using only one type of case-mix adjustment measure. Rankings on performance measures may be significantly affected by the choice of case-mix adjustment measure. These effects may reflect in part differences in comparability of data capture between two data sources. In addition, the relatively small numbers of patients at the facility level, given the wide variances in these measures, may affect the stability of estimates using each of the risk adjustment methods. The effect of even small fluctuations in either numbers of diagnoses assigned, or of numbers of drugs prescribed for chronic conditions, may be larger than expected. Both diagnosis coding and prescribing of medications for chronic conditions are susceptible to practice variation caused by provider culture. Further research is needed to examine the variation across case-mix approaches and the validity of using different case-mix adjustment measures to assess performance among facilities.

This study also shows that the ability of the RxRisk-V model to predict provider visits is slightly less than the DCG/HCC and CDPS models and comparable to the ACG model. In predicting bed days of care, the DCG/HCC model performs best, followed by ACG and CDPS, while the RxRisk-V has the least predictive power. It is not surprising that the diagnosis-based models have better predictive power in bed days of care than the pharmacy based-model, because the diagnosis-based models use same-year inpatient diagnoses to predict same year bed days of care.

The study also shows that combining diagnosis- and pharmacy-based models improves the predictive power for provider visits. The VHA is an

excellent environment for testing the combined diagnostic and pharmacy model because of the availability of VHA national databases containing pharmacy data as well as inpatient and outpatient diagnoses. The VHA inpatient pharmacy databases are currently in development. Addition of inpatient pharmacy records could improve the predictive power of the RxRisk-V model, although more work would need to be done to refine the RxRisk-V for inpatient risk adjustment.

There are three notable limitations in this study. First, our study sample may not generalize to the full VHA population. Compared to the national VHA sample (Rosen et al. 2001), VISN 20 users are younger and a greater proportion of them are female. Second, this study did not include prescription fills outside the VHA system. Veterans with private insurance, or those who are eligible for Medicaid or Medicare, are likely to fill their prescriptions from non-VHA sources. The problem may be limited because of low copayment for drugs in the VHA, but no recent studies have updated findings of research conducted prior to changes in the pharmacy copayment structure. Third, this study did not include utilization outside the VHA system.

It would be infeasible to attempt to include non-VHA use in this analysis, but two observations may ameliorate this concern. First, the lack of information on non-VHA use affects the diagnosis-based risk adjusters more than the pharmacy-based risk adjuster. Previous studies have found that nearly all VHA patients' outpatient prescriptions were filled at the VHA pharmacies (Steiner et al. 1998; Elixhauser et al. 1990). Second, other researchers within VHA have argued that accumulating diagnoses over inpatient and outpatient encounters for a substantial period of time, and availability of up to ten diagnoses per outpatient encounter and fifteen per inpatient episode of care, adequately capture the full range of diagnoses presented by a patient (C. M. Ashton, personal communication 2001). We agree with our colleagues that these two factors are likely to account for a substantial proportion of the care that is received outside of VHA as well as the care delivered within VHA. In addition, we conducted a sensitivity analysis on veterans under age 65 and found that the impact due to lack of information in non-VA care is minimal.

Despite these limitations, this study has raised some important issues regarding performance measurement at the facility level. Risk adjustment, and the type of risk adjustment, does affect the ranking of facilities. Type of risk adjustment affects different performance measures differently. Researchers applying risk-adjustment methods to performance measures, and managers assessing the relative performance of facilities or other types of units, should use caution in applying only one type of risk adjustment. Further research on

the optimal sample size, and the effect of even minor fluctuations in the elements underlying the risk adjustment methods (e.g., diagnosis coding and/or prescribing practices), needs to be conducted.

REFERENCES

- 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 (3): 7–28.
- Ashton, C. M., N. J. Petersen, N. P. Wray, and H. J. Yu. 1998. "The Veterans Affairs Medical Care System: Hospital and Clinic Utilization Statistics for." *Medical Care* 36 (6): 793–803.
- Au, H. A., M. B. McDonnell, D. C. Martin, and S. D. Fihn. 2001. "Regional Variations in Health Status." *Medical Care* 39 (8): 879–88.
- Berlowitz, D. R., A. S. Ash, E. C. Hickey, B. Kader, R. Friedman, and M. A. Moskowitz. 1998. "Profiling Outcomes of Ambulatory Care: Casemix Affects Perceived Performance." *Medical Care* 36 (6): 928–33.
- Cameron, A. C., and P. K. Trivedi. 1998. *Regression Analysis of Count Data*. New York: Cambridge University Press.
- Chang, W., and S. B. McCracken. 1996. "Applying Case Mix Adjustment in Profiling Primary Care Physician Performance." *Journal of Health Care Finance* 22 (4): 1–9.
- Ciesco, E., and M. Greenblatt. 1999. "Implementing a Performance Measurement System at the Veterans Health Administration." Paper presented at the International Quality and Productivity Center Ninth National Performance Measure Conference, September 14, Arlington, VA.
- Clark, D. O., M. Von Korff, K. Saunders, W. M. Baluch, and G. E. Simon. 1995. "A Chronic Disease Score with Empirically Derived Weights." *Medical Care* 33 (8): 783–95.
- Duan, N., W. Manning, C. Morris, and J. Newhouse. 1983. "A Comparison of Alternative Models for the Demand for Medical Care." *Journal of Business and Economics Statistics* 1 (2): 115–26.
- . 1984. "Choosing between the Sample Selection Model and the Multi-part Model." *Journal of Business and Economics Statistics* 2 (4): 283–9.
- Elixhauser, A., S. A. Eisen, J. C. Romeis, and S. M. Homan. 1990. "The Effects of Monitoring and Feedback on Compliance." *Medical Care* 28 (10): 882–93.
- Ellis, R. P., and A. Ash. 1995. "Refinements to the Diagnostic Cost Group (DCG) Model." *Inquiry* 32 (4): 418–29.
- Ellis, R. P., G. C. Pope, L. Iezzoni, J. Z. Ayanian, D. W. Bates, H. Burstin, and A. S. Ash. 1996. "Diagnosis-Based Risk Adjustment for Medicare Capitation Payments." *Health Care Financing Review* 17 (3): 101–28.
- Fishman, P. A., M. Goodman, M. Hornbrook, R. Meenan, D. Bachman, and M. O'Keefe Rossetti. 2003. "Risk Adjustment Using Automated Pharmacy Data: The RxRisk Model." *Medical Care* 41 (1): 84–99.

- Fishman, P. A., and D. K. Shay. 1999. "Development and Estimation of a Pediatric Chronic Disease Score Using Automated Pharmacy Data." *Medical Care* 37 (9): 874–83.
- Franks, P., G. C. Williams, J. Zwanziger, C. Mooney, and M. Sorbero. 2000. "Why Do Physicians Vary So Widely in Their Referral Rates?" *Journal of General Internal Medicine* 15 (3): 163–8.
- Gilmer, T., R. Kronick, P. Fishman, and T. G. Ganiats. 2001. "The Medicaid Rx Model: Pharmacy-Based Risk Adjustment for Public Programs." *Medical Care* 39 (11): 1188–202.
- Hornbrook, M. C., and M. J. Goodman. 1991. "Health Plan Case Mix: Definition, Measurement, and Use." *Advances in Health Economics and Health Services Research* 12: 111–48.
- Hughes, J. S., and A. S. Ash. 1997. "Reliability of Risk-Adjustment Models." In *Risk Adjustment for Measuring Healthcare Outcomes*, edited by L. Iezzoni, pp. 263–86. Chicago: Health Administration Press.
- Johnson, R. E., M. C. Hornbrook, and G. A. Nichols. 1994. "Replicating the Chronic Disease Score (CDS) from Automated Pharmacy Data." *Journal of Clinical Epidemiology* 47 (10): 1191–9.
- Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T. -C. Lee. 1985. *The Theory and Practice of Econometrics*. New York: Wiley.
- Kanji, G. K. 1993. *100 Statistical Tests*, new edition. London: Sage.
- 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., X. S. Ren, A. Lee, K. Skinner, W. Rogers, J. Clark, and D. R. Miller. 1999. "Health Status in VA Patients: Results from the Veterans Health Study." *American Journal of Medical Quality* 14 (1): 28–38.
- Kizer, K. W. 1996. *Prescription for Change*. Washington, DC: Department of Veterans Affairs, Veterans Health Administration.
- Kizer, K. W., J. G. Demakis, and J. R. Feussner. 2000. "Reinventing VA Health Care: Systematizing Quality Improvement and Quality Innovation." *Medical Care* 38 (6, Supplement 1): I 7–16.
- Kronick, R., T. Gilmer, T. Dreyfus, and L. Lee. 2000. "Improving Health-Based Payment for Medicaid beneficiaries: CDPS." *Health Care Financing Review* 21 (3): 29–64.
- Lamers, L. M. 1999. "Pharmacy Costs Groups: A Risk-Adjuster for Capitation Payments Based on the Use of Prescribed Drugs." *Medical Care* 37 (8): 824–30.
- Newman, C., S. White, and D. Burman. 1996. "Physician Profiling: Applications for the Robert Wood Johnson Profiling Project Database." *Journal of Ambulatory Care Management* 19 (4): 49–57.
- Parente, S. T., J. P. Weiner, D. W. Garnick, J. Fowles, A. G. Lawthers, and R. H. Palmer. 1996. "Profiling Resource Use by Primary-Care Practices: Managed Medicare Implications." *Health Care Financing Review* 17 (4): 23–42.
- Phibbs, C. S., R. W. Swindle, and B. Recine. 1997. "Does Case Mix Matter for Substance Abuse Treatment?" *Health Services Research* 31 (6): 755–71.

- Pietz, K., M. Byrne, and N. Petersen. 2000. Results of Profiling Methodology Working Group." Paper presented at the CHIPS Annual Meeting, August, Seattle, Wash.
- Pope, G. C., R. C. Ellis, A. S. Ash, C. F. Liu, J. Z. Ayanian, D. W. Bates, H. Burstin, L. I. Iezzoni, and M. J. Ingber. 2000. "The Principal Inpatient Diagnostic Cost Group Model for Medicare Risk Adjustment." *Health Care Financing Review* 21 (3): 93-118.
- Roblin, D. W. 1994. "Patient Case Mix Measurement Using Outpatient Drug Dispense Data." *Managed Care Quarterly* 2 (2): 38-47.
- . 1998. "Physician Profiling Using Outpatient Pharmacy Data as a Source for Case Mix Measurement and Risk Adjustment." *Journal of Ambulatory Care Management* 21 (4): 68-84.
- Rosen, A. K., A. Ash, J. Rothendler, and S. Loveland. 1999. "Implementing Ambulatory Care Case-Mix Measures in the VA: from Theory to Practice." Paper presented at the VA HSR&D annual meeting, Washington, DC, February.
- Rosen, A. K., S. Loveland, J. J. Anderson, J. A. Rothendler, C. S. Hankin, C. C. Rakowski, M. A. Moskowicz, and D. R. Berlowitz. 2001. "Evaluating Diagnosis-Based Case-Mix Measures: How Well Do They Apply to the VA Population?" *Medical Care* 39 (7): 692-704.
- Salem-Schatz, S., G. Moore, M. Rucker, and S. D. Pearson. 1994. "The Case for Case-Mix Adjustment in Practice Profiling: When Good Apples Look Bad." *Journal of the American Medical Association* 272 (11): 871-4.
- Sales, A. E., C. F. Liu, K. L. Sloan, J. Malkin, P. Fishman, A. K. Rosen, S. Loveland, W. P. Nichol, N. T. 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.
- Selim, A., D. Berlowitz, G. Fincke, A. K. Rosen, X. Ren, C. Christiansen, Z. Cong, A. Lee, and L. Kazis. 2002. "Risk-Adjusted Mortality Rates as a Potential Outcome Indicator for Outpatient Quality Assessments." *Medical Care* 40 (3): 237-45.
- Sloan, K. L., A. E. Sales, C. F. Liu, P. Fishman, W. P. Nichol, N. T. Suzuki, and N. D. Sharp. 2003. "Construction and Characteristics of the RxRisk-V: A VA-Adapted Pharmacy-Based Case-Mix Instrument." *Medical Care* 41 (6): 761-74.
- Starfield, B., and L. Mumford. 1991. "Ambulatory Care Groups: A Categorization of Diagnoses for Research and Management." *Health Services Research* 26 (1): 53-74.
- Steiner, J. F., T. D. Koepsell, S. D. Fihn, and T. S. Inui. 1988. "A General Method of Compliance Assessment Using Centralized Pharmacy Records: Description and Validation." *Medical Care* 26 (8): 814-23.
- Szeto, H., and M. Goldstein. 1999. "Accuracy of Computer Identified Diagnoses in a VA General Medicine Clinic." Unpublished paper.
- Veterans Health Administration. Office of Quality and Performance. 2000. *FY2001 VA Performance Measurement System Technical Manual*. Washington, DC: Department of Veterans Affairs, Veterans Health Administration.

- Von Korff, M., E. H. Wagner, and K. Saunders. 1992. "A Chronic Disease Score from Automated Pharmacy Data." *Journal of Clinical Epidemiology* 45 (2): 197–203.
- Weiner, J. P., B. H. Starfield, D. M. Steinwachs, and L. M. Mumford. 1991. "Development and Application of a Population-Oriented Measure of Ambulatory Care Case-Mix." *Medical Care* 29 (5): 452–72.
- Wilson, N. J., and K. W. Kizer. 1997. "The VA Health Care System: An Unrecognized National Safety Net." *Health Affairs* 16 (4): 200–4.