

Profiling Resource Use: Do Different Outcomes Affect Assessments of Provider Efficiency?

Amy K. Rosen, PhD; Carter C. Rakovski, MS, MA; Susan A. Loveland, MAT; Jennifer J. Anderson, PhD; and Dan R. Berlowitz, MD, MPH

Objectives: To examine whether 2 outcome measures result in different assessments of efficiency across 22 service networks within the Department of Veterans Affairs (VA).

Study Design: A retrospective analysis using VA inpatient and outpatient administrative databases.

Methods: A 60% random sample of veterans who used healthcare services during fiscal year 1997 was split into a 40% sample (n = 1 046 803) for development and a 20% sample (n = 524 461) for validation. Weighted concurrent case-mix models using adjusted clinical groups were developed to explain variation in 2 outcomes: "days of care"—the sum of a patient's inpatient and outpatient annual visit days, and "average accounting costs"—the sum of the average service costs multiplied by the units of service for each patient. Two profiling indicators were calculated for each outcome: an unadjusted efficiency index and an adjusted efficiency index. These indices were compared to examine network efficiency.

Results: Although about half the networks were identified as "efficient" before and after case-mix adjustment, assessments of individual network efficiency were affected by the adjustment. The 2 outcomes differed on which networks were efficient. For example, 4 networks that appeared as efficient based on days of care appeared as inefficient based on average costs.

Conclusions: Assessments of provider efficiency across the 22 networks depended on the outcome measure used. Knowledge about the extent to which assessments of provider efficiency depend on the outcome measure used is an important step toward improved and more equitable comparisons across providers.

(*Am J Manag Care* 2002;8:1105-1115)

are being allocated and what outcomes are being achieved. Information disseminated through report cards regarding the performance of provider entities (eg, managed care plans, groups of hospitals, individual providers, and medical facilities) is useful because "benchmarking" provides meaningful comparisons of provider performance against some internal or external standards.^{1,2} Incorporating case-mix adjustment into performance profiles improves "benchmarking" because it allows for more accurate assessments of provider performance by addressing the disease burden of patients within a provider's panel.³⁻⁵

Most initial efforts to compare provider performance in the hospital setting have focused on outcomes relevant to inpatient care: mortality, complications, length of stay, hospital charges, and readmissions. These outcomes, although limited, are important.⁶ Outcomes that are generally more applicable for examining provider efficiency in managed care organizations are related to outpatient and preventive services, or to total resource use. Currently, a variety of resource use indicators (eg, primary care visits, specialty visits, total costs or

From the Center for Health Quality, Outcomes and Economic Research, Bedford Veterans Affairs Medical Center, Bedford, MA (AKR, CCR, SAL, DRB); Boston University School of Public Health, Department of Health Services, Boston, MA (AKR and DRB); the Clinical Epidemiology Research and Training Unit, Boston University School of Medicine (JJA), and the Section of General Internal Medicine, Evans Memorial Department of Medicine (DRB), Boston Medical Center, Boston, MA.

A version of this paper was presented at the 3rd International Conference of the International Health Economics Association (IHEA), York, England, July 22, 2001.

This work was funded by the VA Health Services Research and Development Service, grant MPC 97-009.

Address correspondence to: Amy K. Rosen, PhD, Center for Health Quality, Outcomes, and Economic Research, Bedford VAMC (152), 200 Springs Road, Bedford, MA 01730. E-mail: akrosen@bu.edu.

Efficient management of healthcare service delivery is paramount for managed care organizations in today's competitive environment. Greater constraints imposed with respect to accountability, performance monitoring, and quality improvement are forcing health plan administrators to examine opportunities for reducing costs without compromising quality of care. Health plan administrators also need to understand how their resources

charges, number of ancillary tests, specialty services) are being used as outcome measures for evaluating provider performance, and several studies have found variations in physician performance after case-mix adjustment.^{2,4,7-9}

Despite the consensus about the need for case-mix adjustment when comparing providers, little attention has focused on whether different outcome measures affect assessments of provider performance. A few studies evaluating quality of care in coronary artery bypass graft surgery have found that assessments of hospital performance generally differed depending on whether mortality or complication rates was used as the outcome measure.¹⁰⁻¹² As increasing numbers of organizations are profiling providers on their efficiency, accurate information has become critical for making informed decisions about provider efficiency.¹³ Knowledge about the extent to which assessments of provider efficiency depend on the specific outcome measure used is an important step toward ensuring that profiling is conducted appropriately and equitably.

The Department of Veterans Affairs (VA) is the largest integrated healthcare system in the United States, currently offering a full continuum of care to approximately 4 million veterans.¹⁴ The VA has been undergoing a major transformation in healthcare delivery, with increased emphasis on ambulatory care, primary care, and implementation of managed care practices.¹⁵ The VA was recently restructured into 22 geographically defined Veterans Integrated Service Networks to replace 172 independent medical centers.¹⁶⁻²⁰ Each of these networks comprises approximately 5 to 10 acute care facilities, and a varying number of clinics, nursing homes, and other programs. The purpose of this organizational transformation was to provide more accessible, cost-efficient, and high quality care to increasing numbers of veterans through better integration of resources. The basic budgetary, planning, and operating decisions for VA patients living within each geographic area are now coordinated at the network level.¹⁶⁻²⁰ As a result of this restructuring, the VA is now faced with the task of allocating resources equitably across the 22 networks.

Numerous managed care, state-level, and commercial organizations currently use diagnosis-based case-mix measures to determine the disease burden of their population, assess provider efficiency, and examine resource allocation.²¹⁻²⁵ These case-mix measures classify individuals into groups based on their expected need for healthcare services.²⁶ In previous work, we found that 2 diagnosis-based

case-mix measures, adjusted clinical groups (ACGs) and diagnostic cost groups, performed moderately well in classifying veterans into clinically homogeneous groupings based on expected resource needs.²⁷ In this study, we built on our prior work and related research to examine variation in resource use across the 22 VA networks using the ACG case-mix system.^{2,9,27-33} Specific objectives were to examine whether 2 different outcome measures—days of care and average costs—gave different assessments of network efficiency; to determine any differences across the networks in efficiency with and without case-mix adjustment; and to assess the relationship between networks' disease burden and efficiency.

This study will provide the VA with important insights about the relative efficiency of individual networks, and inform individual networks as to how they compare with other networks. The use of 2 outcome measures to assess efficiency should lead to more accurate interpretation of study findings. For example, agreement between the 2 outcomes on network efficiency would strengthen our conclusions about a network's efficiency. Disagreement, on the other hand, would indicate that further investigation is needed to understand the discrepancy between the 2 outcomes. Further, by applying a standardized methodology for profiling efficiency, we will provide useful information that other managed care organizations may need to improve the profiling of their own providers. Therefore, although we studied 1 large, integrated delivery system, our results can be directly applicable to other healthcare systems faced with similar issues.

METHODS

Databases

Our primary data sources were 2 inpatient and 1 outpatient administrative databases: the patient treatment file (acute care), the extended care file (long-term care), and the outpatient clinic file (OPC). The inpatient files contained records on all individuals discharged from or residing in VA facilities on September 30 of each fiscal year. Demographics and *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)* diagnosis codes from each episode of care are available.³⁴ Diagnoses included the primary diagnosis; up to 9 secondary diagnoses; the primary bed-section diagnosis (ie, setting of hospitalization where the diagnosis was assigned, such as the surgical or medical care unit); and up to 4 secondary bed-section diagnoses.

The OPC file describes all outpatient care provided at VA facilities during a fiscal year. Each outpatient visit may consist of 1 to 15 “clinic stops,” a VA term indicating the variety of clinical and nonclinical encounters that are delivered for patient care. Available information included demographics, eligibility for care, site and purpose of clinic stop(s), CPT-4 procedure codes, and 1 primary and up to 9 secondary diagnoses associated with each clinic stop. Veterans Affairs facilities were mandated to collect specific information on every patient treated in an outpatient clinic beginning in fiscal year 1997 (FY97).³⁵

We used a third database, developed by the VA Allocation Resource Center (ARC), for cost information. The ARC database is a national, integrated database of all veterans receiving care in a given fiscal year, and includes demographic, utilization, and accounting cost information for each patient. The ARC database is derived from the VA cost distribution report (CDR), which reports facility-level estimates of expenditures, such as staff time, salary, and supplies that are allocated to a clinical program. These cost estimates are then summed and divided by the annual units of care, producing the average accounting cost for a specific clinical program.^{36,37}

To create 1 complete file containing diagnostic, demographic, and cost information on patients, we merged records from all these files using patient identifiers (ie, veterans’ social security numbers). We then merged this file with the Beneficiary Identification and Record Locator Subsystem file, to obtain information on the deaths of any veterans during the study period. We used this information for constructing annualized utilization outcomes.

Sample

We selected a 60% random sample of veterans from the inpatient and outpatient FY97 (October 1, 1996, to September 30, 1997) files. We included all veterans who used acute, long-term, or outpatient care services during this period, excluding those individuals with dental or telephone use only and nonveterans. A split-sample technique was used to develop models on a 40% sample ($n = 1\,046\,803$) and validate them on a 20% sample ($n = 524\,461$).

Overview of the Adjusted Clinical Group System

Adjusted clinical groups are a computerized software system that uses age, sex, and diagnoses generated from patient encounters over a 1-year period to describe the medical problems of patients and their likely effect on healthcare resource consump-

tion.³⁸ The system is both patient- and population-based, and describes and predicts a population’s concurrent or future healthcare utilization and costs.³⁹⁻⁴² Adjusted clinical groups were initially developed to predict the number of annual ambulatory care visits in selected health maintenance organizations and the Maryland Medicaid population, although their primary application at present is provider profiling.²⁷ Since their original development, ACGs have undergone numerous revisions, and the system currently assigns all inpatient and outpatient ICD-9-CM codes to 1 of 32 adjusted diagnosis groups (ADGs). Individuals often have multiple ADGs assigned. Adjusted clinical groups, which represent 93 mutually exclusive categories based on a patient’s age, sex, and total number of ADGs, are generated after several subdivisions of ADGs.^{2,8,25,26,28,43}

Measures of Resource Use

Typically, measures of resource use involve costs. Unlike other settings, however, the VA does not routinely generate a patient bill with explicit charges, so that the VA has no systemwide information on the costs of care incurred at a detailed patient-specific level.^{36,37} Because of the limitations in VA micro-cost accounting, we used 2 outcome measures that we consider to be proxies of total resource use in the VA. Our first outcome measure, days of care, was developed from available inpatient and outpatient data to reflect the “typical” utilization patterns of VA patients. This approach is similar to the revenue ratio adjustment method used by the American Hospital Association that summarizes hospital output into a single utilization measure.^{44,45} For each patient, this outcome (days of care) represents the total number of days of contact with the healthcare system during the study. Days of care was constructed by combining the number of ambulatory visit days with the number of inpatient days for each patient during the 12-month study period. Although an individual could possibly have both a hospitalization and an outpatient visit recorded on the same day, we considered this situation to represent 1 “day of care.” As a result, the maximum possible number of days of care for an individual was 365.

The second outcome measure was based on the VA macro-accounting system. Average accounting costs (referred to hereafter as average costs), obtained from the ARC data, represent the sum of the *average* service costs multiplied by the units of service for each patient. This outcome was patient-specific in that it was based on the actual number of services each patient received. In calculating the

ARC costs, the VA also makes some adjustment for the inpatient diagnosis-related group that accounted for the hospital stay. The average accounting costs per unit of service are site- or program-specific, however, because they are the same for each inpatient day or clinical visit in a given VA hospital program.^{36,37} Therefore, the ARC data used in this study represent a measure of workload that is used in the VA budget allocation process.^{36,37}

Explanatory Variables

Independent variables included patients' demographic information (age and sex) and diagnostic information (*ICD-9-CM* diagnosis codes). We obtained diagnoses from all inpatient and outpatient provider encounters. Diagnoses from ancillary services, such as laboratory and x-ray visits, were excluded to eliminate "rule-out" diagnoses.

Data Analysis

Analyses were performed using the Statistical Analysis System (SAS) software package, version 6.12 (SAS Institute, Cary, NC). We constructed 2 analytic files required for applying the ACG system (version 4.5) to the data. We obtained means and standard deviations (SDs) of the 2 dependent variables: days of care and average costs. We developed 2 weighted least-squares regression models on the 40% sample to explain variation in annualized resource use (as measured by days of care and average costs). These models included 19 age/sex categorical variables and the 32 ADGs from the ACG case-mix system as indicator variables. We weighted all observations by eligibility (ie, months alive in FY'97); all reported means are weighted. We validated all models to verify that our models were not overfit to our particular sample.⁴⁶ This involved applying the fitted model to the 20% validation sample, then refitting the model on the 20% sample, and applying it to the 40% sample.

We used these models to calculate the disease burden of patients in a network, which we obtained from the expected resource use of patients in each network. For each measure of resource use (eg, days of care), we defined a networks' disease burden (case-mix) as the ACG-expected mean network days of care divided by the mean observed overall VA days of care. A network's disease burden indicates whether the patients in that particular network can be expected to use more or less resources than the average VA patient. We then calculated 2 profiling indicators recommended by the ACG system to examine variation in resource use across

the networks: an unadjusted and an adjusted efficiency index.^{2,47} These indicators are being applied increasingly by managed care organizations for profiling.^{13,47} Using these profiling indicators, we compared networks in their relative unadjusted and adjusted resource use, and examined differences between days of care and average costs in their assessments of network efficiency. Because we tested 2 outcome measures, we calculated a total of 4 indicators: 2 relative to days of care and 2 relative to average costs. We ranked networks from lowest to highest based on mean observed costs for all analyses and classified them into regions based on US census definitions.⁴⁸

The first indicator, the unadjusted efficiency index, is the mean observed (O) network resource use divided by the mean observed overall VA resource use. The second indicator, the adjusted efficiency index, is the mean observed network resource use divided by the mean expected (E) network resource use. The adjusted efficiency index (O/E) reflects a network's resource intensity, representing the over- or underuse of each network adjusted by its patients' disease burden. An adjusted efficiency index of 1.0 indicates that a network's resource use is equal to its expected resource use, whereas an adjusted efficiency index of 1.3 indicates that a network used 30% more resources than expected given its disease burden.²

Finally, to broaden our understanding of the relationships among selected indices, we explored the association between disease burden and efficiency at the network level. We calculated a Pearson correlation coefficient to examine the association between networks' disease burden and efficiency using days of care as the outcome measure, and then repeated this analysis using average costs to compare the 2 outcome measures on the strength of this association. We also calculated Pearson correlation coefficients to explore associations between networks' observed costs and efficiency; the number of patients enrolled in a network and network efficiency; and days of care and average costs at the network level.

RESULTS

Descriptive Results

Our 40% sample was 95.5% male, with a mean age of 59.1 years (SD = 15.2). Most patients (97.4%) had 12 months of eligibility. The average unweighted number of days of care was 17.6 (SD = 41.1, range = 1-365 days); the average number of outpatient visit days was 10.5 (SD = 18.4). Among the 15.8% of

patients who were hospitalized, the average number of hospitalizations was 1.7. The average cost per patient was \$6240.02 (SD = \$19 934.00, range = \$1-\$1 534 508). At the network level, the average number of days of care was 17.00 (range = 13.34-22.60 days), and the mean average cost was \$5653.31 (range = \$4111.22-\$7688.82).

Models' Ability to Explain Concurrent Resource Use

Table 1 summarizes the performance of the 2 ACG models in explaining variation in annualized resource use. The models performed consistently across development, validation, and cross-validation samples, indicating overall stability. The ACG model explained slightly more of the variation in average costs ($R^2 = 0.249$) than in days of care ($R^2 = 0.232$). Overall, the ACG model showed moderate ability in explaining concurrent resource use.

As shown in Table 2, about half the networks were classified as efficient by both outcomes. The unadjusted efficiency indices for the 2 outcomes had similar ranges and were generally concordant in direction; if a network was less efficient (unadjusted efficiency index < 1.0) based on days of care, it was also less efficient based on average costs (Table 3). Three networks (D, H, and M), however, were classified differently on efficiency. Networks D and H appeared as less efficient than average based on days of care (both unadjusted efficiency indices = 1.04) and as more efficient based on average costs (unadjusted efficiency indices = 0.93 and 0.96, respectively).

The opposite was true for network M, which appeared as more efficient than average on days of care (unadjusted efficiency index = 0.92) and as less efficient on average costs (unadjusted efficiency index = 1.03). Ten networks differed by 5% or more when the ratio of the 2 unadjusted efficiency indices was calculated. The greatest difference between the 2 indices was 12%, indicated by a ratio of 1.12.

Networks varied in efficiency even after case-mix adjustment, although the variation across networks was reduced. Differences were noted between the 2 outcome measures in assessments of individual network's efficiency based on adjusted efficiency indices (Table 3). Four networks (I, M, R, and T) appeared as more efficient based on days of care and as less efficient on average costs (eg, the adjusted efficiency indices for network I were 0.97 and 1.07, based on days of care and average costs, respectively). Thirteen networks differed by 5% or more when the ratio of the 2 adjusted efficiency indices was calculated. The greatest difference between the 2 indices was 10%, indicated by a ratio of 0.90.

Although about half the networks were still identified as efficient after case-mix adjustment by both outcomes, incorporating case-mix into the unadjusted efficiency index altered assessments of individual network's efficiency in several cases. For example, networks D, H, and I appeared as more efficient than average based on their unadjusted cost-based efficiency indices (0.93, 0.96, and 0.96, respectively), and as less efficient based on their adjusted cost-based efficiency indices (1.04, 1.04, and

Table 1. Performance of the ACG Model in Explaining Concurrent Resource Use* (FY'97)

Case-Mix Model	Number of Parameters	Days of Care [†] R^2			Average Costs [‡] R^2		
		Development	Validated	Cross-validated [§]	Development	Validated	Cross-validated [§]
ACG model: ADGs used as additive dummy variables, 19 age/sex categories	51	0.232	0.231	0.231	0.249	0.250	0.249

Development sample (n = 1 017 302); validation sample (n = 509 693).

*Outcomes are annualized based on whether a patient had died during the 12-month period.

[†]Days of care = sum of the number of ambulatory visit days and number of inpatient days in FY'97 (maximum = 365 days).

[‡]Average costs = sum of the average accounting costs multiplied by the units of service for each patient.

[§]The cross-validated R^2 represents the summary cross-validation measure of each sample fit to the other.

^{||}Age/sex categories are female aged 18–34, 35–44, 45–54, 55–59, 60–64, 65–69, 70–74, 75–79, 80–84, 85+ and male aged 18–34, 35–44, 45–54, 55–59, 60–64, 65–69, 70–74, 75–79, 80–84, 85+.

FY'97 indicates fiscal year 1997; ACG, adjusted clinical group; ADGs, adjusted diagnosis groups.

RESOURCE UTILIZATION

Table 2. Unadjusted Efficiency Indexes Based on Days of Care and Average Costs

Network*	N	Geographic Region [†]	Unadjusted Efficiency Index		Unadjusted Efficiency Days-based Index/ Cost-based Index
			Days of Care [‡]	Average Costs [§]	
A	47 839	Mountain	0.80	0.73	1.08
B	93 882	South Central	0.82	0.85	0.97
C	82 736	South Atlantic	0.90	0.88	1.02
D	57 276	North Atlantic	1.04	0.93	1.12
E	59 743	South Atlantic	0.93	0.94	0.99
F	43 792	West North Central	0.92	0.95	0.97
G	44 815	Pacific	0.98	0.95	1.04
H	38 306	East North Central	1.04	0.96	1.08
I	28 135	Mountain	0.88	0.96	0.91
J	52 134	South Central	0.91	0.98	0.93
K	51 931	South Atlantic	0.97	0.98	0.99
L	45 429	South Central	1.08	1.00	1.08
M	19 664	West North Central	0.92	1.03	0.90
N	44 549	East North Central	1.02	1.05	0.97
O	28 947	North Atlantic	1.13	1.05	1.07
P	27 957	West North Central	1.00	1.07	0.93
Q	28 299	North Atlantic	1.20	1.09	1.11
R	43 168	Pacific	1.02	1.10	0.93
S	51 902	North Atlantic	1.17	1.10	1.06
T	58 151	Pacific	1.02	1.10	0.92
U	47 356	East North Central	1.20	1.22	0.98
V	50 792	North Atlantic	1.35	1.37	0.98

All outcomes are annualized and means are weighted.

*Networks are listed in increasing order of mean observed costs.

[†]Regions are based on the US Census regions, although North Central is now called Midwest, and North Atlantic combines Mid-Atlantic and New England. Networks are classified into regions based on where the major portion of its designated area is located.

[‡]Unadjusted Efficiency Index (days of care) = mean observed days of care for network/mean observed days of care for overall VA. Overall VA mean weighted days of care = 16.77 (N = 1 046 803).

[§]Unadjusted Efficiency Index (average costs) = mean observed cost for network/mean observed cost for overall VA. Overall VA mean weighted cost = \$5594.21 (N = 1 046 803).

1.07, respectively). The opposite was true for networks R and T, which appeared as less efficient than average based on the unadjusted days-based efficiency index (both indices were 1.02), but as more efficient based on the adjusted days-based efficiency index (0.97 and 0.96, respectively).

We found little association between disease burden and efficiency at the network level, regardless of which outcome measure was used. The correlation was slightly stronger ($r = 0.25$) using days of

care than average costs ($r = 0.07$). Networks with higher disease burden are therefore not necessarily less efficient; conversely, networks with lower disease burden are not necessarily more efficient. We found a much stronger association between network costs and efficiency (Figure 1). Although a general trend was toward decreasing efficiency with increasing mean observed costs, some networks showed different patterns. For example, network D had a relatively low mean cost per patient, but its adjusted efficiency index was greater than 1.0 for both outcomes (1.10 and 1.04, days of care and average costs, respectively). Correlation coefficients were significant between average costs and efficiency as measured by average costs and days of care at the network level (0.73 and 0.88, respectively).

The association between the number of patients enrolled in a network and network efficiency based on average costs was significant but negatively cor-

related ($r = -0.45$). This finding suggests that networks with larger numbers of patients tend to spend less on the care of these patients given their disease burden (ie, their expected resource use is higher than what was spent). Interestingly, when we examined networks with larger panel sizes that appeared efficient based on both outcome measures, we found that they were located in the South (eg, networks B, C, and J). Long-standing regional funding imbalances exist in the VA, and the South has

traditionally been allocated fewer resources than other parts of the country.⁴⁹

Finally, we found a strong and significant relationship between days of care and average costs at the network level ($r = 0.87$). This relationship, depicted in **Figure 2**, indicates that our derived measure of resource use, days of care, is strongly correlated with the VA's budget numbers.

DISCUSSION

The primary purpose of this study was to examine whether 2 different outcome measures gave different assessments of network efficiency. This is an important issue faced by any healthcare system that needs information with which to make knowledgeable decisions about provider efficiency. Three important findings are evident from this study: (1) different outcome measures result in varying assessments of individual network efficiency; (2) case-mix adjustments affect which networks are identified as efficient; and (3) networks' disease burden and efficiency are weakly correlated.

The 2 outcome measures, although highly correlated, provided different assessments of individual network efficiency, suggesting that they may capture different aspects of resource use. A greater number of networks appeared as more efficient using days of care as compared with average costs.

Table 3. Adjusted Efficiency Indexes Based on Days of Care and Average Costs

Network*	N	Geographic Region [†]	Adjusted Efficiency Index		Adjusted Efficiency Days-based Index/ Cost-based Index
			Days of Care [‡]	Average Costs [§]	
A	47 839	Mountain	0.85	0.79	1.09
B	93 882	South Central	0.88	0.88	1.01
C	82 736	South Atlantic	0.86	0.81	1.05
D	57 276	North Atlantic	1.10	1.04	1.06
E	59 793	South Atlantic	0.98	0.99	0.99
F	43 792	West North Central	0.89	0.93	0.96
G	44 815	Pacific	1.00	0.97	1.03
H	38 306	East North Central	1.06	1.04	1.02
I	28 135	Mountain	0.97	1.07	0.91
J	52 134	South Central	0.91	0.94	0.97
K	51 931	South Atlantic	1.00	0.99	1.01
L	45 429	South Central	1.16	1.07	1.08
M	19 664	West North Central	0.92	1.02	0.90
N	44 549	East North Central	1.03	1.11	0.93
O	28 947	North Atlantic	1.10	1.04	1.06
P	27 957	West North Central	0.93	0.99	0.94
Q	28 299	North Atlantic	1.16	1.06	1.09
R	43 168	Pacific	0.97	1.08	0.90
S	51 902	North Atlantic	1.10	1.08	1.02
T	58 151	Pacific	0.96	1.01	0.94
U	47 356	East North Central	1.11	1.10	1.01
V	50 792	North Atlantic	1.23	1.31	0.94

All outcomes are annualized and means are weighted.

*Networks are listed in increasing order of mean observed costs.

[†]Regions are based on the US Census regions, although North Central is now called Midwest, and North Atlantic combines Mid-Atlantic and New England. Networks are classified into regions based on where the major portion of its designated area is located.

[‡]Adjusted Efficiency Index (Days of Care) = mean observed days of care for network/mean expected days of care for network.

[§]Adjusted Efficiency Index (Average Costs) = mean observed cost for network/mean expected cost for network.

Differences in networks' adjusted efficiency indices between days and costs varied as much as 10%; more than half the networks had differences of 5% or more in the 2 indices.

We found variation in resource use across the 22 VA networks, both before and after case-mix adjustment. Case-mix adjustment reduced variation in resource use by a factor of about 3. For example,

Figure 1. Comparison of Adjusted Efficiency Indices by Network: Days of Care Versus Costs

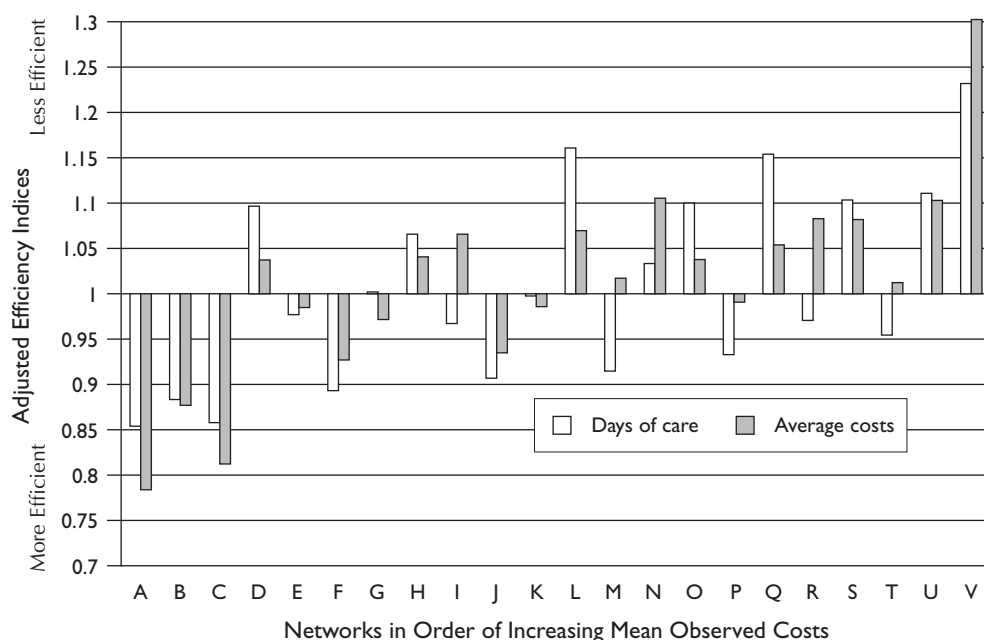
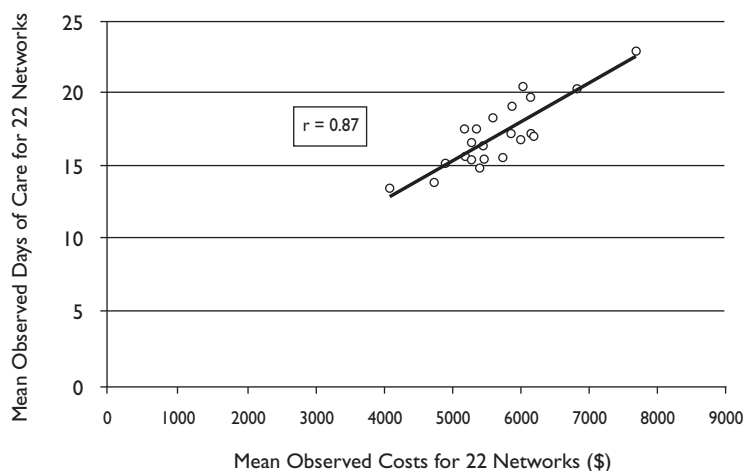


Figure 2. Association of Observed Costs and Days of Care at the Network Level



mean observed costs ranged from \$4111 to \$7689 across the networks; after case-mix adjustment, adjusted mean costs across the networks ranged from \$5012 to \$6195. Adjusting for case-mix helped to explain some of the variation in concurrent resource use for both outcomes. For both ACG models, validated R^2 were comparable to develop-

ment R^2 , suggesting that the models were stable and not overfit. Our results corroborated those of other studies, which have shown that the ACG system is a validated method of case-mix adjustment.^{2,26,30,40,41,50}

Similar to other studies, our research provides further evidence that a case-mix system is a necessary component for accurate measurement of provider efficiency.^{2,30,51-54} Assessments of individual network efficiency were affected by case-mix adjustment. For example, assessments of network efficiency based on average costs differed between unadjusted and adjusted efficiency indices across 5 of the networks. Nonetheless, networks with higher disease burden were

not necessarily less efficient. Efficiency was more strongly associated with a network's observed costs than with its disease burden.

Resource use varied considerably, a finding that was unexplained after case-mix adjustment.^{2,47} As noted in several VA studies, geographic variation may have accounted for some of the variation across

the networks.^{55,56} Differences in patient characteristics might not have been captured by the ACG system. Because *ICD-9-CM* codes do not consistently perform well in specifying disease severity, ACGs may not adequately capture the disease burden of the most severely ill patients within a specific disease category. In addition, coding practices may have varied within and across networks, affecting definition of disease burden.^{6,57-59}

Several limitations of the study should be noted. First, in FY'97, the VA did not have detailed patient-specific cost data available at the national level. Combining ambulatory visit days with inpatient days of care to create "days of care" may have introduced some bias because the intensity of services delivered in the inpatient setting per day typically exceeds that delivered in 1 outpatient visit day. The VA's shift to managed care beginning in 1995 may have offset some of this imbalance because of a dramatic increase in the intensity of outpatient services, resulting in a higher proportion of outpatient costs to total costs by 1997.⁶⁰

Second, "average costs" do not capture patient-specific variation in resource use within visits or days of care, which are important for more accurate prediction of a population's resource use. This may have introduced some "noise" into measurement of efficiency. Researchers currently estimating costs in the VA have several more options, including the average cost methods being developed by the Health Economics Research Center and using costs obtained from the Decision Support System.^{37,61}

Third, we examined diagnostic data from the VA only, because our foremost goal was to examine efficiency within the VA. The inclusion of diagnostic data from other systems, such as Medicare, would have provided a better estimate of the disease burden of VA users who are dually enrolled as well as a more complete picture of veterans' use of services. Fourth, we used only 1 year of data to examine assessments of efficiency across the networks. It would be interesting to explore whether these same profiles would be generated over a longer period.⁴⁷ Finally, the use of the term "efficiency" to represent a network's resource use, given the disease burden of its patients, may be misleading. We found that some of the networks who appeared relatively efficient were those who historically were allocated fewer funds per patient than other networks.

Our study also had a number of strengths. We had a large number of cases for analysis, reducing concerns about statistical instability.² We used profiling

indicators that are well established and that are widely used by many organizations. Also, VA administrative databases have important qualities that make them useful for case-mix adjusted profiling studies. These include the ability to link patients across settings and time; the lack of financial incentives for providers to "upcode" diagnoses; and a high level of data element completion, particularly important in developing clinical profiles.⁶²

This study has broad implications beyond the VA healthcare system. Researchers applying specific case-mix measures to their own populations/settings can benefit from our finding that different outcome measures may produce varying assessments of provider performance. Heightened competition among providers, as well as the need for health plans to improve efficiency through optimal management of provider's time, suggests that the outcomes used to assess performance could have a substantial impact on both providers and patients.²⁸ A better understanding about the extent to which assessments of provider performance depend on the outcome measure used is an important step toward ensuring that profiling is conducted appropriately and equitably.²

As health plans adapt to the current environment of managed care, equitable allocation of resources becomes critical. As suggested by VA data, a provider allocated fewer resources than necessary to deliver care based on its patients' disease burden may appear relatively efficient, but may in fact be limited in the quality of care it can provide to its patients. To date, little evidence exists linking efficiency to quality of care, but more studies are needed in this area.⁹ Or, as shown in the literature, providers who are not reimbursed adequately for patients who have above-average levels of need may either avoid enrolling ill, high-cost patients ("dumping") or resort to enrolling only healthy, low-cost patients ("skimming").^{63,64} Despite the caveats associated with "efficiency," the profiling indicators used in this study appear to be reliable and valid tools for benchmarking resource use across providers. Development of standardized, reliable, and validated outcome measures for profiling will be an important step toward assuring that providers who look inefficient on costs, for example, are not misclassified because other outcome measures were not used. Finally, researchers who conduct profiling should be cautious when interpreting assessments of provider efficiency so as not to penalize those providers who may appear inefficient due to other reasons.

ACKNOWLEDGMENTS

The authors are grateful to Laurie Todd for assistance in manuscript preparation, to Chad Abrams, MA (Johns Hopkins University) for expert guidance on issues related to software implementation of the ACG system, and to Ann Hendricks, PhD, and Bill West, PhD, for sharing their expertise on VA cost issues. Amy Rosen is presently a Senior Research Scientist at the Center for Health Quality, Outcomes, and Economic Research, ENRM Veterans Hospital, Bedford, MA and Associate Professor of Health Services, Boston University School of Public Health. The VA's Institutional Review Board approved this study. The views expressed are solely those of the authors.

REFERENCES

1. Hofer TP, Hayward RA, Greenfield S, et al. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *JAMA*. 1999;281:2098-2105.
2. Tucker AM, Weiner JP, Honigfeld S, Parton RA. Profiling primary care physician resource use: Examining the application of case mix adjustment. *J Ambul Care Manag*. 1996;19:60-80.
3. Berlowitz DR, Rosen AK, Moskowitz MA. Ambulatory care casemix measures. *J Gen Intern Med*. 1995;10:162-170.
4. Cave DG. Analyzing the content of physicians' medical practices. *J Ambul Care Manag*. 1994;3:15-36.
5. Cave DG. Profiling physician practice patterns using diagnostic episode clusters. *Med Care*. 1995;33:463-486.
6. Iezzoni L, ed. *Risk Adjustment for Measuring Health Outcomes*. Ann Arbor, MI: Health Administration Press; 1997.
7. Welch HG, Miller ME, Welch WP. Physician profiling: An analysis of inpatient practice patterns in Florida and Oregon. *N Engl J Med*. 1994;330:607-612.
8. Weiner JP, Starfield BH, Powe NR, Stuart ME, Steinwachs DM. Ambulatory care practice variation within a Medicaid program. *Health Serv Res*. 1996;30:751-770.
9. Powe NR, Weiner JP, Starfield B, Stuart M, Baker A, Steinwachs DM. Systemwide provider performance in a Medicaid program: Profiling the care of patients with chronic illnesses. *Med Care*. 1996;34:798-810.
10. Hartz AJ, Kuhn EM. Comparing hospitals that perform coronary artery bypass surgery: The effect of outcome measures and data sources. *Am J Public Health*. 1994;84:1553-1712.
11. Silber JH, Rosenbaum PR, Schwartz JS, Ross RN, Williams SV. Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. *JAMA*. 1995;274:317-323.
12. Ghali WA, Hall RE, Ash AS, Rosen AK, Moskowitz MA. Evaluation of complication rates after coronary artery bypass surgery using administrative data. *Methods Inf Med*. 1998;37:192-200.
13. DxCG User Conference Program 2001. Boston, MA. Available at: <http://www.dxcg.com>. Accessed October 29, 2001.
14. Department of Veterans Affairs. VHA Annual Performance Report: VA-at-a-glance. In: Annual Accountability Report Fiscal Year, 1999. Available at: <http://www.va.gov/budget/report>, pp. 2-17. Accessed March 10, 2000.
15. VA HSR&D Service. Primary care: Providing continuity and accessibility. *Management Brief*. 2001;19:1-4.
16. Kizer KW. *Vision for Change: A Plan to Restructure the Veterans Health Administration*. Washington, DC: Department of Veterans Affairs, Veterans Health Administration; 1995.
17. Kizer KW. *Prescription for Change*. Washington, DC: Department of Veterans Affairs, Veterans Health Administration; 1996.
18. Kizer KW. Transforming the veterans health care system—the "new VA." *JAMA*. 1996;275:1069.
19. Kizer KW, Fonseca ML, Long LM. The veterans healthcare system: Preparing for the twenty-first century. *Hosp Health Serv Adm*. 1997;42:283-298.
20. Kizer KW. The "New VA": A national laboratory for health care quality management. *Am J Med Qual*. 1999;14:3-20.
21. Ellis RP, Pope GC, Iezzoni L, et al. Diagnosis-based risk adjustment for Medicare capitation payments. *Health Care Financ Rev*. 1996;17(3):101-128.
22. Dunn DL. Applications of health risk adjustments: What can be learned to date? *Inquiry*. 1998;35:132-147.
23. Knutson D. Case study: The Minneapolis buyers health care action group. *Inquiry*. 1998;35:171-177.
24. Wilson VM, Smith CA, Hamilton JM, et al. Case study: The Washington state health care authority. *Inquiry*. 1998;35:178-192.
25. Reid R, MacWilliam L, Roos N, Bogdanovic B, Black C. Measuring morbidity in populations: Performance of the Johns Hopkins Adjusted Clinical Group (ACG) case-mix adjustment system. In: *Manitoba Report*. Manitoba, Canada: Manitoba Centre for Health Policy and Evaluation, Department of Community Health Services, Faculty of Medicine, University of Manitoba; 1999.
26. Smith NS, Weiner JP. Applying population-based case mix adjustment in managed care: The John Hopkins Ambulatory Care Group System. *Manag Care Q*. 1994;2(3):21-34.
27. Rosen AK, Loveland S, Anderson J, et al. Evaluating diagnosis-based case-mix measures: How well do they apply to the VA population? *Med Care*. 2001;39:692-704.
28. Fowles J, Weiner J, Knutson D, et al. Taking health status into account when setting capitation rates: A comparison of risk adjustment methods. *JAMA*. 1996;276:1316-1321.
29. Briggs LW, Rohrer JE, Ludke RL, Hilsenrath PE, Phillips KT. Geographic variation in primary care visits in Iowa. *Health Serv Res*. 1995;30:657-671.
30. Salem-Schatz S, Moore G, Rucker M, Pearson S. The case for case-mix adjustment in practice profiling. *JAMA*. 1994;272:871-874.
31. Parente ST, Weiner JP, Garnick DW, et al. Profiling resource use by primary-care practices: managed Medicare implications. *Health Care Financ Rev*. 1996;17(4):23-42.
32. Selby J, Grumbach K, Quesenberry C, Schmittiel J, Truman A. Differences in resource use and costs of primary care in a large HMO according to physician specialty. *Health Serv Res*. 1999;34:503-518.
33. Chang W, McCracken SB. Applying case mix adjustment in profiling primary care physician performance. *J Health Care Finance*. 1996;22(4):1-9.
34. Lamoreaux J. The organizational structure for medical information management in the Department of Veterans Affairs: An overview of major health care databases. *Med Care*. 1996;34:MS31-MS44.
35. *Ambulatory Care Data Capture*. Washington, DC: Department of Veterans Affairs, Veterans Health Administration; 1996. VHA directive 96-057.
36. Swindle R, Lukas CV, Meyer DA, Barnett PG, Hendricks AM. Cost analysis in the Department of Veterans Affairs: Consensus and future directions. *Med Care*. 1999;37:AS3-AS8.

37. Barnett P. Review of methods to determine VA healthcare costs. *Med Care*. 1999;37:AS9-AS17.
38. Ash AS, Ellis RP, Pope GC, et al. Using diagnoses to describe populations and predict costs. *Health Care Financ Rev*. 2000;21(3):7-28.
39. Starfield B, Weiner J, Mumford L, Steinwachs D. Ambulatory care groups: A categorization of diagnoses for research and management. *Health Serv Res*. 1991;26:53-74.
40. Weiner JP, Starfield B, Steinwachs D, Mumford LM. Development and application of a population-oriented measure of ambulatory care case-mix. *Med Care*. 1991;29:452-472.
41. Weiner JP, Starfield BH, Lieberman RN. The John Hopkins Ambulatory Care Groups (ACGs). A case-mix system for UR, QA and capitation adjustment. *HMO Pract*. 1992;6(1):13-19.
42. Weiner JP, Dobson A, Maxwell SL, Coleman K, Starfield BH, Anderson GF. Risk-adjusted Medicare capitation rates using ambulatory and inpatient diagnoses. *Health Care Financ Rev*. 1996;17(3):77-99.
43. Stuart ME, Steinwachs DM. Patient-mix differences among ambulatory providers and their effects on utilization and payments for Maryland Medicaid users. *Med Care*. 1993;31:1119-1137.
44. Carey K, Stefos T. Measuring inpatient and outpatient costs: A cost-function approach. *Health Care Financ Rev*. 1992;14(2):115-124.
45. Carey K. Cost allocation patterns between hospital inpatient and outpatient departments. *Health Serv Res*. 1994;29:275-292.
46. Hornbrook MC, Goodman MJ. Assessing relative health plan risk with the RAND-36 health survey. *Inquiry*. 1995;32:56-74.
47. The Johns Hopkins ACG Case-Mix System: Documentation & Application Manual. PC and Unix Version 4.5. February 2000. Available at: <http://acg.jhsph.edu>. Last accessed May 15, 2002.
48. US Census Bureau Web site. Available at: http://www.census.gov/www/reg_div.txt. Accessed May 15, 2002.
49. Hendricks A. *Assessing the Veterans Equitable Resource Allocation Budget (Final Report)*. Washington, DC: Department of Veterans Affairs; September 2000.
50. Weiner JP. Ambulatory case-mix methodologies: Application to primary care research. In: Grady M, ed. *Primary Care Research: Theory and Methods*. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services; 1991:75-82. AHCPR publication 91-0011.
51. Berlowitz DR, Ash AS, Brandeis GH, Brand HK, Halpern JL, Moskowitz MA. Rating long-term care facilities on pressure ulcer development: Importance of case-mix adjustment. *Ann Intern Med*. 1996;124:557-563.
52. Rosen AK, Wu J, Chang BH, et al. Risk adjustment for measuring health outcomes: An application in VA long-term care. *Am J Med Qual*. 2001;16:118-127.
53. Wu AW. The measure and mismeasure of hospital quality: Appropriate risk-adjustment methods in comparing hospitals [editorial]. *Ann Intern Med*. 1996;122:149-150.
54. Kassirer JP. The use and abuse of practice profiles [editorial]. *N Engl J Med*. 1994;330:634-635.
55. Ashton CM, Petersen NJ, Soucek J, et al. Geographic variations in utilization rates in Veterans Affairs hospitals and clinics. *N Engl J Med*. 1999;340:32-39.
56. Thomas JW, Bates EW, Hofer T, Perkins A, Foltz-Murphy N, Webb C. Interpreting risk-adjusted length of stay patterns for VA hospitals. *Med Care*. 1998;36:1660-1675.
57. Hannan EL, Kilburn H Jr, Lindsey M, Lewis R. Clinical versus administrative databases for CABG surgery: Does it matter? *Med Care*. 1992;30:892-907.
58. Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T. Comorbidities, complications, and coding bias: Does the number of diagnosis codes matter in predicting in-hospital mortality? *JAMA*. 1992;267:2197-2203.
59. Romano PS, Roos LL, Luft HS, Jullis JG, Doliszny K. A comparison of administrative versus clinical data: Coronary artery bypass surgery as an example. *J Clin Epidemiol*. 1994;47:249-260.
60. Carey K. A multilevel modeling approach to analysis of patient costs under managed care. *Health Econ*. 2000;9:435-446.
61. Health Economics Resource Center. *VA Health Economics Bulletin*. 2001;1(2). Available at: <http://www.herc.research.med.va.gov/Newsletters/Dec2001.pdf>. Accessed November 15, 2002.
62. Kashner TM. Agreement between administrative files and written medical records: A case of the Department of Veterans Affairs. *Med Care*. 1998;36:1324-1336.
63. Ettner SL, Frank RG, Mark T, Smith MW. Risk adjustment of capitation payments to behavioral health care carve-outs: How well do existing methodologies account for psychiatric disability? *Health Care Manag Sci*. 2000;3:159-169.
64. Ettner SL, Frank RG, McGuire TG, Hermann RC. Risk adjustment alternatives in paying for behavioral health care under Medicaid. *Health Serv Res*. 2001;36:793-811.