Fairness in Prospective Payment: A Clustering Approach

Theodore Stefos, Nicole LaVallee, and Frank Holden

Problems of fairness in prospective payment have existed since the inception of this regulatory method in the early 1980s. While prospective payment ostensibly has sought to reward efficient producers and provide disincentives for inefficient producers of health care, many hospitals have been penalized financially as a consequence of facing systematic factors beyond their control. This article defines homogenous peer groups of Department of Veterans Affairs providers for the purpose of establishing competitive prospective reimbursement rates. An econometric analysis classifies hospitals into six categories: small affiliated, small general, midsize affiliated/tertiary, large affiliated/tertiary, midsize general, and psychiatric. The Department of Veterans Affairs adopted this classification to alter its prospective payment system in 1988.

Prospective payment systems (PPS) implemented by the Health Care Financing Administration (HCFA) to reimburse the treatment of Medicare patients, and by the Department of Veterans Affairs (VA) in the allocation of federal budget dollars, have proved to be important policy tools in the effort to contain health care costs. The superiority of PPS to the historical cost-plus reimbursement methodology has been argued elsewhere, notably in Ellis and McGuire (1986). Their analysis

---

Theodore Stefos, Ph.D. is Director of the Management Science Group, at Department of Veterans Affairs (VA) Medical Center, Bedford, MA. At the time of submission, Dr. Stefos was a doctoral candidate at Boston University. Nicole LaVallee, M.A. is a statistician at Parexel International Corporation, Waltham, MA. Frank Holden, M.D. is Director of the Boston Development Center, Department of Veterans Affairs, Braintree, MA. Address correspondence and requests for reprints to Theodore Stefos, Ph.D., Director, Management Science Group, VA Medical Center, Bldg. 12, Room 220, 200 Springs Road, Bedford, MA 01730. This article, submitted to *Health Services Research* June 11, 1990, was revised and accepted for publication May 16, 1991.

sees PPS as supply side cost sharing as opposed to demand side (i.e., patient insurance) cost sharing. While insurance coverage for the patient attempts to control costs by affecting the demand for medical care, prospective payment directly affects a provider's incentives to supply health care.

The performance of prospective payment systems with respect to cost containment in private sector hospitals is mixed. Initial evidence presented in Guterman, Eggers, Riley, et al. (1988), Altman and Rodwin (1988), and Holahan and Palmer (1988) showed sharp declines in average hospital length of stay in 1984, the phase-in year of HCFA PPS. Not surprisingly, since shorter stays are correlated with lower costs, margins of all U.S. hospitals increased dramatically. These initial gains appeared, however, to be temporary. The growth in hospital expenditures caused a resumption of their pre-PPS growth rates, and margins decreased somewhat in the 1985–1987 period. Further evidence in Guterman, Eggers, Riley, et al. (1988) indicated that margins did not appear to be evenly distributed: almost all teaching and urban institutions showed positive margins in 1985, while only two-thirds of small rural hospitals showed any surplus.

Data from VA have shown a similar pattern since the initiation of its own prospective payment system in 1984 (see the next section, "Description of VA PPS"). Data provided in VA annual reports (1983–1988) indicate that lengths of stay declined by over 11 percent per year for the 1984–1988 period and occupancy rates declined an average of 4.5 percent per year, while hospital discharges grew at an average of 2 percent per year. Hospital expenditures per discharge fell 1 percent between 1984 and 1986, and this decrease was accompanied by a rapid increase of expenditures and workload in ambulatory care.

While neither the VA nor the HCFA PPS addresses the need to expand the VA or the Medicare/Medicaid share of the federal budget, an effective PPS should incorporate the principles of horizontal and vertical equity, at least, in distributing available public funds to health care providers. Horizontal equity requires that a prospective payment system reimburse hospitals in similar settings the same amounts for the same types of patients. Vertical equity requires that the variance in payments to hospitals reflect accepted differences in output. If hospitals provide care in different settings and produce different types of output, then a prospective payment system should consider fairness to the provider in its design.

Taken in context, fairness has been characterized by Jencks et al. (1984) and Ellis and McGuire (1988) as the control of a provider's systematic risk. Systematic risk arises from factors unaccounted for in

the PPS that affect a hospital's costs and, in the short run, are beyond the hospital's control. Examples of sources of systematic risk include variation in within-DRG severity, higher labor costs in rural areas adjacent to urban areas, and varying nonlabor costs. High severity within DRGs is thought to occur, for example, in large, urban, teaching facilities, which treat more complicated and costly patients. Competitive factors may also substantially differ across hospitals. For example, VA urban hospitals face higher costs from labor shortages in their respective market areas than do more rural VA hospitals. Since these market factors are beyond a hospital's control, VA hospitals could have been placed at some financial risk.

The current HCFA PPS does incorporate some aspects of fairness; for example, reimbursement partially depends on the urban/rural designation of the hospital's service area. However, hospital costs vary tremendously, even when normalized for case-mix and factor-price differences. The range from the 5th and 95th percentile of the hospital Medicare standardized cost per case in 1985 was $1,566 to $3,795 (Pope 1990). The intuition among health care analysts is that at least some of this variation results from legitimate differences in hospital costs and not from differences in efficiency. Under this circumstance, constraining prices to a national average cost will not strictly result in rewards for efficient producers and penalties for inefficient producers: efficient producers with higher legitimate costs will be penalized as well. Ellis and McGuire (1988) have therefore recommended an even greater control of systematic risk to the hospital. They advocate using the principal of experience rating, that is, conditioning DRG payments partly on the basis of hospital characteristics, to modify PPS. Consequently, they recommend defining homogenous peer groups and using them in calculating average cost reimbursement.

This article examines the alteration of the VA PPS in 1988 to account consistently for systematic risk and, therefore, to incorporate the principle of fairness in allocating funds to VA hospitals. Under "Conditioning VA PPS Reimbursement," the principle of grouping or clustering hospitals in order to control for systematic risk is discussed. The subsequent section outlines the selection of appropriate grouping criteria and, using 1986 data, applies clustering principles and criteria to the calculation of homogenous hospital groups. The degree to which clustering improves fairness in the VA PPS is also examined. The final section briefly summarizes the implications of our work. First, however, is a brief description of VA and the VA PPS.

DESCRIPTION OF VA PPS

The Department of Veterans Affairs, Department of Health Services, is the largest single provider of health services in the United States. Its annual appropriation from Congress is currently in excess of $10 billion. It is the primary educator of health professionals since it has affiliations with most of the nation's medical schools, all of its schools of dentistry, and literally hundreds of schools that train nurses, psychologists, social workers, and other health professionals.

Each year, the federal VA medical care appropriation is distributed to 159 medical care facilities. These facilities are responsible for the management of 172 hospitals, over 110 nursing homes, and 6 independent outpatient clinics.

Until 1984, the distribution of funds to these institutions had been based primarily on the historic budget; that is, each facility received its prior year's budget adjusted for inflation. New programs or program cancellations as well as projected demand changes also influenced the budget to some extent. This methodology came to be viewed as inadequate and obsolete for several reasons. The historic budget could not respond to shifting needs for resources within the VA system that were emerging from changes in the veteran population and new demands for health care services, and it did not accurately identify cost or measure productivity. Increased external review and criticisms of the VA system by Congress, the Office of Management and Budget (OMB), and various other groups, such as the National Academy of Sciences (1977) and the American Enterprise Institute (Lindsay 1975), presented demands for measures of cost and efficiency comparable to measures used in the private sector.

In 1984, the VA introduced a national average cost-based prospective budgeting approach, VA PPS, as a tool for distributing funds to its medical facilities. This was a case-mix system, based on diagnosis-related groups (DRGs), to measure and redistribute acute inpatient care resources including all general medical, surgical, rehabilitation, neurological, and psychiatric services. These comprised some 36 percent of the total recurring VA operating budget.

In 1985, the VA PPS was expanded to control 60 percent of the total VA operating budget by including ambulatory and extended care services. Ambulatory care is funded using an age-adjusted capitation scheme with six price groups determined by the type and amount of utilization recorded over the year. Special outpatient services such as ambulatory surgery and chemotherapy continue, however, to be

funded on a per visit basis. Extended care, that is, intermediate and skilled nursing home care, is funded through a resource utilization group system. This classifies long-term patients according to the amount of direct nursing they require, as determined by their physical status and their measured functional level of activities of daily living (ADL).

The VA PPS is somewhat different from the HCFA PPS in that it uses costs rather than charges in calculating prices/reimbursement rates and is designed more as a budget allocation system than as a true payment system. VA allocation or reimbursement rates can be thought of as expost prices since VA is given an *a priori* fixed total budget, while HCFA prospective rates are true *exante* prices. A VA hospital is prospectively given a budget based on its budget for the previous period, adjusted for the case-mix-predicted expenditures in the VA PPS model. To mitigate the financial risk to the institution, the adjustments are capped at predetermined levels. These levels are currently set so that a hospital cannot gain or lose more than 1 percent of its previous-period budget. Such a capping system finds support in the insurance propositions of Arrow (1963) and has been recently advocated for use in the HCFA PPS by Ellis and McGuire (1988). Using arguments based on marginal utility of income, these authors state that welfare improvements essentially can be made by redirecting budgeted dollars away from hospitals with the largest gains or profits to those with the greatest losses.

Like the HCFA PPS, the VA PPS has always incorporated some measures of fairness to adjust its payments from national cost averages. Education was recognized as an important VA mission and, consequently, VA considered teaching to be a separate but unmeasured output. A teaching adjustment was made to the VA PPS to fund this output by reimbursing resident input costs. In 1986, these costs were based on a fixed stipend per resident, based on a national VA average cost per resident. Competitive factors in labor markets also were thought to affect a hospital's costs, and an adjustment to account for labor cost differentials across hospitals was also attempted. This adjustment was made as a function of the hospital's average salary cost in proportion to the VA national average salary cost. A third adjustment was made for nonteaching facilities, which were thought to draw patients largely below average cost. Therefore, to account for systematic bias in the patient panel, the VA PPS contained adjustments for low-cost workload present at small, unaffiliated hospitals. In order to alter patterns of care, incentive adjustments in the treatment of psychiatry patients were also made. Essentially, each facility was taxed for

each psychiatry inpatient. These funds were then used to create a premium for the treatment of psychiatry patients in ambulatory care settings.

Note that these four nonhomogenous adjustments may have considered only a subset of factors affecting systematic risk. Since each adjustment was considered in isolation and not simultaneously with others, the VA in 1987 considered developing a new model to condition VA prospective payments to its health care facilities.

CONDITIONING VA PPS REIMBURSEMENT

The VA PPS is an allocation scheme whose single price for each product is based (with some adjustments as outlined in the previous section) on a national average cost. The adequacy of the single price to reimburse a hospital's true costs accurately depends on the ability of the PPS to predict variations in costs, that is, differences in resource use, among patients. Variations in costs could arise from several possible sources: (1) different hospitals may have different practice patterns and therefore may arbitrarily choose different levels of factor inputs or services to treat identical patients; (2) as explored earlier, since a case-mix system such as DRGs does not measure output well, a hospital may be treating more numerous (or fewer) complex or severely ill patients; (3) competitive factors affecting the price of medical labor and other factor inputs, as well as patient demand, may be unevenly distributed across facilities; (4) facilities may be subject to diseconomies of scale and scope; and (5) hospitals with different missions, for example, teaching institutions, may face different costs.

A single price for a medical product would appear to be appropriate in the case where hospitals have arbitrary practice patterns and differences reflect inefficiencies alone. If no other sources of inefficiency exist, then it is assumed that a hospital, by altering its practice patterns, can lower its costs of treating patients without affecting the quality of its product.

If hospital costs are systematically dependent on institutional or market characteristics, then a single price will penalize inefficient health care producers. It will also penalize efficient hospitals saddled with characteristics that affect their costs and are beyond their control.

It would appear, therefore, that the single price for each medical product should be conditioned by hospital characteristics in order to improve the fairness of the reimbursement.

An additional reason to alter the VA PPS is outlined in Ellis and McGuire (1988). Over time, actions of a prospective pricing formula are expected to induce hospitals to improve their operating efficiency (Vogelsang and Finsinger 1979). Any remaining differences in average costs among similar hospitals, then, should largely reflect systematic factors.

While VA PPS had attempted to include some degree of fairness, the previous nonhomogenous adjustments were considered to have several shortcomings. For example, the small-hospital adjustment focused on unaffiliated hospitals and did not directly address economies of scope at other types of facilities. The educational adjustments assumed a linear relationship between the costs of teaching and the number of residents without regard to possible economies of scale in education. Wage adjustments did not account for the possibility that VA wages are relatively fixed on a national basis and are limited in their usefulness for measuring the effects of labor shortages on unit costs.

In addition, all nonhomogenous adjustments were calculated in isolation without giving consideration to simultaneous effects with other variables. The next section draws upon the experience gained from using these adjustments and uses 1986 VA data to estimate a simultaneous model to control for systematic risk.

Empirical Estimation of VA Hospital Groups

The most popular methods used to incorporate fairness in prospective reimbursement are evaluating hospitals subjectively, using regression approaches, and clustering hospitals into homogenous groups. Subjective approaches to classifying hospitals have been used since 1974, when the Social Security Administration arbitrarily divided U.S. hospitals into 70 groups. A hospital was classified according to the per capita income of its state, the size of its standard metropolitan statistical area (SMSA), and its bed size. The problem with subjective or Delphic approaches, as Trivedi (1978) pointed out, was that they considered the optimal division of each classification variable in isolation and not simultaneously with other variables. Even if the division of each variable was optimum in isolation, it probably would not be optimum if all classification variables were considered simultaneously. Further, as the number of variables increased, the number of groups

would at least double with each variable, quickly leading to an unmanageable result.

A variation of the fairness concept in conditioning hospital payments on hospital characteristics is presented in Shleifer (1985). His concept of "yardstick" competition pays each hospital the average cost of a set of identical firms. If the firms are not identical, the PPS, in order to be equitable, has to identify characteristics that make them different and correct for any heterogeneity. Shleifer controls for heterogeneity by using a multiple regression to set prices equal to predicted marginal costs derived from a firm's characteristics. A subsidy is also estimated in the same fashion to fund fixed costs. For this exercise it appears that using Shleifer's approach and predicting aggregated average VA hospital costs with multivariate regression may have several deficiencies.

First, and most importantly, Shleifer's methodology is acceptable in the event that the firms under observation are profit maximizers or cost minimizers. Since the objectives of VA managers are to meet budgets rather than to maximize profits or minimize costs, it would be inappropriate to estimate a cost function derived from standard economic theory. Second, under budget-attaining objectives, observed VA costs or expenditures could include elements of wasteful production. A multivariate regression procedure to estimate costs or expenditures would then perpetuate any operating inefficiencies. A more serious drawback in using regression to estimate a cost or expenditure function is the lack of exogeneity of VA output. While this also poses a serious problem in estimating non-VA cost functions (Zwanziger and Melnick 1988), private sector hospitals are thought to control output indicators such as length of stay more easily than other indicators such as discharges. In contrast, VA hospitals can control length of stay, and by controlling the queue for services, they can also control the number of discharges. Since budgets are set ahead of time, these actions would not directly affect the revenue stream at least for that year.

Classifying VA hospitals into peer groups for reimbursement is also an alternative for modifying the VA PPS. Among the several statistical techniques that can be used to classify multivariate data are analysis of variance (ANOVA), discriminant analysis, automatic interactive detection (AID), and cluster analysis. ANOVA and discriminant analysis require a dependent variable that is previously grouped; therefore, they are not appropriate for this analysis. AID requires the designation of an appropriate single dependent variable: in this case, observed average expenditure. However, observed average expenditure would be an inappropriate grouping variable, since inefficient,

high-cost producers would be rewarded by being placed in high-reimbursement groups. AID, therefore, also appears to be inappropriate.

Sorting VA hospitals into groups and then applying the VA PPS within each group appears to be a practical alternative in controlling for systematic risk. Cluster analysis does not require a single dependent variable; it considers all variables simultaneously for grouping. Clustering also offers several advantages over the subjective and regression methods. Unlike subjective procedures and previous VA methods to incorporate fairness, clustering generates homogenous groups of hospitals defined by the simultaneity of the effects of all endogenous variables. Clustering, strictly a forecasting procedure, sidesteps the problems that stem from the absence of profit or cost-minimizing motivations in VA operations. Finally, the clustering technique allows management to continue using the basic VA PPS methodology. The system of reimbursing a hospital the average cost of all hospitals for any given product can simply be reapplied within groups of similar hospitals, or peers. This allows a hospital to compete with a similar group of hospitals for resources while it carries on the cost reduction pressures of the original PPS scheme.

Previous attempts to form groups using a clustering technique included those made by Phillip and Iyer (1975) to group AHA member hospitals in 1975; by Trivedi (1978) to group Washington state hospitals; and by Alexander, Evashwick, and Rundall (1984) to group AHA hospitals according to the geriatric services they offered. The clustering technique we use to group similar VA hospitals is outlined next.

METHODS

Cluster analysis is a multivariate procedure used to sort a sample of entities into distinct groups so that entities within the same group share similar characteristics. A number of techniques can be used to find groups, but the two major types of clustering algorithms are hierarchical methods and iterative partitioning.

A hierarchical method begins with all n objects in separate groups, and then sequentially combines the most similar. The hierarchical method used in this analysis, Ward's method, measures similarity by joining objects so that within-cluster sums of squares are minimized. In the first level of the hierarchy, the two entities most nearly alike are joined to form a group, resulting in n-1 groups.

Next, the two most similar of the n-1 groups are joined to form n-2 groups. This process continues until all n cases are joined in a single

group. This results in n clustering options, ranging from a one-group solution to an n-group solution. The analyst must then decide, based on the properties of the clusters at each level, which solution and how many clusters are "best." It is sometimes viewed as a drawback of cluster analysis that no statistical tests exist to confirm that the selected number of groups is optimal; however, reasonable approaches to choosing the number of clusters do exist. An appropriate number of clusters can be determined from the R^2 statistic, which is simply the proportion of variance in the data explained by the clusters. R^2 will always decrease as the number of groups decreases. A large drop in R^2 occurring after the merger of two clusters implies that the two groups joined were relatively dissimilar: the appropriate number of groups therefore existed before their merger. In general, hierarchical methods are faulted because they make only one pass through the data. That is, once a case has been placed in a cluster, it is not allowed to move to another cluster in any of the subsequent steps. This is a drawback because, as clusters are joined and updated, it may be appropriate to move some cases from their original groups.

The second major type of clustering algorithm is iterative partitioning. In this procedure, an initial partition is made either randomly or by using the groups found from another clustering scheme. The centroid of each group is then computed. Next, the Euclidean distance from each case to each cluster centroid is calculated, cases are moved to the cluster with the nearest centroid, and then each centroid is recalculated. This process is repeated until each case is closest to its own cluster's centroid. The primary objections raised about iterative partitioning are that the number of groups is decided in advance, and that the procedure is very sensitive to the initial partition. Monte Carlo studies have shown that the primary cause of a suboptimal grouping from an iterative procedure is inadequate starting clusters. One advantage of iterative procedures over hierarchical methods, however, is that the former allow cases to move in and out of clusters until all cases are closest to their own cluster's centroid.

In order to overcome the problems associated with using either hierarchical or iterative clustering by itself, a two-stage analysis suggested by Aldenderfer (1984) was performed. As a first stage, Ward's hierarchical method was used to select the appropriate number of groups and the initial hospital partition to use as input into the second-stage K-means iterative method. The number of groups in the first stage was selected based on the point where the R^2 statistic changed by less than 5 percent.

Variables with larger magnitudes and variances have a greater

influence on cluster outcomes. Therefore, variables in this study, which had different units of measure and widely different scales, were recomputed to standard normal form before Ward's clustering was performed. There is a danger that if the variables have different distributions, standardization may alter their relationships and bias the groupings. However, Aldenderfer (1984) suggests the need to standardize, especially if Euclidean distance-based grouping methods are used.

For the second stage *K*-means procedure, the variables were standardized within each cluster, and any correlation between them was controlled for by weighting each distance calculation by the within-cluster covariance matrix. This weighting prevented any correlated variables from having a disproportionate amount of influence on the cluster outcomes.

DATA

The objective in creating hospital groups in this analysis is to identify hospitals with different, legitimate cost structures. Therefore we draw heavily from the cost function literature to determine the set of variables to be used as grouping criteria.

This literature contains three basic categories of empirical models. One type estimates "behavioral" cost functions that arbitrarily use any variables thought to affect cost. This literature, however, has been criticized as lacking any economic foundation. Another set of literature more consistent with the standard economic theory of production, follows the work of McFadden (1978), Christensen, Jorgenson, and Lau (1973, 1975), Lau and Yotopoulos (1972), and Spady and Friedlaender (1978). This work employs a "flexible" functional form that regresses costs on output levels and input prices, but it has been criticized for excluding many factors significant in explaining cost variation in hospitals. More recent work, by Friedman and Pauly (1978), Thorpe (1988), Grannemann, Brown, and Pauly (1986), Vita (1990), and Carey (1991), employ "hybrid" cost functions that incorporate features from both the behavioral and flexible approaches. We draw from this later body of work to delineate factors crucial in explaining hospital cost variation for the cluster analysis. The hybrid literature suggests that appropriate variables to include in a cluster analysis for prediction of true cost variation include output indicators and quality indicators, plus variables measuring factor prices. Factor prices include the price of labor and capital. Note that these factors (capital and labor) implicitly or explicitly measure the role (or mission) of the hospital, the

influence of market competition on labor, and control for the quality of the medical product. Recent works by Trivedi (1978, 1979) and Vertrees and Manton (1986) also indicate these types of variables as necessary to include in the estimation of similar hospitals.

Work by Hornbrook and Monheit (1985) stresses the need in any analysis of cost to measure the interrelationships among hospital volume of output, case mix, scope and complexity of services, factor input prices, capacity, and quality. They indicate that the relationship between average cost and these components critically depends on the relative magnitudes of case-mix diversity, scope and complexity of services, and the volume of patients treated. For example, larger hospitals are expected to have greater diversity of services and therefore are hypothesized to have higher average costs.

The data used in these analyses to group VA hospitals mathematically were hospital aggregates and not information on individual patients. Variables measuring the factor price of capital and the quality indicators were generally unavailable in a reliable form and were therefore dropped from consideration. The potential for strategic behavior also led to our elimination of occupancy rates from the analysis. They had been suggested by Hornbrook and Monheit to estimate the effects of excess capacity on unit costs.

Since hospitals are multipurpose, multiproduct institutions providing ambulatory care and extended care as well as services to inpatients, measuring output depends not only on the number of patients seen but also on their medical characteristics. One would expect two hospitals with identical numbers of discharges to be very different institutions if one hospital treats primarily long-term rehabilitation patients and the other tertiary care patients. These types of differences in the private sector are reflected in HCFA's case-mix index, which measures the relative weights of each hospital's case mix. The relative weights of DRGs in VA are reflected in a relative value scale: the weighted workload unit. These weights are the preassigned resources credited to each DRG. Patients are credited for inpatient weighted workload units based on the discharge DRG and length of stay. The Department of Veterans Affairs has extended this concept by calculating similar relative value scales for ambulatory care and extended care services. The patient totals for each of these three types of weighted workload unit scale are used within each hospital as grouping or classification variables to indicate hospital output. The interested reader is referred to Hartke (1983) for a description of how disease classification takes place and how relative resource scales are calculated.

Use of the sum total of patient-weighted workload units (hereafter referred to as weighted workload units) incorporates the concepts of case mix, volume of patients, and scope of services offered by hospitals. Larger volumes of patients and a broader distribution of available services, such as those found in larger community or teaching hospitals, are likely to be correlated with higher weighted workload unit values.

An additional five variables, the proportion of discharges from medicine, surgery, psychiatry, neurology, and rehabilitation bed sections, are used as indicators for the scope of services supplied by a hospital. For example, an equal distribution of discharge percentages across all five bed sections, as well as high weighted workload unit values across all three workload variables, should indicate a broad scope of available services at a hospital. This should be the case at larger general or metropolitan hospitals, and one can reasonably hypothesize that this type of distribution would be correlated with higher unit costs.

Two variables, average inpatient weighted workload units and average ambulatory care weighted workload units, are used in this analysis as indicators of the average patient's acuity or severity of illness. Average inpatient weighted workload units is computed by dividing inpatient weighted workload units by the total number of inpatients treated by the hospital. Similarly, average ambulatory care weighted workload units is simply the hospital outpatient measure—ambulatory care weighted workload units—divided by the total number of outpatient visits.

Severely ill patients who require intensive services for complex health problems should have much higher weighted workload unit values assigned to their discharge than those less acutely ill, such as long-term patients. Therefore, if the distribution of hospital patients is heavily skewed toward more severely ill cases, the average weighted workload units generated per case should be much higher than if a hospital sees predominantly long-term care patients. More acutely ill patients require more specialized staff and more sophisticated capital equipment than do those less acutely ill. This higher degree of intensity is hypothesized to raise the average cost of a hospital with a high severity value above that of a hospital with a lower degree of complexity.

The last grouping variable is the average salary per employee and is used as an imperfect proxy for the hospital-specific price of labor. It should be reflective of the degree of specialization among hospital staff and, to a lesser extent, of local labor market conditions. For example, a

long-term psychiatric facility would be expected to have a lower degree of labor specialization and lower average salary than a major medical/surgical teaching facility. However, average salary per employee is an imperfect indicator of labor supply and demand conditions since wage scales within VA are fixed on a national basis and are therefore not determined by local labor markets.

All data are from the 1986 fiscal year. Hospital aggregate data from a total of 159 VA hospitals were used in this analysis. The six independent VA outpatient clinics were not included.

RESULTS

The results of the first-stage grouping (not detailed here) using Ward's methodology indicated that a partition of the 159 VA facilities into six hospital groups, determined from the R^2 criterion, would be appropriate for the *K*-means second stage. The profiles of the six groups from the *K*-means second stage with a comparison of group means for several variables are presented in Tables 1, 2, and 3. The tables appear to show a marked difference in hospital groups on the basis of size, scope of services provided, and educational responsibilities. A comparison of the group characteristics in Tables 2 and 3, variables that were not used to generate the clusters, lends support to the validity of the proposed groups. The pairwise mean comparisons in Table 3 clearly show how the groups separate on particular variables.

Groups 1 and 2 are facilities that, on average, have fewer than 550 employees and less than $25 million FY 1986 budget dollars. Group 1 facilities have significantly higher teaching responsibilities than the Group 2 facilities. Group 1 facilities generally handle a larger proportion of surgical caseload than Group 2 (probably a reflection of the affiliation levels). Group 2 facilities appear to have lower caseloads as measured by inpatient weighted workload units and ambulatory care weighted workload units, but much higher long-term responsibilities measured by extended care weighted workload units. Group 1 facilities also tend to have significantly higher-cost staff than Group 2 hospitals.

The 45 facilities in Group 3 are highly affiliated hospitals with workload and staffing in the midrange of the six groups. These facilities tend to have much higher than proportional educational and research responsibilities. The hospitals in Group 3 (and Group 5) also tend to have the most expensive staffing of the six groups. Group 3 facilities are also characterized by high rates of discharge for surgery and low workloads in inpatient psychiatry.

Table 1: Means for the 11 Grouping Variables by Cluster

	Small Teaching, Group 1 N = *36*	*Small General, Group 2* N = *14*	*Midsize Teaching, Group 3* N = *45*	*Midsize General, Group 4* N = *16*	*Metropolitan Teaching, Group 5* N = *26*	*Psychiatric, Group 6* N = *22*	F-*Ratio*
Percentages of Discharges from							
Medicine	60	82	45	44	43	33	92.67
Surgery	31	10	37	26	32	02	123.18
Psychiatry	08	05	12	29	20	64	159.12
Neurology	003	000	053	012	036	003	97.66
Rehabilitation	003	032	008	009	013	007	4.84
Weighted Workload Units							
Inpatient	183,532	141,996	435,950	387,074	708,259	295,912	112.77
Ambulatory care	3,460,181	2,362,410	8,795,145	5,777,530	14,436,065	3,057,175	107.36
Extended care	31,731	77,190	54,275	125,113	113,125	201,620	24.70
Average salary per employee	29,037	27,991	30,338	27,751	29,995	26,824	35.94
Average Weighted Workload Units							
Inpatient	46.44	42.86	53.40	50.12	54.30	51.63	54.93
Ambulatory care	72.05	68.04	65.33	66.02	67.21	55.97	4.39

Table 2: Hospital Characteristics by Cluster

	Small Teaching, Group 1 N = *36*	*Small General, Group 2* N = *14*	*Midsize Teaching, Group 3* N = *45*	*Midsize General, Group 4* N = *16*	*Metropolitan Teaching, Group 5* N = *26*	*Psychiatric, Group 6* N = *22*
Beds (mean)	235	289	518	699	950	803
Inpatient discharge	4,047	3,345	8,142	6,476	12,523	3,496
Inpatient days of care	61,026	79,417	130,351	181,626	238,462	229,375
Average length of stay	15	29	16	28	20	70
Education (*N*)						
Heavy*	3	0	42	3	24	0
Light†	17	2	3	8	1	2
Nonteaching	16	12	0	5	1	20
Demography (*N*)						
Large urban	0	2	21	0	21	4
Urban	25	0	24	10	5	12
Rural	11	12	0	6	0	6

*Members of the Council of Teaching Hospitals (COTH).

†Hospitals affiliated with medical schools but not members of COTH.

Group 4 is a collection of 16 facilities that tend to offer an extensive scope of services across all hospital departments as reflected by the distribution of inpatient, ambulatory care, and extended care weighted workload units, and the distribution of major bed section proportions. These hospitals are also characterized by a much higher than proportional workload in intermediate care than all but the psychiatric group and Group 2.

Group 5 is a cluster of 26 teaching hospitals, generally located in metropolitan areas. These tend to have staffing levels 50 percent higher than other hospital groups and significant programs in most disciplines. They also have very large educational and research programs.

The 22 facilities of Group 6 are largely psychiatric facilities, with very small or nonexistent surgical programs, small medical programs, and large psychiatric and long-term care programs. These facilities tend to have the lowest average staffing costs, probably reflective of the case mix and complexity of the patients at these facilities.

EX-POST VALIDATION

In this section we examine the empirical significance of fairness. In this context, a reduction in systematic risk and a consequent improvement

Table 3: Comparison of Group Means for Other Variables (Significant mean differences in parentheses as measured at the .10 Level)*

	Small Teaching, Group 1 N = *36*	*Small General, Group 2* N = *14*	*Midsize Teaching, Group 3* N = *45*	*Midsize General, Group 4* N = *16*	*Metropolitan Teaching, Group 5* N = *26*	*Psychiatric, Group 6* N = *22*	F-*Ratio**
Total budget ($M)	24.2 (3,4,5,6)	19.9 (3,4,5,6)	63.5 (1,2,4,5,6)	49.6 (1,2,3,5)	101.9 (1,2,3,4,6)	41.3 (1,2,3,5)	109.38
VA PPS controlled budget ($M)	13.6 (3,4,5,6)	11.5 (3,4,5,6)	35.1 (2,3,4,5,6)	29.0 (1,2,3,5)	55.9 (1,2,3,4,6)	24.9 (1,2,3,5)	118.60
Total staffing	524 (3,4,5,6)	507 (3,4,5,6)	1,249 (1,2,5)	1,239 (1,2,5)	2,112 (1,2,3,4,6)	1,144 (1,2,5)	86.18
Average budget per full-time equivalent employee (FTE)	44,890 (2,3,4,6)	39,211 (1,3,5)	47,270 (1,2,4,6)	39,427 (1,3,5,6)	45,337 (2,4,6)	35,899 (1,3,4,5)	34.91
Average number of residents	19.4 (3,5)	1.4 (3,5)	109.5 (1,2,4,5,6)	29.4 (3,5)	151.5 (1,2,3,4,6)	6.5 (3,5)	105.34
Average number of FTE research employees	4.08 (3,5)	.57 (3,5)	37.87 (1,2,4,5,6)	8.81 (3,5)	64.62 (1,2,3,4,6)	4.55 (3,5)	21.94

*Significance indicated by use of analysis of variance with Tukey's multiple comparisons test.

in fairness is measured by a reduction in the deviation of a hospital's observed unit costs from the reimbursement rate.

Under a grouping scheme, this reimbursement rate is the average expenditure of the group in which the hospital is classified rather than the national average cost. If the VA grouping scheme is more fair than a system based on the previous nonhomogenous fairness adjustments outlined in Section 2, then the sum of the squared deviations of each hospital's observed unit expenditures from the group reimbursement rate should be less than the sum of the squared deviations of hospital unit expenditures from the previous nonhomogenous adjustment model relative to a national reimbursement rate.

Data to validate the adequacy of an allocation based on this grouping scheme can be found in Table 4. Table 4 shows a comparison of the six VA hospital groups in an analysis of variance on the basis of several average expenditure variables: mean inpatient expenditures per inpatient weighted workload unit; mean ambulatory care expenditures per ambulatory care weighted workload unit; and mean extended care expenditures per extended care weighted workload unit.

The R^2 statistic in Table 4 represents the reduction in the sum of squared deviations of hospital unit expenditures from reimbursements by using the group average rather than the national average. This reduction in overall squared deviations for the grouping scheme is disaggregated in Table 4 into the three components of average VA hospital cost. Using a group mean reduces by 24 percent the squared deviations of hospital average inpatient expenditures from the national average, 6 percent for mean ambulatory care expenditures, and 6 percent for extended care expenditures. The relative size of these reductions reflects the proportions of these components of the overall budget and also reflects the dominance of inpatient variables used in the grouping model.

In results not reported here, an analysis of variance also compared the relative improvements of the previous nonhomogenous adjustment model to a model using a national average reimbursement rate. The aggregated squared deviations are reduced by amounts virtually identical to those of the grouping scheme through use of the small-hospitals, psychiatry, and education adjustments rather than the national average expenditure reimbursement rate. However, the relative improvement of the VA grouping model probably represents a lower-bound estimate. Observed cost-outlier hospitals, which are placed into the correct groups on efficiency grounds, account for a greater-than-average proportion of the within-group deviations. This would occur if inefficient hospitals (with higher observed unit expenditures and which should

Table 4: Means ($) and Standard Errors (in parentheses) for Average Expenditure Variables

	Small Teaching, Group 1	*Small General, Group 2*	*Midsize Teaching, Group 3*	*Midsize General, Group 4*	*Metropolitan Teaching, Group 5*	*Psychiatric, Group 6*	F-*Ratio**	p-*Value*	R^2
Average inpatient expenditures per average inpatient weighted workload unit	46.35 (.85) (3)	42.79 (2.5) (3,5)	52.32 (.99) (1,2,4,6)	44.72 (1.0) (3)	49.04 (.93) (2)	45.22 (1.2) (3)	9.46	.001	.24
Average ambulatory care expenditures per average ambulatory care weighted workload unit	1.01 (.004)	1.01 (.004)	1.01 (.013)	1.00 (.015)	.98 (.006)	1.02 (.005)	1.85	.11	.06
Average extended care expenditures per average extended care weighted workload units	46.85 (2.4)	42.21 (1.6)	48.42 (2.2)	43.98 (1.6)	45.70 (1.3)	42.37 (.84)	1.58	.17	.06

*Significance of means values indicated by use of analysis of variance with Tukey's multiple comparisons test.

have lower true unit costs) were placed with hospitals with lower observed unit expenditures. The grouping model therefore should have at least the same improvement in fairness as the previous nonhomogenous VA adjustment model. The adjustment model also does not account for other factors that affect hospital costs. It appears, therefore, that systematic risk should be reduced by applying the VA PPS within each group.

A second estimate of the validity of the VA grouping model was made by contrasting the groups along the average expenditure variables in Table 4. If the methodology and the 11 criteria are correctly predicting the separation and the composition of hospital peers, then the true average cost variables should be significantly different across the six VA hospital groups. In Table 4, mean inpatient expenditures show significant differences across the six hospital groups according to the *F*-ratio. Mean ambulatory care and mean extended care expenditures show little variation and are insignificantly different across the six groups. Some degree of overlapping also occurs in the distribution of the mean inpatient expenditures, a drawback of the clustering methodology characterized by Vertrees and Manton (1986, 288) as the "boundary problem." This occurs because the hospitals, grouped with multiple characteristics that vary continuously, are forced into discrete categories. Boundaries separating groups of hospitals may not be clear and distinctive across all variables, as is the case in Table 4. Here at least two distinct inpatient expenditure groups exist. The first is the low-unit expenditure group composed of groups 1, 2, 4, and 6. The small teaching facilities, Group 1, appear to form a high-unit expenditure group within these hospitals. The midsize and metropolitan teaching hospitals seem to form a high-cost group, as would be expected since they have the most extensive educational responsibilities.

DISCUSSION

While the clustering approach appears to offer the only practical approach to altering the VA PPS, the methodology has one drawback. The groups suggested from the application of the clustering methodology cannot be shown, based on statistical tests, to be the absolute optimal partition of the hospitals in terms of promoting fairness. This article does show, however, that the average cost prices resulting from the six VA peer groups do improve fairness in a relative sense. Using a definition that characterizes a fairness improvement as a reduction in the variation between a hospital's costs and its reimbursement price,

the clustering results offer at least the same improvements as the previous nonhomogenous adjustment model. In addition, the validity of the groups is supported by analysis of variance results that show that the groups separate on variables external to the clustering algorithm.

In 1988, VA adopted the hospital grouping methodology for use in the 1989 fiscal year, with several administrative alterations. First, the two small hospital groups were altered by a management review panel to better support the educational affiliations at those facilities. The 50 hospitals in the small teaching and small general clusters were redivided according to the number of residents per million dollars of VA PPS controlled budget. A break point arbitrarily set at one resident per million dollars resulted in 15 hospitals above this ratio being placed into the small teaching category, and 35 hospitals below this ratio being placed in the small general category. An administrative appeals mechanism was also established to allow facilities the opportunity to change groups under extenuating circumstances. Further refinements are intended to improve the discriminating power of the variables and methods. Additional research is also planned to examine the stability of the results over time and to estimate the relative fairness of this model with other competing methodologies.

ACKNOWLEDGMENTS

The authors are grateful to James Burgess, Gordon DeFriese, Randall Ellis, Thomas McGuire, Joseph Newhouse, Elizabeth Stefos, Ingo Vogelsang, and two anonymous referees for their helpful suggestions and comments during the preparation of this text. The views expressed in this article are the authors' and do not necessarily represent those of the Department of Veterans Affairs.

REFERENCES

Alexander, J., C. Evashwick, and T. Rundall. "Hospitals and the Provisions of Care to the Aged: A Cluster Analysis." *Inquiry* 21 (Winter 1984): 303–14.

Aldenderfer, M., and R. Blashfield. *Cluster Analysis.* Beverly Hills, CA: Sage Publications, 1984.

Altman, S., and M. Rodwin. "Halfway Competitive Markets and Ineffective Regulation: The American Health Care System." *Journal of Health Politics, Policy and Law* 13, no. 2 (Summer 1988): 323–39.

Arrow, K. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review* 53, no. 5 (1968): 941–73.

Carey, K., and T. Stefos. "Measurement of the Unit Costs of Hospital Outpatient Services." Unpublished paper, Management Science Group, December 1991.

Christensen, L., D. Jorgenson, and L. Lau. "Transcendental Logarithmic Production Frontiers." *Review of Economics and Statistics* 55, no. 1 (February 1973): 28–45.

———. "Transcendental Logarithmic Utility Functions." *American Economic Review* 65, no. 3 (1975): 367–83.

Ellis, R., and T. McGuire. "Insurance Principles and the Design of a Prospective Payment System." *Journal of Health Economics* 7, no. 3 (1988): 215–37.

———. "Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply." *Journal of Health Economics* 5, no. 2 (June 1986): 129–52.

Friedman, B., and M. Pauly. "Cost Functions for a Service Firm with Variable Quality and Stochastic Demand: The Case of Hospitals." *The Review of Economics and Statistics* 3, no. 4 (November 1981): 620–24.

Grannemann, T., R. Brown, and M. Pauly. "Estimating Hospital Costs: A Multiple-Output Analysis." *Journal of Health Economics* 5, no. 2 (1986): 107–27.

Guterman, S., P. Eggers, G. Riley, T. Green, and S. Terrell. "The First Three Years of Medicare Prospective Payment: An Overview." *Health Care Financing Review* 9, no. 3 (Spring 1988): 67–77.

Hartke, L. *DRGs Concept and Use.* Madison, WI: Institute for Health Planning, 1983.

Holahan, J., and J. Palmer. "Medicare's Fiscal Problems: An Imperative for Reform." *Journal of Health Politics, Policy and Law* 13, no. 1 (Spring 1988): 53–81.

Hornbrook, M., and A. Monheit. "The Contribution of Case Mix Severity to the Hospital Cost Outpatient Relation." *Inquiry* 22 (Fall 1985): 259–71.

Jencks, S., A. Dobson, P. Willis, and P. Feinstein. "Evaluating and Improving the Measurement of Hospital Case Mix." *Health Care Financing Review* (Annual Supplement, November 1984): 1–11.

Lau, L., and P. Yotopoulos. "Profit Supply and Factor Demand Functions." *American Journal of Agricultural Economics* (1972): 11–18.

Lindsay, C. *Veterans Administration Hospitals.* Washington, DC: American Enterprise Institute for Public Policy Research, 1975.

McFadden, D. "Cost, Revenue, and Profit Functions." In *Production Economics: A Dual Approach to Theory and Applications.* Edited by M. Fuss and D. McFadden. Amsterdam: North-Holland, 1978.

National Academy of Sciences. *Study of Health Care for American Veterans.* Washington, DC: U.S. Government Printing Office, 1977.

Phillip, J., and R. Iyer. "Classification of Community Hospitals." *Health Services Research* 10, no. 4 (Winter 1975): 349–68.

Pope, G. C. "Using Hospital Specific Costs to Improve Fairness of Prospective Reimbursement." *Journal of Health Economics* 9, no. 3 (1990): 237–51.

Shleifer, A. "A Theory of Yardstick Competition." *The RAND Journal of Economics* 16, no. 3 (1985): 319–27.

Spady, R., and A. Friedlaender. "Hedonic Cost Functions for the Regulated Trucking Industry." *Bell Journal of Economics* 9, no. 1 (Spring 1978): 159-79.

Thorpe, K. "Why Are Urban Hospital Costs So High? The Relative Importance of Patient Source of Admission, Teaching, Competition, and Case Mix." *Health Services Research* 22, no. 6 (February 1988): 821-36.

Trivedi, V. "Classification of Short Term General Hospitals for Control and Equity." *Inquiry* 15 (September 1978): 255-64.

———. Commentary. *Inquiry* 16 (Winter 1979): 363-67.

Vertrees, J., and K. Manton. "A Multivariate Approach for Classifying Hospitals and Computing Blended Payment Rates." *Medical Care* 14, no. 4 (1986): 283-300.

Vita, M. "Exploring Hospital Production Relationships with Flexible Functional Forms." *Journal of Health Economics* 9, no. 1 (June 1990): 1-21.

Vogelsang, I., and J. Finsinger. "A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms." *Bell Journal of Economics* 10, no. 1 (1979): 157-71.

Zwanziger, J., and G. Melnick. "The Effects of Hospital Competition and the Medicare PPS Program on Hospital Cost Behavior in California." *Journal of Health Economics* 7, no. 4 (December 1988): 301-20.