A Statistically Based Method for Identifying Hospital Classification Criteria

BRUCE R. NEUMANN, PhD

A MAJOR ISSUE IN HEALTH SERVICES RESEARCH is the comparability and homogeneity of providers of care. Any evaluation of performance across institutions requires a determination as to whether the services or outputs are so dissimilar as to void the analysis. Yet in many analyses of hospital costs in which ad hoc comparisons are made between institutions, the researcher does not explicitly state the basis of the comparisons. Comparability is required both in pure research studies and for studies concerned with establishing cost limitations for reimbursement purposes.

A variety of models for classifying hospitals have been suggested in the health services research literature, for example, by Berry (*1,2*), Hess and Srikantan (*3*), Kaluzny and co-authors (*4*), and Veney (*5*). In each of these studies, the authors relied on a statistical model of one form or another to confirm their intuitive judgments regarding the variables by which sets of hospitals could be differentiated. That is, the classification process was based on judgments that were validated by statistical testing and the construction of models. The research reported here relies more heavily on a statistical model and less on judgment.

Several statistically based classification models for hospitals have been recently suggested. The State of Washington has developed a set of classification criteria based on cluster analysis (*6*). Phillip and Iyer (*7*) suggest a similar approach to classifying all U.S. hospitals. In its pure form, cluster analysis is used to establish peer groups and not to establish classification criteria. In my study, however, a statistical model was used to establish the classification criteria. Kosnick (*8*) also suggests that such an approach is feasible.

In a report entitled "Hospital Costs in Colorado" (*9*), the authors suggested that each hospital be assigned to a unique peer group that may not overlap any other peer grouping. That is, the classification criteria change, depending on the individual hospital of interest. Again, these criteria are based solely on judgment and precedent. In similar fashion, many other classifications of hospitals rely solely on the researcher's judgment and previous experience, for example, the classifications of Agnew (*10,11*), Carr and Feldstein (*12*), Francisco (*13*), and the Department of Health, Education, and Welfare (*14*).

The Federal Government is becoming more involved in the specification of mandatory peer groups. The Department of Health, Education, and Welfare (DHEW) now requires that hospitals be grouped by size and economic environment (urban versus nonurban and by groups of similar States). The DHEW criteria result in a three-dimensional classification of all U.S. hospitals. Reimbursement criteria based on maximum daily costs are applied to all hospitals in the same stratum. That is, all hospitals in a given cell of the classification model are eligible for no more than the allowed standard per diem costs.

Given the effect of this type of legislation on hospital reimbursement levels, it is essential that similar hospitals be grouped in the same stratum and that each level of the classification model be significantly

An abstract of this paper was presented at the Southwestern Regional Meeting of the American Accounting Association, held in Houston, Tex., March 14–17, 1979. Tearsheet requests to Dr. Bruce R. Newmann, Associate Professor of Accounting, School of Business, State University of New York, Albany, N.Y. 12222.

differentiated from any other groups (strata) of hospitals. The classification model that I have developed can be used to differentiate groups of hospitals that actually have different performance characteristics. An automatic interaction detector (AID) is used to select attributes and characteristics by which large community hospitals can be classified.

Data Base

The data base for my research was a subset of 92 large community hospitals that were reporting members of Hospital Administrative Services (HAS)—a division of the American Hospital Association—in each month of 1970 and 1971. The hospitals were located throughout the United States. Moreover, the subset comprised more than 25 percent of the large community hospitals in the United States. Although HAS membership may represent some self-selection bias, I was unable to discern any major differences between the subset of 92 hospitals with such membership and the population of large community hospitals in general.

All hospitals in the AID subset had 400 or more beds and had the following additional characteristics in common in both 1970 and 1971:

General medical and surgical services (service code = 10)
Short-term lengths of stay (LOS = 1)
A full- or part-time pharmacy
An intensive care unit
A physical therapy department
Radioisotope facilities
Approval as Medicare participants
Accreditation by the Joint Commission on Hospital Accreditation

Consequently, these variables or service characteristics were not used as inputs for the AID analyses.

Raw data for the subset of hospitals, which all remained anonymous, were obtained from Hospital Administrative Services. The data were carefully edited and purged of obvious errors in reporting. The monthly HAS data reports were all transformed by simple addition into annual data. After 5 hospitals were entirely deleted, the usable data base consisted of 87 large community hospitals in urban locations.

In the following table, the characteristics of the AID subset of hospitals are compared with those of all the community hospitals with more than 399 beds that responded to the American Hospital Association's 1972 annual survey. All large U.S. community hospitals are well represented in the AID subset:

	Averages per hospital	
Characteristic	*AID subset of hospitals*	*Large community hospitals*
Total hospitals	87	339
Total beds	543	555
Total admissions	18,947	18,999
Daily census	459	467
Cost per patient day	$112.26	$97.47
Percent occupancy	83.3	84.1
Full-time equivalent employees per bed	2.65	2.78

SOURCE: Hospital statistics, 1971. American Hospital Association, Chicago, Ill., 1972.

The Automatic Interaction Detector

The automatic interaction detector algorithm was the classification model that I used in identifying the relationships between a hospital's services and facilities and its total operating costs. AID is a flexible search technique designed specifically for use on qualitative data in which the relationships may be nonlinear, interactive, or multicollinear. With AID, as updated to AID 3, a complex prestated strategy is followed in searching for the predictors that will increase the variance accounted for by the dependent variable (*15*). Assael (*16*) provides the following definition of AID:

AID is a multivariate technique for determining what variables and categories within them continue to produce the greatest discrimination in group means by the dependent variable. The program divides the sample through a series of binary splits into mutually exclusive subgroups.

With AID, each group is partitioned into subgroups in such a way that the difference between the two new group means accounts for more of the total sum-of-squares than the difference between the group means of any other pair of subgroups (*17*). This least-squares evaluation of predictive error is analogous to a regression least-squares criterion. However, unlike stepwise regression or analysis of variance, AID measures the effect of each predictor on each subgroup, rather than over the whole sample. Since my research was based on dichotomous variables, I needed a statistical technique, such as AID, with which such variables could be analyzed.

Hospital services affect the cost structure (the fixed and variable relationships) of the hospital. For example, certain services have higher fixed costs relative to other services. This relationship between the services available and the cost structure within the hospital is the justification for using the presence or absence of services and facilities as classification criteria. Hospitals with the same facilities and services are expected to have cost structures similar and comparable to other hospitals with similar services and facilities. Dichotomous or qualitative variables may then be used as classification criteria and can serve as a surrogate indicator of the hospital's cost functions. Given this desire to classify based on the facilities and services provided by a hospital, AID is one of the few statistical techniques that can identify these types of variables. Consequently, average monthly costs were used as the dependent variable. Identical results, however, would be obtained if average annual costs were used as the dependent variable.

Since the possible combinations of predictors and predictor classes in AID is vast, the usual notion of degrees of freedom explodes, and therefore statistical tests cannot be applied to the results. However, I evaluated the AID classification in terms of configuration scores, since these scores may be used as an index of variation.

The purpose of the configuration score is to indicate what part of the total variation is explained by a particular set of groups that has been identified under the AID algorithm. Sonquist and his co-authors (*17*) have stated that if in a two-stage analysis involving forced splits, the configuration score does not decrease markedly from the original case, the analyst can have some confidence in the stability and reproducibility of the results. These authors describe the configuration score as follows:

Computation of the configuration score provides the analyst with some indication of what his predictors are worth in explanatory power when all the "stops are pulled." It is suggested that if the variation explained by the configuration is undesirably small, the analyst had best spend his time obtaining hints as to what other variables he might undertake to include in subsequent investigations.

Instead of operating sequentially it [the configuration score] subdivides the sample factorially (into all possible combinations of the predictor set), even though some combinations have few or no cases, and does a one-way analysis of variance components indicating what fraction of the total variance (around the mean) is accounted for by the subgroup means . . . Basically this option calculates a single one-way analysis of variance asking what fraction of the total variance is accounted for by the subgroup means if one defines a subgroup for each combination of predictor classes.

The configuration score is thus the percentage of variance in the dependent variable that is attributable to the split. In my research, if a variable was to be used as a classification criterion, it had to explain 3 percent of the total (original) variation. Each split also had to result in a group size in excess of five hospitals. Use of configuration scores increases the reliability of the AID method.

Evaluation of the stability of the AID results also requires replication, evaluation of competing variables, and the possible reduction of end-points (terminal branches) of the AID tree. With AID a specified pattern can be forced onto a second data set. Such forcing of a prespecified pattern is necessary because if the AID algorithm is permitted to freely select a new structure from a second data set, the probabilities of identifying two identical trees are infinitesimal (*17*). Therefore, I implemented a two-stage application of AID, using it first on the 1970 data and then forcing it onto the 1971 data. If the classification process was stable, one would expect to find most of the same variables being selected independently from both data sets. Since both the 1970 and 1971 data exhibited such stability, I forced the 1970 classification pattern onto the 1971 data to determine whether the forced splits would explain almost as much variance as was explained by the original splits that had not been forced.

There is another interpretation of the data that lends credence to the results achieved with AID. In most of the studies cited, the researchers identified hospital size (as measured by the number of beds for adults) as a crucial variable in the classification of hospitals. My empirical research could be viewed as a detailed analysis of one branch of a more global classification model in which only the initial classifiction variable would be hospital size. Using all available 1970 and 1971 HAS data for the subset of 87 large community hospitals, I sought to determine which additional variables could be used to further stratify the test hospitals.

The characteristics that I used as input to AID are listed in table 1, along with the number of hospitals having the given characteristic or service. Descriptions of each of the characteristics can be found in any annual American Hospital Association Guide to the Health Care Field or in the definitional instructions for the association's annual survey. Kimball and Lorant (*18*) and Berry (*19*) used similar service characteristics to classify hospitals.

The educational service characteristics (intern and residency programs, medical school affiliation, and membership in the Council of Teaching Hospitals) have been redefined into several classes of a new variable called medical education programs. The rationale for doing this is based on research by Hess and Srikantan (*3*) and on the educational variables defined by the Blue Cross Association. Lave and Lave (*20*) reported that the Blue Cross Association used the following indicators of teaching status: "An advanced teaching hospital is one with a medical school affiliation and with three or more approved residency programs; and a teaching hospital is one with an approved nursing school or internship program or with at least one residency program." I identified the following three sets of medical education characteristics in the test hospitals:

• At least one residency program, an internship program, and membership in the Council of Teaching Hospitals.

• At least one residency program and an internship program.

• Other medical education programs.

The first set of medical education characteristics represents advanced teaching programs. The second set is termed teaching programs. The remaining set is called other programs. A hospital-based nursing school is not included as a criterion in any of the three categories; nor does nursing education enter into any of the AID classification criteria.

In a sense, AID is used to replicate the thought processes of other researchers, all of whom have selected a set of classification criteria and have intuitively chosen a set of characteristics to use in their research designs. I expected the AID algorithm to closely parallel "the activity of a researcher investigating a body of data with a basic theory of which variables are important" (*17*). Moreover, with AID I hoped to obtain a classification method so structured and reproducible that the resulting classification characteristics could be justified by the AID criterion of least-squares variance reduction. In this context, reproducibility is a measure of reliability.

The 26 characteristics listed in table 1, composed of 24 dichotomies and two multicategory variables, would

Table 1. Characteristics of test hospitals that were input to automatic interaction detector (AID) algorithm, 1970 and 1971

No. and descriptions of characteristics	*Number of hospitals* 1970	1971
1. Control (all nonprofit, non-Federal)	87	87
State (governmental)	5	5
County (Governmental)	5	5
City (governmental)	1	1
Hospital district (governmental)	5	5
Church-operated (nongovernmental)	14	12
Community (nongovernmental)	57	59
2. Postoperative recovery	86	86
3. Inhalation therapy	85	86
4. Histopathology	86	82
5. Intensive cardiac care	79	81
6. X-ray therapy	79	82
7. Electroencephalograph	86	86
8. Radium therapy	78	75
9. Social work	76	80
10. Emergency psychiatric	62	59
11. Inpatient psychiatric	69	69
12. Occupational therapy	59	58
13. Renal dialysis (inpatient and outpatient)	56	56
14. Cobalt therapy	61	62
15. Open-heart surgery	59	59
16. Organ bank	18	17
17. Blood bank	81	80
18. Premature nursery	78	79
19. Extended care unit	8	6
20. Rehabilitation unit	42	41
21. Outpatient department	77	73
22. Emergency department	85	86
23. Cancer program (American Cancer Society)	54	55
24. Professional nursing school (American Nurses Association)	43	37
25. Blue Cross participant (Blue Cross Association)	86	85
26. Medical education	87	87
Advanced teaching programs	62	61
Teaching programs	14	13
Other programs	11	13

form an astronomically large number of possible classification structures. If all these characteristics were applied jointly, the resulting set of possible classifications would have 201,326,592 cells: $(2^{24})(4^{1})(3^{1}) = 201,326,592$. By use of the AID algorithm, the number of characteristics needed to evaluate the costs for the test hospitals is reduced, and possible subdivisions that would be unreliable (insufficient data) or irrelevant (no explanatory ability) are discarded or avoided. By using AID, the list of 26 characteristics in table 1 is reduced to a much more reliable and manageable set of classification criteria.

Results With AID Algorithm

The information obtained in the AID analyses is shown in table 2. The tree diagrams indicate the branch points and terminal subdivisions for each phase of the analysis. The diagrams are restricted to those splits that meet the configuration score criteria. As can be seen, the terminal branches and their associated characteristics ultimately serve as the criteria for determining groups of homogeneous hospitals.

Separate AID analyses performed on the 1970 and 1971 data resulted in an explanation of 32.9 percent of the variation in costs for 1970 and 35.0 percent of

Table 2. Results of application of automatic interaction detector (AID) algorithm to average monthly costs for test hospitals, 1970 and 1971

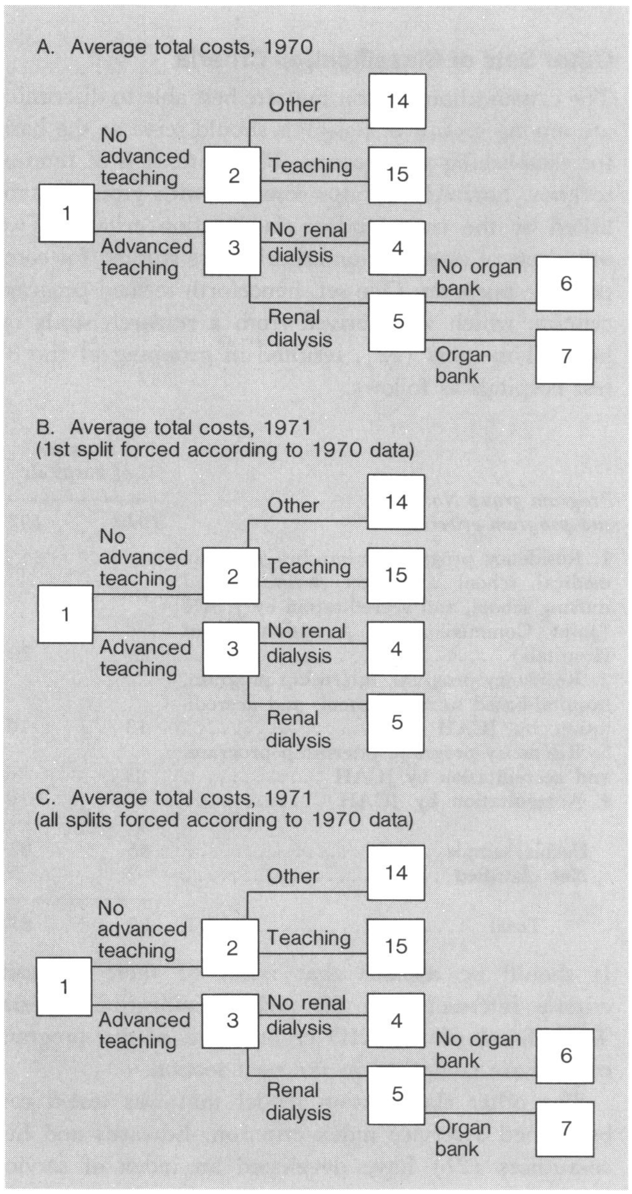
Split on characteristics	*AID group No.*[1]	*Number of hospitals*	*Average monthly costs*	*Means or variance ratio*	*Percent of variation explained*
	A. Results of application to 1970 costs				
	1.	87	$1,365,468	.0042	
Advanced teaching	2.	25	1,012,340	.0096	15.8
	3.	62	1,507,858	.0044	
Renal dialysis	**4.**	19	1,223,244	.0049	8.0
	5.	43	1,633,615	.0049	
Organ bank	**6.**	31	1,519,050	.0053	5.3
	7.	12	1,929,577	.0053	
Teaching	**14.**	11	780,738	.0183	3.8
	15.	14	1,194,312	.0146	
Total					32.9
	B. Results of application to 1971 costs (1st split forced according to 1970 data)				
	1.	87	$1,526,712	.0036	
Advanced teaching	2.	26	1,159,291	.0068	13.8
	3.	61	1,683,317	.0037	
Renal dialysis	**4.**	18	1,184,730	.0063	17.5
	5.	43	1,892,028	.0046	
Teaching	**14.**	18	968,836	.0137	6.1
	15.	8	1,587,812	.0125	
Total					37.4
	C. Results of application to 1971 costs (all splits forced according to 1970 data)				
	1.	87	$1,526,712	.0036	
Advanced teaching	2.	26	1,159,291	.0069	13.8
	3.	61	1,683,317	.0037	
Renal dialysis	**4.**	18	1,184,730	.0031	17.5
	5.	43	1,892,028	.0046	
Organ bank	**6.**	30	1,808,923	.0041	1.9
	7.	13	2,083,806	.0064	
Teaching	**14.**	8	968,836	.0137	5.9
	15.	18	1,587,812	.0125	
Total					39.1

[1] Boldface type denotes final groups.

the variation for 1971 (not shown). Three of the characteristics that were identified as criteria for the 1970 total costs (table 2, part A) also appeared in the 1971 results (not shown), namely, advanced teaching programs, renal dialysis facilities, and teaching programs.

Since the AID structure based on 1970 data met the stability criteria previously discussed, this structure was forced onto the 1971 data. To further test the impact of advanced teaching programs, which had been identified in the analysis of the 1970 data (table 2, part A), this characteristic was forced to serve as the first split of the 1971 data (table 2, part B). This forced split had the unexpected result of increasing the percentage of variation for 1971 that was explained—from 35.0 percent (not forced) to 37.4 percent (forced). This result indicates that the pattern or structure obtained by using the advanced teaching characteristic reveals more information than can be uncovered with the AID algorithm alone (without intervention). It also indicates that the latter structure (table 2, part B) has more predictive ability than does the "unforced" analysis of 1971 data. Additionally, the basic structure is almost identical with that of the 1970 data, even though only the first split was forced. This similarity suggests that a basic structure underlies both data sets and that teaching programs and renal dialysis services are both essential to the underlying pattern or structure.

Tree diagrams showing branch points and terminal subdivisions for the analyses in table 2



The final step in this iterative procedures was to force the entire 1970-based structure onto the 1971 data (part C of table 2). It is of interest that the total variation explained in this final step increased slightly over the preceding step, from 37.4 to 39.1 percent. This slight increase in predictive ability indicates that the structure is consistent and meaningful over both years.

The implication of these analyses is that hospitals can be classified by heretofore unidentified criteria that will facilitate the evaluation of their comparative efficiency. Table 2, part B, shows that the classification structure obtained is partially reproducible. In fact, when exactly the same structure is forced upon the 1971 data, its predictive ability actually increases. Thus, with AID as the classification method, a unique set of statistically based classification criteria can be identified. As was expected, total hospital costs were found to increase as more services were provided (table 3). This evidence supports the general contention of Edwards and co-authors (*21*) that as hospital services increase in scope, hospital costs also increase.

The ranking of hospitals based on the AID service criteria closely approximates their ranking based on total costs, costs per patient day, and costs per admission, as the following table shows. In each of the cost columns, the rank of each group is almost identical across the 2 years.

AID group No. and classification criteria	*Rank based on cost measures*		*Rank based on AID criteria*
	1970	*1971*	*1970 and 1971*
14. Other teaching	1	1	1
15. Teaching	2	3	2
4. Advanced teaching	3	2	3
6. Advanced teaching, renal dialysis	4	4	4
7. Advanced teaching, renal dialysis, organ bank	5	5	5

Table 3. Cost comparisons for AID groups, 1970 and 1971

AID group No. and classification criteria	Average monthly costs		Cost per patient day		Cost per admission	
	1970	1971	1970	1971	1970	1971
14. Other teaching	$ 780,738	968,836	$ 66.41	$ 73.70	$537,58	$580.03
15. Teaching	1,194,312	1,587,817	80.82	94.30	724.62	852.72
4. Advanced teaching	1,223,244	1,184,730	84.44	85.09	761.62	732.88
6. Advanced teaching, renal dialysis	1,519,050	1,808,923	94.42	111.85	874.01	1,029.57
7. Advanced teaching, renal dialysis, organ bank	1,929,577	2,083,806	103.73	117.06	993,79	1,013.26

The data in this table provide an inherent measure of validity, in that an increase in services corresponds to an increase in costs. These rankings indicate an inherent correspondence between the structure of hospital services and cost behavior patterns. Given the limited number of groups, even nonparametric statistics cannot be used to test the divergence in rankings in 1971 (a divergence that is consistent for several different cost measures). The one divergence in the rankings (group 15 and group 4 flip-flop between 1970 and 1971) indicates that the classification structure may need to be periodically revised. However, changes in group rankings do not invalidate the use of peer groups for reimbursement purposes, since the analyst is primarily interested in annual trends within groups, rather than among groups. That is, cost comparisons and trends in costs are of major interest for a particular group of hospitals. The observation that in 1971 the costs for hospitals in group 15 exceeded the costs for those in group 4 should have little impact on the evaluation of either of these groups of hospitals.

The rankings of the several different cost ratios according to hospital size (number of beds) is not the same as the AID ranks:

Hospital groups according to number of beds	*Rank based on cost measures 1970*	*1971*	*Rank based on AID criteria 1970 and 1971*
Group 1	1	1	1
Group 2	3	4	2
Group 3	2	2	3
Group 4	4	3	4
Group 5	5	5	5

The AID results have a better relationship to cost behavior patterns than do classification criteria based on size alone. To this extent, the AID results are valid. I have already documented their reliability in conjunction with the reproducibility and stability of the analyses shown in table 2. What is more important, the criteria used to identify the five AID groups can serve as the basis for further analyses of hospital costs and efficiency (for example, to determine whether there are identifiable dimensions of differences in the components of total costs).

Other Sets of Classification Criteria

The classification criteria that are best able to discriminate among groups of hospitals should serve as the basis for establishing peer groups. Therefore, it is of interest to know how AID groups compare with groups established by the use of other classification criteria. Two other sets of classification criteria were studied for comparative purposes. One set, henceforth termed program criteria, which was derived from a research study on hospital mergers (*22*), resulted in grouping of the 87 test hospitals as follows:

Program group No. and program criteria	*Number of hospitals 1970*	*1971*
1. Residency program, internship program medical school affiliation, hospital-based nursing school, and accreditation by JCAH (Joint Commission on Accreditation of Hospitals)	29	20
2. Residency program, internship program, hospital-based nursing school, and accreditation by JCAH	13	16
3. Residency program, internship program, and accreditation by JCAH	33	36
4. Accreditation by JCAH	10	10
Usable sample	85	82
Not classified	2	5
Total	87	87

It should be obvious that many of these program criteria intersect with the AID classification criteria. The efficacy of the AID criteria and of the program criteria are compared in the next section.

The other classification model that was tested can be termed a service index criterion. Edwards and her co-authors (*21*) have developed an index of service capabilities, using a count of the presence or absence of certain facilities or services. Their service index ranges between 1 and 20, and a service index (score) can be calculated for each hospital. These authors demonstrated a positive relationship between an increase in services and an increase in total operating costs. For research purposes, the sample of 87 hospitals in my study can be partitioned according to the criteria in the following table, hereafter called service index criteria.

Service index group No. and service index criteria	*Number of hospitals*	
	1970	*1971*
1. 13–14	8	8
2. 15–16	23	21
3. 17–18	34	33
4. 19–20	17	19
Usable sample	82	81
Insufficient data	5	6
Total	87	87

When the test hospitals were classified into different groups according to each of the different classification criteria, many of the groups contained similar subsets. The purpose of my research was to test whether any of the classification criteria were effective in differentiating between groups of hospitals.

Since the smallest group of hospitals identified under any of the other sets of classification criteria comprised only eight institutions, use of a large number of performance variables in the comparison would have been inappropriate. This degrees of freedom problem may be encountered whenever sample sizes are small. To counter it, a principal component factor analysis was used to identify 8 principal components among the 92 different performance variables that were available in the data base. Factor analysis was applied to each separate category of variables.

Categories of variables	*Variables in each category*	*Principal components*	*Variables in each principal component*
Salaries	18	1	1
		1	2
Overhead	20	1	1
		1	3
Labor hours	17	1	1
Use of services	22	1	1
Size	9	1	2
Financial position	6	1	1
Total variables	92	8	12

The complete factor results are beyond the scope of this paper, but I will supply them upon request. To reduce a large and cumbersome data base to manageable proportions, I restricted the domain of my investigation to the following eight performance variables:

Performance variables	*Categories of variables*
1. Nursing salaries	Salaries
2. Medical staff and general service salaries	Salaries
3. Employee health and welfare	Overhead
4. General services, administration and fiscal, and depreciation	Overhead
5. Nursing man-hours	Labor hours
6. Holiday and vacation paid	Financial position
7. Laboratory tests	Use of services
8. Square footage capacity serviced by plant and housekeeping departments	Size

The performance variables identified by principal components analysis retained most of the predictive ability of the larger set (92) of original performance variables. Therefore, they were the inputs to the multiple discriminant analysis of the other sets of classification criteria.

To evaluate the overall significance of the discrimination between groups, Rao's F ratio (*23*) was used to test whether the group centroids were equal or unequal. The purpose of this test was to determine the discriminatory ability of the principal component predictor variables.

In addition, the significance of each successive discriminant function was tested with Bartlett's V as an approximate chi-square statistic: that is, if the overall discriminatory ability is significant under Rao's F ratio, Bartlett's V is used to determine how many separate dimensions of difference are significant. Each successive discriminant function was evaluated with a variant of Bartlett's V statistic (*23*):

> As soon as the residual after removing the first s discriminant functions becomes smaller than the prescribed centile point . . . of the appropriate chi-square distribution, we may conclude that only the first s discriminant functions are significant at the alpha level.

This test can be used to identify the dimensions along which significant differences among groups can be found. Its implications are that if groups of hospitals differ solely on one dimension (for example, factor prices), then administrators or government regulators only need to take into account certain variables and not the entire universe of possible control variables.

High statistical significance may not be a sufficient indication of discriminatory power. Therefore, Tatsuoka (24) suggests a multivariate test (ω^2) of group differences that will indicate what part of the total variability of the discriminant functions can be attributed to group differences. The ω^2 is the percentage of the variability in the discriminant space that is relevant to group differentiation. If there are no significant discriminant functions, then the discriminatory power is essentially zero. Therefore, three tests (Rao's F ratio, Bartlett's V as an approximate chi-square statistic, and Tatsuoka's ω^2) are all used to evaluate various aspects of the differences between groups of hospitals.

The discriminant analyses of each set of classification criteria are summarized in table 4. The results of the tests with Rao's F ratio indicate that only the groups based on AID criteria are different from each other. Use of the program criteria and the service index criteria did not result in significant differences in the groups of test hospitals. When Bartlett's V was used as an approximate chi-square test of the successive (AID) discriminant functions, only the first discriminant function was significant. Thus, the differences between AID groups can be defined in a single dimension.

The ω^2 measure of group variability indicates that about one-half (1970 = 49.9 percent, 1971 = 52.9 percent) of the total variability in the AID criteria discriminant space is due to group differentiation. That is, AID was able to capture about half of the total variability in the eight performance variables (eight-dimensional discriminant space). However, the program criteria and the service index criteria captured none of this variability. Table 5 displays the complete discriminant functions and structure matrices for 1970 and 1971 for the groups of hospitals based on AID criteria.

Limitations of AID Application

The AID model has only been applied to 87 large hospitals, and the relevance of these classification criteria to a larger set of hospitals, or to hospitals of different sizes, remains untested. The generalizability of the classification methods described should be tested in future studies seeking to obtain groups of comparable hospitals. Such groups might permit the establishment of baseline measures of the relative performance or efficiency of health care organizations and be use-

Table 4. Results of application of tests of significance to AID and to other sets of classification criteria

Tests of significance	*AID criteria* [1]		*Program criteria*		*Service index criteria*	
	1970	*1971*	*1970*	*1971*	*1970*	*1971*
Tests of overall significance with Rao's F ratio	**1.947**	**2.124**	1.364	1.785	1.091	1.406
Tests of successive discriminant functions with Bartlett's V:						
1st function	**43.898**	**45.365**	19.037	21.525	21.147	24.577
2nd function	10.691	10.716	9.932	13.261	4.258	6.226
Tests of discriminatory power with Tatsuoka's ω^2 (percent)[2]	49.9	52.9	0	0	0	0

[1] Numbers in boldface are significant at 0.01.
[2] Significant discriminant function only.

Table 5. Discriminant functions and structure matrices for AID criteria

Performance variable	*Significant discriminant function*		*Stducture matrices*	
	1970	*1971*	*1970*	*1971*
Nursing salaries	−.0974	.3585	.70	.79
Medical staff and general service salaries	.1743	.4081	.64	.71
Employee health and welfare	−.2742	−.4348	.38	.42
General services, administration and fiscal, and depreciation	.5852	.2901	.69	.74
Nursing man-hours	.5948	−.5937	.72	.65
Holiday and vacation paid	−.0812	−.0393	.08	.20
Laboratory tests	.4226	.2786	.56	.55
Square footage capacity serviced by plant and housekeeping departments	.0589	−.0036	.70	.50
Eigenvalues	.7370	.7694	...	...
Percentages	78.3	75.5	...	...

ful in evaluating changes in the ways health care is provided.

With an AID model, classification criteria can be selected from a much broader group of potential classification variables than with other such models. Researchers desiring to establish comparative groups of hospitals have previously had only limited evidence regarding the variables that could serve as classification criteria. My research, however, provides a pragmatic and tested statistical tool that will enable other researchers to establish comparative criteria. This tool applies explicitly to situations in which nominal variables are used as classification criteria. Other techniques are better suited to interval scaled data. Additional research will be needed to determine classification criteria for smaller hospitals. My research will also need to be extended to include comparisons of AID with other, more current sets of classification criteria. For example, the hospital classification criteria used by the State of Washington (*6*) could be subjected to such analysis.

Discussion

In any comparative study of hospitals, the proposed classification models should be evaluated on the basis of whether the resultant groups are homogeneous. The classification model based on AID, whose derivation I have described, can be empirically validated. Similar models currently being proposed as the basis for evaluating hospital costs also must be evaluated for consistency and efficacy.

As noted, one of the primary uses of classification criteria is in the establishment of peer groups or control groups of hospitals. Control groups are established so that each hospital's performance may be evaluated relative to the group's performance. Group norms or averages are often used to set the standard or expected level of performance. For example, when evaluating th effects of a hospital merger or consolidation, the only way to anticipate what might have happened in the absence of the consolidation is to evaluate the relative performance of a group of comparable, but unmerged, hospitals. Classification models can be used to establish the criteria that define each of the control groups which would be comparable to the consolidated institutions. Through statistical validation methods, an agency that is mandating such classification criteria can determine whether the proposed classification process results in real differences in the groups by maximizing inter-group variance and intragroup homogeneity. The AID classification criteria, but not the other criteria tested, met this standard. This result does not suggest that AID criteria will necessarily provide the most powerful discrimination among groups in all applications. It does indicate that any classification method should be tested against the standard of maximizing homogeneity within groups and minimizing similarities among groups.

The term "peer groups" often applies to groups of hospitals that operate under similar reimbursement procedures. Peer groups have been used by the Social Security Administration to set standards on maximum limits of hospital reimbursements per day of service (*14*). Many States have implemented, or are contemplating, the use of peer groups as a basis for reimbursement (*6,8*). The statistical methods described in this paper could be applied by a Federal or State agency that is considering the adoption of hospital classification criteria for reimbursement purposes.

One caveat must be noted. Many reimbursement procedures often incorporate perverse effects that are so pervasive as to vitiate the incentive effects of the reimbursement model. For example, implementing the AID classification model described in this paper might motivate hospitals to add organ banks or renal dialysis units. Two controls are necessary in this regard. One is that the classification model must be periodically updated. It is doubtful whether hospital managers could predict whether the updated model would be more or less beneficial as services were added or deleted. In other words, hospital managers might make some short-term adjustments in response to the classification model, but those adjustments can be expected to be minimal in the face of long-run uncertainties regarding revisions in the classification criteria. This policy question is one of the major ones that must be addressed in deciding how frequently the classification criteria should be revised. A major benefit of periodic revision is that it would discourage a hospital manager from trying to finesse peer groupings by adding or subtracting services merely in order to move the hospital into what might be perceived as a more advantageous peer group. As the revision process becomes more frequent, it should discourage these short-run perturbations in services. On the other hand, revision of the classification criteria is expensive, and these costs must be weighed against the benefits of periodic revisions.

Secondly, existing controls on new facilities and services, such as certificate of need and the approval requirements for health service agencies, preclude hospital managers from automatically adding new services to change their peer group status. As these external controls are strengthened, or as controls are implemented that mitigate against changes in classification status, the tendency to add new services will be minimized. In any event, the need for such controls must be recognized as an integral part of any classification model.

References

1. Berry, R.E.: Returns to scale in the production of hospital service. Health Serv Res 2: 123–139, summer 1967.
2. Berry, R.E.: On grouping hospitals for economic analysis. Inquiry 10: 5–12, December 1973.
3. Hess, I., and Srikantan, K.S.: Recommended variables for the multiple stratification of hospitals. Health Serv Res 5: 12–24, spring 1970.
4. Kaluzny, A.D., Veney, J.E., Gentry, J.T., and Sprague, J.B.: Scalability of health services: an empirical test. Health Serv Res 6: 214–223, fall 1971.
5. Veney, J.E., et al.: Implementation of health programs in hospitals. Health Serv Res 6: 350–362, winter 1971.
6. Baker, R.D.: Prospective rate setting in Washington State. Topics in health care financing: prospective rate setting. Aspen Systems Corporation, Germantown, Md.
7. Phillip, P.J., and Iyer, R.N.: Classification of community hospitals. Health Serv Res 10: 349–368, winter 1975.
8. Kosnick, K.R.: Methodology and empirical validation of multi-dimensional hospital classification structure. Doctoral thesis, Northwestern University, Evanston, Ill., 1975.
9. O'Donoghue, P., editor: Hospital costs in Colorado. Spectrum Research, Inc., Denver, Colo., 1977.
10. Agnew, H.: Point system of fixing hospital service rates. Hospitals (Journal of the American Hospital Association) 18: 36–41, 1944.
11. Agnes, H.: A units of credit system for the payment of hospitals. Canad Hosp 12: 43–49, November 1948.
12. Carr, W.J., and Feldstein, A.J.: The relationship of cost to hospital care. Inquiry 4: 45–65, June 1967.
13. Francisco, E.W.: Analysis of cost variations among short-term general hospitals. *In* Empirical studies in health economics—Proceedings of the Second Conference on Economics of Health, edited by H. Klarman. Johns Hopkins Press, Baltimore, Md., 1979.
14. Social Security Administration: Federal health insurance for the aged and disabled. Limitations on coverage of costs under Medicare. 20 CER Part 405, Regulation No. 5, Section 223 of Public Law 92-602. Federal Register, Vol. 39, No. 54, March 19, 1974.
15. Institute for Social Research: OSIRIS III, Release 1, August 1973. Institute for Social Research Monitor System, Ann Arbor, Mich., 1973.
16. Assael, H.: Segmenting markets by group purchasing behavior: an application of the AID technique. J Marketing Res 7: 153–158, May 1970.
17. Sonquist, J.A., Baker, E.A., and Morgan, J.H.: Searching for structure. Institute for Social Research, Ann Arbor, Mich., 1973.
18. Kimbell, L.J., and Lorant, J.H.: Methods for systematic and efficient classification of medical practices. Health Serv Res 8: 46–49, spring 1973.
19. Berry, R.W.: Product heterogeneity and hospital cost analysis. Inquiry 7: 67–75, March 1970.
20. Lave, J.R., and Lave, L.R.: The extent of role differentiation among hospitals. Health Serv Res 6: 15–38, spring 1971.
21. Edwards, M., Miller, J.A., and Schumacher, R.: Classification of community hospitals by scope of service: four indexes. Health Serv Res 5: 301–313, winter 1972.
22. Samaritan Health Service: Demonstration and evaluation of integrated health care facilities, Samaritan Health Service, Phoenix, Arizona, final report, June 28, 1972. Health Services Research Center, Chicago, 1972.
23. Tatsuoka, M. M.: Multivariate analysis: techniques for educational and psychological research. John Wiley & Sons, Inc., New York, 1971, p. 165.
24. Tatsuoka, M. M.: Discriminant analysis. Institute for Personality and Ability Testing, Champaign, Ill., 1970, pp. 37–38.

SYNOPSIS

NEUMANN, BRUCE R.: (State University of New York at Albany): *A statistically based method for identifying hospital classification criteria. Public Health Reports, Vol. 95, May-June 1980, pp. 232–242.*

The development of valid classification criteria for U.S. hospitals. A number of bills recently introduced in the U.S. Congress call for the linkage of classification criteria to cost limits for hospitals. Such proposals have not indicated how the classification criteria should be validated or tested.

A research project was therefore undertaken to determine whether 87 large community hospitals could be classified into interpretable and reproducible homogenous groups. By means of an automatic interaction detector (AID), a set of unique classification criteria were identified. These included residency and internship education programs, medical school affiliation, renal dialysis, and organ bank facilities. Application of the criteria to 1970 and 1971 data for the 87 hospitals resulted in five reproducible and stable groups of hospitals. The criteria were validated by several tests involving different types of cost comparisons and ratios.

The research results indicate that an AID-based classification structure is a feasible model for grouping or classifying large hospitals for comparative purposes. Only a small number of variables are necessary to classify large hospitals, and the criteria do not need to be overly complex. Many of the variables traditionally used may be omitted.