DOCUMENT RESUME

ED 136 738 HE 008 824

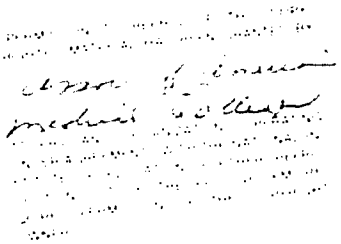
AUTHOR McShane, Michael G.
TITLE An Empirical Classification of U.S. Medical Schools by Institutional Dimensions. Final Report.
INSTITUTION Association of American Medical Colleges, Washington, D. C.
SPONS AGENCY Health Resources Administration (DHEW/PHS), Bethesda, Md. Bureau of Health Manpower.
PUB DATE Mar 77
CONTRACT 231-76-0011
NOTE 51p.; For related documents, see HE 008 822 and HE 008 823 ; Tables and appendices may be marginally legible due to small print of the original

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS *Classification; Cluster Analysis; Higher Education; *Medical Schools; *Private Colleges; Private Schools; Public Schools; *State Universities; *Statistical Analysis; Statistical Studies; Typology

ABSTRACT

In a related study, factor analysis was applied to reduce a selected set of medical school characteristics to their principal dimensions. In this study, the results were then used as input to a series of multivariate cluster analyses that isolated clusters of medical schools that were similar to each other and different from schools in other clusters on the dimensions depicted by the factor analysis. The eight clusters in the final solution each had distinctive profiles on the six factor scores. There were five clusters that consisted completely or predominantly of public schools. Three of these clusters consisted of established schools with varying profiles, while the other two were composed of new and developing schools. Of the remaining three clusters, two were predominantly private schools and one was an equal mix of public and private schools. Each cluster was also described in terms of variables selected from the original data. (Author/MSE)

# AN EMPIRICAL CLASSIFICATION OF U.S. MEDICAL SCHOOLS BY INSTITUTIONAL DIMENSIONS



**FINAL REPORT**



**Association of American Medical Colleges**
**One Dupont Circle, N.W., Washington, DC 20036**



**U.S. Department of Health, Education and Welfare**
**Public Health Service**
**Health Resources Administration**
**Bureau of Health Manpower**
**Contract No. 231-76-0011**

# AN EMPIRICAL CLASSIFICATION OF U.S. MEDICAL SCHOOLS BY INSTITUTIONAL DIMENSIONS

Michael G. McShane Ph.D.

FINAL REPORT

*RELATED STUDIES*

*A Second Exploratory Analysis Of The Relations Among Institutional Variables*

*A Multidimensional Model Of Medical School Similarities*

Division of Operational Studies

ASSOCIATION OF AMERICAN MEDICAL COLLEGES

March 1977



*The work upon which this publication is based was supported in part by the Bureau of Health Manpower, Department of Health, Education, and Welfare pursuant to contract number 231-76-0011. However, any conclusions and/or recommendations expressed herein do not necessarily represent the views of the supporting agency.*

TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# EXECUTIVE SUMMARY

This report, An Empirical Classification of U.S. Medical Schools by Institutional Dimensions, describes one of five studies performed by the Association of American Medical Colleges (AAMC) in 1976 examining the characteristics of U.S. medical schools and the interrelationships among the schools and among variables that describe them. Two of the five studies were replications of earlier work. The other three studies, including this one, used multivariate statistical methods--factor analysis, cluster analysis, and multidimensional scaling--to explore the extensive body of data on the institutions maintained by AAMC in the Institutional Profile System (IPS). In 1976, factor analysis was applied to reduce a selected set of variables to their principal dimensions. The variables used represented the data found most interesting in earlier studies and new data which showed a potential for revealing interesting new areas of institutional variability. The results of the factor analysis were then used as input to a series of multivariate cluster analyses which isolated clusters of medical schools that were similar to each other and different from schools in other clusters on the dimensions depicted by the factor analysis.

The original data on which the study is based were selected from the more than 8,000 data elements in IPS. A total of 140 variables were selected from four categories of measures: (1) institutional, (2) student, (3) faculty, and (4) curriculum. Through a series of correlational studies this variable set was reduced to 33 variables which represented the most complete, representative, and interesting data available. The 33 variables were factor analyzed, eight factors were rotated using a varimax criterion, and factor scores were computed on the eight factors for 110 medical schools.

The cluster analysis described in this report was performed in two stages. Initially, the 110 schools were clustered hierarchically using a technique developed by Ward. The result of this analysis was used as input to a non-hierarchical cluster analysis to refine the final groupings of schools. A number of combinations of factor

scores and numbers of clusters were produced and the resulting clusters compared. A final solution of eight clusters based on six factor scores was selected reflecting the best groups of medical schools on the most meaningful dimensions.

The eight clusters in the final solution each had distinctive profiles on the six factor scores. There were five clusters which consisted completely or predominantly of public schools. Three of these clusters consisted of established schoosl with varying profiles, while the other two were composed of new and developing schools. Of the remaining three clusters, two were predominantly private schools and one was an equal mix of public and private schools. Each cluster of schools was also described in terms of variables selected from the original data which was factor analyzed. This information provided an added dimension of distinctiveness to the clusters described in the study.

# Chapter I

## INTRODUCTION

This report describes the third in a series of studies performed by the Association of American Medical Colleges (AAMC) in which multivariate cluster analysis was used to group medical schools on the basis of quantitative data contained in the Association's Institutional Profile System (IPS). The purpose of the series of studies is to empirically derive groups or clusters of medical schools such that the schools in each cluster are similar to each other and different from schools in other clusters. The basis on which the clusters were formed included several measurable aspects of the institutions such as general institutional, financial, faculty, student, and curricular characteristics. In other words, the goal of the analyses was to isolate groups of medical schools which are similar to one another on a number of dimensions.

### Previous AAMC Cluster Analysis Studies

The first AAMC cluster analysis study, Classification of Medical Education Institutions (Nunn and Lain, 1976), was performed in 1975. That study was patterned after one conducted by the Rand Corporation in 1972 (Keeler, et al, 1972). In the Rand study, 31 institutional variables for 94 U.S. medical schools were factor analyzed, six factors were rotated, and factor scores for each factor were generated for the 94 schools. These schools were then formed into 10 groups using cluster analysis. In the 1975 AAMC study, 23 of the 31 variables used in the Rand study were factor analyzed. Using data primarily from 1973-74, six factors were extracted from the 23 variables, the factors were rotated, and factor scores were calculated for 99 medical schools. These 99 schools were then clustered into 16 groups based on their similarities on the six factor scores. In the 1975 AAMC study the cluster analysis was performed in two stages. First, a hierarchical cluster analysis was used to assess the number of clusters and potential cluster centers. The second step was to use a non-hierarchical cluster analysis to refine the membership of the 16 clusters.

The 1975 AAMC cluster analysis study was replicated one year later (McShane, 1977a) using the same variables and essentially the same methods. There were two principle differences between the 1975 study and the 1976 replication:

(1) the data used in the replication were primarily from 1974-75, and (2) due to the availability of more complete data, 109 schools were included in the cluster analysis. In both the factor analysis and the cluster analysis, a number of similarities in the results from the two years were found. However, the discrepancies in the findings and the apparent increased clarity of the replication results seemed to indicate that further development of the methods and further exploration of the data would increase understanding of similarities and differences among medical schools.

Overview of the Present Study

As a result of the 1976 cluster replication, a number of recommendations were made for further studies in this area. Among the recommendations were the following:

1. The selection of variables should be altered to include new variables which would describe a research emphasis dimension on which medical schools could be compared.

2. Special attention should be paid to the "control" (public vs. private) dimension, and a way should be sought to either eliminate or statistically control the effects of this dimension.

3. The number of clusters should be determined through the analytic process rather than specified a priori.

4. The changes in the membership of the clusters over time should be examined to ascertain whether there are some schools which group together in several studies.

5. The potential for basing the cluster analysis on the original data as opposed to factor scores should be given further consideration, and the effects of missing data and outlying schools on the analysis should be investigated.

All of the recommendations listed above were taken into consideration during the course of the study described in this report. A new set of 33 variables, including 7

variables used in the previous clustering studies and 26 additional variables, were selected through a series of factor analyses (Sherman, 1977b). These 33 variables were factor analyzed, and 8 factors were extracted and rotated. Factor scores were then computed on the eight factors for each of the 110 medical schools that had data for more than 80 percent of the variables.

In the cluster analysis stage of the study, the second and third recommendations listed above were incorporated. The effects of the control dimension on the solution were taken into account through the selection of the factor scores used in the cluster analyses. Since factor scores represent independent composite measures of dimensions on which medical schools vary, they replace the raw data as the basis for the cluster analysis. As such, one or more of the factor scores may be deleted from the analysis and the effect of removing the variables may be assessed.

A number of combinations of factor scores were used as input in a hierarchical cluster analysis, and the effects of the exclusion of variables on the resultant clusters were assessed in these solutions. In addition, there was no preconception of the number of clusters which would emerge from the analysis. The number of clusters was determined by the analysis of the data at hand and by comparing solutions involving varying numbers of clusters. Finally, the memberships of the clusters were refined by using one school from each cluster in the hierarchical solution as a starting point for a non-hierarchical cluster analysis. In the non-hierarchical cluster analysis, a number of solutions involving varying numbers of clusters were derived. The solutions which optimally satisfied the criterion of minimizing differences among the schools in each cluster while maximizing the differences among the clusters are presented in this report.

## Chapter II

## M E T H O D

The study described in this report was conducted in five stages; (1) selection of variables, (2) factor analysis, (3) computation of similarities, (4) hierarchical cluster analysis, and (5) non-hierarchical cluster analysis. In this chapter each step in the analysis will be described. Further explication of the first two steps in the analytic process can be found in a companion report by Sherman (1977b).

### Selection of Variables

AAMC's Institutional Profile System (IPS) is the repository for most of the institutional data collected by the Association. In August, 1976, there were over 8,000 data elements from over 60 different sources in IPS. Many of the data were longitudinal repetitions of the same variable for as many as 15 years (1959-60 through 1974-75). The data in IPS come from a number of different kinds of sources. The major sources are annual surveys such as Parts I and II of the Liaison Committee on Medical Educations (LCME) Annual questionnaire which provide a wealth of information on medical school finances and detailed counts of students, faculty, and facilities; the Fall Enrollment Questionnaire which provides additional student counts; and information on types of programs and electives gathered to be published in the AAMC Curriculum Directory. Additional data are taken from special-purpose surveys and questionnaires, such as the 1973 Health Services Delivery and Primary Care Education questionnaire, the 1975 AAMC questionnaire on student affairs resources, and the 1973 questionnaire on medical school facilities; other AAMC information systems such as the Faculty Roster System (FRS), the Medical School Applicant file, and the Medical Student Information System (MSIS); and other organizations' information systems such as AMA's Medical School Alumni file and the IMPAC file maintained by the Division of Research Grants (DRG) of the Department of Health, Education, and Welfare (DHEW) which contains information on grant applications to NIH and selected other agencies within DHEW. All of the data transmitted from other AAMC information systems and other agencies are aggregated by institution prior to being stored in IPS.

To facilitate the use of data from IPS in the studies using institutional data, a Researchable Data Base was constructed. Data elements were selected for inclusion in the Researchable Data Base if they were the most recent repetition of a particular variable and were potentially useful in one or more of the studies specified in the contract. A total of 399 variables, including institutional, faculty, student, and curriculum measures, were transferred from IPS to the Researchable Data Base. In addition, 201 variables were computed from the original data and stored in the Researchable Data Base. The computed variables described attributes of the medical schools within (e.g. the percentage of females among undergraduate medical students) and across (e.g. the ratio of undergraduate medical students to full time medical school faculty members) the four categories noted above. A complete discussion of the 1976 IPS Researchable Data Base and a list of the variables included may be found elsewhere (McShane, 1977b).

From the total of 600 variables in the Researchable Data Base, 139 were selected for consideration in this study. A series of correlational studies was conducted within each of the ad hoc categories described above to select a final set of variables which had recorded values for nearly all schools and were representative of the principal dimensions within each of the categories. The final set of variables factor analyzed and used to produce the factor scores which were the basis of this study are presented in Table 1. (A Glossary of abbreviations is presented in Appendix A). From the information presented in Table 1 it is evident that 17 of the 33 variables used in the final factor analysis on which this study was based were new variables. These new variables were either not available for earlier studies in the series, or replaced similar variables for reasons of completeness or representativeness discussed above. In addition, since part of the intent of Sherman's (1977b) study was to expose previously undisclosed relationships among variables, when two variables were approximately equivalent in completeness and representation, a previously unused variable was selected over one used in earlier studies.

The final set of 33 variables contained 14 student variables, 13 institutional variables, 4 faculty variables, and 2 curriculum variables. There are a number of reasons for the disproportionate selection of variables from the four categories. First, most of the data in IPS are either institutional or student descriptors. Secondly, the curriculum data in IPS are predominantly qualitative and as such are

TABLE 1

VARIABLES USED IN FACTOR ANALYSIS OF INSTITUTIONAL DATA, 1976

| VARIABLE |
|---|
| VAR388 AV SALARY - SFT ASSOC PROF BASIC SCIENCE[2] |
| STC043 RAT: HOUSESTAFF TO UNDERGRAD MD-STUD[1,2] |
| INC058 RAT: MD STUDENTS TO FT FAC[3] |
| STC105 % LIVING MD-ALUMNI IN GENERAL PRACTICE |
| FAC001 & PT & FT SAL FAC WITH MD |
| VAR016 # MD-STUDENTS[1,2] |
| INC048 LOG AGE OF MEDICAL SCHOOL[1] |
| STC112 % LIVING MD ALUM BOARD CERTIFIED |
| VAR002 CONTROL: 0 = PUBLIC, 1 = PRIVATE[1,2] |
| VAR394 1975-76 RESIDENT MD-STUDENT TUITION |
| STC029 % IN-STATE 1ST-YR MD-STUD[1] |
| STC084 RAT: APPLICANTS PER 1ST-YR MD-STUD[2] |
| INC007 % REV FROM FED SOURCES & RCOV IND COSTS |
| INC012 % REV FROM ALL GIFTS |
| STC082 % UNDERREP MINORITY 1ST-YR MD-STUD |
| FAC004 % PT & FT SAL FAC FROM ETHNIC MINORITIES |
| STC008 % NON US-CANADIAN 1ST-YR MD-STUD[4] |
| VAR093 1ST-YR MD-STUD: MEAN MCAT SCIENCE SCORE[2] |
| INC040 NIH-NIMH R01 $ AWARD AS % OF $ APP SBMITTED |
| VAR352 IMPAC: MEAN STD P-SCR - R01 APP |
| INC045 IRG APPROVAL RATE OF NIH R01 COMP APPS |
| STC003 % FEMALE MD STUDENTS |
| VAR273 REL ELECTIVES: ALCOHOLISM |
| CRC002 % OF RELATED ELECTIVES OFFERED |
| FAC019 RAT: VOL FAC TO FT FAC[1,2] |
| INC003 DRG FED SPON RES CON$ %CHG 67-9 to 72-4[2] |
| STC114 PROJTD ANNL % 1ST-YR ENROLL CHG: 1974-79 |
| VAR384 DRG GRANTS - # R01 APPS APPROVED |
| INC026 % EXPD FOR ADMIN & GENL EXPENSE |
| INC017 % TOTAL EXPD FOR SPON RESEARCH[2] |
| STC045 RAT: BMS GRAD-STUD TO UNDERGRAD MD-STUD[1] |
| INC004 ADJUSTED TOTAL REVENUE[2] |
| STC013 % 1ST-YR MD-STUD: PRE-MED GRA 3.6-4.0[2] |

1. Variable used in 1975 and 1976 cluster analysis studies (Nunn and Lain, 1975; McShane, 1977a)

2. Variable used in exploratory analyses of the relations of institutional variables (Sherman, 1976 and 1977 a).

3. The inverse of this variable, the ratio of full time faculty to the number of medical students was used in both 1 & 2.

4. A similar variable, the percentage of non-U.S.-Canadian medical students was used in 2 above.

of limited utility in studies of this type. Finally, computed variables which crossed categories (e.g. the ratio of medical students to full time faculty - INC058) were classified as institutional measures for the purpose of the development of the IPS Researchable Data Base.

In addition, the final data set contained predominantly computed variables (ratios and percentages) rather than the original variables taken from IPS. Only 8 of the 33 variables were IPS data elements; the other 25 measures were computed from IPS data. The reason for selecting predominantly computed variables was that computed variables allow for comparisons of emphasis rather than extensiveness and for illumination of institutional characteristics other than overall "size".

## Factor Analysis

The second step of the analysis performed in the course of this study was the factor analysis of the 33 selected variables described above. The data reduction technique actually employed was principal components analysis. One of the assumptions underlying most factor analytic techniques is that the variance in each variable in a set can be broken down into common variance (the variance shared by the other variables in the set) and the variance that is unique to the particular variable. In principal components analysis, however, no assumptions are made about the structure underlying the variables in the analysis. Instead, the variables are mathematically transformed so that the first component extracted accounts for as much of the variance in the data as possible and each subsequent component extracted accounts for as much of the remaining variance in the data as possible (Mulaik, 1972). In this manner it is possible to determine whether a large proportion of the variance in a set of variables can be explained by a relatively small number of dimensions (components).

In the current study, the first 9 components extracted accounted for 74.4 percent of the variance in the data. A number of varimax rotations were performed in which different numbers of the components, ranging from 9 down to 4, were rotated. These six solutions were then compared, and the 8 component solution was selected as the most interpretable and intuitively appealing. The eight components were explained in some detail by Sherman (1977b) and served as the basis for the cluster analyses described in this report.

## Computation of Similarities

An important conceptual step in conducting a cluster analysis, and one which is often transparent to both user and consumer, is the computation of indices of similarity. Since the goal of cluster analysis is to construct clusters containing objects that are as similar as possible, some measure of similarity (or its converse, dissimilarity or distance) is necessary. Measures of similarity include coefficients of association and correlation; measures of dissimilarity or distance include weighted and unweighted Euclidean distance coefficients, the "city-block" metric, and the Mahalanobis generalized distance coefficient. The various methods of computing similarity indices are discussed in many of the texts on cluster analysis including those by Anderberg (1973), Everitt (1974), and Bailey (1975).

In this study, distances were computed between each of the pairs of schools using the Euclidean distance coefficient. For a given pair of schools, the Euclidean distance is equal to the square root of the sum of the squared differences between the two schools on each variable. One advantage of this type of distance coefficient is that it has an easily interpretible and unique zero point. The distance between two schools can be zero if and only if they have identical values on all variables. Negative distance is undefined and larger coefficients imply that schools are farther apart on one or more variables.

It is important to note that in the computation of the Euclidean distance described above, each variable is equally important in determining the distance coefficient between pairs of schools. Important variables may be given added impact in an analysis by weighting those variables. Alternatively, variables which have little heuristic importance may be dropped from the analysis.

## Hierarchical Cluster Analysis

The cluster analysis performed in this study was actually a two-step process. Initially, hierarchical cluster analysis was performed using a technique developed by Ward (1963). The results of the hierarchical cluster analysis were then used to give indications of the number of clusters of schools present, based on the factor scores used as input, and the schools which could be used as starting points for the non-hierarchical cluster analysis.

Generally speaking, hierarchical cluster analysis is a class of empirical methods of forming objects into groups, through a series of stepwise merges. At first, each object is in a group of its own. Two groups are joined to form a larger group. Then, again, two of the remaining groups are merged. This continues until all objects are combined into a single group. At each step of the merging process, the two most similar of the groups are combined, and once combination has taken place the groups remain intact for the duration of the analysis. By forcing all objects to be combined, hierarchical cluster analysis allows for distortion of natural clusters by the inclusion of outlying objects.

Ward's hierarchical cluster analysis method defines the distance between clusters as the distance between the centers of the clusters (the cluster centroids) and uses as its criterion the increase in the sum of the squared distances from the objects in the cluster to the cluster centroid. At each step of the analysis, the two clusters that cause the least increase in the sum of squared distances within clusters are combined. Stated another way, the Ward method attempts to minimize differences within clusters and maximize differences among clusters.

In the study described in this report, 110 U.S. medical schools were hierarchically clustered on the basis of their values on 8, 6, and 5 factor scores. These three analyses were conducted to assess the impact of selected factor scores on the hierarchical solution, and specifically to determine whether the omission of the control factor would have beneficial results in the interpretation of the clusters. It should be noted that, unlike previous AAMC clustering studies, no variable was given disproportionate weight in determining the distance index between pairs of groups.

Non-Hierarchical Cluster Analysis

The information provided by the hierarchical cluster analysis was used to initiate a refinement of the derived clusters through non-hierarchical cluster analysis. Non-hierarchical cluster analysis places all objects into a predetermined number of clusters in such a way that a specified criterion is optimized. This kind of procedure avoids the problem of objects necessarily remaining together once they have been combined and reduces the effects of outlying objects on cluster memberships. However, in order to use a non-hierarchical cluster analysis it is preferable to have some idea of the number of clusters of groups of objects that exist based on the data at hand, and to be able to provide some indication of the

approximate location of the "centers" of the clusters. In this study the result of the hierarchical cluster analyses was used to provide a range of the number of clusters present and initial cluster "centers", one school from each cluster in the hierarchical solution, for the non-hierarchical cluster analysis.

The non-hierarchical cluster analysis method used in this study was developed by Forgy (1965) and is known as the K-means technique. Using the number of clusters and cluster centroids specified by the user, each object is assigned to the cluster with the closest centroid. After all objects have been initially assigned to a cluster, new cluster centroids are computed for each cluster based on the objects assigned to the cluster. A cluster centroid is a point in p dimensional space (where p is the number of variables) defined by the mean of the objects in the cluster on each variable. The distance of each object from each of the cluster centroids is then computed and objects are reassigned, if necessary, to the cluster which now has the closest centroid. After the reassignment of objects, the cluster centroids are recomputed, and a new cycle of computing distances, reassigning schools and recomputing cluster centroids is begun. This cycle is repeated until no objects are reassigned after cluster centroids have been calculated. This procedure, like the Ward technique, minimizes the differences of objects within the clusters but without the artificial permanence of cluster membership inherent in the hierarchical approach.

In this study several non-hierarchical cluster analyses were performed using the Forgy method. Numbers of clusters ranging from 12 down to 6 were derived using both 5 and 6 factor scores as input. From the variety of possible clusterings, an 8 cluster solution based on 6 factor scores was selected for presentation in this report based on its representation of the schools and their similarities. The rationale through which this solution was selected and a description of the clusters in terms of both the factor scores and the original variables are presented in the following chapter.

## Chapter III

## RESULTS AND DISCUSSION

The results of this study were derived at three different stages of the analytic sequence. The factor analysis and hierarchical cluster analysis each produced results which were utilized at later stages; and the non-hierarchical cluster analysis produced the final clusters of medical schools. The results of each step of the analysis will be presented in this chapter.

### Factor Analysis of 33 Measures of Institutional Characteristics

As described in Chapter II, the first step in the analysis for this study was the factor analysis of 33 variables selected from IPS. The 33 variables were selected to represent several measurable aspects of medical schools in the U.S. including institutional, financial, faculty, student and curricular characteristics.

The rotated factor pattern matrix which resulted from the factor analysis is presented in Table 2. The matrix was the result of a study by Sherman (1977b) and is discussed in detail in the report of that study. For the purposes of this report, the factor pattern matrix will be interpreted only briefly.

Factor 1 provides a means for assessing the graduate medical education program emphasis among medical schools. Schools which are strong in this area would typically have a high ratio of interns and residents to undergraduate medical students, proportionally more faculty who hold MD degrees, higher faculty salaries, and fewer undergraduate medical students per full time faculty member. Interestingly, schools with these qualities have in the past produced a relatively small proportion of graduates who went into general practice.

Factor 2 measures the size and age of the medical schools. This factor bears out the common assertion that older schools tend to have greater numbers of undergraduate medical students and larger proportions of alumni who have achieved board certification. Secondary loadings on this factor indicate that older medical schools are experiencing less growth in enrollment and federally sponsored research funding than newer schools. While these findings are not particularly startling, it is interesting to note that these measures do form an

TABLE 2

EIGHT COMPONENT VARIMAX FACTOR PATTERN RESULTING FROM
PRINCIPAL COMPONENTS ANALYSIS OF 33 VARIABLES
DESCRIBING U.S. MEDICAL SCHOOLS, 1976

| | VARIABLE | 1 Graduate Medical Program | 2 Size, Age | 3 Control | 4 Minority | 5 Research Funding Success | 6 Curriculum Electives | 7 Development Stage | 8 Research Emphasis | $h^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | VAR388 AV SALARY - SFT ASSOC PROF BASIC SCIENCE | .84 | -.02 | .03 | -.00 | -.01 | .05 | .16 | -.03 | .73 |
| 2 | STC043 RAT: HOUSESTAFF TO UNDERGRAD MD-STUD | .79 | -.03 | .19 | -.03 | -.00 | .04 | .05 | .07 | .68 |
| 3 | INC058 RAT: MD STUDENTS TO FT FAC | -.67 | .22 | -.02 | -.05 | -.14 | .05 | .23 | -.36* | .71 |
| 4 | STC105 % LIVING MD-ALUMNI IN GENERAL PRACTICE | -.54 | .33 | -.13 | .04 | -.42* | -.24 | .14 | -.14 | .70 |
| 5 | FAC001 % PT & FT SAL FAC WITH MD | .40 | .27 | .33* | .06 | .14 | -.02 | -.01 | -.03 | .37 |
| 6 | VAR016 # MD-STUDENTS | -.09 | .88 | -.06 | .03 | -.04 | .08 | -.04 | .16 | .83 |
| 7 | INC048 LOG AGE OF MEDICAL SCHOOL | -.28 | .75 | .21 | .01 | .15 | .03 | -.33* | .15 | .83 |
| 8 | STC112 % LIVING MD ALUM BOARD CERTIFIED | .14 | .71 | .32 | -.17 | .04 | .07 | -.28 | .09 | .75 |
| 9 | VAR002 CONTROL: 0 = PUBLIC, 1 = PRIVATE | .15 | .14 | .87 | .09 | -.00 | .03 | -.13 | -.01 | .83 |
| 10 | VAR394 1975-76 RESIDENT MD-STUDENT TUITION | .05 | .13 | .86 | -.07 | .13 | .10 | -.14 | -.02 | .82 |
| 11 | STC029 % IN-STATE 1ST-YR MD-STUD | .01 | -.06 | -.81 | -.23 | -.14 | .00 | .18 | -.16 | .79 |
| 12 | STC084 RAT: APPLICANTS PER 1ST-YR MD-STUD | .10 | -.09 | .79 | -.03 | .01 | -.07 | .27 | -.03 | .72 |
| 13 | INC007 % REV FROM FED SOURCES & RCOV INC COSTS | .05 | -.01 | .48 | .05 | .22 | .28 | -.27 | .48* | .66 |
| 14 | INC012 % REV FROM ALL GIFTS | .20 | .08 | .38 | .11 | -.32 | .10 | -.06 | .13 | .33 |
| 15 | STC082 % UNDERREP MINORITY 1ST-YR MD-STUD | -.04 | -.09 | .06 | .94 | -.10 | .02 | .06 | -.03 | .91 |
| 16 | FAC004 % PT & FT SAL FAC FROM ETHNIC MINORITIES | -.11 | -.06 | -.04 | .87 | .03 | .03 | -.14 | -.14 | .82 |
| 17 | STC008 % NON US-CANADIAN 1ST-YR MD-STUD | .19 | .17 | .25 | .67 | -.08 | -.03 | .06 | .10 | .60 |
| 18 | VAR093 1ST-YR MD-STUD: MEAN MCAT SCIENCE SCORE | .43* | .23 | .35* | -.44 | .26 | .04 | .08 | .36* | .75 |
| 19 | INC046 NIH-NIMH R01 $ AWARD AS % OF $ APP SBMT | -.01 | .11 | .14 | -.10 | .84 | .04 | .07 | .13 | .77 |
| 20 | VAR352 IMPAC: MEAN STD P-SCR - R01 APP | -.35 | .04 | -.09 | .14 | -.73 | -.05 | .21 | -.05 | .74 |
| 21 | INC045 IRG APPROVAL RATE OF NIH R01 COMP APPS | -.04 | .29 | -.05 | .01 | .70 | -.03 | .22 | .38* | .78 |
| 22 | STC003 % FEMALE MD STUDENTS | .20 | -.13 | .18 | .31 | .48 | .24 | .02 | -.28 | .56 |
| 23 | VAR273 REL ELECTIVES: ALCOHOLISM | .07 | .03 | -.01 | .02 | -.03 | .88 | .02 | .03 | .79 |
| 24 | CRC002 # OF RELATED ELECTIVES OFFERED | .03 | .14 | .12 | .01 | .14 | .82 | .01 | .24 | .78 |
| 25 | FAC019 RAT: VOL FAC TO FT FAC | -.12 | -.02 | -.02 | -.11 | .08 | .10 | .74 | -.30 | .68 |
| 26 | INC003 DRG FED SPON RES CONS %CHG 67-9 TO 72-4 | .14 | -.44* | -.12 | .15 | -.01 | .00 | .73 | .17 | .82 |
| 27 | STC114 PROJTD ANNL % 1ST-YR ENROLL CHG: 1974-79 | .09 | -.43* | -.17 | -.04 | -.01 | -.09 | .58 | -.17 | .60 |
| 28 | VAR384 DRG GRANTS - # R01 APPS APPROVED | .41* | .41* | .05 | -.01 | .27 | .05 | -.03 | .67 | .87 |
| 29 | INC026 % EXPD FOR ADMIN & GENL EXPENSE | .19 | -.13 | .02 | -.13 | .15 | -.02 | .25 | -.64 | .57 |
| 30 | INC017 % TOTAL EXPD FOR SPON RESEARCH | .24 | .13 | .45* | -.02 | .20 | .26 | -.04 | .63 | .78 |
| 31 | STC045 RAT: BMS GRAD-STUD TO UNDERGRAD MD-STUD | -.05 | .03 | .09 | -.09 | .19 | .23 | .01 | .61 | .48 |
| 32 | INC004 ADJUSTED TOTAL REVENUE | .43* | .52 | -.01 | .04 | .16 | .05 | -.08 | .57 | .82 |
| 33 | STC013 % 1ST-YR MD-STUD: PRE-MED GPA 3.6-4.0 | .23 | .02 | -.06 | -.19 | -.04 | -.05 | -.04 | .55 | .40 |
| | COLUMN SUM OF SQUARES | 3.36 | 3.21 | 4.01 | 2.63 | 2.71 | 1.85 | 2.13 | 3.36 | |
| | PERCENT OF VARIANCE | 14.44 | 13.81 | 17.23 | 11.30 | 11.65 | 7.95 | 9.15 | 14.46 | |

independent dimension empirically unrelated to the other seven factors derived in this analysis.

Factor 3 measures the control dimension among medical schools. The variables which have their highest loadings on this factor are control (in which public schools were represented by a '0', private by a '1'), and other variables which are related to the degree to which a school resembles public or private medical schools: resident medical student tuition, the percent of in-state medical students, the number of applicants per first year medical student, the percent of the school's revenue which comes from federal sources, and the percent of revenue from gifts. Schools which have high values on this factor tend to resemble most private schools in that they have relatively high resident tuition, few resident students, and high numbers of applicants per first-year medical student. These schools also tend to receive a greater proportion of their revenues from the federal government and from gifts than do schools which are more similar to public medical schools.

Factor 4 assesses the medical schools' involvement with ethnic minority faculty and students. It is evident from the variables loading on this factor that schools with high proportions of ethnic minorities among their faculty and students and proportionally high enrollments of foreign medical students would have high values in the fourth factor. Closer inspection of the data revealed that the inclusion of data from two historically Black medical schools, Howard and Meharry, and the University of Puerto Rico probably had a great deal of influence on the emergence of this factor.

Factor 5 measures the research funding success of the medical schools on applications for new single-investigator research (R01) grants from NIH. Schools with high approval rates also have the "best" priority scores (where a lower score reflects a higher priority) and are awarded a higher percentage of the sum of dollars requested on all reviewed R01 proposals. Interestingly, schools which possess these qualities also tend to have a relatively high proportion of female medical students. It is also interesting that this dimension of institutional differences is apparently independent of other measures of research emphasis which combined to form a separate factor.

Factor 6, which was formed by the only two curriculum variables in the variable set, measured the degree to which curriculum electives were used by the medical schools.

The isolation of these two variables indicates that the curriculum information available, in addition to being scarce and not readily amenable to studies of this type, is independent of other dimensions on which medical schools were observed to vary.

Factor 7, which measures the developmental stage of the medical schools, illustrates the tendency of schools to grow simultaneously in all areas. The three variables which have their highest loadings on the seventh factor, one each from the student, faculty, and institutional domains, are all potential indicators of institutional growth. Thus, this factor may distinguish developing from established schools.

The final factor, Factor 8, measures the research emphasis of medical schools. The variables which have high loadings on these factors are primarily related to the extent and emphasis of sponsored research activity. Schools with a strong research emphasis have relatively high percentages of their budgets expended for sponsored research, large numbers of research grants approved, high ratios of basic medical science graduate students to undergraduate medical students, high percentages of students with superior undergraduate grade point averages, and low percentages of expenditures for administration and general expenses.

To summarize, the factor analysis of 33 variables selected from IPS to represent the complete range of medical school activities resulted in the following eight factors: (1) graduate medical education emphasis, (2) size and age, (3) control, (4) minority participation, (5) research funding success, (6) curriculum electives, (7) developmental stage, and (8) research emphasis. Only three of the factors, numbers 1, 3, and 8, were similar to factors derived in earlier AAMC studies (Sherman 1976 and 1977a; McShane, 1977a). Factor 2 which was labelled "Size and Age" here, is similar in content to factors labelled "Undergraduate Medical Education" elsewhere (Keeler, 1972; McShane, 1977a). Factors 4 through 7 represent new dimensions of medical schools which have previously been unexplored.

## Hierarchical Cluster Analysis

Based on the factor analysis described in the preceding section, eight factor scores were computed for 110 medical schools. Factor scores were computed for those schools which were missing values for less than 20 percent of the 33 variables (fewer than six variables). The amount of missing data allowed in this study was based on the proportion of missing data

allowed in other studies of this type (Nunn and Lain, 1976; McShane, 1977a).* The seven schools dropped from the analysis due to insufficient data at this point were Baylor University, University of North Dakota, University of Hawaii, Eastern Virginia Medical School, Wright State University, University of South Carolina, and the Uniformed Services University of the Health Professions. Only the first three of these schools, however, were in complete operation at the time the data on which this study is based were collected.

The eight factor scores were used as input to Ward's hierarchical cluster analysis. Three separate hierarchical clusterings were performed based on different sets of factor scores. The effects of using different combinations of factor scores in analyses is similar to using various combinations of variables of any type in an analysis. The results of cluster analysis are inherently sensitive to the data on which the distances between pairs of schools are computed, and the resultant clusters may be very different when different variables are used. Since one of the goals of this study was to delineate clusters of medical schools which vary on meaningful dimensions, a limited number of combinations of factor scores were used and the results compared for interpretability.

The first hierarchical cluster analysis performed was based on all eight factor scores. The results of this analysis (presented in Appendix B-1) seemed to indicate that the major element on which the clustering was based was the minority factor, and did not appear readily interpretable in terms of the clusters which were derived. At this point, therefore, two factor scores, Minority and Curriculum Electives, were dropped from the variable set. These two dimensions were considered less important than the remaining six in determining clusters of medical schools.

The second hierarchical cluster analysis, based on six factor scores, resulted in potentially interesting groupings of schools, and will be discussed in more detail below. However, to assess the impact of the control dimension on the clustering, the control factor score was dropped and the schools were clustered a third time on five factor scores

* An investigation of the effects of missing and distorted data on solution involving similarities among schools and alternative methods for compensating for such effects is anticipated during the next phase of this series of studies.

(factors 1, 2, 5, 7, and 8). The results of the hierarchical clustering based on five factor scores (presented in Appendix B-2) did not appear to have any compelling qualities which made it inherently more meaningful than the six factor clustering. A non-hierarchical cluster analysis based on the five factor scores was also performed. The results of that analysis are presented in Appendices C-1 and C-2 for comparison with the results of the analysis based on six factor scores, the principal analysis described in this study.

The results of the Ward hierarchical cluster analysis on six factor scores are presented in Figure 1. This tree-diagram, or dendrogram, depicts the merge sequence that developed in the analysis in 25 equal intervals. Each interval represents four percent of the total within cluster sum of squared distances at the final merge (when all schools were merged into a single cluster). From the diagram it is apparent that the majority of the combinations produce relatively little within cluster deviations from cluster centroids. The first 92 merges, principally combinations of small groups of schools, accounted for only 25 percent of the total sum of within cluster deviations. By contrast, the final five merges accounted for over 40 percent of the increase in the criterion. On the basis of this information it was determined that the medical schools could probably be best represented by somewhere between 5 and 17 groups. In other words, based on the information contained in Figure 1, representing the schools by as many as 17 clusters would leave schools which are relatively very similar in different clusters, but representing the schools by as few as 5 clusters would force some schools into clusters in which they do not belong.

Non-hierarchical Cluster Analysis

Based on the results of the six factor hierarchical cluster analysis, an optimal solution was sought using Forgy's non-hierarchical cluster analysis method. The results of the hierarchical clustering were used to give an indication of the number of clusters which would represent the schools, and schools were selected as seedpoints for the non-hierarchical cluster analysis based on the hierarchical clusters. In the hierarchical cluster analysis and on further inspection of the data, one school, the Mayo Medical School, appeared so dissimilar from the other 109 schools that it was not included in further comparisons.

The Forgy non-hierarchical cluster analysis technique complements the Ward hierarchical method by optimizing the

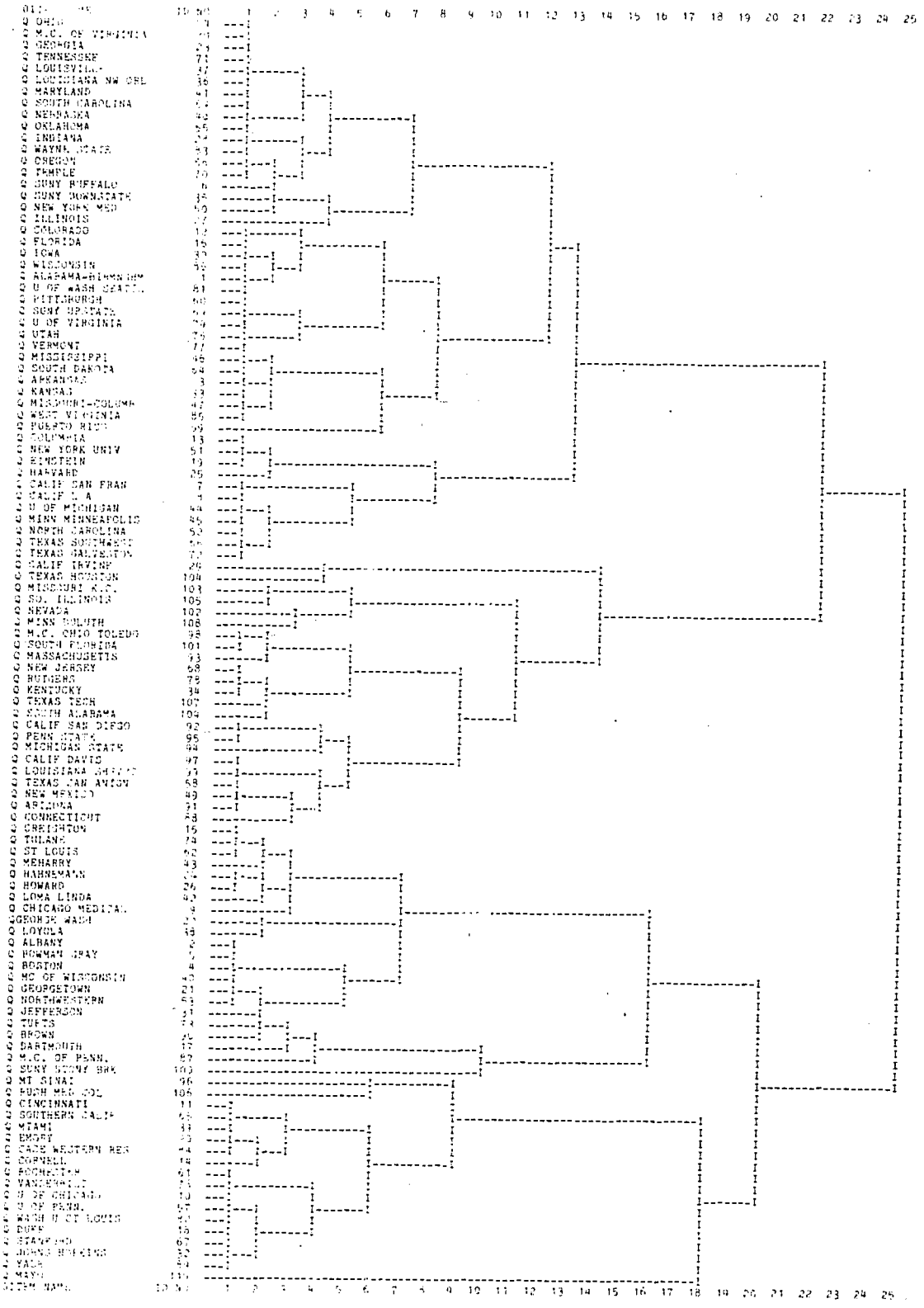


FIGURE 1

WARD HIERARCHICAL CLUSTER ANALYSIS OF 110 U.S. MEDICAL SCHOOLS

BASED ON 6 FACTOR SCORES, 1976

same criterion, the sum of the squared distances of the schools from the cluster centroids, but does not maintain the permanence of cluster membership inherent in the hierarchical methods. Several non-hierarchical solutions were obtained using varying numbers of clusters (12, 10, 8 and 6) and different sets of seedpoints (initial cluster centroids). The eight cluster solution was selected for presentation and description in this report.

Figure 2 presents the composition of the eight clusters derived in the Forgy analysis and the profile of each cluster centroid on the six factor scores used in the analysis. The schools in each cluster are listed in the left hand column of the table, and the mean scores for the schools in the cluster on the six factors are graphed as cluster profiles.

To aid in the interpretation and understanding of the clusters, the means of the schools in each of the clusters on selected variables from the factor analysis are presented in Table 3. In consideration of Table 3 it must be remembered that the factor scores were computed for some schools which were missing data on some variables. As a result of that process, the means are computed based on the number of schools in a given cluster that had data for that particular variable.

Cluster 1, the first cluster depicted in Figure 2, is made up of 17 public medical schools, which are all established schools, but which, as a group, have no other distinguishing characteristics that can be seen in their cluster profile. The schools in Cluster 1 are below the average for all medical schools in emphasis on graduate medical education, development, research funding success, and research emphasis. The schools which form the cluster have an average enrollment of slightly over 500 undergraduate medical students, 95 percent of whom are from the state in which the school is located. These schools tend to be among the least expensive to attend (average tuition $1,166), and they have the smallest ratio of applicants per first year medical student of any of the eight clusters.

The schools which combined to form Cluster 2 are, as a group, the oldest and largest of the 109 medical schools. Six of the 8 schools in the cluster are public schools with an average enrollment of 883 undergraduate medical students. These schools resemble the schools in Cluster 1 in that they do not place much emphasis on either graduate medical education or research, and their research funding success is slightly below average. The schools which make up Cluster 2

FIGURE 2

CLUSTER MEMBERSHIP AND PROFILES OF CLUSTER CENTROIDS ON SIX FACTOR SCORES

| | MEMBERSHIP | | GRADUATE MEDICAL EDUCATION PROGRAMS | SIZE AND AGE | CONTROL | RESEARCH FUNDING SUCCESS | DEVELOPMENT STAGE | RESEARCH EMPHASIS |
|---|---|---|---|---|---|---|---|---|
| CLUSTER 1 | MARYLAND, OREGON, LOUISIANA NEW ORL, M.C. OF VIRGINIA, TENNESSEE, OHIO, GEORGIA, ARKANSAS, MISSISSIPPI, SOUTH CAROLINA, KENTUCKY, NEBRASKA, RUTGERS, SOUTH ALABAMA, SOUTH DAKOTA, OKLAHOMA, LOUISVILLE | HI<br>MD<br>LO | | | | | | |
| CLUSTER 2 | TEMPLE, WAYNE STATE, SUNY-BUFFALO, SUNY-DOWNSTATE, TEXAS-GALVESTON, INDIANA, ILLINOIS, JEFFERSON | HI<br>MD<br>LO | | | | | | |
| CLUSTER 3 | UTAH, FLORIDA, KANSAS, COLORADO, NORTH CAROLINA, WISCONSIN, IOWA, NEW MEXICO, PUERTO RICO, CONNECTICUT, U OF VIRGINIA, MISSOURI-COLUMB, ALABAMA-BIRMINGHM, ARIZONA, WEST VIRGINIA, U OF WASH SEATTL | HI<br>MD<br>LO | | | | | | |
| CLUSTER 4 | PITTSBURGH, NEW YORK UNIV, MIAMI, EINSTEIN, COLUMBIA, SUNY-UPSTATE, TEXAS-SOUTHWEST, CALIF-LA, U OF MICHIGAN, NEW YORK MED, NEW JERSEY, HARVARD, CALIF-SAN FRAN, MINN-MINNEAPOLIS | HI<br>MD<br>LO | | | | | | |

FIGURE 2

(Continued)

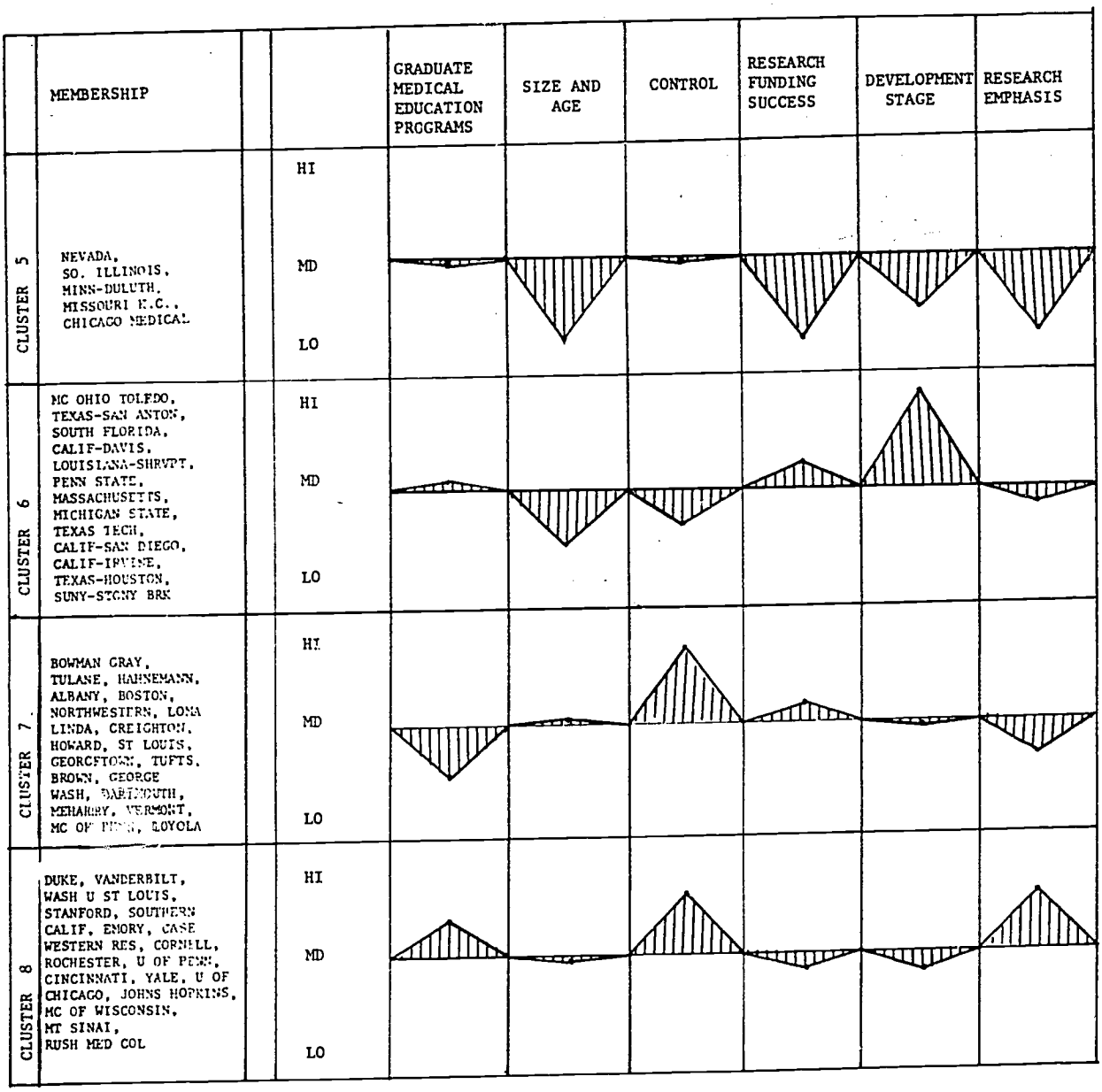


TABLE 3

MEAN VALUES FOR EIGHT CLUSTERS OF U. S. MEDICAL SCHOOLS ON SELECTED VARIABLES
FROM THE 33-VARIABLE FACTOR ANALYSIS OF INSTITUTIONAL DESCRIPTIVES, 1976

| VARIABLE DESCRIPTIONS | CLUSTER 1 (N=17) | CLUSTER 2 (N=8) | CLUSTER 3 (N=16) | CLUSTER 4 (N=14) | CLUSTER 5 (N=5) | CLUSTER 6 (N=13) | CLUSTER 7 (N=19) | CLUSTER 8 (N=17) | TOTAL |
|---|---|---|---|---|---|---|---|---|---|
| AV SALARY - SFT ASSOC PROF BASIC SCIENCE | 23.53 | 24.63 | 23.74 | 27.08 | 24.06 | 26.42 | 23.11 | 26.35 | 24.76 |
| RAT: HOUSESTAFF TO UNDERGRAD MD-STUD | .48 | .48 | .70 | 1.24 | .08 | .94 | .56 | 1.35 | .80 |
| RAT: MD STUDENTS TO FT FAC | 2.16 | 2.86 | 1.40 | 1.24 | 2.02 | 1.83 | 2.34 | .99 | 1.78 |
| % LIVING MD-ALUMNI IN GENERAL PRACTICE | 21.00 | 17.11 | 13.14 | 10.63 | 9.94 | 12.47 | 16.68 | 10.05 | 14.23 |
| % PT & FT SAL FAC WITH MD | 59.43 | 64.86 | 59.71 | 69.39 | 51.82 | 56.35 | 67.26 | 68.00 | 63.13 |
| # MD-STUDENTS | 504.24 | 882.63 | 443.25 | 649.36 | 165.80 | 247.67 | 493.11 | 449.47 | 485.08 |
| LOG AGE OF MEDICAL SCHOOL | 1.82 | 1.99 | 1.66 | 1.85 | .96 | .95 | 1.97 | 1.82 | 1.70 |
| % LIVING MD ALUM BOARD CERTIFIED | 44.25 | 54.18 | 40.26 | 58.58 | 22.78 | 1.83 | 47.62 | 60.11 | 45.94 |
| CONTROL: 0=PUBLIC, 1=PRIVATE | 0 | .25 | 0 | .50 | .20 | 0 | .95 | .94 | .40 |
| 1975-76 RESIDENT MD-STUDENT TUITION | 1166.29 | 1754.86 | 1027.94 | 2445.14 | 1933.40 | 972.42 | 3828.53 | 3491.77 | 2207.74 |
| % IN-STATE 1ST-YR MD-STUD | 94.49 | 90.03 | 90.03 | 82.07 | 86.86 | 93.62 | 38.14 | 44.27 | 74.06 |
| RAT: APPLICANTS PER 1ST-YR MD-STUD | 10.83 | 20.15 | 13.20 | 22.34 | 23.93 | 29.18 | 40.57 | 39.92 | 25.85 |
| % REV FROM FED SOURCES & RCOV IND COSTS | 31.88 | 29.76 | 40.34 | 37.88 | 29.40 | 22.73 | 42.57 | 48.85 | 37.20 |
| % REV FROM ALL GIFTS | 3.44 | 7.06 | 3.89 | 7.10 | 7.45 | 4.25 | 8.01 | 11.93 | 6.70 |
| IRG APPROVAL RATE OF NIH R01 COMP APPS | 65.82 | 77.85 | 75.21 | 78.68 | 0 | 77.14 | 71.53 | 74.08 | 70.35 |
| % FEMALE MD STUDENTS | 14.96 | 17.12 | 17.87 | 19.78 | 16.97 | 21.15 | 21.50 | 18.79 | 18.73 |
| RAT: VOL FAC TO FT FAC | 1.91 | 3.33 | 1.38 | 1.87 | 3.09 | 5.26 | 2.30 | 1.04 | 2.33 |
| DRG FED SPON RES CONS % CHG 67-9 to 72-4 | 5.44 | -2.03 | 44.67 | 13.36 | -51.28 | 317.60 | 20.88 | 25.26 | 43.50 |
| PROJTD ANNL $ 1ST-YR ENROLL CHG: 1974-79 | 1.26 | .22 | 2.29 | -.39 | 3.00 | 9.76 | .45 | 1.25 | 2.09 |
| DRG GRANTS - # R01 APPS APPROVED | 17.59 | 30.00 | 31.81 | 53.86 | 0 | 17.92 | 15.42 | 46.76 | 28.65 |
| % EXPD FOR ADMIN & GENL EXPENSE | 10.82 | 11.95 | 6.53 | 7.68 | 15.07 | 14.33 | 10.96 | 7.49 | 9.99 |
| % TOTAL EXPD FOR SPON RESEARCH | 14.10 | 17.82 | 21.96 | 27.00 | 6.95 | 14.66 | 23.19 | 36.74 | 22.04 |
| RAT: BMS GRAD-STUD TO UNDERGRAD MD-STUD | .17 | .19 | .34 | .20 | .05 | .17 | .17 | .35 | .22 |
| ADJUSTED TOTAL REVENUE THOUSANDS OF $ | 19,241.3 | 33,464.9 | 29,693.4 | 53,017.6 | 5,344.9 | 16,264.4 | 14,913.1 | 39,997.9 | 27,648.0 |
| % 1ST-YR MD-STUD: PRE-MED GPA 3.6 - 4.0 | 33.43 | 38.82 | 44.09 | 35.19 | 26.25 | 35.49 | 25.91 | 55.99 | 37.85 |

may be characterized as having a high ratio of undergraduate medical students per full-time faculty, slightly below average resident tuition rates and ratios of applicants per first year medical students, and slightly above average amounts of total revenue.

The 16 schools which comprise Cluster 3 are public schools which have a high degree of research emphasis and research funding success as opposed to graduate medical education emphasis. These schools are of moderate size and age and are not in the process of development, except in the area of research emphasis. The schools in this cluster experienced a 45 percent growth in DRG research support between 1966-67 and 1972-74, and they devote a relatively low percentage (6.5%) of their expenditures to administration and general expense.

Cluster 4 consists of 14 medical schools which are well established and have a strong graduate medical education program in addition to their undergraduate medical education program. These schools have an average of almost 650 undergraduate medical students, but have a low ratio of undergraduate medical students per full-time faculty member. The comparative strength of the medical schools in this cluster is illustrated by the fact that Cluster 4 has highest mean values of the eight clusters on the following variables: average salary (strict full-time basic science associate professor), percent of faculty with an M.D.-degree, R01 application approval rate, number of R01 applications approved by Initial Review Groups, and total revenue. In addition, the schools in Cluster 4 have second highest mean values of the eight clusters on ratio of housestaff (interns and residents) to undergraduate medical students, percent of living alumni who are board certified, and percent of total expenditures devoted to sponsored research. It is interesting to note, however, that an average of only 10 percent of the living alumni of the schools in Cluster 4 were in general practice.

Cluster 5 is a group of primarily new medical schools which either are two-year schools or were not operating with full student bodies in 1974-75. The schools in this cluster had the lowest average enrollments, the lowest ratio of housestaff to undergraduate medical students, the lowest percentage of faculty holding M.D.-degrees, and correspondingly, the highest proportion of expenditures devoted to administration and general expenses. It may very well be that these five schools, as well as the Mayo Medical School,

may be so distinct that they are not representative of the general population of medical schools at the current time. The development of these schools and their changing patterns of similarity to the rest of the population may merit special consideration as the schools become established.

By comparison, the schools in Cluster 6 are relatively new, public medical schools which are currently experiencing rapid development. While they are below average in size and age and research emphasis, the schools have had a moderate degree of research funding success and have slightly above average emphasis on graduate medical education. The most notable characteristic of the schools in this cluster is that they have the highest average values for both change in federal research support and projected change in enrollment. These schools have the lowest average in-state tuition, enroll over 93 percent in-state undergraduate medical students, and have the third-highest ratio of applicants per first year medical student. In addition, they utilize relatively more volunteer faculty than any other group of schools and devote almost as much of their total expenditures to administration and general expenses (14.33 percent) as to sponsored research (14.66 percent).

The final two clusters are composed of established, largely private schools with almost complementary profiles in other respects. The schools in Cluster 7 are slightly above average in size and age and have a moderately high degree of research funding success, but place low emphasis on graduate medical education and research compared to other medical schools. The schools in this cluster tend to be of average size, but have the lowest average total revenues of the clusters of established schools. As a group, these schools are the most expensive to attend, enroll the fewest undergraduate medical students from the states in which they are located, and have the highest number of applicants per enrolled first year medical students of any of the clusters.

The schools in Cluster 8, by way of contrast, have strong emphasis for both research and graduate medical education, but tend to have slightly fewer undergraduate medical students and slightly less research funding success than the average school. The schools in Cluster 8 have the highest ratio of housestaff to undergraduate medical students and the lowest ratio of undergraduate medical students to full-time faculty of all clusters. They also have the second highest average total revenue of all clusters and receive the highest proportion of their revenues from the federal government of any of the clusters.

The preceding paragraphs describe the eight clusters which were derived in the course of the current study. However, not every school fits equally well into the cluster in which it is a member. One measure of how well a school fits into a cluster is the distance from the school to the cluster centroid. The membership of the clusters and the distance of each school from the cluster centroid is presented in Table 4. In examining Table 4 it should be remembered that the schools are in the cluster with the closest centroid and that one of the basic assumptions of cluster analysis is that all objects (schools) are placed into one of the clusters. As a result, the clusters vary in the degree of homogeneity, or similarity, of the schools which they contain. Three of the clusters (numbers 1, 2, and 4) appear from the information in Table 4 to be reasonably homogeneous. The remaining five clusters each tend to have several schools close to the cluster centroid with a smaller number of the periphery of the cluster.

In general it is evident that the clustering described in this report reflects principally the size, age, and control of the schools. These characteristics were also evident in earlier studies (Nunn and Lain, 1976; McShane, 1977a). Differences in the composition of the clusters were the result of the changes in the variables selected, changes in the quality of the data used, and changes in the schools over time. It should be remembered that the current series of studies are exploratory in nature, and while the study described in this report represents an advancement over the previous studies it is only one of an infinite complex of possible solutions.

TABLE 4

MEMBERSHIP OF EIGHT CLUSTERS OF U.S. MEDICAL SCHOOLS IN ORDER OF DISTANCE FROM CLUSTER CENTROID RESULTING FROM CLUSTER ANALYSIS OF SIX FACTOR SCORES, 1976

| School | Distance |
|---|---|
| **CLUSTER 1** | |
| MARYLAND | .1386 |
| LOUISIANA NW ORL | .2262 |
| M.C. OF VIRGINIA | .4092 |
| TENNESSEE | .4639 |
| LOUISVILLE | .5242 |
| OHIO | .5584 |
| GEORGIA | .5696 |
| ARKANSAS | .5788 |
| MISSISSIPPI | .7469 |
| OREGON | .7777 |
| SOUTH CAROLINA | .8160 |
| KENTUCKY | 1.0929 |
| NEBRASKA | 1.3232 |
| SOUTH ALABAMA | 1.9552 |
| RUTGERS | 2.0372 |
| OKLAHOMA | 2.1762 |
| SOUTH DAKOTA | 2.6781 |
| **CLUSTER 2** | |
| TEMPLE | .8109 |
| WAYNE STATE | 1.3371 |
| SUNY BUFFALO | 1.4050 |
| SUNY DOWNSTATE | 1.4777 |
| TEXAS GALVESTON | 1.7632 |
| INDIANA | 2.2552 |
| ILLINOIS | 2.6031 |
| JEFFERSON | 2.6975 |
| **CLUSTER 3** | |
| UTAH | .3168 |
| FLORIDA | .5255 |
| KANSAS | .7351 |
| COLORADO | .9751 |
| NORTH CAROLINA | 1.1698 |
| WISCONSIN | 1.2245 |
| NEW MEXICO | 1.3239 |
| U OF VIRGINIA | 1.3523 |
| IOWA | 1.3950 |
| MISSOURI-COLUMB | 1.4144 |
| WEST VIRGINIA | 1.7121 |
| ALABAMA-BIRMNGHM | 1.9385 |
| ARIZONA | 1.9873 |
| U OF WASH SEATTL | 2.4595 |
| CONNECTICUT | 4.5207 |
| PUERTO RICO | 8.0911 |
| **CLUSTER 4** | |
| PITTSBURGH | .7754 |
| MIAMI | .8758 |
| TEXAS SOUTHWEST | 1.1703 |
| NEW YORK UNIV | 1.2367 |
| EINSTEIN | 1.3062 |
| COLUMBIA | 1.3952 |
| SUNY UPSTATE | 1.6207 |
| MINN-MINNEAPOLIS | 1.6357 |
| U OF MICHIGAN | 1.9033 |
| CALIF SAN FRAN | 2.2119 |
| HARVARD | 2.2875 |
| NEW JERSEY | 2.6667 |
| NEW YORK MED | 3.2606 |
| CALIF L A | 3.8145 |
| **CLUSTER 5** | |
| NEVADA | 2.0703 |
| SO. ILLINOIS | 2.2886 |
| MINN-DULUTH | 2.5629 |
| MISSOURI K.C. | 3.1050 |
| CHICAGO MEDICAL | 4.7853 |
| **CLUSTER 6** | |
| M.C. OHIO TOLEDO | 1.0706 |
| TEXAS SAN ANTON | 1.2566 |
| SOUTH FLORIDA | 1.5081 |
| CALIF DAVIS | 1.6373 |
| LOUISIANA SHRVPT | 2.6802 |
| PENN STATE | 2.9570 |
| TEXAS TECH | 3.9943 |
| MASSACHUSETTS | 4.0861 |
| MICHIGAN STATE | 5.0520 |
| CALIF SAN DIEGO | 5.0919 |
| CALIF IRVINE | 7.0285 |
| TEXAS HOUSTON | 17.7744 |
| SUNY STONY BRK | 20.8547 |
| **CLUSTER 7** | |
| BOWMAN GRAY | .5787 |
| TULANE | .7235 |
| ALBANY | .7534 |
| HAHNEMANN | .9137 |
| BOSTON | 1.0506 |
| NORTHWESTERN | 1.3305 |
| CREIGHTON | 1.6183 |
| HOWARD | 1.6467 |
| ST LOUIS | 1.6894 |
| GEORGETOWN | 1.9770 |
| MEHARRY | 2.1976 |
| VERMONT | 2.4396 |
| TUFTS | 2.7253 |
| BROWN | 3.4968 |
| GEORGE WASH | 4.0908 |
| LOMA LINDA | 5.9505 |
| LOYOLA | 6.9853 |
| DARTMOUTH | 8.5947 |
| M.C. OF PENN. | 11.0754 |
| **CLUSTER 8** | |
| DUKE | .4389 |
| STANFORD | .6789 |
| WASH U ST LOUIS | .7999 |
| SOUTHERN CALIF | .8573 |
| ROCHESTER | .8654 |
| EMORY | .9160 |
| CASE WESTERN RES | 1.3179 |
| U OF PENN. | 1.5335 |
| VANDERBILT | 1.6652 |
| CORNELL | 1.7411 |
| JOHNS HOPKINS | 1.8317 |
| CINCINNATI | 2.0827 |
| YALE | 2.1735 |
| MC OF WISCONSIN | 2.1912 |
| U OF CHICAGO | 2.3764 |
| MT SINAI | 6.4374 |
| RUSH MED COL | 8.2129 |

## Chapter IV

## SUMMARY AND CONCLUSIONS

The study described in this report applied methods developed in earlier clustering studies performed at AAMC to a different set of variables and produced eight clusters of medical schools. A total of 140 variables were extracted from the IPS Researchable Data Base. Through a series of correlational studies this number was reduced to a final data set of 33 variables representing several of the measurable dimensions in the data maintained in IPS. The 33 variables were factor analyzed, eight factors were rotated using a varimax criterion, and factor scores for the eight factors were computed for 110 U.S. medical schools. The schools were then grouped using two techniques sequentially; Ward's hierarchical cluster analysis was used initially to give an indication of the potential number of clusters and initial cluster centers, and a non-hierarchical cluster analysis technique developed by Forgy was subsequently used to refine the cluster memberships. A number of cluster analyses of both types were performed, varying the set of factor scores input and the number of clusters derived. A final solution which produced eight clusters based on six factor scores was selected as the most representative solution based on the selected data.

The factor analysis resulted in eight factors describing the following dimensions: (1) graduate medical education emphasis, (2) size and age, (3) control, (4) minority participation, (5) research funding success, (6) curriculum electives, (7) development stage, and (8) research emphasis. Clustering 110 medical schools on six of these factor scores (factors 1, 2, 3, 5, 7, and 8) yielded eight clusters that represented reasonably homogeneous groupings of schools. However, each cluster retained distinguishing characteristics that allowed for representation of the group of schools as different from the other seven groups.

### Conclusions and Recommendations

While the study described in this report does represent a step forward from the earlier AAMC cluster analysis studies, it represents only one possible solution based on a particular selection of variables. There are a number of methodological and application issues that need to be given consideration before a solution representing the most coherent possible

set of clusters of medical schools can be obtained. Among the issues which merit consideration are the limitations of the data, the impact of missing and non-representative data on the solutions, the selection of variables, and the criteria for including schools in the analysis.

The first of these issues, the limitations of the data, may be the final limitation on the utility of studies of this type in this area. The data in IPS are largely self-reported by the schools or extracted from other systems where it is self-reported by faculty members, students, applicants, or alumni. As a result, the data are only useful if they are reported completely and accurately, if the data collected are meaningful, and if the data are collected in such a way that they are comparable across institutions. Efforts aimed at enhancing the quality of some of the data in IPS in the ways noted above are being undertaken. If the data in the system are to be optimally useful in the context of studies of this type, or any other context, these efforts must be maintained where they now exist and increased wherever possible.

The second issue, the impact of missing and non-representative data on clustering and scaling solutions, is a methodological problem the effects of which have not been completely ascertained. By using factor analysis and computing factor scores, the effects of missing data are somewhat obscured and may be compensated adequately. However, in the course of these studies it has become apparent that the effects of missing and non-representative data may be greater than previously anticipated. The degree of impact of missing and distorted data on cluster analysis and scaling solutions, especially when the original variables are used, should be determined and a method of compensating for these effects should be developed.

The selection of variables also plays an important role in studies of this type. As noted earlier, the measures of similarity used in studies of this type are extremely sensitive to the variables on which they are based. Small changes in the variables selected may have considerable effect on the solutions. Variable selection is even more important in a situation, such as that with medical schools, in which all of the members of the population of interest are included in the analysis. The problem is one of sampling variables for analysis from the universe of variables, rather than sampling subjects. In studies of this type, the ramifications

of sampling variables should be investigated. It may be that some alternative analytic technique, such as Alpha factor analysis (Kaiser and Caffrey, 1965), would serve better in the selection of variables.

In this study, variables were selected on the basis of their potential for revealing dimensions not previously described among medical schools in addition to their completeness and representativeness. The effects of the variable selection are apparent in both the factor analysis and in the results of the cluster analysis. It would seem to be an appropriate next step to combine the knowledge gained from these results with that of previous studies to select, possibly with the aid of Alpha factor analysis, a new set of variables representing the universe of data in IPS.

The final issue, that of the criteria for including schools in the analysis, is one which affects the application of these techniques to institutional data. There are two possible reasons for excluding schools from analysis, one being a high proportion of missing data, and the other being that a school is highly dissimilar to all other schools in the analysis. In this study, seven schools were excluded for the former reason, one for the latter. While the amount of missing data that exists primarily affects the degree to which data are representative of a school, the inclusion of schools which are highly unlike other schools affects the analysis itself. One of the underlying assumptions of cluster analysis is that all of the objects submitted to analysis will be placed in one of the clusters. The inclusion of outlying objects causes some distortion in the clusters and could possibly affect the cluster solutions in other ways. The desire to cluster as many schools as possible should be weighed against the effects of outlying schools on the cluster solution.

In addition to the issues discussed above, there is also a need for further investigation into the clustering techniques themselves. These issues are more methodological than those discussed above, and include alternative methods of computing similarity among schools and of ascertaining the starting points for non-hierarchical cluster analyses. In the studies performed by AAMC to date the Euclidean metric has been used to compute similarities. While this method may be accurate and robust, there may be an alternative method, such as the "city-block" metric, which would depict the particular data under consideration differently and render more meaningful results. Similarly, for the starting

points for non-hierarchical clustering, the use of representative schools may be an adequate method of specifying initial starting points, but it is also possible that some other method, such as using randomly selected schools or specific centroid coordinates, would be more applicable in the current context.

In conclusion, the results of this study achieve the goals of clustering medical schools on the selected data. They provide a new and different basis for looking at medical schools and how they are similar to one another. There are several factors, however, that place limits on the universality of these results.

BIBLIOGRAPHY

Anderberg, M. R. Cluster Analysis for Applications. New York: Academic Press, 1973.

Bailey, K. D. "Cluster Analysis." In Sociological Methodology: 1975. San Francisco: Jossey-Bass Publishers, 1975.

Everitt, B. Cluster Analysis. London: Heinemann Educational Books, Ltd., 1974.

Forgy, E. W. "Cluster Analysis of Multivariate Data: Efficiency Versus Interpretability of Classifications." Paper presented at the Biometric Society Meeting, Riverside, California, 1965.

Kaiser, H. F. and Caffrey, J. "Alpha Factor Analysis." Psychometrika, 30: 1-14, 1965.

Keeler, E., Koehler, J.E., Lee, C., and Williams, Jr., A.P. Finding Representative Academic Health Centers. Working Note prepared for NIH/DHEW. Santa Monica, California: The Rand Corporation, 1972.

McShane, M.G. Classification of U. S. Medical Schools: A Replication. Washington, D.C: AAMC, 1977a.

McShane, M.G. Medical Schools in the United States: A Descriptive Study. Washington, D.C: AAMC, 1977b.

Mulaik, S.A. The Foundations of Factor Analysis. New York: McGraw-Hill, 1972.

Nunn, R. and Lain,L.L. Classification of Medical Institutions. Washington, D.C: AAMC, 1976.

Sherman, C.R. Study of Medical Education: Interrelationships Between Component Variables. Bethesda, Maryland: Bureau Of Health Manpower, DHEW Publication No. (HRA) 76-98, 1976.

Sherman, C.R. Exploratory Analyses of the Relations of Institutional Variables: A Replication. Washington, D.C: AAMC, 1977a.

Sherman, C. R. A Second Exploratory Analysis of the Relations Among Institutional Variables. Washington, D.C: AAMC, 1977b.

Ward, J. H. "Hierarchical Grouping to Optimize an Objective Function." Journal of the American Statistical Association, 58:236-244, 1963.

## APPENDIX A

### Abbreviations Used in 1976 Researchable Data Base Variable Labels

| | |
|---|---|
| $ | Dollars |
| # | Number |
| % | Percent |
| % Chg | Percent Change |
| A-Health | Allied Health |
| Accel | Accelerated |
| Act | Avcite, Activity |
| Adm | Administration |
| Admin & Genl | Administration & General |
| Admt | Admitted |
| Adm-Pref | Admittance-Preference |
| Adu Stdg | Advanced Standing |
| AEC | Atomic Energy Commission |
| Affil | Affiliated |
| Agrmt | Agreement |
| Alum | Alumni, Alumnae |
| Amer | American |
| Amt | Amount |
| Annl | Annual |
| App | Applications, Applicant |
| Applicnts | Applicants |
| Apply | Applying |
| Appr | Appropriations |
| Assist | Assistant (ASST) |
| Assoc | Associate |
| Avail | Available |
| Av | Average |
| BA | Bachelor of Arts |
| Bas | Basic (Sciences) |
| Bal | Balance |
| BHRD | Bureau of Health and Resources Development |
| BMS | Basic Medical Sciences |
| BS | Bachelor of Science |
| Budg | Budget(ed) |
| Bus & Ind | Business and Industry |
| Ch | Choice |
| Chg | Change |
| Clin | Clinical (Sciences) |

APPENDIX A (Continued)

| | |
|---|---|
| Coll | College |
| Comm | Committee |
| Comp | Competing |
| Con$ | Constant Dollars (adjusted for inflation) |
| Curr | Curriculum |
| Def | Deficit |
| Deg | Degree |
| Dept | Department (al) |
| DHEW | Dept. of Health, Education and Welfare |
| Diff | Difference |
| Dir | Direct |
| Disadv | Disadvantaged |
| Dist | Distributed |
| DOD | Dept of Defense |
| DRG | Division of Research Grants (NIH) |
| Ed | Education, Educational (Educ) |
| Elec | Electives |
| Emerg-Med | Emergency Medicine |
| Endow | Endowments |
| Enroll | Enrollment |
| Equivs | Equivalents |
| Exp | Expenditures (Expd) |
| Fac | Faculty |
| Facil | Facility |
| Fed | Federal |
| Fem | Female |
| Fin | Financial |
| Fin-Yr | Final Year |
| FMG | Foreign Medical Graduate |
| Fr | From |
| FT | Full Time |
| Gen | General |
| Govt | Government |
| GPA | Grade Point Average |
| Grad | Graduate |
| GT | Greater than |
| HMO | Health Maintenance Organization |
| IMPAC | DRG's computer file of grants & contracts |

APPENDIX A (Continued)

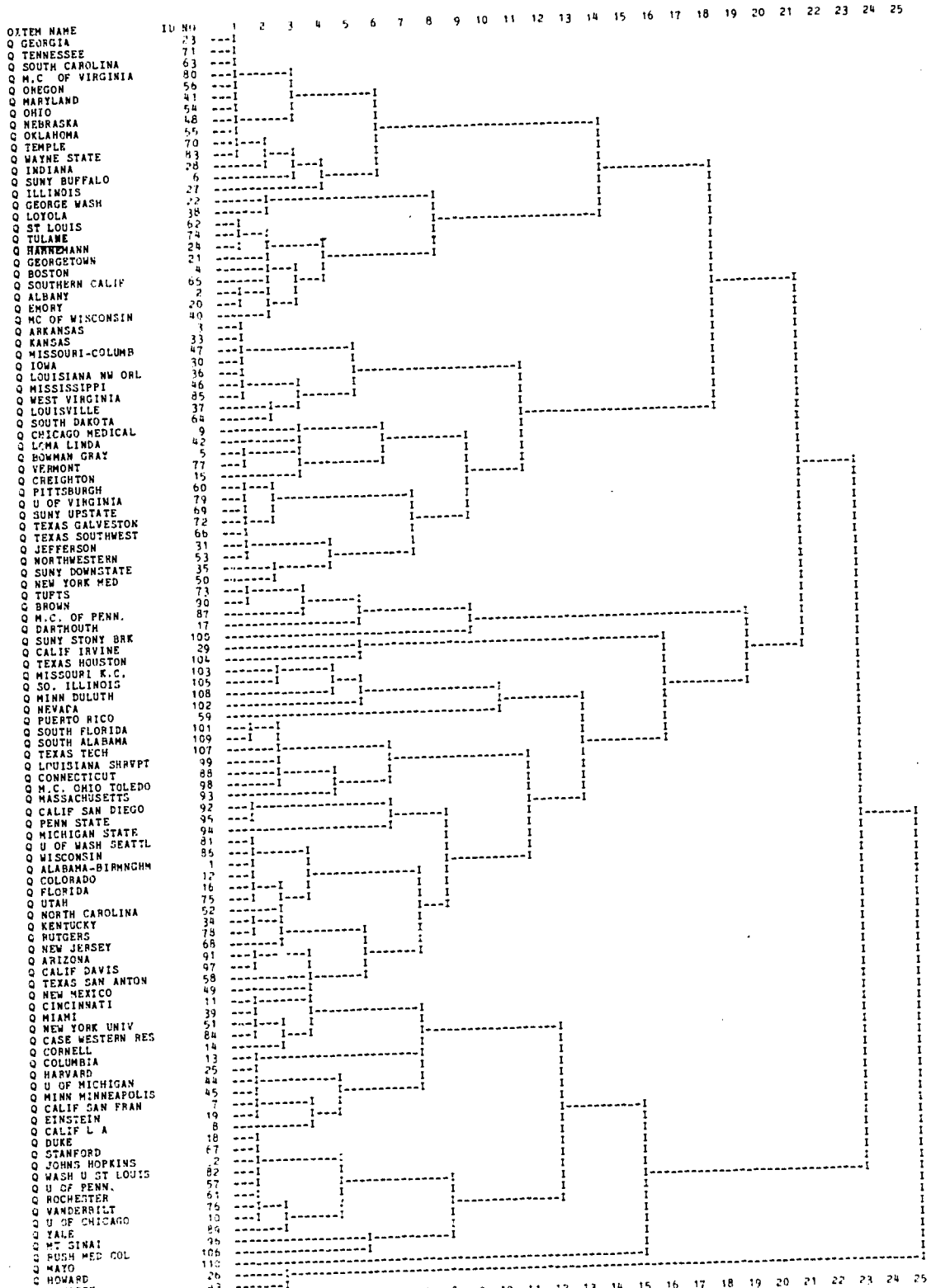
| | |
|---|---|
| Incl | Including |
| Indir | Indirect (Ind) |
| Innov | Innovations |
| Instr | Instructor |
| Instrct | Instructional |
| Intrn | Interns |
| IRG | Initial Review Group (study section) |
| LCME | Liaison Committee on Medical Education |
| Liv | Living |
| Log | Logarithm |
| LT | Less Than |
| Matric | Matriculant |
| MCAT | Medical College Admissions Test |
| MD-Stud | Medical Student |
| Med | Medical |
| Med-Sch | Medical School |
| Mid-Yr | Middle Year |
| Min | Minority |
| Mnlnd | Mainland |
| MS | Master's degree |
| Multi-Purp | Multi-Purpose (MP) |
| Multi-Serv | Multi-Service |
| NBME-1 | National Board Medical Examiners (test) - Part I |
| NBME-2 | National Board of Medical Examiners - Part II |
| NIH | National Institutes of Health |
| NIMH | National Institute of Mental Health |
| Non-Govt | Non-Governmental |
| Non-Res | Non-Resident |
| NSF | National Science Foundation |
| Oper & Maint | Operation and Maintenance |
| Org | Organized, Organizational |
| Outpat | Out patient |
| P-Scr | Priority Score |
| PØl | Program and Project Grants |
| Phys | Physical |
| Pop | Population |
| Pos | Position |
| Post-Docs | Post-Doctorates |
| Post-Grad | Post-Graduates |
| Prac | Practice |

APPENDIX A (Continued)

| | |
|---|---|
| Pre-Med | Pre - Medical |
| Priv | Private |
| Prof | Professional |
| Prog | Program (Pgm) |
| Projtd | Projected |
| PT | Part Time |
| Pub | Public |
| Quant | Quantitative |
| RØl | Traditional Research Grants |
| Rat | Ratio |
| Rec | Received |
| Recov | Recovery (RCOV) |
| Reg Oper Expd | Regular Operating Expenditures |
| Rel | Related |
| Res | Research |
| Resrv | Reserves |
| Ret | Retention |
| Rev | Revenues |
| Rsdnt | Resident |
| Sal | Salary |
| SBMT | Submitted |
| Sch | School |
| Sci | Science |
| SD | Standard Deviation |
| Sep | Separately |
| Serv | Service |
| SFT | Strict Full Time |
| SMSA | Standard Metropolitan Statistical Area |
| Spec | Special, Specialty |
| Spons | Sponsored |
| Sq | Square |
| St & Loc | State and Local (S&L) |
| St Rel | State Related |
| Std | Standardized |
| Stud | Student |
| Tch-Trn | Teaching and Training |
| Tchng | Teaching |
| Tot | Total |
| Undergrad | Undergraduate (Ungrad, UG) |

APPENDIX A (Continued)

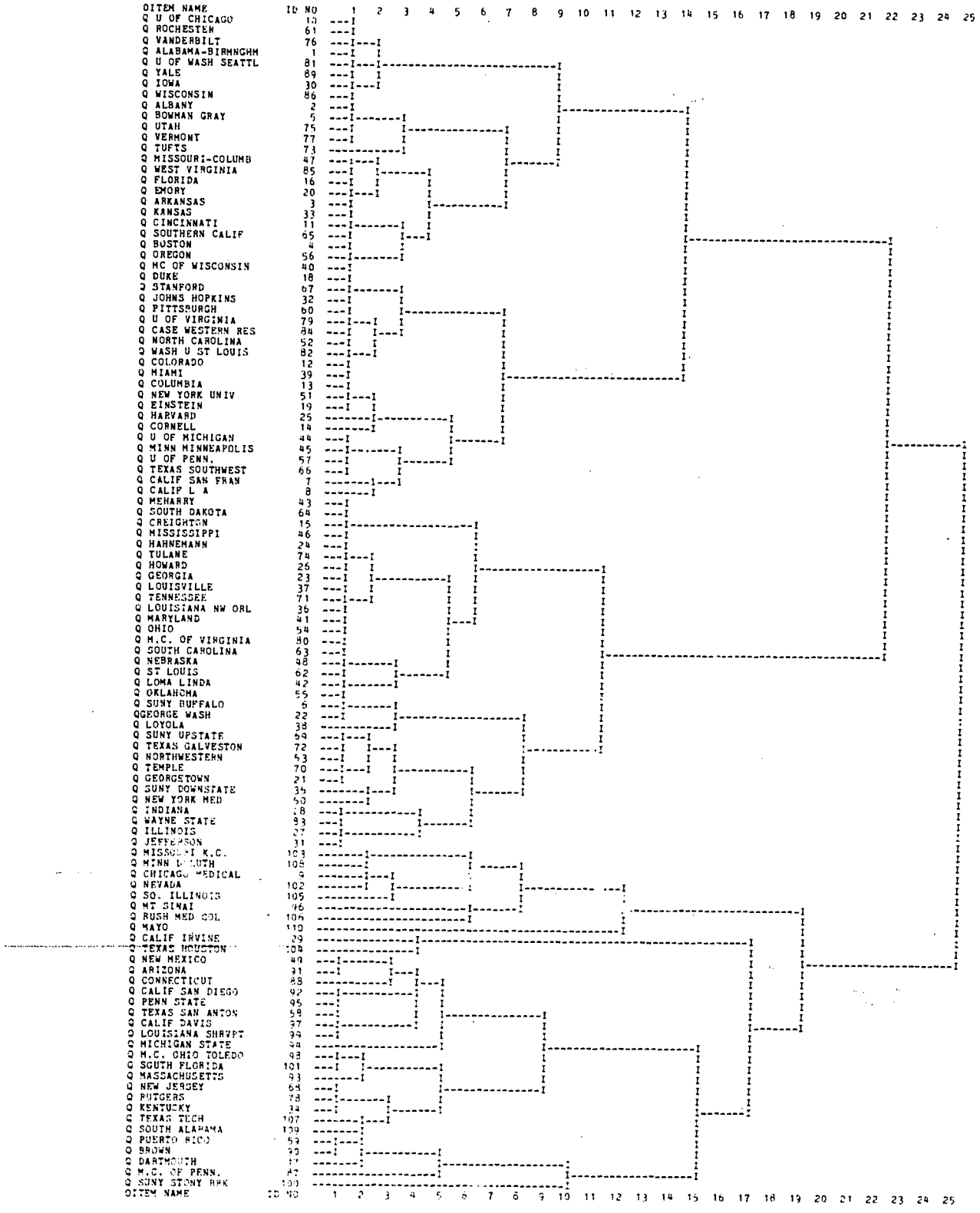
| | |
|---|---|
| Underrep | Under-represented |
| Unk | Unknown |
| Unrestr | Unrestricted |
| US-Can | United States and Canadian |
| Vol | Volunteer |
| Yr | Year |




APPENDIX B-1

WARD HIERARCHICAL CLUSTER ANALYSIS OF 110 U.S. MEDICAL SCHOOLS

BASED ON 8 FACTOR SCORES, 1976




APPENDIX D-2

WARD HIERARCHICAL CLUSTER ANALYSIS OF 110 U.S. MEDICAL SCHOOLS

BASED ON 5 FACTOR SCORES, 1976

APPENDIX C-1

CLUSTER MEMBERSHIP AND PROFILES OF CLUSTER CENTROIDS FROM CLUSTER ANALYSIS OF FIVE FACTOR SCORES. 1976

| | MEMBERSHIP | | GRADUATE MEDICAL EDUCATION PROGRAMS | SIZE AND AGE | RESEARCH FUNDING SUCCESS | DEVELOPMENT STAGE | RESEARCH EMPHASIS |
|---|---|---|---|---|---|---|---|
| CLUSTER 1 | BROWN, TUFTS, MASSACHUSETTS, SOUTH FLORIDA, M.C.OHIO TOLEDO, CONNECTICUT, M.C. OF PENN., DARTMOUTH, PUERTO RICO, MICHIGAN STATE, SUNY-STONY BRK | HI<br>MD<br>LO | | | | | |
| CLUSTER 2 | COLORADO, FLORIDA, ROCHESTER, EMORY, VANDERBILT, UTAH, U. OF CHICAGO, YALE, CINCINNATI, DUKE, WISCONSIN, VERMONT, NEW MEXICO, KANSAS, STANFORD, SOUTHERN CALIF., BOSTON, ARIZONA, NORTH CAROLINA, MISSOURI-COLUMBIA, WASH. U. ST. LOUIS, ALABAMA-BIRMINGHAM, IOWA, MC OF WISCONSIN, U OF WASH. SEATTL | HI<br>MD<br>LO | | | | | |
| CLUSTER 3 | U.OF PENN., NEW YORK UNIV., TEXAS-SOUTHWEST, MIAMI, EINSTEIN, U. OF MICHIGAN, MINN-MINNEAPOLIS, COLUMBIA, CALIF.-SAN FRAN., CORNELL, JOHNS HOPKINS, CALIF. L.A. | HI<br>MD<br>LO | | | | | |
| CLUSTER 4 | HAHNEMANN, SOUTH DAKOTA, ST. LOUIS, CREIGHTON, TULANE, LOUISIANA ST. U., ARKANSAS, MARYLAND, MISSISSIPPI, HOWARD, LOUISVILLE, GEORGIA, BOWMAN GRAY, SOUTH CAROLINA, HEHARRY, OKLAHOMA, SOUTH ALABAMA, LOMA LINDA, WEST VIRGINIA, NEBRASKA | HI<br>MD<br>LO | | | | | |

APPENDIX C-1

(Continued)

| | MEMBERSHIP | | | GRADUATE MEDICAL EDUCATION PROGRAMS | SIZE AND AGE | RESEARCH FUNDING SUCCESS | DEVELOPMENT STAGE | RESEARCH EMPHASIS |
|---|---|---|---|---|---|---|---|---|
| CLUSTER 5 | NORTHWESTERN, OHIO, TEMPLE, GEORGETOWN, M.C.OF VIRGINIA, TENNESSEE, JEFFERSON, SUNY-BUFFALO, ORECON, INDIANA, WAYNE STATE, SUNY-DOWNSTATE, ILLINOIS, LOYOLA, GEORGE WASH. | | HI<br>MD<br>LO | | | | | |
| CLUSTER 6 | PITTSBURGH, SUNY-UPSTATE, CASE WESTERN RES, U. OF VIRGINIA, RUTGERS, ALBANY, KENTUCKY, TEXAS-GALVESTON, HARVARD, NEW JERSEY, NEW YORK MED. | | HI<br>MD<br>LO | | | | | |
| CLUSTER 7 | SO. ILLINOIS, LOUISIANA-SHRVPT, NEVADA, TEXAS TECH, MISSOURI K.C., MINN-DULUTH, CHICAGO MEDICAL, RUSH MED COL, MT.SINAI. | | HI<br>MD<br>LO | | | | | |
| CLUSTER 8 | TEXAS SAN ANTON, CALIF-DAVIS, PENN STATE, CALIF-SAN DIEGO, CALIF-IRVINE, TEXAS HOUSTON | | HI<br>MD<br>LO | | | | | |

APPENDIX C-2

MEMBERSHIP OF EIGHT CLUSTERS OF U.S. MEDICAL SCHOOLS IN ORDER OF DISTANCE FROM CLUSTER CENTROID BASED ON CLUSTER ANALYSIS OF FIVE FACTOR SCORES

| School | Distance |
|---|---|
| **CLUSTER 1** | |
| BROWN | 1.2914 |
| TUFTS | 1.3060 |
| MASSACHUSETTS | 1.6796 |
| SOUTH FLORIDA | 1.8369 |
| MC. OHIO TOLEDO | 2.3469 |
| CONNECTICUT | 2.7556 |
| M.C. OF PENN. | 4.0012 |
| DARTMOUTH | 4.1798 |
| PUERTO RICO | 5.1041 |
| MICHIGAN STATE | 7.0551 |
| SUNY STONY BRK | 14.2849 |
| **CLUSTER 2** | |
| COLORADO | .3258 |
| FLORIDA | .4116 |
| ROCHESTER | .4451 |
| EMORY | .5268 |
| NORTH CAROLINA | .5641 |
| UTAH | .6734 |
| U OF CHICAGO | .6734 |
| WASH U ST LOUIS | .8177 |
| DUKE | .8249 |
| WISCONSIN | .8673 |
| KANSAS | .9193 |
| IOWA | .9201 |
| VANDERBILT | .9998 |
| STANFORD | 1.0871 |
| BOSTON | 1.1410 |
| ALABAMA-BIRMNGHM | 1.2461 |
| MC OF WISCONSIN | 1.5048 |
| SOUTHERN CALIF | 1.5793 |
| U OF WASH SEATTL | 1.6304 |
| MISSOURI-COLUMB | 1.8118 |
| NEW MEXICO | 1.8349 |
| YALE | 1.8425 |
| CINCINNATI | 1.9762 |
| VERMONT | 2.3603 |
| ARIZONA | 2.5346 |
| **CLUSTER 3** | |
| U OF PENN. | .1021 |
| NEW YORK UNIV | .4211 |
| TEXAS SOUTHWEST | .5455 |
| MIAMI | .6465 |
| EINSTEIN | .8548 |
| U OF MICHIGAN | .9073 |
| MINN-MINNEAPOLIS | 1.2214 |
| COLUMBIA | 1.2418 |
| CALIF SAN FRAN | 1.3610 |
| JOHNS HOPKINS | 1.5516 |
| CORNELL | 1.6007 |
| CALIF L A | 2.0233 |
| **CLUSTER 4** | |
| HAHNEMANN | .3818 |
| ST LOUIS | .4429 |
| TULANE | .4914 |
| CREIGHTON | .5009 |
| ARKANSAS | .5022 |
| LOUISIANA NW ORL | .5370 |
| HOWARD | .5607 |
| LOUISVILLE | .6337 |
| MARYLAND | .6714 |
| MISSISSIPPI | .7541 |
| GEORGIA | .8673 |
| NEBRASKA | .8938 |
| BOWMAN GRAY | 1.0283 |
| MEHARRY | 1.1973 |
| SOUTH CAROLINA | 1.2167 |
| SOUTH ALABAMA | 1.5677 |
| OKLAHOMA | 1.7851 |
| WEST VIRGINIA | 1.8382 |
| SOUTH DAKOTA | 1.8838 |
| LOMA LINDA | 2.9749 |
| **CLUSTER 5** | |
| NORTHWESTERN | .2354 |
| OHIO | .3756 |
| GEORGETOWN | .4439 |
| TEMPLE | .4599 |
| M.C. OF VIRGINIA | .4974 |
| TENNESSEE | .7677 |
| JEFFERSON | .9759 |
| SUNY BUFFALO | 1.0207 |
| OREGON | 1.0271 |
| WAYNE STATE | 1.2812 |
| INDIANA | 1.6136 |
| SUNY DOWNSTATE | 2.1479 |
| ILLINOIS | 3.6354 |
| GEORGE WASH | 4.9173 |
| LOYOLA | 5.6293 |
| **CLUSTER 6** | |
| PITTSBURGH | .1649 |
| SUNY UPSTATE | .4070 |
| CASE WESTERN RES | .5244 |
| U OF VIRGINIA | .7098 |
| RUTGERS | 1.0170 |
| KENTUCKY | 1.0641 |
| TEXAS GALVESTON | 1.1260 |
| ALBANY | 1.1471 |
| NEW JERSEY | 1.2880 |
| HARVARD | 1.6775 |
| NEW YORK MED | 1.8277 |
| **CLUSTER 7** | |
| SO. ILLINOIS | 1.4077 |
| LOUISIANA SHRVPT | 2.1350 |
| NEVADA | 2.6368 |
| TEXAS TECH | 2.7573 |
| MISSOURI K.C. | 2.9510 |
| MINN-DULUTH | 3.3004 |
| CHICAGO MEDICAL | 3.7453 |
| RUSH MED COL | 5.6961 |
| MT SINAI | 6.8062 |
| **CLUSTER 8** | |
| TEXAS SAN ANTON | 1.0116 |
| CALIF DAVIS | 2.2052 |
| PENN STATE | 2.6666 |
| CALIF SAN DIEGO | 2.8082 |
| CALIF IRVINE | 2.8530 |
| TEXAS HOUSTON | 11.7976 |