THE DEMAND FOR GENERAL HOSPITAL FACILITIES: A PRINCIPAL COMPONENTS ANALYSIS

Gerald Rosenthal - Harvard University

In an earlier study,[1] the use of general hospital facilities and its components, the admissions rate and the average length of stay in terms of patient-days per 1000 population were expressed as a function of the influence of the following social, demographic and economic characteristics:

X_1	Price	Average 2-bed room rate
X_2	Income	% incomes over $5999
X_3	"	% incomes under $2000
X_4	Insurance	% hospital insurance coverage
X_5	Age	% over 64 years of age
X_6	"	% under 15 years of age
X_7	Marital Status	% Females 14 and over who are married
X_8	Sex	% 14 years and over that is male
X_9	Urbanization	% residing in urban area
X_{10}	Education	% 25 years and over with 13 or more years of education
X_{11}	Race	% non-white
X_{12}	Family size	Population per dwelling unit

The relationship was estimated by means of a cross-sectional multiple regression analysis with the individual states in the Continental United States as units of observation.[2] The model was estimated for the census years 1950 and 1960 and displayed corrected multiple correlation coefficients (R^2's) of .7195 and .6832 respectively for patient-days per 1000 population. The model proved to be quite useful for the purposes of the study, but certain difficulties arise in the use of twelve characteristics. An ultimate goal of this research is to develop time-series analyses. Since there are few data prior to 1948, it seems unlikely that a period greater than 15 or 16 years can be used. This limitation on observations places a high premium on reducing the dimensionality of the model. In addition, there is considerable evidence that many of the twelve characteristics in the original model are collinear and reflect similar influences. To the extent that this statistical redundancy can be reduced, the model can be both statistically and conceptually more useful.

This investigation was supported by Public Health Service Research Grant HM00302 from The Division of Hospital and Medical Facilities.

1. Rosenthal, Gerald, <u>The Demand for General Hospital Facilities</u>, American Hospital Association Monograph Series No. 14.

2. District of Columbia, Maryland and Virginia were aggregated into a single unit leaving 47 observations. The correlation matrix for all variables is presented in Appendix 1.

Derivation of Independent Influences

In the current paper a procedure is established whereby interrelated characteristics can be grouped into sets of independent influences on hospital utilization. The procedure used is principal components analysis. By means of this statistical technique, it is possible to derive a set of linear combinations of the twelve variables, each of which is independent of the other. Each of the combinations is known as a principal component. Each component will reflect, to some degree, all of the twelve variables. However, the degree to which individual variables will be associated with each component will differ significantly.

The first step is to ascertain how much of the total variation in the original twelve variables can be explained by each principal component. If there are fewer than twelve independent influences and it is possible to represent each independent influence by a single variable, a basis is provided for the development of a smaller and more meaningful model. The percentage of variation in the original twelve variables which is explained by each principal component is presented for the years 1950 and 1960 in Table 1. In each case the principal components are listed in order of the percentage of total variation they explain. In both years 4 or 5 principal components explain approximately 90% of the variation in the twelve variables as a whole.

Interpretation of Independent Influences

After the principal components have been derived, one can examine the degree to which each of the components is related to each of the twelve original variables. These relationships provide the basis for ascribing meaningful interpretations to the statistically-created principal components. This is done by correlating the observations for each variable with the statistically-created observations for each principal component. The correlation coefficients measuring the relationships between each principal component and each of the original variables are known as factor loadings. They represent the degree to which each individual characteristic is embodied in the general influence represented by each principal component. The factor loadings for each variable in each of the seven principal components are also presented in Table 1.

The factor loadings show that in each principal component certain of the original twelve variables have much higher factor loadings than others. This, then, suggests a way of distributing the original twelve variables into the basic independent influences with which they are most closely related. When this is done, it becomes possible to ascertain whether any meaningful interpretation can be given to these principal components.

Table 1

FACTOR LOADINGS AND % VARIATION EXPLAINED OF TWELVE ORIGINAL VARIABLES ON PRINCIPAL COMPONENTS 1950 and 1960

1950

			1	2	3	4	5	6	7
% Variation Explained		Cumulative	48.33	73.04	83.68	89.07	93.10	95.39	97.17
		Individual	48.33	24.71	10.64	5.39	4.03	2.29	1.78
Characteristics	1		-.7893	-.0833	.2934	-.2090	-.3272	-.3281	-.1477
	2		-.8208	.2156	.3804	.0893	.2260	-.0516	.2323
	3		.8940	-.2262	-.1684	-.2985	-.0348	.0266	.0092
	4		-.7277	-.4987	.1379	.2130	.1728	.1742	-.2384
	5		-.5667	-.2837	-.7527	-.0622	-.0206	-.0255	.0362
	6		.8956	.2496	.1765	.1634	-.1676	.0690	-.1381
	7		.1818	.8814	.0208	-.2090	.2839	-.0019	-.2102
	8		.0207	.9427	-.0256	.1432	.0801	-.1423	.0342
	9		-.8364	-.2416	.3638	-.1922	.0086	.1400	-.0168
	10		-.4514	.7300	.0590	-.0719	-.3936	.2888	.0756
	11		.7139	-.3045	.3976	-.4248	.1138	.0347	.0850
	12		.7776	-.3456	.3263	.3775	-.1061	-.0625	.0431

1960

		Cumulative	43.41	69.12	81.88	88.39	92.18	94.91	96.83
		Individual	43.41	25.71	12.76	6.51	3.79	2.73	1.92
Characteristics	1		.8626	.0242	.1985	.0325	.2610	.3368	.0443
	2		.8612	.2451	.3535	.0554	-.1258	-.0233	.0335
	3		-.8743	-.2383	-.2048	-.2320	.1423	.0570	-.1080
	4		.6675	-.4882	.0820	.3956	-.2234	-.0392	-.2930
	5		.3573	-.6036	-.6529	-.0157	.1321	-.1402	.0525
	6		-.6734	.5932	.1912	.2761	.1354	-.1606	.0363
	7		.0621	.8520	-.2876	-.1833	-.3402	-.1004	.0256
	8		.0406	.8574	-.3607	.0824	-.0510	.2901	-.0534
	9		.7463	.0109	.4909	-.3653	.0303	-.1454	.1083
	10		.4872	.7172	-.0215	-.0878	.3586	-.1931	-.2583
	11		-.7126	-.1739	.4636	-.3953	-.1100	.1053	-.1964
	12		-.7904	.0866	.4307	.3917	.0652	-.0243	.0682

The most important variables in the first principal component in both 1950 and 1960 are the economic variables. In 1950, however, more of the original twelve variables showed up strongly in this first factor. The proportion of population under 15 is not directly an economic variable, but can perhaps be interpreted as indicative of economic ability to raise a family. The low income, urban, high income, and price variables, all of which reflect economic conditions, show up quite strongly in the first principal component. The two income variables and the price variable also show up strongly in the first component in 1960. It seems quite evident that this principal component represents the influence of the general level of economic prosperity, or, perhaps, the ability to pay.

In both years marital status, the sex distribution, education, and the proportion of aged in the population appear to be very unimportant in the first principal component. However, in the second principal component, the sex distribution and marital status show much higher factor loadings than do any of the other variables in both years. In the third principal component, both sex and marital status show little association. The aged variable, a variable which shows little relative association with the other principal components, is most highly associated with this factor. This suggests that the influence of the proportion of population 65 and over is largely independent of both the sex-marital status factor and the economic prosperity factor.[3]

In the fourth factor, race and population per dwelling unit seem to be fairly highly associated in both 1950 and 1960, though in 1960 both the urban-rural distribution and insurance coverage are much more highly associated than in 1950. In the original study, however, the significance of the insurance coverage variable was considered to be quite different in 1950 and 1960 because of differences in benefit levels. It is possible that it can be interpreted as being a different variable in each year. In 1950, the proportion of population with incomes under $2000 also shows up in this factor. This combination of characteristics might be interpreted as representing the "crowdedness" of housing since both the population per dwelling unit and the urban-rural distribution as well as the low-income measure reflect this influence. The influence of the insurance variable in 1960 is difficult to account for.

In the fifth principal component, the level of education seems to show the highest association. At this point, however, the degree of association between the variables and the principal components is rather small and differentiation among various characteristics are relatively insignificant.

3. This finding also suggests that the present model is more meaningful than the original 12-variable models, which yielded results showing no significant influence for this factor.

The analysis suggests that there are 4 or 5 independent influences which might be related to the utilization of short-term general hospitals: a general economic prosperity factor, a sex-marital status factor,[4] an aged factor, a housing factor, and possibly an education factor.

<u>A New Model of Hospital Utilization</u>

The principal components analysis made it possible to express the total variation of the twelve variables as a set of four or five basic influences. These independent influences, or factors, provided the basis upon which a new model of hospital utilization could be developed. Two aspects of the factors merit attention. First, the individual principal components provide a basis for ranking the influences in the order of their importance and second, the factor loadings within each principal component provide a basis for selection of individual characteristics which can then be used to represent the factor in an analytic model. A model which says that the utilization of short-term general hospitals is a function of the first, second, third, fourth, and fifth influences is desired. Since these influences are statistical creations, however, they cannot be measured directly. Thus it is necessary to represent these influences by a single variable selected from among the variables associated with each principal component.

For the first, third, and fifth variables, the selection was fairly straightforward. In general, the variables with the highest factor loadings also satisfied the requirement of being meaningful on <u>a priori</u> grounds. For the first variable, the low income measure was used, <u>a priori</u> reasoning serving to eliminate the proportion of population under 15 years of age from consideration.[5] For the third and fifth factors, the aged variable and the education variable were used respectively.

For the second and fourth variables, an experimental criterion was added. The marital status variable was, in all cases, preferable to the sex variable for the second principal component, both on grounds of <u>a priori</u> reasoning

4. In the original study it was suggested that the marital status variable and, more tenuously, the sex variable were indicators whether a household existed in which an individual could receive some degree of medical care without consuming hospital facilities. To the extent that this reasoning holds, this influence could be characterized as an alternative-to-hospitalization factor.

5. Some valid objections were raised concerning the desirability of using a fixed-dollar cutoff for income in two different time periods. However, the results using per capita income or median income do not differ significantly from the results obtained here.

and statistical stability. For the fourth variable, the urban-rural measure was selected. Even though it had a lower factor loading in 1950 it showed more consistent statistical behavior.[6] The case for using population per dwelling unit also had appeal although it seemed to reflect an influence similar to the marital status. The arbitrary selection here is primarily in the interests of brevity. The race variable proved of no usefulness while the insurance variable merits some special, and more detailed attention.

The results of the new regression analyses, using only five variables, are presented in Table 2. As in the original analysis, these are cross-section linear multiple regressions expressing patient-days per 1000 population in short-term general hospitals and its components, admissions per 1000 population and average length of stay, as a function of the five selected independent variables.

Evaluation of Statistical Results

A comparison of the statistical results obtained using the four-and five-variable models with the results obtained using a twelve-variable model is the first way of evaluating the performance of the new models. The basis for comparison is the efficiency of prediction of the models; that is, the amount of total variation in hospital utilization which is "explained" by the model as presented by the corrected R^2's obtained in the original study for 1950 and 1960 which are presented in Table 3.

Table 3

Original Twelve-Variable Models Corrected Multiple Correlation Coefficients

	1950	1960
Patient-Days per 1000 population	.7195	.6832
Admissions per 1000 population	.3742	.5693
Average length of stay	.6348	.7097

It is evident that the models using fewer variables do not "predict" as well as did the twelve-variable model. However, the degree of loss is not so great as to render the five-variable models statistically unsatisfactory. In 1950, the smaller model showed a R^2 corrected close to .6 which is certainly significant and indicates a relatively high degree of prediction. As will be discussed later, ability of the model to predict patient-days per thousand was significantly lower for 1960, although its ability to predict the admissions rate and the average length of stay was similar for both 1950 and 1960. It seems reasonable to suggest that, even though the estimating equations do not perform as well as the original model, they have sufficiently high multiple R^2's to serve as a base upon which to estimate utilization. Far more important is the indication that the degree of multi-collinearity among the variables has been reduced, and, in turn, that coincidently the statistical problems associated with this multi-collinearity have also been reduced (but not eliminated).[7]

If estimation were the sole desideratum, it would be possible to create a model using five of the twelve original variables which would have a higher multiple R^2 than do the models that have been presented here. Such models were estimated by selecting, from the twelve original variables, the five variables which, for each dependent variable, would yield the highest R^2. The variables were selected in order of their contribution to the R^2. The most highly correlated variable was taken first, and then, of those that were left, the next most significant variable, and so on. The results of this exercise are presented in Table 4 and compared to the earlier results in Table 5. The purpose of this experiment was to evaluate the degree of "prediction" loss incurred by insisting on a more meaningful selection of variables.

Table 5

Comparison of Corrected R^2's-
Two Alternative Estimating Procedures

	Principal Components	
	1950	1960
Pat.Days/1000	.5838	.3901
Adm./1000	.3106	.3678
Length of Stay	.5409	.5736
	Maximum "Prediction"	
	1950	1960
Pat.Days/1000	.6043	.6182
Adm./1000	.3669	.3846
Length of Stay	.6334	.6888

For the 1950 estimates, no housing variables showed up in the patient-days per 1000 population model, no education variable showed up in the length of stay model. In the latter case, proportion of population under 15 years of age did appear to be strongly related. This variable is highly associated with the economic variables. For 1960 none of the models included the aged variable or an education variable. There is considerable difference between these two statistical approaches with respect to the actual variables used in the estimating process. However, with the exception of the 1960 estimates for patient-days per 1000, the principal component models do not perform significantly worse than the maximum prediction models. The cost of using consistent and meaningful models of utilization based on sound a priori analysis is not,

6. Professor Rothenberg has suggested, rightly, that the urban variable has a significant potential for reflecting supply conditions and is, therefore, a poor choice for a demand equation. This observation alters the merits of the variables and makes population per dwelling unit a preferable variable for this analysis.

7. Obviously, collinearity is only removed if the principal components are used to estimate demand. However, this requires all 12 variables. The 5 variable models are a compromise and contain some collinearity, albeit less than the original.

Table 2

Regression Estimates - Principal Components Models

1950

Constant	Income		Marital		Aged		Urban		Educ		R^2	R^2 Corr.
					Patient Days per 1000 Population							
2574.19	-11.87	(-3.75)	-24.73	(-2.63)	42.26	(2.74)	-3.94	(-2.06	23.13	(2.53)	6369	.5838
					Admissions per 1000 Population							
146.90	-1.24	(-2.26)	.04	(.02)	3.60	(1.35)	-1.00	(-3.03)	3.20	(2.03)	.3986	.3106
					Average Length of Stay							
21.42	-.04	(-1.46)	-.22	(-2.99)	.12	(.98)	.03	(2.14)	-.01	(-.14)	.5995	.5409

1960

Constant	Income		Marital		Aged		Urban		Educ		R^2	R^2 Corr.
					Patient Days per 1000 Population							
2950.20	-17.21	(-4.08)	-24.48	(-2.30)	33.61	(2.64)	-4.66	(-2.55)	2.15	(.27)	.4680	.3901
					Admissions per 1000 Population							
213.86	-1.50	(-3.36)	-.39	(-.34)	2.28	(1.69)	-.97	(5.03)	1.27	(1.51)	.4485	.3678
					Average Length of Stay							
18.30	-.06	(-3.07)	-.16	(-3.27)	.12	(2.06)	.02	(2.06)	-.06	(-1.53)	.6280	.5736

Table 4

Regression Estimates - Maximum "Prediction" Models

1950

Equation	R^2 Corr.
$Y_1 = -75 - 6.9X_3$ (-2.9) $+ 50.6X_5$ (3.4) $+ 14.1X_{10}$ (1.5) $- 36.8X_7$ (-3.2) $+ 62.4X_8$ (2.6)	.6043
$Y_2 = -690.0 + 19.1X_8$ (4.1) $+ 10.0X_5$ (2.8) $- 3.0X_7$ (-1.5) $+ 1.2X_{11}$ (1.7) $- .8X_3$ (-1.5)	.3669
$Y_3 = 51.5 - .19X_6$ (-1.7) $- .38X_7$ (-3.6) $- .49X_5$ (-2.9) $- .03X_{11}$ (-1.5) $- 2.6X_{12}$ (-1.5)	.6334

1960

Equation	R^2 Corr.
$Y_1 = -1475.7 + 9.8X_4$ (5.9) $+ 134.1X_8$ (5.3) $- 52.8X_7$ (-4.5) $- 20.1X_1$ (-3.5) $- 287.4X_{12}$ (-2.7)	.6182
$Y_2 = 296.7 - .43X_{11}$ (-1.3) $- .80X_9$ (-3.4) $- 1.45X_3$ (-2.4) $- 1.29X_1$ (-1.4) $- 16.02X_{12}$ (-1.1)	.3846
$Y_3 = 8.69 + .03X_4$ (3.7) $- .11X_6$ (-2.3) $- .18X_7$ (-3.1) $+ .02X_9$ (2.4) $+ .22X_8$ (1.7)	.6888

where:

Y_1 = Pat. Days/1000
Y_2 = Admissions/1000
Y_3 = Average Length of Stay

X_1 = Price
X_3 = Lo Income
X_4 = Insurance
X_5 = Age 65+
X_6 = Age 15-

X_7 = Marital Status
X_8 = Sex
X_9 = Urban
X_{10} = Education
X_{11} = Race

X_{12} = Pop/DU

in this case, very great.[8] The primary purpose of this analysis is to provide some insight into the factors which influence hospital utilization and to find a form in which this knowledge can be used to "predict" utilization. The procedure described in this report has the advantage of providing this meaningful form.

Some Observations On the Variables

The analytic results for the first variable indicate that, in a statistical sense, the low income variable performs well.[9] In both 1950 and 1960 it is significant in almost all cases. There is, however, one interesting point. The first principal component represented something over 40% of the total variation in the twelve original variables, and the degree of association between the low income variable and the first principal component was quite high in both years. The degree of explanation contained in that variable, however, was much lower in 1960 than it was in 1950. Thus, to the extent that the first variable represents the ability to pay component, it can be said that the ability to pay was a less important determinant of overall hospital utilization in 1960 than it was in 1950.

This is true in spite of the fact that the ability to pay factor was almost as strongly represented in the twelve variables in 1960 as it was in 1950. This could lead to the conclusion that by 1960 there were other factors, not represented in the original twelve variables, which had begun to show a significant association with hospital utilization. This would suggest that the specification of the model must be changed and that new characteristics reflecting influences not now contained in the model must be incorporated into it. Although this type of change may arise from many circumstances, conceivably the explanation is that the organizational structure of medical care has become a more important determinant of the utilization of short-term hospitals than it was in the past.

There is, however, another explanation suggested by the data. In 1960, the first principal component no longer represented all of the ability to pay influence. The insurance variable, which is included in the fourth principal component, appeared to be an independent influence in 1960. It is possible that both an insurance and a housing variable should be incorporated in the prediction model.[10] This hypothesis is corroborated by the fact that in the maximum prediction models for 1960, the insurance variable was the first chosen for both patient-days per 1000 and average length of stay. It did not appear in the admissions rate models. In 1950, it was not of any significance. In the maximum prediction models, the 1960 estimates had higher R^2's, relative to 1950, for all utilization measures. This suggests that the twelve original variables do contain the same explanatory value but that the structure of their relationship to utilization has changed during the 10 year period.

Both of these hypotheses require additional analysis before the degree to which each can be considered a realistic interpretation of the true circumstances can be determined. When the time series analyses are developed, it is hoped that individual geographic areas can be separated into groups roughly similar in organization of medical care and that the relationship between hospital utilization and an area's characteristics can be examined more explicitly. The differences between 1950 and 1960 may reflect changes in types of practice rather than different responses to the same types of organization. This suggestion is appealing on a priori grounds, but must be subjected to empirical investigation.

The marital status variable, in general, shows a positive association to admission rates, which reflects the impact of child bearing, and a negative association with the average length of stay, in part because the average length of stay for obstetrical admissions is shorter than that for all admissions. To the extent that higher obstetrical admissions yield an increase in the number of admissions of shorter duration, the average length of stay should be diminished.

Even though the association of marital status with each of the sub-components of utilization is consistent with the impact of child-bearing, the overall association of marital status with patient-days per thousand is negative. This suggests that while an increase in the proportion of the population that is married adds to utilization by increasing child-bearing, some other influence, associated with marital status, functions in the opposite direction because the overall net effect is a lowering of utilization. Since we know that the negative influence operates primarily through lowering length of stay, our analysis suggests that a tentative description of the second factor as representing the existence of a substitute for hospital services might not be unreasonable. The existence of a household may permit shorter hospital stays by providing a form of convalescent facility outside of the hospital.

In almost all cases, the proportion of population 65 and over had a positive association with utilization. It is interesting to note that in most cases the association was more strongly operative through the admissions rate than through the length of stay. This suggests that the total incidence of all disease among the aged is more important in affecting total utilization than the fact that the disease mix is more heavily chronic than for other age groups.

---

8. The corrected R^2's are still significant, although less of the variation in demand is explained.
9. See Footnote 5.
10. Subsequent analyses indicate that this is true.

This latter circumstance would serve to increase the average length of stay rather than the admissions rate while the higher incidence of all disease would be reflected in the admissions rate. There is some indication that the influence of the proportion of aged on hospital utilization was greater in 1960 than in 1950. This may stem in part from the dilution over the ten-year period of the impact of economic circumstances. Since there is some basis for assuming that the physiological requirements for hospitalization are greater for the aged, the results suggest that technical rather than economic influences are becoming relatively more important in determining hospital utilization.

The impact of the urban-rural distribution appears primarily in a reduction in the admissions rate which more than offsets the significant, but quantitatively less important, positive association with the average length of stay.[11] In this case the negative association with admissions rate might suggest that the urban-rural distribution reflects proximity to medical care outside of the hospital. Heavily urban areas are likely to have a greater array of substitutes for hospitalization. The behavior of the urban-rural distribution seems to support this interpretation. Its positive association with the average length of stay might indicate that those who go to the hospital are sicker, since less ill patients utilize other sources of medical care, and therefore remain in the hospital longer. It should be noted that this interpretation differs considerably from the usual hypothesized relationship between hospital utilization and the urban-rural distribution. In general, it is held that in a more rural area there will be a higher length of stay since the length of time needed to reach the facility makes it unlikely that two short stays will be used when one single long stay may provide the desired medical services.

The results with respect to the education variable are quite different from those obtained in the original models and they provide an excellent demonstration of the advantages of limiting the interrelationships among the data. In most cases education has a positive association with the admissions rate and a negative association (insignificant in 1950) with the average length of stay. To the extent that the educational level measures the ability to perceive illness and to seek medical care, these findings are quite consistent. This would indicate that a higher level of education indicates a population group likely to seek care more frequently, and that the net result of this early medical treatment may be a reduction in the severity of illness, indicated by the reduced length of stay in 1960.

Another observation might be made with regard to the impact of the educational level variable. Some people have suggested that the increase in the admissions rate which stems from awareness and perception of the usefulness of early entrance into the hospital ought to be compensated for by diminution in the length of stay. The evidence gathered here indicates that the net effect of a high level of education, given other characteristics, is an increase in total utilization as measured by patient-days per thousand. This finding is consistent with what might be called an increasing propensity to consume care, and it has been suggested by at least one researcher in this area that constant exposure to medical attention will, in the long run, merely result in the finding of more disease rather than in an overall diminution in the amount of medical care that is required.

11. See Footnote 6.

Admissions Rate Influences Vs.Length of Stay Influences

One significant feature of this analysis is the demonstration of the differing impact of the various influences on the two components of hospital use. These influences were summarized rather briefly as the ability to pay or the economic circumstances, the existence of a family, the impact of the aged, the degree of crowdedness in housing, and finally the level of education. Two of these factors seem to have a highly significant positive association with the admissions rate. These factors, the proportion of the aged and the educational level, apparently represent, respectively, technical requirements for medical care and perceptions of the desirability of medical care. Two other factors, the existence of a family and the degree of crowdedness, each of which reflect the availability of a substitute for hospital utilization, operate mainly on the average length of stay. These factors, while not necessarily significantly affecting the likelihood that a patient will go to the hospital do affect the length of time he is likely to remain. The impact of the income variable, or the ability to pay measure, is apparent in both these aspects of hospital utilization, but significantly less so in 1960 than in 1950.[12]

In an effort to test the hypothesis that certain influences operate by affecting the admissions rate and that others operate by affecting the length of stay, one additional experiment was attempted. In the regression analysis that has been presented the order for introducing the variables was determined by the degree to which the principal component that they represented reflected the total variation of all twelve variables from which they were selected. Another set of relationships was derived statistically in an attempt to determine experimentally which variables showed up most strongly in the admissions rate equations and which variables showed up more strongly in the average length of stay equations. In each case the five variables were identical to those which had been chosen on the basis of the principal components analysis. However, the computation of these regression relationships was executed in such a way that each variable would be selected in order of its contribution to the

12. The insurance variable, however, is associated almost entirely, and positively, with the length of stay.

R^2. In one case, the dependent variable was admissions per thousand and in the other the dependent variable was the average length of stay.

The variables showing up most strongly in the average length of stay for 1950 were urban-rural proportion, marital status, and income, with education and age contributing least to the "prediction" ability of the relationship. For the admissions rate, on the other hand, the variable with the largest contribution to the R^2 was education, with the aged variable falling somewhere in the middle and the marital status variable being least contributory to the relationship.

In 1960, the results are similar but not so clear-cut. The degree of total prediction derived from the admissions rate relationship is particularly small, and very little meaningful information can be obtained from it. The educational variable made a greater contribution to estimations of the admissions rate than it did to estimates of the average length of stay. The aged variable was reflected similarly, but the marital status variable was most important in its contribution to explaining the average length of stay and least important in terms of explaining the admissions rate.

The empirical observations support the suggestion that economic constraints and the availability of substitutes, as we have defined them, affect utilization by affecting the average length of stay. Apparently, the characteristics of the population which reflect physiological requirements for hospitals, such as age distribution,or perception of the need for care, such as educational level, operate by affecting the admissions rate, and, once the admissions rate is set, the other characteristics determine the actual duration of stay within a significant range.

Some Implications

The early results justify some tentative observations that might be of some interest. One, while the economic factors have a demonstrably strong effect, they are clearly not overridingly important. Indications are that other factors are of considerable importance in determining utilization. Indeed, the economic circumstances demonstrate relatively little impact on the admissions rate as compared to other population characteristics.

A second point relates to the existence of substitutes for hospitalization. Two of the variables measure to some degree the availability of substitutes for hospitalization. It might be suggested that the marital status variable represents the existence of an individual to perform hospital services outside the hospital, and the population crowdedness variable represents the existence of physical facilities outside the hospital for receiving this care. The behavior of these variables gives additional encouragement to the notion that utilization of general hospitals can be reduced by the provision of certain hospital-type services in other kinds of facilities.

A third and final observation relates to the association between hospital utilization and the characteristics of the population as between 1950 and 1960. There are many indications that the degree of association between the five influences as represented by the five characteristics was significantly less in the later period. This strongly suggests that either the organizational characteristics of medical care or other influences within the population characteristics not included in the analysis have become more important. In some earlier discussions of organizational characteristics it was hypothesized that the physician was the determining factor. The results of the research presented here suggest that there might be a considerable measure of truth in this a priori observation, since our ability to predict is poorest when dealing with the admissions rate. In addition, the degree to which the substitute measures were of importance might indicate that measures of substitutes for hospital facilities which reflect the existence of other types of medical care facilities and which are subsumed under our category of organization of medical care also should provide a fruitful area for further research. It is hoped that the observations presented here prove a stimulus to the conduct of research on this aspect of utilization.

APPENDIX 1

CORRELATION MATRIX FOR INDEPENDENT VARIABLES

1960

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
X_1												
X_2	.78											
X_3	.75	-.91										
X_4	.52	.52	-.56									
X_5	.16	-.06	-.02	.44								
X_6	-.52	-.35	.37	-.64	-.66							
X_7	-.08	.20	-.20	-.39	-.32	.35						
X_8	.05	.12	-.16	-.37	-.31	.38	.79					
X_9	.68	.78	-.66	.37	-.03	-.47	-.02	-.19				
X_{10}	-.14	.53	-.52	-.07	-.19	.13	.55	.57	.40			
X_{11}	-.25	-.49	.66	-.44	-.46	.33	-.22	-.34	-.22	-.45		
X_{12}	-.57	-.49	.50	-.42	-.60	.78	-.19	-.09	-.51	-.35	.57	