Development of the Index of Medical Underservice

by the Health Services Research Group, Center for Health Systems Research and Analysis, University of Wisconsin

A mathematical model was developed to predict experts' relative assessments of scarcity of personal health services. This model provides, quickly and inexpensively, estimates of the relative assessments experts would make of any area in the country, in the form of an Index of Medical Underservice. The index is being used by the Bureau of Community Health Services in the preliminary designation of medically underserved areas for the federal HMO program.

Recent federal health programs have been aimed at promoting innovation in health care delivery and at securing minimum levels of health and health services for citizens. Consistent with these ends, Congress passed the Health Maintenance Organization Act (P.L. 93-222) in Dec. 1973 to support the development of health maintenance organizations (HMOs), a type of health delivery organization with widely publicized potential for providing high-quality, comprehensive, efficient health care.

As has been the case with most recent federal legislation intended to promote innovation and expansion of health care, the HMO act included provisions requiring that priority in funding be given to HMOs that would serve members of "medically underserved" populations. Specifically, the act provided that (1) within three months of its enactment, the Secretary of Health, Education, and Welfare was to report to Congress the criteria to be used in discriminating medically underserved from well-served areas and populations; (2) within 12 months, Congress was to receive a list of the areas and populations designated as underserved; and (3) priority was to be given to applications for federal HMO funding that claimed a plan to serve memberships 30 percent or more of which would come from medically underserved areas or populations. The act defined a medically underserved population as a population living in an area designated by the Secretary of Health, Education, and Welfare as having a shortage of personal health services.

Research supported by Grant No. HSA-105-74-138 from the Bureau of Community Health Services, U.S. Department of Health, Education, and Welfare.

Address communications and requests for reprints to Peter Adams, Center for Health Systems Research and Analysis, University of Wisconsin, 1225 Observatory Drive, Madison WI 53706.

Within DHEW, the Bureau of Community Health Services (BCHS) was assigned the responsibility for designating underserved areas. The University of Wisconsin Health Services Research Group (HSRG) had already been working with BCHS on the problem of designating such areas, and the two groups had both concluded that designation on the basis of anticipated improvements in the health status of the population (or other goals such as improved access or equity) was still beyond the state of the art. For example, even the most widely accepted measure of health care outcome, the function status index of Bush et al. [1], has been studied in relation to changes in the health delivery system only in a very limited context. Further, the time constraints imposed by the HMO act precluded the development of a functional model or index to relate the development and utilization of HMOs to changes in health status or any other complex objective.

During 1973, efforts of HSRG and BCHS to define the concept of medical underservice were frustrated by disagreements among health experts about the nature and sources of medical underservice. The informal observation that experts could agree in their *assessments* of relative medical underservice of actual communities, while simultaneously disagreeing in their *definitions* of medical underservice, led to the further observation that, if experts agreed in their assessments, then either they all used the same definition or each definition resulted in approximately the same conclusions about medical underservice. Since the HMO act provided essentially no restrictions on the designation of medically underserved areas and no requirement for an explicit definition, it was considered that experts' consensus assessments (if such could be proved to exist) would represent an acceptable practical standard for designation of medically underserved areas. To utilize this standard it would be necessary to develop some method that could predict experts' consensus assessments quickly, inexpensively, and on a common scale for any area in the country.

Work toward such a method led to the development of the current Index of Medical Underservice. The specific task of HSRG was to determine the validity of two assumptions underlying the proposed approach: (1) that experts from different disciplines and geographic areas tend to agree in their assessments of the relative scarcity of community health services and (2) that consensus assessments of the relative scarcity of health services can be predicted by a mathematical model using readily available data. The following sections of this article describe the efforts of HSRG to examine these assumptions—that is, to determine whether experts really would agree on their assessments and then to determine whether a mathematical model could be developed that could be used to predict expert assessments reliably enough that the model could be safely used to identify areas of health services scarcity.

Establishing the Existence of Consensus Assessments

The existence of a "natural" consensus without systematic differences in experts' perceptions of scarcity would obviate a decision about which experts' perspectives on scarcity are "correct." Further, the absence of systematic differences would make it less important that all or nearly all of the variance in assessments should be attributable to site differences: if variance not attributable to site differences were simply replication error, then it would not much matter whether 55 percent or 65 percent of the variance were attributable to site differences. The remaining variance would be noise and would be largely eliminated by averaging the assessments of several experts for any site.

To establish empirically the existence of consensus assessments, relative assessments of scarcity of personal health services were obtained from groups of experts in three states. First, a total of 62 communities in Michigan, Arizona, and Wisconsin were identified that seemed likely to be perceived as ranging from a great degree of scarcity of health services to a small degree of scarcity. Next, HSRG obtained the cooperation of local health experts in the same three states, who were for the most part staff members of state regional medical programs, comprehensive health planning agencies, and departments of health. These persons were selected because they were presumed to be familiar with the communities selected for the study. In individual interviews, they were requested to consider the communities (identified only by name and geographic boundaries) in their respective states. Then, using their first-hand experience and any data sources they wished, they were asked to rank the identified communities according to degree of scarcity of health services and to compare differences among the communities on a ratio basis, producing relative assessments of scarcity on an interval scale.

These initial assessments were part of a series obtained between Nov. 1973 and Oct. 1974. In all, seven panels of experts were asked to provide assessments of the relative scarcity of health services of a number of sites. The panels, which ranged in size from six to 16 people, were asked to evaluate from 13 to 31 sites. Thirty-three local health authorities participated. These local experts made assessments of communities, identified by name and geographic boundaries only, in their states. Another 24 experts, from 12 states, made assessments of some of these areas' profiles with four, seven, or nine variables but without place name identification. In all cases the experts made their assessments independently.

Thus assessments of relative scarcity of health services were obtained for 62 counties, towns, cities, and groups of census tracts in Arizona, Michigan, and Wisconsin. Under a variety of circumstances, a total of 57 experts provided a total of 1,662 assessments of scarcity of health services. Analysis of these assessments bore out the initial indications of existence of consensus in the degree necessary to support the HMO index.

The results of tests for consensus among members of all six panels are shown in Table 1. Two-way analysis of variance was used to estimate the proportions of the variation in the experts' assessments of sites that could be attributed respectively to differences among sites and differences among judges. The analysis showed that an average of 68 percent of the variation in site assessments was attributable to differences among sites and less than 8 per-

Table 1. Statistical Tests for Consensus Regarding Relative Scarcity of Health Services: Four Expert Panels Judging Communities within Their Own States

Test	Expert panel*						
	1	2	3	4	5	6	7
Mean assessment	53.6	48.4	54.3	54.3	57.8	56.2	56.7
ANALYSIS OF VARIANCE							
Proportion of the variance in assessments attributable to:							
Sites	0.749	0.327	0.573	0.577	0.758	0.733	0.736
Experts	0.054	0.055	0.056	0.077	0.097	0.044	0.021
Total explained (corrected R^2)	0.794	0.364	0.631	0.643	0.849	0.771	0.751
KENDALL'S COEFFICIENT OF CONCORDANCE							
Proportion of variance in rank sums accounted for (W)	0.787	0.385	0.721	0.692	0.792	0.695	0.754
Average rank order correlation across all pairs of experts in each panel	0.744	0.262†	0.693	0.658	0.766	0.673	0.736
MEAN STANDARD DEVIATION							
Average of the standard deviations about the mean assessments of the areas	12.4	25.8	16.5	20.4	11.8	14.1	14.6

* Panel 1 consisted of six experts from Wisconsin who assessed relative scarcity of health services in 18 Wisconsin counties; panel 2 consisted of eight experts from Michigan who assessed service scarcity in 13 areas (counties, towns, cities, and groups of census tracts) in Michigan; panel 3 consisted of 11 experts from Arizona who assessed 13 areas in Arizona; and panel 4 consisted of nine different experts from Arizona who assessed the same 13 Arizona areas.

Panel 5 was a group of nine experts from five states who assessed 22 Wisconsin counties described by 9-variable profiles without place name or other geographical identification. Panel 6 was composed of 15 experts from ten states who assessed 31 areas (towns, cities, and groups of census tracts) in Arizona, Michigan, and Wisconsin described by 4-variable profiles without place name or other geographical identification. Panel 7 was composed of the same 15 experts; they assessed the same 31 areas in Arizona, Michigan, and Wisconsin but used 7-variable profiles.

† Not significant at the 0.05 level of confidence.

cent to differences among experts. (In a separate test of data from three panels in which the members could be clearly grouped by disciplines, it was found that experts differed almost as much with members of their own discipline as with members of other disciplines.) In order to compare assessments on a rank-order basis, Kendall's coefficient of concordance (W) was computed for each panel. (This is also shown in Table 1.) The average W across panels was 0.689. Thus the proposition that experts are in substantial agreement in their assessments is supported by both the parametric and nonparametric measures. In Arizona, where two different groups of local experts provided assessments of the same 13 sites, the groups' mean assessments for each site were correlated at 0.92 (product-moment) and 0.88 (Spearman rank-order), again supporting the consensus assumption.

The assessments made by the panel of Michigan local experts (Table 1, col. 2) at first appeared to contradict the consensus assumption. However, further analysis revealed that these experts did agree about the relative scarcity of services in the eight sites (towns, cities, and counties) outside of Detroit for which they provided assessments. An analysis of variance for these eight sites attributed almost 64 percent of the variance in the assessments to differences among sites, with no variance attributed to judge differences. The coefficient of concordance for assessments of the eight sites was 0.642. Thus the Michigan data were in part consistent with the consensus assumption. However, they indicated that consensus may not exist for certain sections of large metropolitan areas, as will be discussed under the heading "Limitations."

Developing Models to Predict Expert Assessments

The most direct method for the development of a model would have been to obtain assessments for a large number of randomly selected or representative communities throughout the nation from a group of randomly selected health experts and then to develop a model by regressing the mean assessments on available socioeconomic, geographic, health service, and health status data for the communities assessed. Developing a model in this manner was, like the cause-and-effect model discussed earlier, infeasible in terms of time and money. (HSRG is currently attempting to construct a regression model using subjective and empirical data from 14 states. If these efforts succeed, this new model may replace the current model in the federal HMO program.)

HSRG therefore decided to develop a self-explicated multiattribute utility (MAU) model [2]. Such a model differs from mathematical regression models or other statistical techniques in being prescriptive rather than descriptive. Data are weighted and combined according to rules prescribed by informed judgment as to what will yield a useful outcome. Therefore the robustness of the model is not necessarily limited, as that of a regression model would be, by the relatively small and nonrandom sample of sites.

The objective of an MAU model is to compute a single index number based on values for a number of the variables commonly used to describe the phenomenon in question. To develop an MAU model, respondents are first ranked to select a small subset of commonly used variables such that the subset would be the most useful combination for assessing the phenomenon if no other information were available. It is assumed that the variables selected are not necessarily equally useful as indicators, so the respondents are asked to provide estimates of their relative usefulness. If the variables are measured on different scales, the respondents are asked to convert all variable measurements to a common scale through a process called utility estimation [3]. In making the utility estimates, the respondents consider each variable independently and select the raw scores for each variable that represent the most and least desirable points; a utility value of 100 is assigned to the most desirable level and a utility value of zero is assigned to the least desirable level. These points are plotted on utility graphs, and the respondents are asked

Table 2. Variables and Weights Selected by the Experts at the Nov. 1973 Conference

Variable	Weight*
Practicing physician equivalents per 1000 population	100.0
Infant mortality rate	91.8
Preventable deaths as percentage of all deaths	82.2
Percentage of population age 65 and over	70.9
Percentage of population with incomes below poverty level	69.1
Average travel time to regular source of primary care	67.9
Per capita expenditures on personal health care	59.9
Average travel time to emergency care	58.5
General acute hospital beds per 1000 population	48.3

* Each average weight was multiplied by the same constant to inflate the highest-rated variable to 100.

to establish intermediate values for the variables by drawing utility curves connecting the extreme points. A composite score is obtained by converting the raw values for each variable to a common scale by means of the utility curves, weighting each utility value by the estimated relative usefulness of its variable, and summing the weighted scores for all variables.

A panel of nine "national" experts from five states was selected to generate the MAU model. The panel consisted of three practicing physicians, two physicians teaching in medical schools, a physician administrator, a professor of economics, and two professors of health administration. (This panel did not include any of the local experts who made the initial assessments that were tested for consensus.) These experts first identified variables that they considered to be useful indicators of relative scarcity of health services. This was done initially using a DELPHI procedure through mailings and finally at a conference held in Nov. 1973. After the number of indicators had been reduced to a manageable number, the experts were asked to rank them and then rate them on a ratio basis according to importance. Nine variables (shown with their weights in Table 2) were chosen as most useful and most likely to be represented by available data.

The rankings (i.e., weights) and utility curves for the variables were refined in the context of judging relative service scarcity among 30 actual areas, each represented by a coded profile sheet showing that area's data on the nine variables. Twenty-two Wisconsin counties were included, with eight duplications to provide an estimate of replication error. The experts were told that the areas described were actual Wisconsin counties and were provided with the overall data range for each variable. First they were asked to review qualitatively all 30 profiles in terms of scarcity of health services. Second, they were asked to place the areas (not the variables) in rank order from the least to greatest degree of scarcity. Third, they were asked to consider the judgments they had made of the areas and to rate each area's scarcity relative to other areas, producing subjective interval scale assessments with the site having least scarcity anchored at 100 and the site having greatest scarcity anchored at 0.

After the experts made these "global" site assessments, using all nine variables simultaneously to make judgments of scarcity of health services, they were asked to consider each of the nine variables separately. Specifically, they were asked to view each variable as an independent indicator of scarcity of health services and to rank the variables according to their usefulness in assessing scarcity. Experts made these initial rankings independently and then discussed differences among their rankings. After this discussion they again individually ranked the variables and rated them on a ratio basis according to usefulness as indicators of health services scarcity. The experts, using a similar procedure, were finally directed to draw utility curves for each of the nine variables as discussed earlier. By averaging variable weights and utility curves across experts for each of the nine variables, an aggregate MAU model was obtained. The overall scarcity of services in an area could then be estimated as the sum of the weighted utility values for the nine variables. Table 3 shows how the health services scarcity score was calculated for a community profile on the basis of the nine-variable MAU model.

Table 3. Aggregating the Scarcity Score for Community T Using the Rules of the Nine-Variable MAU Model

Variable	Actual value (from community profile)	Utility value (from utility curve)	×	Variable weight*	=	Weighted score for variable
Physicians per 1000 pop.	0.62	34		0.154		5.27
Infant mortality rate × 1000 . .	24.7	32		0.134		4.29
Preventable death rate	11.7	44		0.126		5.54
Percent pop. age 65 and over . . .	16.8	50		0.106		5.30
Percent pop. below poverty level	21.1	57		0.111		6.33
Travel time to primary care (min)	26	75		0.105		7.88
Per capita health care expenditures ($)	79	45		0.098		4.41
Travel time to emergency care (min)	30	70		0.090		6.30
Hospital beds per 1000 pop. . . .	3.0	50		0.076		3.80
				Index value		49.12

* Weights normalized to sum to 1.

During the initial evaluation of the results of the Nov. 1973 conference it became apparent that the nine-variable model would not meet the operational needs of BCHS. BCHS officials did not want to use an index that would force HMO applicants and local health planners to undertake major data collection efforts if an index using only available data would serve as well. Further, it became apparent that at least two of the original nine variables, preventable death rate and per captia health expenditures, were not collectable in a reliable manner. Therefore a second conference was held in Apr. 1974 to develop an index function using only the variables for which data were readily available and to test the consensus assessment process with more diverse sites

Table 4. Models Developed at the Nov. 1973 and Apr. 1974 Conferences

Model	Physicians per 1000	Percent pop. below poverty level	Infant mortality rate	Percent pop. age 65 and over	Travel time to emergency care	Travel time to primary care	Hospital beds per 1000 pop.
MEAN MAU WEIGHTS—NOV. 1973							
7-variable model							
Normalized weight*	19.7	13.7	18.1	14.0	11.5	13.4	9.5
Standard error†	8.2	6.2	5.6	4.7	6.2	7.3	7.3
4-variable model							
Normalized weight*	30.1	20.8	27.7	21.4			
Standard error†	12.5	9.4	8.5	7.2			
MEAN MAU WEIGHTS—APR. 1974							
7-variable model							
Normalized weight*	20.3	17.7	18.8	13.4	12.2	12.1	5.3
Standard error†	2.7	7.2	3.8	6.2	4.6	4.2	6.0
4-variable model							
Normalized weight*	28.7	25.1	26.0	20.2			
Standard error†	4.9	9.3	7.2	9.3			
MEAN REGRESSION MODELS—NOV. 1973							
7-variable model							
Raw coefficient	11.40	–0.72	–1.36	–0.74	–1.17	0.24	1.38
Normalized‡ weight	22.27	14.38	20.51	6.37	23.18	3.30	9.99
Standard error†	3.26	3.00	4.07	3.36	4.16	4.54	3.11
4-variable model							
Raw coefficient	17.73	–1.43	–1.14	–1.02			
Normalized‡ weight	38.82	31.97	19.37	9.84			
Standard error†	4.80	4.25	4.42	4.44			
MEAN REGRESSION MODELS—APR. 1974							
7-variable model							
Raw coefficient	47.22	–1.18	–0.59	–0.01	0.20	0.20	–0.47
Normalized‡ weight	33.21	36.79	10.97	0.12	8.12	7.92	2.86
Standard error†	9.47	7.79	8.55	8.28	8.93	9.90	7.91
4-variable model							
Raw coefficient	58.37	–0.48	–0.78	–0.10			
Normalized‡ weight	57.35	21.01	20.20	1.43			
Standard error†	6.98	7.00	7.25	6.29			

* MAU weights were normalized to sum to 100 while preserving the ratios of the weights.

† All standard errors are of normalized weights.

‡ Regression coefficients were first normalized to reflect the variance of the raw variables, where $B_i = b_i(S_i/S_y)$ and b_i is the raw coefficient. Then the regression coefficients were renormalized to sum to 100 while preserving the ratios of the weights.

and experts. The Apr. 1974 conference involved 15 health experts from ten states, including three practicing physicians, five DHEW officials, two consumer advocates, two state/local health planners, and three academic researchers. These experts were asked to assess 31 urban and rural towns, counties, cities, and groups of census tracts from Michigan (13 sites), Arizona (12 sites), and Wisconsin (six sites). The communities were described, as before, by profile sheets. At this conference the experts were asked to assess the 31 communities, first with four-variable descriptions and then with seven-variable descriptions. (The four variables were physicians per 1,000 population, percentage of the population below poverty level, percent of the population age 65 and over, and infant mortality rate; the seven variables were the basic four plus travel time to emergency care, travel time to primary care, and beds per 1,000 population.) The 15 experts were then asked to provide the weights and utility curves needed for the four-variable and seven-variable MAU models by the same procedures used at the first conference.

Validation of the Models

MAU models were constructed based on computational rules explicated as has been described, by two separate panels of experts. In addition, these panels' assessments of profiles for sites in Michigan, Wisconsin, and Arizona served as the basis for regression models using different subsets of independent variables in the regression equations. In all, then, eight models were constructed, two MAU models and two regression models, each for a set of four variables and a set of seven variables for which data were thought to be available nationwide (see Table 4, p. 175).

To show that the models could predict local experts' assessments of relative scarcity of health services (assumption two), the ideal approach would have been to obtain a nationwide random sample of sites and ask all experts familiar with those sites to provide assessments. But local health experts cannot be expected to be familiar with sites outside their own states. Therefore, the models' predictions could only be evaluated against mean local assessments calculated within each state. Time and budget constraints allowed such validation tests in three states. (Work in progress will provide assessments in more than 15 states.) In each state, the mean assessment provided by *selected* local experts was used as an estimate of the mean assessment that *all* relevant local experts would have provided for the sites considered.

The ability of each model to predict the mean assessments of sites made by four groups of local experts is shown by the correlation results in Table 5. The models were all able to account for approximately 60 percent of the variance in local experts' mean assessments; all correlations were significant at the 0.05 confidence level, except for the four-variable regression model compared to the rank ordering of sites assessed by Michigan experts.

The assumption that the MAU models can properly place local relative assessments within different states on a common national scale is supported in two ways. First, two independent groups of experts provided essentially identical computational rules for construction of MAU models (see Table 5 and figure on p. 178). Second, these same two groups of experts made assessments of a set of sites drawn from three states, identified only by profiles containing the variables used in the models. Significant correlations were observed between these global assessments, the assessments predicted

Table 5. Ability of the Mathematical Models to Predict Experts' Mean Assessments of Scarcity of Health Services

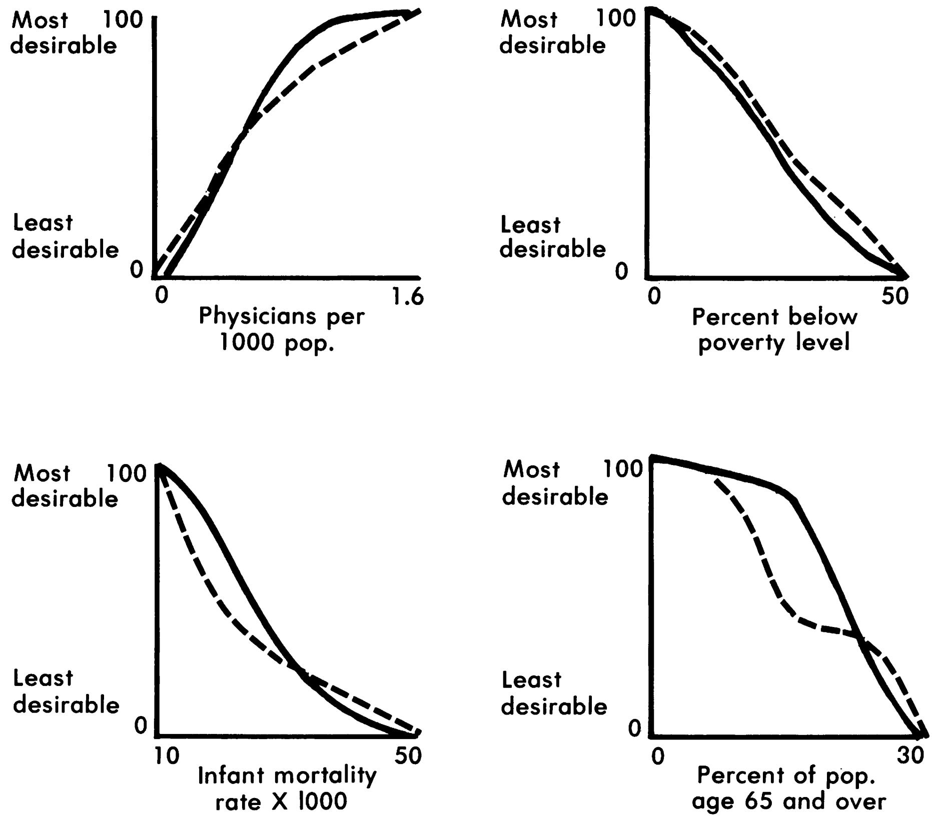
Panel, profile, and areas assessed	4-variable MAU model		7-variable MAU model		4-variable regression model		7-variable regression model	
	r*	R†	r*	R†	r*	R†	r*	R†
MODELS DEVELOPED AT THE NOV. 1973 CONFERENCE								
9 "national" experts using 9-variable profiles of 22 areas	0.92	0.93	0.92	0.89	0.97	0.97	0.99	0.99
6 WI experts assessing 18 WI counties	0.80	0.78	0.78	0.73	0.93	0.96	0.91	0.85
MODELS DEVELOPED AT THE APR. 1974 CONFERENCE								
15 "national" experts using:								
4-variable profiles of 31 areas	0.92	0.91	0.88	0.87	0.95	0.99	0.91	0.86
7-variable profiles of 31 areas	0.96	0.92	0.94	0.92	0.87	0.85	0.90	0.86
11 AR experts assessing 13 AR areas	0.82	0.86	0.83	0.85	0.78	0.69	0.79	0.69
9 AR experts assessing 13 AR areas	0.80	0.78	0.80	0.87	0.74	0.72	0.77	0.63
8 MI experts assessing 13 MI areas	0.58	0.59	0.60	0.62	0.60	0.37§	0.57	0.58
6 WI experts assessing 18 WI counties	0.86	0.84	0.87	0.85	0.82	0.74	0.88	0.87

* Pearson's product-moment correlation coefficient.
† Spearman's rank order correlation coefficient.
§ Not significant at the 0.05 confidence level.

by the mathematical models, and the assessments of local experts. In combination, these results support the use of the MAU models to provide a common national scale for indexing relative scarcity of health services.

Limitations of the MAU Models in Measuring Medical Underservice

Validation work reported here indicates that the use of an MAU model is a reasonable means to meet the designation requirements of the HMO act. However, it is important to note three limitations in the methodology. The first limitation to the generality of the results is due to nonrandom selection of sites and experts to validate the models. The fact that some experts declined to make assessments may have biased the results. For example, true consumers



Mean utility curves produced at the Nov. 1973 conference (broken line) and the Apr. 1974 conference (solid line).

(as opposed to consumer-advocates) probably would not have much knowledge about scarcity of health services outside their own neighborhoods and therefore probably would not be able to make relative assessments of scarcity for ten counties or 15 urban health areas. Similarly, selection of the 62 study sites may have affected the estimates of experts' consensus and the models' predictive abilities. The study sites used do not constitute a random sample of all 3,141 counties in the United States, and some statistical evidence suggests that the sites are not representative. Therefore, if the principal concern were to compare the health services scarcity of counties, the correlations calculated to estimate the extent of consensus and predictive ability may be artificially inflated [4].

In the same way, the nonrandom selection of sites may have seriously undermined the reliability of the regression models. Unlike the prescriptive MAU models, regression models are largely a function of the observations used in their construction. The differences in the signs and magnitudes of the coefficients of the regression models—developed with two different groups of sites and experts—raises questions about which of the regression models in Table 4 would be appropriate for the entire country. That variations in the regression model parameters may have been due to the idiosyncracies of the sites is reflected by the fact that MAU models developed independently by two different groups of experts were essentially identical although the corresponding regression models differed. Fortunately, the four-variable MAU model was at least as consistent and precise as any of the regression models. Although some researchers have found the predictive ability of linear models (of which both MAU and regression models are examples) to be largely insensitive to variation in variable weights, the instability of the regression parameter estimates led to the adoption of the four-variable MAU model as the index to be used throughout the country [5]. Another factor supporting selection of the four-variable MAU model (apart from its relative predictive power) over seven-variable models is that the four variables are available for all 3,141 counties in the country and for many subcounty areas as well.

The second limitation, and perhaps more significant, on the use of the models as an index of relative scarcity is that expert consensus, a critical building block in this methodology, appears to be less strong in some large metropolitan areas. The lack of consensus regarding groups of census tracts in Detroit was noted earlier. Considerable agreement was demonstrated by experts in their assessments of whole towns, counties, cities, and for sections of Phoenix and Tucson, but comments made independently by Michigan experts indicated the possibility of substantive disagreements in assumptions underlying their assessments of sections of metropolitan Detroit.

To further study this phenomenon, HSRG asked 11 local health experts in New York City to provide estimates for a number of health constructs, including scarcity of health services, for 15 of New York City's 33 comprehensive health planning districts. For scarcity assessments, an analysis of variance attributed 44.0 percent of the variance to differences among sites and none to differences among experts. Kendall's coefficient of concordance was 0.49. If the unexplained variance were noise, and not principled differences in perceptions, such results would not indicate a significant limitation on the use of experts' mean assessments, for the reasons given earlier. However, the fact that several of the New York experts refused to make assessments of where additional primary care physicians would most improve health status, arguing that additional physicians by themselves would not improve health status, led to concerns that a portion of the unexplained variance was site-expert interaction. Unfortunately, with only one scarcity assessment per site per expert, it was not possible to obtain separate estimates of replication error and site-expert interaction. (Efforts are presently under way to obtain scarcity assessments and replications in six large metropolitan areas.)

Another limitation on the methodology is the impossibility of evaluation. Although it may be possible to validate the Medical Underservice Index, it will be extremely difficult to evaluate whether additional personal health services did produce more benefits at sites designated as underserved than at undesignated sites. This is because the objectives or benefits were never made explicit. Evaluation of the Index of Medical Underservice can only be done in terms of a standard established apart from the current index.

If one accepts a natural consensus (as opposed to a forced consensus) of judgment of scarcity of health services among experts as an acceptable standard of comparison, the statistical results obtained strongly support the ability of the predictive models described here to meet the goal of ranking medical underservice in any area of the country on the basis of a nationwide standard. The predictive models allow an interval or rank-order comparison of any areas for which the data have been collected. Competing areas could be given priority based on their predicted scores, or a number of areas around the country could be judged medically underserved independently of the models and all those areas with scores as low as or lower than the scores of these underserved areas could be designated as underserved. With the error theory developed for the predictive models, it is possible to estimate false positive and false negative rates for these designation strategies.

References

1. Bush, J. W., W. R. Blischke, and C. C. Berry. Health indices, outcomes and the quality of medical care. Paper presented at Conference on Evaluation in Health Services Delivery, Berwyck, ME, Aug. 1973.
2. Huber, G. P. MAU models. *Manage Sci* 20:1393 June 1974.
3. Kneppreth, N. P., D. H. Gustafson, R. P. Leiffer, and E. M. Johnson. *Techniques for the Assessment of Worth.* Technical Paper No. 254. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, 1974.
4. Warren, W. G. Correlation or regression: Bias or prediction. *Appl Statis* 20(2):148, 1971.
5. Dawes, R. M. and B. Corrigan. Linear models in decision making. *Psychol Bull* 81:95 Feb. 1974.