TESTING FOR VARIATION IN SOCIAL GROUP PREFERENCES FOR FUNCTION LEVELS OF A HEALTH INDEX

C.C. Berry, J.W. Bush, R.M. Kaplan

Division of Health Policy, Department of Community Medicine,
University of California, San Diego

Research supported by Grant No. 2 R18 HS00702, National Center for Health Services Research, USDHEW.

Abstract

Social preferences for Levels of Function represent the value dimension of health status. If substantial differences in preferences exist between social groups, they may pose aggregation problems in creating a general Health Status Index. 806 respondents, representing all socioeconomic groups in a household interview survey, each rated the relative desirability of a set of case descriptions on a scale anchored with "0" for death and "10" for completely well. There were eight sets of case descriptions totalling 343 sampled from a domain of possible health states. Using individual respondents as the units of analysis and demographic attributes as test characteristics, the major canonical analyses revealed no significant differences among social groups for the values associated with health states.

The recurrent need for a comprehensive social indicator for health planning and evaluation has led researchers to again attack the problems of creating a Health Status Index [Sullivan 1966; Torrance 1973; Berg 1973; Goldsmith 1973]. Our research group has undertaken a series of studies to define the problems more precisely and to propose approaches that avoid some of the earlier criticisms [Bush et al. 1972; Patrick et al. 1973a,b; Berry and Bush 1974; Chen et al. 1975; Bush et al. 1975].

This paper addresses a major problem in the construction of a Health Index -- the possibility that different social groups may have different preferences for health states or levels of function. Before examining this problem in detail, however, we will present a general conceptual framework for health indices.

The Health Index

The social construct "health" is composed of two distinct components: Level of Well-being and prognosis. Levels of Well-being are the measurable preferences or weights that society associates with point-in-time Levels of Function. Prognoses, the probabilities of transition among the Levels over time, are implicit in our concept of health and decision-making [Bush et al. 1972], and can be determined in follow-up studies of different patient and population groups. Distinguishing these two components of health status permits separate measurement and explicit combination of the two sets of variables.

Measuring Levels of Well-being requires standardized descriptions of states of function. Using items from standard survey instruments and disability classifications, we constructed three scales with mutually exclusive and collectively exhaustive steps to cover the spectrum of objective disturbances that diseases and disabilities cause in role performance [Exhibit 1, Patrick et al. 1973a]. Differences in Function Levels, defined as possible combinations of steps of the three basic scales, occur because of physical disabilities, symptoms, sensory disturbances, mental retardation, and mental illness. Most of these disturbances were categorized in a set of 35 symptom/problem complexes. When social preferences for each Function Level have been derived from ratings of case descriptions, a health index that incorporates prognoses can be expressed as the product of the preferences and the expected duration of stay in each Level over the life expectancy [Bush et al. 1972; Chen et al. 1975]. This computation produces a weighted or value-adjusted life expectancy in equivalents of completely Well-years:

$$E = \sum_{j=1}^{42} W_j Y_j$$

where j is the index for Function Levels [$j = 1, 2, \ldots, 42$],

E is the value-adjusted or weighted life expectancy in Well-years,

W_j are the weights or social preferences measured for each Function Level, and

Y_j are the expected durations in each Function Level computed from the transition probabilities [Bush et al. 1971].

A major question for Health Index research is whether social groups differ significantly on the value dimension of health, that is, on the preferences or Levels of Well-being (W_j) that they assign to Levels of Function on a continuum from death (0.0) to optimum function (1.0).

Method

With collaboration from the University of Michigan Survey Research Center, we fielded a household interview survey of the San Diego metropolitan area in the spring of 1974. A major portion of the interview time was devoted to measuring the preferences that community members associate with each of the Function Levels.

From a matrix representing all possible combinations of 42 Function Levels, 4 age groups and 35

symptom/problem complexes, we had selected a stratified random sample of 343 case descriptions (items) and divided them among eight sets of computer generated booklets. Each page in the booklets described the age group, the Mobility, Physical Activity, and Social Activity steps, and a Symptom or Problem in the day of a potentially observable person [Exhibit 2]. All respondents were assigned randomly to one of the eight booklets, creating eight subgroups of approximately 100 respondents each.

Each booklet, containing 56 case descriptions, began with the same eleven warm-up items, representing the full spectrum of Function Levels. Four additional items, also chosen to represent a spectrum of scale steps, were placed in all eight booklets to test the reliability of the rating procedure between groups. The bulk of each booklet consisted of one of the eight sets of 41 "regular" case descriptions. To avoid possible spurious order effects, all the case descriptions except the warm-ups were placed in a computer generated random sequence unique for each respondent. Each respondent was instructed to rate the relative desirability of the cases on an equal interval category scale anchored at "0" for dead and "10" for the well state.

The 867 respondents were a probability sample of the San Diego metropolitan area, including 29 blacks and 52 of Spanish heritage. Forty-five respondents refused to perform the rating procedure, and the responses of an additional 16 subjects were eliminated because of gross confusion or lack of cooperation. No respondents were eliminated because their values differed from other respondents. The ratings from one respondent, who used "0" for well and "10" for dead, were inverted and retained in the data set.

Results

Each respondent's ratings of the case descriptions were related to his own attributes using a series of canonical correlations. The individual respondent, who serves as the unit in all analyses, was described by fourteen socioeconomic characteristics [Exhibit 3]. Three different sets of preference ratings served as the other sets of variables in different canonical analyses.

The most direct analyses were performed for each of the eight sets of 41 regular case descriptions that were given to random subsamples of the study population. Since each randomly created subgroup rated a different set of items, the separate analyses provide eight independent tests of this relationship. With approximately 100 respondents in each subgroup, no statistically significant canonical correlates were observed.

The second analysis used the fifteen warm-ups and reliability cases rated by all respondents. The first canonical root (R = .36) was significant at the .01 level, with by far the largest weight for attributes (-.71) going to age.

A third set of analyses created one variable on the left for each step of the Age, Mobility, Physical Activity, and Social Acitivity scales used in the case descriptions. The value of each variable was the mean of each respondent's ratings of the ten to fifteen items (from his total set of 56) that contained that scale step or age group. Since the same scale-steps and age groups made up all items, using means as a set of variables permitted pooling the data from all respondents (N = 806) for several analyses. One analysis used all 18 such variables and four other analyses used the steps of each scale separately. Exhibit 4 displays the largest canonical correlation from each analysis. All were non-significant.

Discussion

Our previous research indicates that category rating of case descriptions given results equivalent to magnitude estimation and procedures simulating social choice [Patrick et al. 1973b]. In addition, the procedure is easily understood, efficient, and more reliable than the other psychometric methods. Therefore, category rating served as the scaling method for household interviews. Eventually Function Level weights (W_j) will be derived across the individual ratings of the case descriptions, using appropriate statistical models [Chen et al. 1973]. In these analyses, however, we used individual items and means for each respondent to test for social group differences.

The first eight canonical analyses are consistent with the hypothesis of no substantial differences between social groups, but they lack power because of the large number of parameters estimated and the low number of respondents (about 100) in each experimental group. Although the analyses cannot be accepted as definitive, they are important because they are based on individual case descriptions which would be sensitive to substantial correlations between social group attributes and interactions among the scale steps in each item. The absence of significant relationships, even without adjusting alpha levels for multiple comparisons, lends support to the general hypothesis that social groups do not differ systematically in preferences for states of function.

The significant result in the second analysis must be viewed with some suspicion. During the warm-up items -- eleven of the fifteen in this analysis -- the respondents were still learning the rating procedure. The observed correlation is low; the canonical variate for social group attributes explains only 13% of the variance in the canonical variate for the 15 items. Since the canonical variate for items represents only a portion of the item variance, the differences in item ratings attributable to social group characteristics must be very small.

Furthermore, no clear pattern emerges among the coefficients. One possibility considered was whether the single significant root could be attributed to the well-known error of central tendency in which respondents who are more likely to become confused by such procedures place extreme items near the center of the scale

[Guilford 1954]. But several non-extreme items and social group attributes not supportive of the hypothesis also had large weights. Although the central tendency hypothesis cannot be ruled out, acceptance would require a clearer demonstration of its effect across a larger number of cases.

The third set of analyses is the most nearly definitive and therefore the most important. By using the means for each scale step unique to each respondent, we tested directly whether social groups differ systematically in the ratings given to the steps of the Function Levels, that is, to the major components of a Health Status Index. In addition, the total data set (806 respondents) is used to estimate a small number of parameters (from 4 to 18) giving increased power.

Since each respondent rated each of the 35 symptom/problem complexes only once or twice in his set of 56, using the means for symptom/problem complex ratings did not offer a stable test of differences. Symptom/problem effects on the ratings are therefore confounded with the scale step means in this analysis. If a significant effect had occurred that might be attributed to a symptom or problem, the confounding would be more troublesome. Since each scale step was combined with many different symptoms and problems, however, systematic suppression of an effect by the confounding seems unlikely. We are, however, planning analyses to help resolve this remaining question.

Therefore, with a large number of respondents rating a wide array of health states and using the most efficient and sensitive analyses yet brought to bear on the question, we conclude that social groups do not differ significantly or systematically in their ratings of the Function Level steps required to construct a Health Index. This lack of differences should not be surprising. The norms that govern the rating of health states are terminal rather than instrumental preferences, an important distinction made by Rokeach between two classes of human values [1973]. Instrumental values refer to modes of conduct and terminal values refer to the relative desirability of states of existence -- means versus ends. Although terminal values may vary from person to person, the large variations commonly observed among social groups are usually due to differences in beliefs, life styles, political preferences, or other instrumental values.

Consensus has been documented for terminal values such as the ordering and differences in the seriousness of crimes rated on interval scales [Rossi et al. 1974; Sellin and Wolfgang 1964] despite differences of opinion about police or sentencing policies. Rahe has demonstrated a similar consensus on stress associated with life events [Rahe 1969]. Our analyses indicate that the dysfunctions imposed by illnesses and injuries are also equivalently undesirable among different social groups, although the groups may have widely varying modes of treating, responding to, or coping with the disorders.

With a Health Index, differences in the means used (programs, policies) become questions of the relative effectiveness of programs. Once a consensus on the terminal values of a Health Index is established, the expected output of a program depends only on the effect of the program (means used) on prognoses, which are empirical questions for evaluative research [Chen et al. 1975].

The importance of consensus valuations for a Health Index varies according to its proposed uses. As an outcome measure for evaluative research and as a social indicator, the convenience and simplicity of a measure standardized across all social groups probably outweighs the importance of small differences in the weights for an Index of Well-being. Further analysis will attempt to assess the magnitude and importance of possible differences.

As criterion for social decision-making, on the other hand, theoretical problems do exist with aggregating individual and therefore social group differences. Arrow [1963] demonstrated that social preference orderings cannot be logically derived from differing individual preferences without violating one or more of a reasonable set of democratic and rationality assumptions. Given this impossibility, demonstrating that individual differences do not vary systematically across identifiable social groups increases confidence in the basic fairness of the decision criterion, especially when all groups are represented in the measurement.

Even if small differences were detectable in the preferences with large numbers of subjects, highly refined measurements, and precise statistical methods, the importance of the differences for social choice should not be exaggerated. Given the instrumental values in all political decisions and the current range of error in assessing program effectiveness, not developing a practical Health Index with a single set of terminal values because of preoccupation with small social group differences could be a serious error of misplaced precision. Although individual or social group differences on particular cases cannot be ruled out, the analysis above indicates that if social group differences exist, they are very small, especially when compared to the precision for measuring other variables in existing decision and policy processes.

Conclusion

Canonical analysis of data from a probability sample of 806 respondents in a large metropolitan area demonstrates consensus among all major population subgroups on the terminal values associated with restriction in Mobility, Physical Activity, and Social Activity. This empirical finding refutes widely held assumptions and frequently voiced objections to incorporating objective measures of social values in Health Status Indices.

The authors gratefully acknowledge helpful comments from W.R. Blischke, Ph.D., on the analysis and earlier versions of this paper.

Exhibit 1: Scales and Steps for the Classification of Function Levels*

MOBILITY SCALE

5	Drove car and used bus or train without help
4	Unable to drive or needed help to use bus or train
3	In house
2	In hospital
1	In special care unit
0	Death

PHYSICAL ACTIVITY SCALE

4	Walked without physical problems
3	Walked with physical limitations
2	Moved own wheelchair without help
1	In bed or chair
0	Death

SOCIAL ACTIVITY SCALE

5	Did work, school or housework and other activities
4	Did work, school or housework, but other activities limited
3	Limited in amount or kind of work, school or housework
2	Performed self-care but not work, school or housework
1	Needed help with self-care activities
0	Death

*Definitions and sources of scale items available from authors.

Exhibit 2: Example Warm-up Case Descriptions and Their Ratings

	Rating	
	Mean	Median
Adult (18-64 years) In house Walked without physical problems Performed self-care, but not work, school or housework Sick or upset stomach, vomiting, or diarrhea (watery bowel movements)	.636 N = 804	.645
Older adult (65 years and over) Drove car and used bus or train without help Walked without physical problems Worked or did housework and other activities No symptom or problem	.939 N = 805	.982
Small child (below 6 years) In special care unit In bed or chair Needed more help with self-care than usual for age Loss of consciousness such as seizures (fits), fainting, or coma (out cold or knocked out)	.293 N = 804	.262

Exhibit 4: Canonical Correlations (with Statistical Significance) of Respondent Socioeconomic Attributes with Respondent Means for Scale Steps, by Scales Included in Each Analysis (N=806).

Scale Step Means Used	Largest Canonical Correlation	Statistical Significance
All Scales (18 Steps)	.31	($p > .4$)
Age Groups (4 Steps)	.16	($p > .5$)
Mobility Steps (5)	.19	($p > .5$)
Physical Activity Steps (4)	.20	($p > .5$)
Social Activity Steps (5)	.21	($p > .5$)

Exhibit 3: Coding of Respondent Social Group Attributes for Canonical Correlations

Variable	Units, Code, or Reference Group
Age	Years
Sex	Male = 0, Female = 1
Education	Years
Marital Status	Married = 1, Single, Divorced, Widowed, Separated = 0
Annual Personal Earnings	Dollars
Annual Family Income	Dollars
Grew up in West (=1)	0 = Did not grow up in West
Grew up in North (=1)*	0 = Did not grow up in North
Size of Place of Residence Before Age 16	1 = Farm, 2 = Small Town, 3 = Suburb of Large City, 4 = Large City (Over 250,000).
Black (=1)	0 = Non-black
Spanish (=1)	0 = Non-Spanish
Democrat (=1)	0 = Non-Democrat
Protestant (=1)	0 = Non-Protestant
Catholic (=1)	0 = Non-Catholic

* Includes Census Bureau Northeast and North Central Regions.

REFERENCES

Arrow KJ: Social Choice and Individual Values. New Haven, Yale University press, 1963.

Berg RL, ed: Health Status Indexes. Chicago, Hospital Research and Educational Trust, 1973.

Berry CC and Bush JW: Maintaining health in a defined population using a health status index. Presented before the Pacific Division, American Association for the Advancement of Science, Irvine, CA, June 1974.

Bush JW, Blischke WR and Berry CC: Health indices, outcomes, and quality of medical care. In Yaffe R and Zalkind D, eds: Evaluation in Health Services Delivery. Engineering Foundation, New York, 1975.

Bush JW, Chen MM, Patrick DL and Blischke WR: Linear models of social preferences for constructing a health status index. Presented before the Operations Research Society of America, San Diego, CA, November 1973.

Bush JW, Chen MM and Zaremba J: Estimating health program outcomes using a Markov equilibrium analysis of disease development. Am J Public Health 61:2362-2375, December 1971.

Bush JW, Fanshel S and Chen MM: Analysis of a tuberculin testing program using a health status index. J of Socio-Economic and Planning Sciences 6:49-68, February 1972.

Chen MM, Bush JW and Patrick DL: Social indicators for health planning and policy analysis. Policy Sciences 6:71-89, March 1975.

Chen MM, Bush JW and Zaremba J: A critical analysis of effectiveness measures for operations research in health services. In Schuman L, Young J and Speas D, eds: Operations Research in Health -- A Critical Analysis. Baltimore, Johns Hopkins University Press, In Press.

Fanshel S and Bush JW: A health status index and its application to health-services outcomes. Oper Res 18:1021-1066, 1970.

Goldsmith SB: A reevaluation of health status indicators. Health Serv Reports 88: 937-941, 1973.

Guilford JP: Psychometric Methods. New York, McGraw-Hill, 1954.

Nunnally JC: Psychometric Theory. New York, McGraw-Hill, 1967.

Patrick DL, Bush JW and Chen MM: Methods for measuring levels of well-being for a health status index. Health Serv Res 8:228-245, Fall 1973b.

Patrick DL, Bush JW and Chen MM: Toward an operational definition of health. J Health Soc Behav 14:6-23, March 1973a.

Rahe RH: Multi-cultural correlations of life change scaling: America, Japan, Denmark and Sweden. J Psychosom Res 13:191, 1969.

Rokeach M: The Nature of Human Values. New York, The Free Press, 1973.

Rossi PH, Waite E. Bose CE and Berk RE: The seriousness of crimes: normative structure and individual differences. Am Sociol Rev 39:224-237, 1974.

Rothenberg J: The Measurement of Social Welfare. Englewood Cliffs, N.J., Prentice-Hall, 1961.

Sellin T and Wolfgang ME: The Measurement of Delinquency. New York, Wiley, 1964.

Sullivan DF: Conceptual problems in developing an index of health. Vital and Health Statistics, Series 2, No. 17, May 1966.

Torrance GW: Health index and utility models: some thorny issues. Health Serv Res 8:12-14, Spring 1973.