Predicting Self-assessed Health Status: A Multivariate Approach

by Thomas T.H. Wan

Two-stage multivariate analysis was used to examine factors affecting personal perception of health status. In the first stage, sociodemographic variables were used as independent variables in Automatic Interaction Detector (AID) analysis in order to partition the study sample (11,153 civilian noninstitutionalized adults aged 58–63) into subgroups. In the second stage, binary multiple regression analysis was performed on each AID subgroup and on the total sample. Predictors used were indicators of psychological, socioeconomic, and sociomedical well-being. Finally the applicability of these indicators in classifying persons in one of the two categories of perceived health status was examined by discriminant function analysis. Sociomedical health indicators were better explanatory variables of self-assessed health status than socioeconomic or psychological indicators of well-being.

In recent years there has been considerable interest in formulating measures of personal health status based on self-reported data [1–5]. These studies have provided evidence of the validity and reliability of self-assessed health status as an important component of perceived quality of life [6–8]. However, none of these studies has systematically examined the applicability of sociomedical indicators of well-being in predicting perceived health status. Moreover, few previous studies have taken differences in population groups into consideration, so it is not known whether most health indexes can be used for different populations.

This article describes an attempt using two-stage multivariate analyses to examine, at least partially, the adequacy of a composite index of health status developed for this study. In the first stage, variation in self-assessed health status was analyzed with respect to social and demographic variables and subgroups (clusters) were identified that demonstrated homogeneous health behavior and responses. In the second stage, psychological, socioeconomic, and sociomedical indicators of well-being were used in a set of regression equations formulated for the subgroups to determine the relative contribution of these factors in affecting the observed variation in self-assessed health status.

Address communications and requests for reprints to Thomas T.H. Wan, Ph.D., Research Fellow, Intramural Research Section, National Center for Health Services Research, 5600 Fishers Lane, Room 15-68, Rockville, MD 20852.

Stepwise discriminant analysis was used in order to test the applicability of health-related indicators in classifying perceived health status. Results were analyzed using two simple epidemiological tests, sensitivity and specificity [9].

Data and Methods

The present study was based on data obtained from the Longitudinal Retirement History Survey, a ten-year study of the retirement process conducted by the Bureau of the Census for the Social Security Administration [10]. The sample was composed of 11,153 noninstitutionalized civilian adults aged 58–63 in 1969.

The analytic technique used in this study was a two-stage multivariate approach that identified sociodemographic differentials in perceived health status and then verified the relative importance of 12 health-related indicators pertaining to psychological, socioeconomic, and physical well-being of individuals in predicting perceived health status. Indicators of psychological well-being included happiness and general life satisfaction. Socioeconomic well-being was measured by memberships in social or professional organizations, the ability to get along on one's income, and current working status. Sociomedical indicators included hospitalization, reported health conditions, severity of disability (extent of work limitations), duration of disability, and mobility limitations (confinement in the house, difficulties in traveling and moving). Detailed definitions of these indicators are presented in Table 1 (p. 466).

The dependent variable, personal perception of health status, was determined by the respondents' assessments of how their health compared with that of others of the same age. This variable was dichotomized: persons who perceived their health as being worse than that of others were coded as 1, and those who perceived their health as being the same or better than that of others ("not worse") were coded as 0.

In the first stage of analysis, Automatic Interaction Detector (AID) analysis was employed to partition the sample into clusters of individuals whose perceptions of health were more or less similar. This technique splits the sample into a series of nonoverlapping subgroups, which reduces the error in predicting the dependent variable [11]. (See Table 1 for code values for indicators.)

In the second stage of analysis, multivariate analyses were performed within each of the subgroups identified by the AID analysis to investigate the relative influence of the 12 health-related indicators on self-assessed health status, taking all factors into account simultaneously. The probability of considering oneself to be in worse health than others of the same age was estimated for individuals by using a binary-variable multiple regression method. In this analysis the additive effect of independent variables was examined. Each binary-coded variable represented a single subclass of a factor and was assigned a value of 1 if it was in the subclass and 0 if not (see Table 1). Each factor (indicator) was transformed into a number of regressor variables equivalent to the number of subclasses minus 1.

Table 1. Operational Definitions for Health-related Indicators Used in Binary-variable Regression Analysis and Discriminant Analysis

Indicator	Coded value: Regression analysis*	Coded value: Discriminant analysis†
Happiness		
Very happy	1	1
Pretty happy	1	2
Not too happy	1	3
Overall life satisfaction		
Very satisfactory	1	1
Satisfactory	1	2
Unsatisfactory	1	3
Very unsatisfactory	1	4
Number of memberships in professional, social orgs.		
One or more	0	0
None	1	1
Ability to get along on one's income		
I always have money left over (excellent)	1	1
I have enough, a little extra sometimes (good)	1	2
I have just enough (fair)	1	3
I can't make ends meet (poor)	1	4
Work status		
Full-time work	0	0
Part-time/irregular work or no work	1	1
One or more days of hospitalization§ in 1968		
No	0	0
Yes	1	1
Severity of disability		
My disability has no effect on my work/housework	1	1
Health limits amount of work/housework I can do	1	2
Had to change jobs because of health	1	3
Unable to work because of health	1	4
Duration of disability		
None	1	0
Under one year	1	1
One to four years	1	2
Five or more years	1	3
Bedridden or housebound most of the time		
No	0	0
Yes	1	1
Help needed for boarding a bus		
No	0	0
Yes	1	1
Help needed for going outside		
No	0	0
Yes	1	1

* Nonresponse and other subclasses coded as 0 were the omitted subclass for each indicator.

† Nonresponses were treated as missing values and were excluded from the analysis.

§ Hospital, rest home, sanitarium, or nursing home.

For individuals in a given group j, the probability Y of considering oneself in worse health than other members of the group can be systematically estimated by the 12 health-related indicators (F_i). A condensed equation for estimating this probability is

$$\hat{Y} = \text{Intercept} + \sum_{i=1}^{12} F_i$$

The intercept is the estimated probability for persons in the omitted subclass of each factor. The product of the predicting factor (regressor) and its regression coefficient is F_i.

It should be noted that use of a dichotomous dependent variable may create a problem in making precise linear estimations since there is no provision in the estimation procedure to prevent the estimates from going out of the unit interval (0–1) [12–14]. In this study conditional probability estimates greater than 1 were arbitrarily set equal to 1 and estimates less than 0 equal to 0. Logit analysis [15] or probit analysis [16] could have been used to transform the data in order to produce estimates (of the dependent variable) bounded by 0 and 1, but the logit and probit approaches are more expensive and complicated than the regression approach. Furthermore, Knoke [17] indicates that choice of approach probably makes little difference in substantive implications of the estimates if the range in proportions of the dependent dichotomy is between 0.25 and 0.75.

To assess the usefulness of the 12 health indicators in classifying individuals in "worse health" and "not worse health" categories, stepwise discriminant analysis was used. This analytic technique derives a discriminant function and efficiently selects important indicators as discriminating variables to classify an individual in one of two categories—in this case, perceived health status. It can empirically measure how effective the indicators are in discrimination by observing the proportion of correct classification [18]. After the classification is made, one can also measure the sensitivity of the classification (the proportion of true positives, i.e., individuals correctly classified as perceiving themselves to be in worse health than others of the same age) and the specificity of the classification (the proportion of true negatives, i.e., individuals correctly classified as not perceiving themselves to be in worse health than others of the same age).

It is important to note that a priori probabilities selected in discriminating between the two categories of perceived health can determine the proportion of misclassification. Choice of a higher probability of "worse health" in discriminant analysis will result in a higher sensitivity but a lower specificity for making correct classifications of individuals.

Results

First-stage Analysis: Identification of Subgroups

Table 2 presents the percentage and number distribution of persons aged 58–63 by sociodemographic characteristics and self-assessed health status. Of

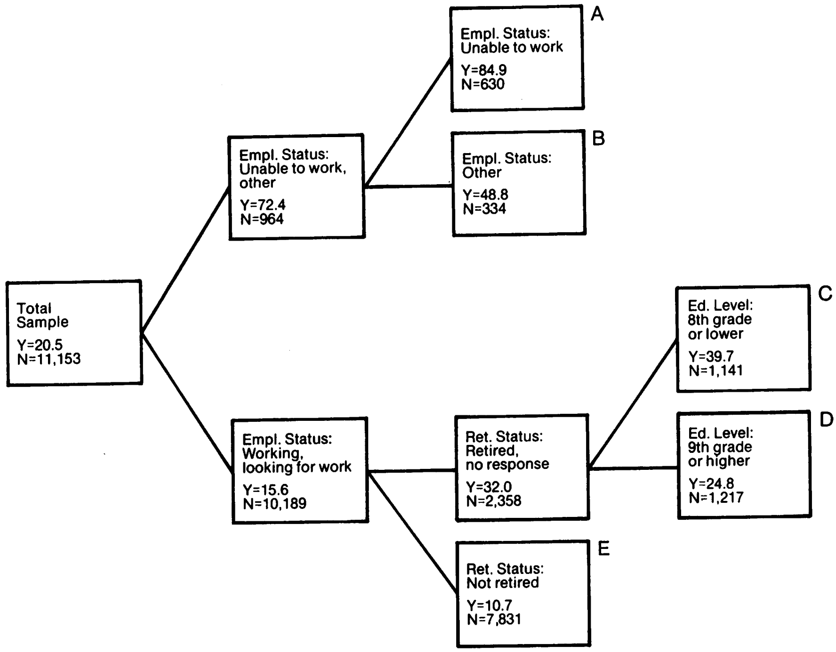
Table 2. Self-assessed Health Status, by Social and Demographic Variables

Variable	Total Number	Total Percent	Percent in worse health
Employment status			
Working	8287	74	11.93
Looking for work	234	2	15.38
Keeping house	1020	9	31.86
Retired	648	6	37.50
Unable to work	630	6	84.92
Other	334	3	48.82
Residential location			
Urban in urbanized area			
3,000,000 or more	2068	19	17.79
1,000,000–2,999,999	1118	10	17.62
250,000–999,999	1570	14	17.83
Under 250,000	1170	10	21.62
Urban not in urbanized area			
25,000 or more	416	4	20.19
10,000–24,999	494	4	19.23
2,500–9,999	827	7	19.59
Rural (under 2,500)	3490	31	24.41
Marital status			
Married	7136	64	19.38
Widowed	2256	20	20.92
Divorced	609	5	28.74
Separated	308	3	28.57
Never married	844	8	20.50
Educational level			
Never attended	167	1	32.34
Elementary 1–8th grade	4415	40	27.86
High school 9–12th grade	4681	42	17.07
College 1–5 years	1571	14	11.27
College 6 or more years	319	3	9.71
Age			
58	2040	18	18.14
59	1814	16	21.22
60	1834	16	20.01
61	1850	17	23.19
62	1872	17	20.51
63	1743	16	20.42
Race			
White	9971	89	19.42
Nonwhite	1182	11	30.03
Sex			
Male	8132	73	20.95
Female	3021	27	20.39
Family size			
One person	2217	20	21.02
Two persons	5342	48	19.37

Table 2. (Continued)

Variable	Total Number	Total Percent	Percent in worse health
Three persons	1918	17	19.50
Four persons	815	7	22.58
Five or more persons	861	8	26.95
Retirement status (subjectively defined)			
Completely retired	1795	16	53.87
Partially retired	934	8	27.41
Not retired	7912	71	11.10
No response	512	5	37.11
Income			
Under 4,000	2208	20	14.72
4,000–9,999	1369	12	7.96
10,000 or more	193	2	6.22
Unknown	7383	66	24.99
Occupation			
Professional, technical & kindred workers	883	8	7.47
Farmers & farm managers	540	5	21.11
Managers, officials, & proprietors	1255	11	11.00
Clerical & kindred workers	836	7	12.32
Sales workers	482	4	14.32
Craftsmen, firemen, & kindred workers	1616	14	15.97
Operative & kindred workers	1711	15	20.10
Private household workers	271	2	21.40
Service workers	995	9	17.99
Farm laborers	175	2	30.29
Laborers, except farm & mine	578	5	22.32
No response	1811	16	43.07

persons who were working or looking for work, three-fifths of the total sample, a substantially lower proportion thought their health was worse than that of others of the same age. More rural respondents assessed their health as worse than that of others their same age than persons living in urban areas. Divorced and separated persons were more likely than married, widowed, or never-married persons to perceive their health as worse than that of others their same age. The distribution of persons who perceived themselves to be in worse health than others their same age was monotonic by educational level with the highest rate among those who never attended school and the lowest among those with six or more years of college education. A slightly higher proportion of persons aged 61 thought their health worse than that of others of the same age as compared with persons of other ages. More nonwhites than whites perceived their health as being worse than that of others the same age. Males and females were equally likely to perceive themselves as having worse health than others their same age. The proportion reporting "worse health" increased as family size increased except for the one-person family. Over one-half of the completely retired and over one-quarter of the partially retired persons



Predictor tree for analysis of self-assessed health status. Y = percentage of persons who perceived themselves as having worse health than others. N = number of persons in group.

perceived their health as worse than that of others the same age. Of those who reported their anticipated retirement incomes, income was inversely related to the proportion who considered themselves as being in worse health. Among the 11 occupational categories, the greatest difference in self-assessed health status was found between professionals and farm laborers: the latter were four times more likely to perceive their health as worse than that of others in their age group.

The accompanying figure shows the results of the AID analysis (a split reducibility criterion of 0.006 and a minimum group size criterion of 50 were used in partitioning the subgroups). The predictor tree shows capacity for employment to be the most important social determinant of self-assessed health status. Thus the first split in the total sample was between those who were employable (working or looking for work) and those who were not (unable to work, other). Only 16 percent of the former group perceived their health as worse than that of others of the same age, whereas 72 percent of the latter group said they were in worse health than others their age. The unemployable group made a further split into those unable to work (Group A) and others (Group B). The percentages who considered themselves in worse health than others their own age were 85 and 49, respectively. For employable persons,

Table 3. Relative Contribution of Social and Demographic Variables in Explaining the Variance in Self-assessed Health Status

Variable	Proportion of variance explained*	
	Gross β^2	Partial β^2
Employment status	0.1559	0.1716
Residential location	0.0044	...
Marital status	0.0036	...
Educational level	0.0239	0.0072
Age	0.0009	...
Race	0.0065	...
Sex	0.0000	...
Family size	0.0021	...
Retirement status	0.1401	0.0452
Income	0.0237	...
Occupation	0.0608	...
	Total variance explained (R^2)	0.2240

* Gross β^2, or the maximum proportion of variance that can be explained by each variable by one split of the first group into two subgroups, denotes the relative importance of one predictor when other variables are not controlled. Partial β^2 is the actual proportion of variance explained by each predictor in AID analysis. Ellipses indicate variables not used in AID splits.

the second split was made on subjectively defined retirement status. Of those who considered themselves as retired or did not report retirement status, 32 percent considered themselves in worse health than others their age, whereas of those who defined themselves as not retired (Group E) 11 percent considered themselves in worse health. The retired group was divided by educational level into two final groups, those with 8th grade education or less (Group C) and those who had gone beyond 8th grade (Group D). The respective proportions who considered themselves in worse health were 40 and 25.

Table 3 shows the gross and net variance in probability of perceiving oneself as being in worse health that was explained by the 11 sociodemographic variables in the AID analysis. The net effect of employment status was greater than that of any of the other predictors, accounting for 17 percent of the total variance in self-assessed health status. Retirement status accounted for 5 percent of the variance, and educational level accounted for less than 1 percent. The other eight variables did not enter the AID splits, so their effects were negligible.

Second-stage Analysis: Multivariate Analysis Within Subgroups

The five final groups or clusters derived from the AID analysis constituted readily identifiable subpopulations with varying sociodemographic characteristics and perceptual differences in evaluating their health. In the second stage of the analysis, multivariate analysis was performed within each of these sub-

Table 4. Regression of Self-assessed Health Status on Health-related Indicators for Total Sample and Five Subgroups Identified by AID Analysis

Self-assessed health status, the dependent variable, was dichotomized: persons who perceived their health as worse than that of others of the same age were coded as 1, and those who did not perceive their health as worse than that of others the same age were coded as 0. Nonresponse and other subclasses coded as 0 were the omitted subclass for each indicator.

Indicator	Regression coefficient (b)					
	Total sample	Group A	Group B	Group C	Group D	Group E
Happiness						
Very happy	–0.050	–0.110	–0.091	0.167	–0.076	–0.023
Pretty happy	–0.027	–0.039	–0.130	0.256	–0.062	–0.006
Not too happy	0.080*	0.082	–0.045	0.363*	0.079	0.087*
Memberships						
One or more	–0.014	0.065	0.017	–0.045	–0.012	–0.012*
Life satisfaction						
Very satisfactory	–0.023	0.302	–0.235	–0.387*	–0.067	0.034
Satisfactory	–0.031	–0.028	–0.132	–0.336	–0.040	0.021
Unsatisfactory	–0.005	–0.025	–0.123	–0.297	0.007	0.041
Very unsatisfactory	0.020	–0.010	0.014	–0.256	0.008	0.048
Ability to get along on income						
Excellent	0.002	–0.291*	–0.043	0.048	0.078	0.022
Good	0.000	0.005	–0.093	0.048	0.096	0.004
Fair	0.020	0.062	–0.066	–0.042	0.125*	–0.024
Poor	0.052	0.126*	–0.073	0.019	0.111	–0.020
Work status						
Part-time/irregular work	0.078*	–0.157	0.249*	0.129*	0.149*	0.114*
Hospitalization						
Yes	0.088*	0.058	0.071	0.099*	0.127*	0.077*
Health condition						
Yes	0.141*	0.113*	0.050	0.157*	0.157*	0.133*
Severity of disability						
No effect	–0.068*	–0.187	–0.183	–0.171*	0.001	–0.063*
Secondary limitation	0.086	0.640	0.405	0.130	0.182	0.017
Occupationally disabled	0.060*	–0.122	0.079	–0.022	0.159*	0.072*
Severely disabled	0.305*	–0.168	0.512*	0.301*	0.382*	0.138*
Duration of disability						
Under one year	0.035	–0.107	–0.368	–0.038	–0.162	0.170*
One to four years	0.092	–0.049	–0.314	–0.072	–0.124	0.222*
Five or more years	0.104	0.009	–0.183	0.008	–0.153	0.180*
Bedridden/housebound						
Yes	0.072*	0.081	–0.010	0.054	0.070	0.096
Help needed boarding bus						
Yes	0.068*	0.029	0.024	0.012	0.038	0.095*
Help needed going out						
Yes	0.070	0.088	0.048	0.067	–0.038	0.082
Constant (intercept)	0.068	0.119	0.305	0.136	0.027	0.016
R^2 (coefficient of multiple determination)	0.436	0.160	0.529	0.360	0.350	0.240

* Significant at 0.05 or lower level.

populations in order to determine (1) the extent to which the sociomedical indicators of physical well-being, such as measures of functional status, could be used as predictors of self-assessed health status, taking into account the effects of psychological and socioeconomic indicators of well-being; and (2) the extent to which these indicators were sensitive and specific enough to be applied in the classification of individuals in "worse health" and "not worse health" categories.

Multiple Regression Analysis. Table 4 displays the results of a binary-variable multiple regression analysis of self-assessed health status using the 12 health-related indicators as predictors. The regression coefficients (*b*s) indicate the direction of the relationships between regressor variables and self-assessed health status. The conditional probability of "worse health" responses for individuals in different subgroups and in the total sample was estimated by using the values obtained in the general equation on p. 467 for each group. For the total sample the value of the intercept used in solving the equation was 0.068. For groups A through E intercept values were 0.119, 0.305, 0.136, 0.027, and 0.016, respectively.

Data in Table 4 reveal that severe disability had a prominent effect on self-assessed health in the total sample and in all the subgroups except group A. However, the relatively low R^2 values (ranging from 0.16 to 0.53) indicate that considerable variance was unexplained for each group. Developing better and more precise measures of well-being and selecting better indicators in future analyses should improve predictability.

Discriminant Analysis. When the 12 recoded discriminatory variables (Table 1) were used in a stepwise discriminant analysis, weights (discriminant coefficients) for the discriminant function were derived. Only a single function was formed for each cluster since there were two subcategories of perceived health status. Table 5 summarizes the results of this analysis applied to the total sample and each of the five groups. The canonical correlation coefficients show that the discriminant function formed in each of the five clusters was moderately associated with the two categories of perceived health status. The coefficients ranged from 0.32 to 0.67. Inspection of standardized function coefficients in Table 5 reveals that the "happiness" variable, a psychological well-being indicator, exerted the strongest effect on perceived health status in group A and a moderate effect in other subgroups and the total sample. General life satisfaction appeared to be a strong discriminator in groups A and B, but it had a relatively weak influence in group E and in the total sample. The discriminatory power of indicators of socioeconomic well-being seemed to be weak compared to that of happiness and overall life satisfaction variables. Review of the relative importance of sociomedical indicators affecting individuals' perception of health showed no clear patterns of influence that could be used for meaningful generalization.

AID analysis, using the 12 indicators as predictors, was performed in order to estimate the overall effects of three major dimensions of well-being on self-assessed health status (Table 6). Since the AID R^2 value is the sum of

Table 5. Relative Importance of Health-related Indicators in Discriminating the Self-assessed Health Status for Total Sample and for Five Subgroups Identified by AID Analysis: Standardized Function Coefficients (d′) and the Order of Their Importance

Indicator	Total sample		Group A		Group B		Group C		Group D		Group E	
	d′	Rank	d′	Rank	d′	Rank	d′	Rank	d′	Rank	d′	Rank
Psychological well-being												
Happiness	0.210	4	–0.725	1	–0.152	5	–0.293	3	–0.245	3	–0.240	5
Life satisfaction	–0.051	11	–0.381	2	–0.223	3	–0.169	6	–0.152	5	–0.66	9
Socioeconomic well-being												
Memberships	0.114	9	–0.170	7			0.049	10	0.067	7	0.050	11
Ability to get along on income	0.052	10			0.086	7					0.053	10
Work status	–0.019	12					0.231	4			–0.173	7
Physical well-being												
Hospitalization	0.130	8	–0.261	6	–0.125	6	–0.140	7	–0.206	4	–0.187	6
Health condition	0.339	2	–0.375	3	–0.170	4	–0.337	2	–0.435	1	–0.390	2
Severity of disability	0.207	5	–0.277	5	–0.407	1	–0.112	9	–0.401	2		
Duration of disability	0.201	6	–0.345	4	–0.368	2	–0.363	1			–0.276	4
Bedridden/housebound	0.199	7					–0.132	8	–0.146	6	–0.129	
Help needed boarding bus	0.315	3									–0.322	
Help needed going outside	0.413	1					–0.196	5			–0.540	1
Canonical correlation coefficient	0.621		0.322		0.667		0.563		0.537		0.467	

Table 6. Partial β^2 and R^2 in AID Analysis of Health-related Indicators in Predicting Self-assessed Health Status for Five Clusters

See footnote to Table 3 for explanation of partial β^2.

Predictor*	Group A (N = 630)	Group B (N = 334)	Group C (N = 1,141)	Group D (N = 1,217)	Group E (N = 7,831)
Psychological well-being					
Happiness	0.0556	...	0.0244	0.0340	0.0272
Life satisfaction	...	0.0164	...	...	...
Socioeconomic well-being					
Memberships	...	...	...	...	...
Ability to get along on income	0.0191	...	...	...	...
Work status	...	0.2343	0.0074	0.0683	...
Physical well-being					
Hospitalization	...	...	0.0075	...	...
Health condition	0.0378	...	0.0249	0.0125	0.0234
Disability					
Severely disabled	...	0.0104	0.0554	...	...
Occupationally disabled	...	0.0193	0.0085	...	...
Secondarily disabled	0.0131	0.3415	0.2012	0.2069	...
No effect	0.0086	0.1059	...	...	...
Duration of disability	0.0115	0.0128	...	...	0.1674
Total variance explained (R^2)	0.1457	0.5306	0.3293	0.3310	0.2180

* Indicators of mobility limitation did not enter the AID splits.

coefficients, the aggregate values of these coefficients for three dimensions of well-being indicators could be calculated. In group A, for example, the proportions of variance explained by the indicators of psychological well-being, socioeconomic well-being, and physical well-being were 0.056, 0.019, and 0.04, respectively. In other subgroups the variables pertaining to work limitations accounted for the majority of the total variance explained. Indicators of mobility limitation did not enter the AID splits.

The discriminant function analysis also provided additional information concerning correct classifications of persons by self-assessed health status. Table 7 shows the percentages of correctly classified cases and the sensitivity and specificity of the classifications. When it was assumed that there was an equal probability for persons to be classified in either category, the respective percentages of correct classification made in the total sample and groups A–E were 83, 70, 80, 73, 83, and 84, respectively. However, when the discriminant analysis was based on the known probabilities (i.e., the actual "worse health" proportions for clusters) in making discriminations, the percentages of correctly classified persons were slightly improved. A computation of the proportion of true positives (those who perceived themselves as having worse health than others) and true negatives (those who did not perceive themselves as having worse health than others) was made after the classification.

Table 7. Tests of the Applicability of the Discriminant Function Formed by Important Health-related Indicators in Discriminating "Worse" from "Not Worse" Perceived Health Status

Probability in discrimination test	Group					
	Total sample	A	B	C	D	E
$P = q$*						
Grouped cases correctly classified (%)	82.9	69.5	80.2	73.3	82.9	83.6
Sensitivity (%)	80.7	71.6	92.0	80.6	80.7	70.6
Specificity (%)	83.5	57.9	69.0	68.5	83.5	85.2
$P \neq q$†						
Grouped cases correctly classified (%)	85.4	85.1	80.5	74.9	80.9	88.4
Sensitivity (%)	63.6	97.8	91.4	69.8	54.6	42.0
Specificity (%)	91.0	13.7	70.2	78.2	89.6	94.0

* A priori probabilities were assumed equal in discriminating between the two subclasses, "worse" and "not worse," of health status.

† When a priori probabilities were known and unequal in each of the five subpopulation groups, these probabilities were applied in the discriminant analysis. A priori probabilities for classifying "worse" state of health (ps) were based on the proportions of persons who perceived themselves as having worse health than others the same age in five subgroups. These were 0.85, 0.49, 0.40, 0.25, 0.11, and 0.20 for groups A–E and for the total sample, respectively.

In most cases (i.e., except for group A) sensitivity was generally reduced and specificity was slightly increased if the second a priori assumption (i.e., $p \neq q$) was chosen in the discriminant analysis. Since selection of a discriminatory criterion is somewhat arbitrary, it is difficult to recommend an optimal set of criteria. However, it is a general rule of thumb in epidemiological research to use an effective measure that produces a relatively low sensitivity but a high specificity when the condition (e.g., disease) is not highly prevalent in the community [19].

Implications and Conclusions

The above analyses indicate the variation in personal perception of health with respect to different social as well as health conditions. The findings of this study provide evidence showing the usefulness and validity of multivariate analysis applied to the assessment of personal health status. In addition, there are some substantive implications of the findings. First, in analyzing the role of sociodemographic variables in self-assessed health status the study goes beyond recognition of social differentials in health status. It reflects how individuals from various population groups vary in their health behavior. Identification of homogeneous subgroups or clusters serves as the first step for further investigation of the relevance and applicability of health indicators in prediction of self-assessed health status. Second, evidence has been provided that sociomedical health indicators are better explanatory variables of perceived health status than socioeconomic and psychological indicators of well-being. This finding implies that the payoff in health status research on index construction is most likely to come from emphasis on measures of functional status and other sociomedical indicators of well-being. A specific and sensitive index of health status should be constructed by taking into account all the important dimensions of well-being. Furthermore, a health index should never be applied in a given population unless it has been systematically evaluated in terms of its reliability, validity, predictability, and applicability in that population.

Finally, the analytic approach proposed in this study has suggested the need for further methodological and conceptual refinement of health status indicators so that future research on health status can realistically assess the usefulness of currently existing data obtained from national health surveys such as the National Health Interview Survey.

REFERENCES

1. Belloc, N.B., L. Breslow, and J.R. Hochstim. The measurement of physical health in a general population survey. *Am J Epidemiol* 93:328 May 1971.
2. Hessler, R.M., P.K. New, P. Kubish, and D.L. Elison. Demographic context, social interaction, and perceived health status. *J Health Soc Behav* 12:191 Dec. 1971.
3. Lawton, M.P. and J. Cohen. The generality of housing impact on the well-being of older people. *J Gerontol* 29:194 Mar. 1974.
4. Maddox, G.L. Self-assessment of health status. *J Chron Dis* 17:449 May 1964.
5. Linn, M.W. Studies in rating the physical, mental, and social dysfunction of the chronically ill aged. *Med Care* 14(Suppl):119 May 1976.
6. Campbell, A., P. Converse, and W.L. Rogers. *The Quality of American Life.* New York: Russell Sage, 1976.
7. Andrews, F.M. and S.N. Withey. Developing measures of perceived life quality: Results from several national surveys. *Soc Indic Res* 1:1 May 1974.
8. Maklan, C., C.F. Cannell, and J.R.P. French. Subjective and objective concepts of health: A background statement for research. Institute for Social Research, University of Michigan, Ann Arbor, MI, 1974.
9. Goldsmith, C.H. and G.S. Layhew. Generalized Measures of Sensitivity, Specificity, and Predictive Value in k × k Contingency Tables. Paper presented at the 9th International Biometric Conference, Boston, MA, Aug. 27, 1976.
10. Irelan, L.M. Retirement history study: Introduction. *Soc Secur Bull* 35:3 Nov. 1972.
11. Sonquist, J.A. *Multivariate Model Building.* Ann Arbor, MI: University of Michigan Institute for Social Research, 1970.
12. Nerlove, M. and S.J. Press. *Univariate and Multivariate Log-Linear and Logistic Models.* Santa Monica, CA: Rand, 1973.
13. Cox, D.R. *The Analysis of Binary Data.* London: Methuen, 1970.
14. Huang, D.S. *Regression and Econometric Method.* New York: Wiley, 1970.
15. Berkson, J. Application of the logistic function to bio-assay. *J Am Stat Assoc* 39:357 June 1944.
16. Finney, D.J. *Probit Analysis*, 3rd ed. Cambridge, England: Cambridge University Press, 1971.
17. Knoke, D.A. Comparison of log-linear and regression models for systems of dichotomous variables. *Soc Methods Res* 3:416 May 1975.
18. Eisenbeis, R.A. and R.B. Avery. *Discriminant Analysis and Classification Procedures.* Lexington, MA: Lexington, 1972.
19. Sartwell, P.E. (ed.). *Preventive Medicine and Public Health*, 9th ed., Chap. 1. New York: Meredith, 1965.