A NEW APPROACH FOR THE MANAGEMENT OF HEALTH

Martin K. Chen

National Center for Health Services Research and Development

A fairly exhaustive review of the available literature reveals that not much progress has been made in the measurement of health since Stouman and Falk (1) developed quantitative indices in three areas: vitality and health, environment, and public health activity. In recent years, the work of Chiang (2) and that of Fanshel and Bush (3) represent the more notable efforts in this area using, respectively, stochastic and deterministic models. The indicators developed by these authors, however, present conceptual and methodological problems that tend to limit their utility in their present state.

Chiang's index is based on the probability distribution of the population at risk with respect to what he assumes to be three independent random variables--the frequency of illness, the duration of illness in number of days, and the time lost due to death in a given year, again in number of days. Conceptually, the duration of illness is purported to be a measure of the severity of illness, but in reality it is more appropriately a measure of the chronicity or acuteness of illness. Furthermore, his assumption of independence of the three random variables is hardly tenable in view of the fact that it is a common phenomenon for many individuals to be sick before their death and that within a given year the duration of illness cannot be independent of the time lost due to death.

The index of Fanshel and Bush, termed the Health Status Index (HSI), is operationally defined as the mean value of 11 weighted functional states of health of the population at a given time. The weights for the different functional states vary from 0 for the lowest state, death, to unity for the highest state, well-being. The other intermediate weights are assigned by a technique similar to Thurstone's paired comparison scaling technique (4), based on the proportions of times a panel of judges rate one attribute (in this case functional state) over the other, pair by pair.

While the paired comparison technique is a rigorous way of estimating personal preferences--indeed, Mosteller (5) has shown that under certain conditions the estimates are identical with the least squares estimates, Fanshel and Bush's technique permits of modifications of the estimated values according to expert prognoses and expected benefits from health program intervention given limited resources. In so doing the authors make their HSI contingent on prognoses and judgmentally-determined optimum resourse allocations, with the result that the HSI is no longer an invariant measure of health status per se.

Philosophically and conceptually, the authors of the two indices implicitly, if not explicitly, define health in absolute terms, when in reality health is more appropriately regarded as a relative concept. Dubos (6), who has written extensively on health from the philosophical, biological and sociological points of view (7,8), states:

"Health ... cannot be absolute and permanent values, however careful the social and medical planning."

The World Health Organization (9), after a lengthy discussion of the concept of health in which it frankly admits that its definition of health as "a state of complete physical, mental and social wellbeing" does not lend itself to objective measurement, comes to the conclusion that health may best be expressed as "a degree of conformity to accepted standards of given criteria in terms of basic conditions of age, sex, community and region, within normal limits of variation."

Health as a relative concept, then, is the basis for the formulations of the two health indicators that are to follow. Two key factors are clearly implied by this concept and both are required for the validity of the formulation. One key factor is that an "ideal norm", however derived, is available in each health dimension. These norms would be the "accepted standards of given criteria" as stated by WHO. The second key factor is that members of a given group or population must vary around, not necessarily symmetrically, the "ideal norm" along each health dimension. This factor is related to a measure of variation as conceived of by WHO.

Before we proceed to derive the indicators, we would like to emphasize that the "ideal norm" of a health dimension, such as age and height adjusted bodily weight, may or may not be the mean value of a group of people. For instance, it is recognized that in our affluent society where food is plentiful, a large number of middle-aged males and females have an overweight problem. In such a case, the mean weight of an age and sex group obviously cannot be the "ideal norm", in the form of a point or in the form of a range. Another norm, less than the mean, must be sought.

Since health is multi-dimensional (See, for example, Sullivan (10)), it is essential that the dimensions, which may be quite disparate, be measured on the same scale and with the same degree of variability. This is, of course, impossible because the units of measurement for different health dimensions, as for example, visual acuity and blood pressure, cannot be the same. The only feasible solution, then, is to transform the raw measures, taken as deviations from the "ideal norms," into standard scores.

Notice that the deviations from the "ideal norm" divided by the "standard deviation" are not standard scores unless the norm happens to coincide with the mean. The sum of the deviations from an origin other than the mean is, of course, not zero, and the "standard deviation" is not a true standard deviation in that the squared deviations summed are not taken from the mean. Thus the transformed scores do not possess the properties of the standard scores. In fact, it is algebraically proved (See appendix) that the variance of these transformed scores is always less than unity.

The real problem posed by using the transformed scores is that neither the means nor the standard deviations of the transformed scores on different health dimensions are comparable, and simple aggregation of the transformed scores would not ensure equal weights to the health dimensions, because the weights would vary directly in proportion to the magnitudes of the means of the health dimensions. The solution to this problem is to treat the deviation scores from the "ideal norm" as raw scores, and where positive and negative deviations connote different degrees of seriousness from the health point of view, differentially weight the positive and negative deviations. Then the positive and negative signs of the scores can be discarded and the absolute values used in transforming them into z scores.

DERIVATION OF -H INDEX

Symbolically, if we let $\pm X_{ij}$ be individual i's deviation score from the "ideal norm" on dimension j, the corrected deviation score after weighting is $X'_{ij} = V_{+,-} \left|\pm X_{ij}\right|$, where $V_{+,-}$ are the differential weights for the positive and negative deviation scores, and $\left|\pm X_{ij}\right|$ is the absolute value of individual i's deviation score on dimension j. $V_{+,-}$ can be any number or zero for deviation scores that are zero, because any number times zero is zero. Then treating the corrected deviation scores as raw scores, we perform the following statistical operations to derive an individual's z score on a single health dimension and his health index, which is the simple aggregation of the weighted health scores in standard form. We shall designate the index as -H to indicate that as an index it has a negative relation with the health level of the individual.

We write $x'_{ij} = X'_{ij} - \bar{X}'_j$, where $\bar{X}'_j = \sum_{i}^{n} X'_{ij}/n$, and n is the size of the group. We further write $S_j = (\sum_{i} x'^2_{ij}/n)^{\frac{1}{2}}$. Then individual i's z score on dimension j is $z_{ij} = X'_{ij}/S_j$. To eliminate negative signs, we add a constant 5 to all the z scores to transform them into Z scores. Finally, we write: $-H = \sum_{j}^{m} w_j Z_{ij}$, where w_j are the weights, however derived, for the component health dimensions. $w_j = 1$ if no particular weights are assigned to the dimensions.

It can be seen from the equation for -H that if an individual's score fell on the "ideal norms" on all the health dimensions, his deviation scores would all be zero, his corrected deviation scores would also be zero, and his z scores would be negative and the highest among all the negative z scores in his group. After transformation, however, his Z scores would be the lowest among all the z scores. The magnitudes of the other Z scores are a function of the magnitudes of the corrected deviation scores. The larger the corrected deviation scores, the larger the Z scores. This phenomenon fully justifies the rationale of -H as an index of "negative health" in relative terms.

Further examination of -H shows, assuming some degree of normality of the distributions of the Z scores, that the lowest -H value that indicates optimum health in the aggregated dimensions would be approximately $1.5 \sum_{j}^{m} w_j$, the produce of 1.5 times the sum of the weights, because practically all the Z scores would fall within 3.5 standard deviations of the means, which is 5 (the constant added to the z scores to derive Z scores). Similarly, the highest -H value would be $8.5 \sum_{j}^{m} w_j$. If no differential weights were assigned, then the range of -H would be approximately 1.5m to 8.5m.

If a measure of variability of -H is desired, its variance can be computed as follows (See, for instance, Ghiselli, 1964): $S^2_{-H} = \sum_{j}^{m} w_j^2 S_j^2 + 2\sum_{j}^{m} w_i w_j S_i S_j r_{ij}$, where $i \neq j$, r_{ij} are the correlations between all possible pairs of health dimensions in the composite and the summation is taken over all m dimensions. However, since all the standard deviations of the Z scores on the m dimensions are unity (if this is not clear, it is recalled that the standard deviation of z scores is unity and that adding a constant to the scores does not affect the standard deviation), we can simplify the formula by writing:

$$S^2_{-H} = \sum_{j}^{m} w_j^2 + 2\sum_{j}^{m} w_i w_j r_{ij}.$$

If it could be safely assumed that the health dimensions are uncorrelated (which would probably be true of hearing acuity and visual acuity), then we could further simplify by writing: $S^2_{-H} = \sum_{j}^{m} w_j^2$, since the last term would vanish.

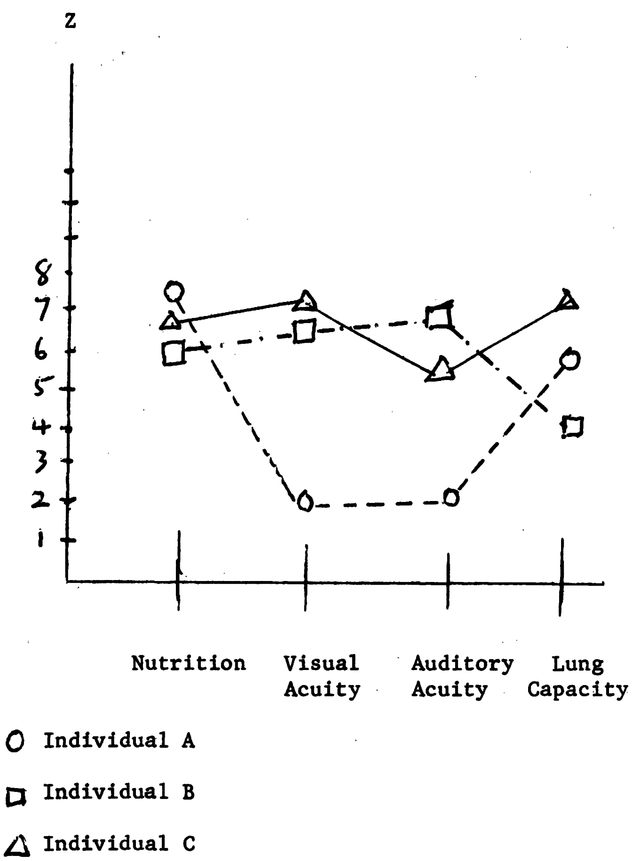
The variance of -H would be useful in computing the standard error of measurement that would provide some information as to the accuracy of -H as an individual index for a particular group. Using the analogy of a single test, the standard error of measurement for -H, which is a composite measure, would approximately be: $S_{e(-H)} \cong S_{-H}(1 - \bar{r}_{xx})^{\frac{1}{2}}$, where $\bar{r}_{xx}$ is the mean of the reliabilities of the component dimensional measures.

Note that once the Z scores for each health dimension is derived, the decision whether or not to aggregate the Z scores into a single index for each individual must hinge upon the purpose of the index. If the purpose is to have a single index as a measure of the outcome of a health program, then the single index should consist of those health dimensions on which the program is designed to have an effect. For example, if a health program had as its goals the elevation of nutrition standards and the improvement of visual acuity of the citizens of a community, the single index would be composed of two dimensions: nutrition level and visual acuity.

On the other hand, the purpose may simply be to compare a group of individuals on several health dimensions that are deemed important to the performance of a particular task or job. In that case, a profile of the Z scores of these individuals may give a better picture than a single aggregated index. A hypothetical profile of Z scores of three individuals on four health dimensions is given in Figure 1.

FIGURE 1

Hypothetical Profile of Three Individuals on Four Health Dimensions



This profile shows that Individual A is fairly high in nutrition, pretty low in visual acuity and auditory acuity, and above average in lung capacity. Individual B is below average in nutrition, average in visual acuity, slightly above average in auditory acuity, and average in lung capacity. Individual C is average in nutrition, slightly above average in visual acuity, slightly below average in auditory acuity, and above average in lung capacity.

It might be argued that in combining the health dimensions perhaps a multiplicative (non-linear) model would be more appropriate than an additive (linear) model that has been suggested. Our response to this argument would be that in our present state of knowledge there is no theoretical or empiric basis for choosing one model over the other. At any rate, if a multiplicative model were indeed found to be more appropriate, we could always perform a logarithmic transformation of the Z scores and still apply the additive model.

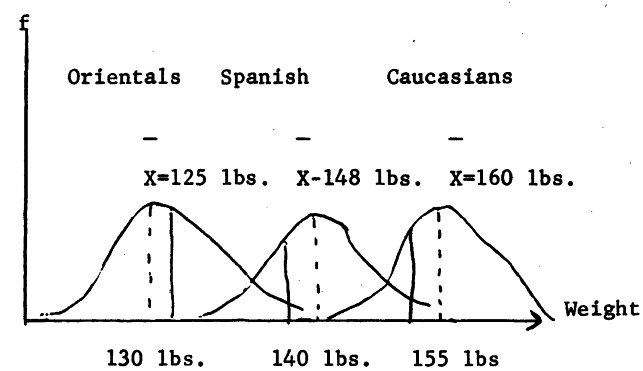
One thing that may not be obvious is whether or not aggregation of the Z scores as a single index makes the index invariant over occasions or communities. Assuming that identical instruments are used in the measurement of the health dimensions, and further assuming that the distributions of the health dimensions are normal or similar in shape across communities, then the single index does posses the property of invariance. This statement is based on the fact that when the scores which are measured on equal intervals, are derived from deviations from comparable norms, the scores take on the property of a ratio scale, with what is psychometrically known as a relative zero point (11,12).

DERIVATION OF -T INDEX

Although the rationale for developing -H for individuals is also applicable to a community or nation, the heterogeneity of the population in a community or nation does not permit a straight forward aggregation of the individual -H values into a single index. The situation may be illustrated in Figure 2.

FIGURE 2

Theoretical Representation of the Heterogeneity of the Male Adult Population in the United States with Respect to a Health Dimension: Weight



In this figure, the numbers along the abscissa are the "ideal normative weights" for the three subpopulations. The mean weights of the three subpopulations are given atop the three curves. In this theoretical example, more people in the Oriental subpopulation are underweight than overweight, more Spanish Americans are overweight than underweight, and more Caucasians are overweight than underweight. This figure is, of course, an oversimplification, because in reality the "ideal" weights within each subpopulation must be contingent upon bone structure or height. Nevertheless, it does serve to indicate the complexity of the problem of generating a single index of health for a population.

Theoretically, however, it should be feasible to derive a single index for a community or a nation by using the same rationale that is used in developing an individual index. Assuming that m dimensions of health are used and that there are k subpopulations in a community, each subpopulation being of size n_l, we could obtain the deviation scores from the "ideal norms" of the subpopulations in each dimension, correct the deviation scores by differential weighting of positive and negative deviations, transform the corrected scores into z scores and further transform the z scores into Z scores by adding the constant 5 to get rid of the negative signs.

Before we proceed to derive the index, we would like to emphasize that while the "ideal norms" for a health dimension may differ from community to community or from nation to nation, the transformation of the corrected deviation scores into z scores is performed by treating the corrected deviation scores from all the communities or nations on one dimension as one variable. As long as the individuals from each subgroup within a community or nation are properly identified by some coding scheme, a mean Z score on a given dimension is obtainable for each of the subgroups. The magnitudes of the mean scores are a function of the magnitudes of the corrected deviation scores from their "ideal norms." For purposes of simplification, we assume there are k number of subgroups within each community or nation.

We write $\bar{Z}_{..l} = \sum_j^m \sum_i^{n_l} w_j Z_{ijl}/n_l$, the mean of subgroup$_k$l for all weighted m dimensions of health, and $\sum_l n_l = N$, the total population of a community or nation. Then the index for a community or nation, which we shall designate as -T, would be:

$$-T = \sum_l^k n_l \bar{Z}_{..l}/N = \sum_l^k \left(n_l \sum_l^k Z_{..l}/n_l\right)/N = \sum_l^k Z_{..l}/N.$$

It is seen that -T is actually a weighted mean of all health dimensions for all subgroups, the weights being the sizes of the subgroups. The variance of -T is:

$$S^2_{-T} = (1/N)^2 \left(\sum_j^m w_j^2 + 2\sum_j^m w_i w_j r_{ij}\right), \text{ where } i \neq j,$$

and the summation is taken over all m dimensions in the aggregate. Where the dimensions are statistically independent, the second term drops out and we have: $S^2_{-T} = (1/N)^2\left(\sum_j^m w_j^2\right)$, for the general case. If no special weights are assigned to the dimensions, then we have:

$S^2_{-T} = (1/N)^2(m + 2\sum_j^m r_{ij})$. For the independent case we have simply: $S^2_{-T} = (1/N)^2 m$.

The standard error of measurement for -T would be: $S_{e(-T)} = S_{-T}(1-\bar{r}_{xx})^{\frac{1}{2}}$, where $\bar{r}_{xx}$ is the mean of the reliabilities of the measured health dimensions.

The assumption that there are k subgroups within a community in no way implies that k is a constant for all the communities to be compared with the -T index. In fact, Community A may have three subgroups and Community B five subgroups. The number of subgroups within a community is determined by the number of "ideal norms" that are required on a health dimension. For example, if the "ideal norms" for weight were the same for blacks and whites in the United States, then there would be no reason to consider blacks and whites as two subgroups on that particular dimension. However, the number of health dimensions used must be the same at all times for -T to be invariant over communities.

PROBLEMS OF ESTABLISHING NORMS

Since in most cases the mean of a measured health dimension would not be the "ideal norm," some means of establishing the norm had to be found. One approach would be to convene a large panel of medical experts and assign to it the task of recommending the "ideal norms" for age-sex specific groups on various health dimensions. This panel could also be charged with the responsibility of determining the weights of the various dimensions for the purpose of aggregating them into a single index when such an index is required. In the absence of objective data for decision making,a consensus of the opinions of these medical experts based on the largest body of available medical evidence could constitute a reasonable approximation to objective criteria.

Where there is any theoretical or empirical basis for believing that the health dimensions are related to mortality and/or morbidity, one can easily determine the points or ranges of these measured dimensions within which mortality and morbidity are at the lowest, given that the appropriate data are either available or obtainable. For purposes of aggregating the health dimensions into a single index, one can apply regression or discriminant analysis to determine the optimum weights for these dimensions that are most predictive of mortality and morbidity.

SOME CLOSING REMARKS

It has been pointed out that the decision to use an aggregated index and the kinds of health dimensions in the index must depend on the purpose which the index is to serve. In terms of mortality and morbidity data, the index composed of such data cannot appropriately serve as an outcome measure unless a health program is aimed at overall reduction of mortality and morbidity from all causes. Also, in using the index for cross-community comparisons, it may be necessary to differentiate mortality and morbidity by cause, so that the index comprises only those dimensions that are applicable to all the communities compared. For example, it would not be fair to compare New York City with a quiet rural town in Wyoming in terms of an index that is composed of mortality and morbidity data that include automobile deaths and injuries, for the simple reason that there may not be any automobile in the Wyoming town.

A look at the formula for standard error of measurement for either of the two indices reveals that it is a function of two computed statistics: S_{-H} (or S_{-T}) and $\bar{r}_{xx}$. If $\bar{r}_{xx}$ is unity; that is, if all measured dimensions have perfect reliability, then $S_{e(-H)}$ or $S_{e(-T)}$ reduces to zero, indicating no measurement error, regardless of the magnitude of S_{-H} or S_{-T}. On the other hand, if $\bar{r}_{xx}$ is zero, then $S_{e(-H)} = S_{-H}$ and $S_{e(T)} = S_{-T}$, the standard deviation of the -H or -T values. The lesson to be learned here is that one can always increase the precision of the index by increasing the reliabilities of the various measures of the component health dimensions, by reducing the variability of the group or subpopulation through judicious selection of "ideal norms," or by doing both.

A couple of caveates are now in order. First, the -H or -T index is meant to be a general index, hopefully useful for research and administrative purposes. It is not meant to be a diagnostic tool for ascertaining the degrees of threat of life from a variety of potential risk factors. For example, the fact that Individual A has a lower -H value than Individual B does not necessarily mean that Individual A has a longer age corrected life expectation than that of Individual B, for the simple reason that Individual B may have undiagnosed cancer of the lung and this information is not incorporated into the -H index. Such information would be useful for the development of indices that are ill-oriented rather than health-oriented, as the -H and -T indices are.

Another caveat is that, like most indices that have been designed, the -H or -T index does not possess inherent validity. In fact, with the exception of a few indicators in the mental health area, most health indicators that are in existence do not provide information on validity and reliability. If these indicators are to win acceptance in research quarters, they must be accompanied by such information, in the same way that standardized aptitude and achievement tests must be accompanied by such information.

APPENDIX

Algebraic Proof that the Variance of Transformed Scores Not Based on First and Second Moments of the Distribution of a Random Variable Is Less than Unity.

Let X be a random variable, $\bar{X}$ the means of this variable, $\bar{X}'$ the "ideal norm" that differs from $\bar{X}$ by d.

We write:

$x = X - \bar{X}$

$x' = X - \bar{X}'$

$d = \bar{X} - \bar{X}'$

Then $x' - x = (X - \bar{X}') - (X - \bar{X}) = \bar{X} - \bar{X}' = d$

$$x' = x + d$$

$$\Sigma x' = \Sigma x + nd = nd \tag{1}$$

$$x'^2 = (x + d)^2 = x^2 + 2dx + d^2$$

$$\Sigma x'^2 = \Sigma x^2 + 2d\Sigma x + nd^2 = \Sigma x^2 + nd^2 \tag{2}$$

$$S^2_{x'} = \Sigma x'^2/n = \Sigma x^2/n + d^2 = S^2_x + d^2 \tag{3}$$

$$z_{x'} = x'/S_{x'}$$

$$\Sigma z_{x'} = \frac{1}{S_{x'}} \Sigma x' \tag{4}$$

Substituting (1) into (4), we ave

$$\Sigma z_{x'} = \frac{1}{S_{x'}} (nd) \tag{5}$$

$$\Sigma z^2_{x'} = \frac{1}{S^2_{x'}} \Sigma x'^2 \tag{6}$$

Substituting (2) into (6), we have

$$\Sigma z^2_{x'} = \frac{1}{S^2_{x'}} (\Sigma x^2 + nd^2) \tag{7}$$

$$S^2_{z_{x'}} = \left(\frac{1}{n} \Sigma z^2_{x'} - \frac{(\Sigma z_{x'})^2}{n}\right) \tag{8}$$

Substituting (3), (5) and (7) into (8), we have

$$S^2_{z_{x'}} = \frac{1}{n}\left(\frac{\Sigma x^2 + nd^2}{S^2_x + d^2} - \frac{n^2d^2}{n(S^2_x + d^2)}\right)$$

$$= \frac{1}{n}\left(\frac{\Sigma x^2 + nd^2 - nd^2}{S^2_x + d^2}\right)$$

$$= \frac{1}{n}\left(\frac{\Sigma x^2}{S^2_x + d^2}\right)$$

If $d^2 = 0$; that is, if $\bar{X}' = \bar{X}$, then

$$S^2_{z_{x'}} = \frac{1}{n}\frac{\Sigma x^2}{S^2_x} = \frac{S^2_x}{S^2_x} = S^2_{z_x} = 1$$

Since for any given sample Σx^2, n, S^2_x are all constants, $S^2_{z_{x'}} < S^2_{z_{x'}} = 1$ whenever $d^2 > 0$.

REFERENCES

(1) Stouman, K. and Falk, I. S., "Health indices-a study of objective indices of health in relation to environment and sanitation," Bulletin of Health of the League of Nations, 8(1939)63.

(2) Chiang, C. L., An Index of Health: Mathematical Models. PHS Publication No. 1000 - Series 2, No. 5. Washington, D.C.: U.S. Government Printing Office, 1965.

(3) Fanshell, S. and Bush, J. W., "A health index and its application to health research outcomes," Operations Research, 18(1970) 1021-1066.

(4) Thurstone, L. L., "A law of comparative judgment," Psychological Review, 34(1927) 273-286.

(5) Mosteller, F., "Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations," Psychometrika, 16(1951)3-9.

(6) Dubos, R. J., Mirage of Health, Utopias, Progress and Biological Change. New York: Harpers, 1959.

(7) Dubos, R. J. and Dubos, J., Bacterial and Mycotic Infections of Man. Philadelphia: Lippincott, 1948.

(8) Dubos, R. J. and Dubos, J., The White Plague: Tuberculosis, Man and Society, Boston: Little, Brown, 1948.

(9) World Health Organization, Measurement of Levels of Health. WHO Technical Report Series No. 137. Geneva: World Health Organization, 1957.

(10) Sullivan, D. F., Conceptual Problems in Developing an Index of Health. PHS Publication No. 1000, Series 2, No. 17, Washington, D.C.: U.S. Government Printing Office, 1966.

(11) Stevens, S. S., "On the theory of scales of measurement," Science, 103(1946)677-680.

(12) Ghiselli, E. E., Theory of Psychological Measurement. New York: McGraw-Hill, 1946.