Journal of Consulting and Clinical Psychology
1983, Vol. 51, No. 5, 730-742

The Structure of Psychological Distress and Well-Being in General Populations

Clairice T. Veit and John E. Ware, Jr.
The Rand Corporation, Santa Monica, California

We describe the development of the Mental Health Inventory (MHI), a new 38-item measure of psychological distress and well-being, developed for use in general populations. The MHI was fielded in four large samples having quite different characteristics (N = 5,089). One data set was used to explore the MHI's factor structure, and confirmatory factor analyses were used for cross-validation. Results support a hierarchical factor model composed of a general underlying psychological distress versus well-being factor; a higher order structure defined by two correlated factors—Psychological Distress and Well-Being; and five correlated lower order factors—Anxiety, Depression, Emotional Ties, General Positive Affect, and Loss of Behavioral Emotional Control. Summated rating scales produced high internal-consistency estimates and substantial stability over a 1-year interval. Our results provide strong psychometric support for a hierarchical model and scoring options ranging from five distinct constructs to reliance on one summary index. This trade-off, which is between the unique information contained in the subscales versus the simplicity of a single score, should be evaluated further.

A review of general population mental-health-survey instruments pointed out important trends in questionnaire content and conceptual issues that deserve empirical attention (Ware, Johnston, Davies-Avery, & Brook, 1979). Early instruments were very heterogeneous in content (Gurin, 1960; Langner, 1962; Macmillan, 1957). They included measures of physical and psychosomatic symptoms, functional status, other health problems or worries, and health habits, in addition to measures of more straightforward psychological constructs (e.g., symptoms of anxiety and depression). More recent instruments seem to focus almost exclusively on the more straightforward psychological constructs (Bradburn, 1969; Cleary, Goldberg, & Kessler, 1982; Dupuy, 1972). Whereas the more heterogeneous measures may be satisfactory for testing hypotheses about health status in general, they do not seem to do well in distinguishing changes in mental health from changes in physical health (Ware et al., 1979; Ware, Brook, Davies-Avery, et al., 1980b).

Another characteristic of more recently developed mental health surveys is their focus on symptoms such as anxiety and depression that reflect the more prevalent kinds of psychological distress in general populations (Ware et al., 1979). Including these measures should increase the sensitivity of a general population survey in detecting changes in mental health. However, a substantial proportion of people in a general population rarely or never report occurrences of even the most prevalent psychological distress symptoms. To increase measurement precision, it may be necessary to extend the definition of mental health beyond the mere frequency or intensity of psychological distress symptoms to include characteristics of psychological well-being (e.g., feeling cheerful, interest in and enjoyment of life). Psychological well-being items have the potential to improve the precision of mental health measurement by distinguishing among persons who receive perfect scores on measures of psychological distress. Whereas inclusion of such items is a feature of more recent instruments (Bradburn, 1969; Dupuy, 1972; Goldberg, 1978; Ware et al., 1979), whether they actually improve the power of hypothesis testing remains to be determined.

In 1975 we began work on a new general population measure, the Mental Health Inventory (MHI). This standardized survey, which was designed to measure general psy-

Requests for reprints should be sent to Clairice T. Veit, The Rand Corporation, 1700 Main Street, Santa Monica, California 90406.

chological distress and well-being, is the primary mental health outcome measure fielded in the Rand Health Insurance Experiment (HIE) and is also being used to predict use of mental and general health care services (Manning, Newhouse, & Ware, 1982; Wells, Manning, Duan, Ware, & Newhouse, 1982).[1] Although the MHI was designed for a particular study, the rationale behind its development and our findings regarding its underlying mental health dimensions should be useful in conceptualizing and measuring mental health for other general population studies.

When work on the MHI began, a number of conceptual issues with theoretical implications had not been resolved for general population measures. The first issue was whether mental health, as defined by instruments like the MHI, has one or more dimensions. If more than one, do items seem to group into two distinct constructs—psychological distress (negative mental health states) and psychological well-being (positive mental health states)—or do they tend toward bipolarity, with factors defined by both negative and positive mental health states? Third, if more than one, are the dimensions correlated or independent?

Recently, Dohrenwend and his colleagues have argued for a single-factor interpretation of screening instruments like those the MHI is based on, calling it "demoralization" (Dohrenwend, Oskenberg, Shrout, Dohrenwend, & Cook, 1981; Link & Dohrenwend, 1980). Contrary to this position, a number of studies have presented evidence for dimensionality within the psychological distress dimension (Costello & Comrey, 1967; Derogatis, Lipman, Covi, & Rickels, 1971; Derogatis, Lipman, Rickels, Chlenhuth, & Covi, 1974; Ware et al., 1979; Edwards, Yarvis, Mueller, Zingale, & Wagman, Note 1).

Arguments for independent positive and negative dimensions have also been offered. Bradburn (1969) has interpreted high interitem correlations within groups of positive and negative items and low correlations between positive and negative items to mean that his Affect Balance Scale defines two distinct and independent (uncorrelated) dimensions—positive and negative affect. Factor analytic evidence for a psychological well-being dimension distinct from dimensions of psychological distress is also abundant from studies of other instruments containing both positive and negative items (Goldberg, Steele, Johnson, & Smith, 1982; Radloff, 1977; Ware et al., 1979).

Both the construction of the MHI and the techniques selected to analyze the data were geared to better understand the dimensionality and independence issues discussed above. Positive and negative items were included to address the issue of whether psychological distress and well-being are distinct dimensions as defined by the MHI and whether these dimensions in turn are multidimensional (i.e., whether they should be considered higher order mental health factors). Confirmatory factor analytic procedures provided us with an objective basis for testing between alternative measurement models and structural hypotheses. For practical applications of the MHI, we were also interested in whether satisfactory reliability could be achieved with simple summated rating procedures for estimating a summary MHI score and scores on the separate factors.

Whereas issues concerning models of mental health must ultimately be resolved in terms of which model best predicts mental health behavior and other important concepts, a psychometric prerequisite for using a multidimensional specification for purposes of prediction is the extraction and cross-validation of distinct mental health factors. Our concern here is with the psychometric basis of such a model.

Method

Construction of the MHI

The MHI was based substantially on the General Well-Being Schedule (GWB) developed by Dupuy (1972, Notes 2–5). The GWB represents several psychological distress constructs and includes items measuring well-being. Its

[1] The HIE is a social experiment designed to estimate the effects of different health-care-financing arrangements on the demand for health care services as well as on health status and patient satisfaction outcomes. Families are assigned randomly to plans differing in coinsurance and deductible arrangements in the fee-for-service system or in a prepaid health plan. Their use of services and changes in health and other variables are monitored over time (Newhouse, 1974; Newhouse, Manning, Morris et al., 1981).

content emphasizes symptoms of psychological distress that are most prevalent in general populations (e.g., anxiety and depression). Hence, the GWB provided a good starting point for constructing the MHI. To address adequately the dimensionality issues left unresolved by previous studies, it was necessary to add numerous other items. These items were added and tested in two phases.

In the first phase, the factor structure of a 22-item version of the GWB was evaluated (Ware et al., 1979; Dupuy, Note 2). The hypothesis of six underlying factors—Anxiety, Depression, General Health, General Positive Affect, Loss of Behavioral/Emotional Control, and Vitality—was clearly supported. Correlations between four of these factors (Anxiety, Depression, General Positive Affect, and Loss of Behavioral/Emotional Control) and numerous other criterion variables further supported their validity. As an inspection of their content anticipated, the GWB General Health and Vitality factors clearly failed discriminant tests of validity. Specifically, they correlated substantially with both mental and physical health factors (Ware et al., 1979, 1980b). Hence, these two factors were eliminated from the MHI.

In the second phase, 15 GWB items showing good discriminant validity were supplemented with 20 additional items based on instruments developed by others (Beck, 1967; Costello & Comrey, 1967; Comrey, 1970; Dohrenwend, Shrout, Egri, & Mendelsohn, 1980) to enhance measurement of these four GWB mental health factors. Three items identified by Ware et al. (1979) to represent a fifth hypothesized factor, Emotional Ties, were also added to the battery. The first column in Table 1 provides a short description of each item (items are grouped according to final conclusions of this article); Column 2 identifies each item's originally hypothesized construct.

Sample Characteristics and Data-Gathering Methods

The MHI was fielded at all six HIE sites: Dayton, Ohio; Seattle, Washington; Fitchburg and Franklin County, Massachusetts; Charleston and Georgetown County, South Carolina.[2] Characteristics of respondents (N = 5,089) are summarized in Table 2. With one exception, data for our analyses came from a self-administered questionnaire that was fielded when the HIE began. The exception was Dayton, where MHI data came from surveys fielded after the experiment began. Longitudinal data for the 1-year stability analyses came from identical versions of the MHI fielded approximately 1 year apart to a subsample at each site (N = 3,525). Questionnaires were mailed out and mailed back and respondents were compensated for completing them (see Ware et al., 1980a for additional details). Interviewer assistance was provided in person or by telephone when needed. These procedures produced high return rates (above 90% at all sites) and very few missing responses.

Data Analysis

Our analyses addressed the following questions: (a) Does a single mental health factor underlie the data? (b) Is there empirical support for a two-factor (Psychological Distress and Psychological Well-Being) higher order model? If so, are the two dimensions correlated or independent? (c) Is there empirical support for the five lower-order factors that were built into the instrument (see Table 1)? (d) Do models of increasing complexity do significantly better than simpler models in accounting for the data? (e) Can the conclusions based on one site be generalized across the other three sites? (f) Is sufficient reliability achieved using a simple summated ratings method to estimate factor scores? and (g) Are scores stable across a 1-year interval?

These analyses were conducted in four stages. First, we selected one site to address questions a through d above. Second, we performed independent cross-validation tests of the derived factor structure in each of the three remaining HIE sites. Because results were very consistent across all four sites, we also tested the alternative factor structures and estimated model parameters on data combined over all four sites. Third, we evaluated the reliability of factor scores obtained by the simple summated ratings procedure. Finally, we estimated the stability of the MHI score and subscale scores over a 1-year interval.

Model development. We selected Seattle for the initial analyses because the MHI was first administered there and because it had the largest number of respondents. A series of analyses was performed on data obtained in this site. First, we performed factor analyses on two separate correlation matrices in Seattle to assess whether the items shown in Table 1 define an underlying mental health dimension. Factor analyses were performed using SPSS software (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975). In one analysis, we used a principal components factor extraction and a Varimax factor rotation to analyze a matrix of correlations among 38 MHI and 19 physical functioning items. The physical functioning items are described elsewhere (Stewart, Ware, & Brook, 1981). The idea behind this analysis was that, if the hypothesis of a distinct underlying mental health dimension is appropriate, items in the MHI should emerge on a separate factor than those selected to define physical functioning. In the other analysis, a principal components solution for just the 38 MHI items provided an assessment of the proportion of variance in the interitem correlation matrix accounted for by the first extracted factor. The proportion should be substantial and the magnitude of loadings roughly consistent across items to support the hypothesis of a general underlying mental health component.

Second, we used factor analyses (principal components factor extraction, Varimax rotation, and a Promax oblique rotation) as an aid in determining item clusters for the two- and five-factor models. Conclusions about item clusters were determined as follows. Generally, an item was hypothesized to define the factor on which it had the highest factor loading. Items that had good convergent validity but poor discriminant validity[3] were still hypothesized to define their originally hypothesized construct (see Table 1) on the strength of item content (face validity) and previous empirical data on the item.[4]

[2] The two sites within Massachusetts and the two within South Carolina were combined to achieve a high subjects-to-variables ratio in the site-specific analysis.

[3] Convergent validity was judged satisfactory when an item's correlation with its hypothesized factor equaled or exceeded .40. The discriminant validity of an item was considered poor when the difference in magnitude between its highest and second highest factor loading was less than .15.

[4] At this stage in the analysis we also studied the effects of two kinds of response set (acquiescence and socially

Table 1
Summary Information About Mental-Health Inventory Items

Factor groupings/item content[a]	A priori scale hypothesis	Item mean[b]	Item *SD*
Anxiety (A)			
Very nervous person	A	4.96	1.12
Bothered by nervousness	A	5.14	.99
Felt tense or high-strung	A	4.79	1.05
Anxious, worried	A	4.78	1.04
Difficulty trying to calm down	A	5.04	.95
Nervous or jumpy	A	4.46	1.12
Restless, fidgety, impatient	A	4.68	1.04
Rattled, upset, flustered	A	4.52	.91
Hands shake when doing things	A	5.46	.90
Relax without difficulty[c]			
Depression (D)			
Moody, brooded about things	D	4.90	.95
Low or very low spirits	D	5.01	.88
Felt downhearted and blue	D	4.93	.97
Felt depressed	D	4.10	.71
Strain, stress, pressure[c]	A	4.48	1.25
Loss of Behavioral/Emotional Control (B)			
Control behavior, thoughts, feelings	B	5.00	.96
Concern about losing control of mind	B	5.54	.91
Felt emotionally stable	B	4.93	1.23
Nothing turns out as wanted	D	4.48	1.10
Felt like crying	D	5.00	1.05
Better off if dead	D	5.76	.66
Down in the dumps	D	5.26	.90
Think about taking own life	D	4.91	.40
Nothing to look forward to	D	5.20	1.07
General Positive Affect (G)			
Happy person	G	4.41	1.03
Happy, satisfied, or pleased	G	4.27	1.03
Daily life interesting	G	4.00	1.17
Felt calm and peaceful	G	4.07	1.19
Felt cheerful, lighthearted	G	4.10	1.14
Generally enjoyed things	G	4.43	1.03
Relaxed and free of tension	A	3.83	1.20
Living a wonderful adventure	G	4.29	1.39
Expect an interesting day	G	4.09	1.23
Wake up fresh, rested	G	3.95	1.14
Future hopeful, promising	G	4.18	1.34
Emotional Ties (E)			
Felt loved and wanted	E	4.69	1.28
Love relations full, complete	E	4.42	1.49
Time felt lonely[c]	E	5.09	1.04

Note. All items in this table are positively scored.
[a] For the two-factor solution, all of the items on the A, B, and D factors and the "time felt lonely" item were hypothesized a priori to load on Psychological Distress; remaining items were hypothesized to load on Psychological Well-Being. This was also the definition of these two factors in the final two-factor solution.
[b] Means and standard deviations are based on a combined sites analysis, $N = 5{,}089$.
[c] These items had secondary loadings in the final five-factor solution and were not used to score subscales.

desirable) by including direct measures of these response tendencies. They did not seem to influence our conclusions about the structure of the MHI and were dropped from further consideration (see Ware, Veit, & Donald, in press, for further details).

Third, the LISREL software for testing structural equation models (Joreskog & Sorbom, 1981) was used to assess how well the one-, two-, and five-factor models accounted for the MHI data; for the two- and five-factor models, we examined both correlated and uncorrelated factor structures. LISREL was also used to test for differences in fit

Table 2
Summary of Respondent Characteristics

Site	N	% male	% nonwhite	Age Range	Age M	No. of school years completed Range	No. of school years completed M	Family income[a] Range ($)	Family income[a] M ($)
Seattle, Wash.	1,755	48	4	13–61	32.0	4–25	12.8	0–22,240	9,536
Dayton, Ohio	1,046	46	10	14–69	33.8	3–24	12.7	0–23,850	12,465
Fitchburg/Franklin County, Mass.	1,041	46	2	13–66	32.6	2–22	12.4	0–18,158	7,876
Charleston/Georgetown County, S.C.	1,247	44	43	13–59	31.0	0–27	11.3	0–45,454	8,862
All sites combined	5,089	46	15	13–69	32.2	0–27	12.4	0–45,454	9,636

[a] Expressed in 1973 dollars for Dayton, and in 1974 dollars for all other sites.

among the most viable models resulting from these assessments.[5] All of the analyses described above were performed using data from Seattle respondents only (N = 1,755).

Cross-validation analyses. The generalizability of the Seattle factor structure to the other three sites was evaluated by testing for significant differences within those sites among the viable models. The idea was that a generalizable conclusion was one where the relative "goodness" of fits of the competing models remained invariant across sites. Such a result would justify using a standard set of scoring rules in future studies. These evaluations were performed using confirmatory factor analysis.

Reliability of summated rating scales. Internal-consistency reliability was evaluated for scale scores by estimating coefficient alpha (Cronbach, 1951). Total scale scores were computed using the summated ratings procedure (Likert, 1932). We also computed Pearson product-moment correlations among the summated scale scores and compared their magnitudes with the reliability estimates. Intercorrelations among the scales (factors) should be substantially lower than the reliability estimates to conclude that the scales actually measure different constructs. Data from all four sites were combined for these analyses (N = 5,089).

Stability of factor structure. Pearson product-moment correlations were used to evaluate the stability of the MHI total score and the scale scores of the final higher and lower order factor solutions over a 1-year interval for a subset of respondents (N = 3,525).

Results

Model Development in Seattle

Using data from the Seattle site, we initially examined the hypothesis that the MHI defined a general mental health factor. Then, we assessed the plausibility of one-, two-, and five-factor orthogonal and oblique structures using factor analysis. After item clusters had been initially determined from the factor analyses, we tested the competing factor solutions using confirmatory factor analysis.

Testing for an Underlying Mental-Health Component

A two-factor solution for physical functioning and mental health items with orthogonal rotation produced a simple structure with the mental health and physical functioning items clustering on separate factors. The smallest difference in magnitude between an item's factor loadings on the physical and mental factors was about .30, indicating good discriminant validity for all of the items. These results, which are documented elsewhere (Ware, Veit, & Donald, in press), support our hypothesized distinction between mental and physical functioning concepts and items.

In the principal components analysis of just the 38 mental health items, about 43% of the variance in the correlation matrix was accounted for by the first extracted factor in Seattle. Loadings on this factor ranged from .42 to .80, indicating that the first factor accounted for a substantial proportion of the variance in each item.[6] This result supports

[5] Decisions on fixed parameters and estimated parameter input values for each confirmatory factor analysis were based on the corresponding exploratory factor analytic solution and the assumption of a simple structure model.

[6] Items with the lower loadings tended to be the more extremely worded items that had highly skewed distributions and restricted variances. They were retained because they were found to have good discriminant validity in other research.

the notion of a general underlying mental health factor; at the same time, it calls attention to the fact that a substantial proportion of the variance would be left unexplained by a unidimensional model.

Two- and Five-Factor Orthogonal and Oblique Solutions

The total matrix variance explained increased from 43% to 50% when a second factor was extracted, and to 60% for a five-factor solution.[7] The item groupings indicated by the factor loadings observed in the two-factor solution corresponded exactly to those predicted by our two-dimensional (psychological distress and well-being) hypothesis. These hypothesized item groupings are defined in a footnote to Table 1. All items hypothesized to measure Anxiety, Depression, and Loss of Behavioral/Emotional Control correlated highest with the psychological distress factor. All items hypothesized to measure General Positive Affect and Emotional Ties correlated highest (with the exception of "time felt lonely") with the psychological well-being factor. Further, all items exceeded our .15 discriminant validity criterion in the two-factor solution, providing support for a two-dimensional specification of mental health based on unipolar psychological distress and well-being factors.

The five-factor solution provided a basis for a five-dimensional specification of mental health. In that factor solution, 13 of the 38 items did not meet our .15 discriminant criterion. These items were grouped with those items defining their originally hypothesized factor for confirmatory analyses. The resultant item groupings are displayed in Column 1 of Table 1. As can be seen by comparing the item groupings in Column 1 with our hypothesis in Column 2 of Table 1, there was a high degree of correspondence between our five-factor hypothesis and the pattern of item correlations for three of the five factors—Anxiety, General Positive Affect, and Emotional Ties. However, it is apparent that the depression and loss of behavioral/emotional control items did not correlate with their respective factors as hypothesized. In previous research, similar items have been grouped into a single depression scale. This led us to retain (for our confirmatory analyses) a four-factor model that combined the depression and loss of behavioral/emotional control factors (Table 1) to get a better idea of which solution—a four- or five-factor oblique model—better defined the MHI lower order factor structure.

The interfactor correlations for oblique solutions were substantial (ranging from .38 to .59 in absolute magnitude), suggesting that models with correlated factors should be tested in the confirmatory analyses. Complete tables of results from these solutions are reported elsewhere (Ware et al., in press).

Confirmatory Tests Among Seven Models

Chi-square goodness-of-fit estimates were obtained for seven models—single-factor, two-factor orthogonal and oblique, four-factor orthogonal and oblique, and five-factor orthogonal and oblique. Goodness of fit indexes for these models in the Seattle site are shown in Table 3.

The chi-squares in Table 3 indicate that none of the models accounts for the data without significant deviations. However, the chi-square statistic is a direct function of sample size and thus, the probability of rejecting a model increases as sample size increases. Since we are dealing with a very large sample, a "poor" fit to almost any model would be expected (Bentler & Bonett, 1980; Joreskog & Sorbom, 1979). Thus, better assessments of the models would result from inspection of the number of noteworthy residual correlations and from differences in the incremental fits of the models.[8]

The number of residual correlations greater than or equal to .10 and the magnitude of the largest residual correlation are noted in Columns 5 and 6, respectively, of Table 3. The

[7] The only difference between the orthogonal and oblique rotations for the two- and five-factor solutions was that the difference in magnitude among an item's factor loadings was greater for the oblique solutions. The sixth, seventh, and eighth factors were relatively small, accounting for only an additional 2% of the variance each. The interpretability of the five-factor solution was clearly more straightforward than solutions with more factors. The effects of these additional factors on the interpretability of the structure is discussed in detail elsewhere (Ware et al., in press).

[8] A residual correlation is obtained by subtracting the correlation between two items predicted by the model from the actual obtained correlation.

Table 3
Goodness-of-Fit Indexes and Residuals for Seven Models: Seattle (N = 1,755)

Model	*df*	χ^2	χ^2 (df)	No. of residuals ≥ .10	Maximum residual correlation
H1	665	10495.00	15.78	88	.42
H2O	665	9353.49	14.07	348	.65
H2C	664	7829.30	11.79	58	.35
H4O	665	9809.81	14.74	513	.65
H4C	659	5685.58	8.63	71	.28
H5O	665	11305.88	17.00	457	.65
H5C	655	5412.6	8.26	69	.28

Note. All differences between models are statistically significant ($p < .001$).
H1 is a single-factor model; H2O is a two-factor orthogonal model; H2C is a two-factor correlated model; H4O is a four-factor orthogonal model that combines items shown under Behavioral/Emotional Control and Depression in Table 1; H4C is a four-factor correlated model; H5O is a five-factor orthogonal model; H5C is a five-factor correlated model.

number of large residual correlations was substantially greater for the orthogonal than the oblique solutions. For example, in the two-factor orthogonal solution, 348 of the 703 residuals were greater than or equal to .10, with a high of .65 (see Row 2). This number was reduced to 58 (with a maximum residual of .35) when the factors were allowed to correlate (see Row 3). Even larger reductions in number and magnitude of noteworthy residuals were observed when factors in the four- and five-factor models were allowed to correlate. (Note that the five-factor uncorrelated model accounted for the data noticeably less well than the single-factor model.) A comparison of residuals between the orthogonal models and their respective correlated solutions points to the importance of estimating correlations among mental health factors. The chi-square difference in goodness of fit between each orthogonal model and its oblique counterpart was highly significant ($p < .001$; see Table 3). These results argue for rejecting all orthogonal specifications.

In Panel 1 of Table 4, we present differences in goodness of fit among the remaining three correlated models. The single-factor model is used as the basis for comparison.[9] Column 4 shows the chi-square difference statistics for these comparisons. Each one of the chi-square differences was highly significant ($p < .001$), indicating significant increases in information with the addition of each correlated factor. The index in the last column of Table 4 provides information about the incremental fit among the models, independent of sample size (Bentler and Bonett, 1980).[10] As can be seen from the last row in Panel 1 of this column, the five-factor oblique model picked up about 48% more information in the correlation matrix than the single-factor model (see Row 4). However, the five-factor oblique model contributed only a 3% increase in amount of information over the four-factor oblique model, which combined the Loss of Behavioral Emotional Control and Depression factors (see Row 3). Despite this small increase, both the four- and five-factor oblique models were retained for cross-validation tests to better evaluate the best lower order MHI factor structure.

Cross-Validation Tests

Comparisons among the four models derived and tested in Seattle (Panel 1 of Table 4) were performed independently in the Dayton, South Carolina, and Massachusetts sites, where respondent characteristics were quite different. A summary of these tests is presented in Panels 2 through 4 of Table 4. In all sites, the order of goodness of fit of the four models was the same. A two-factor oblique model added from 26% (Dayton) to 29% (South Carolina) additional information in the correlation matrix over a single-factor model; from 8% (South Carolina) to 16% (Dayton) was gained with a four-factor oblique model over the two-factor model; and a five-factor oblique model gained from 2% (South Carolina) to 4% (Dayton) more information than a four-factor oblique model. All of the chi-square differences were highly significant ($p < .001$). The five-factor oblique model showed gains of from

[9] The oblique models are all special cases of the more restricted single-factor model. A procedure for distinguishing among a set of nested models is to evaluate the *difference* in goodness of fit (Bentler & Bonett, 1980; Joreskog & Sorbom, 1979).

[10] The statistic shown in Column 6 of Table 4 uses the chi-square associated with the single-factor model (used as a baseline for comparison here) in the denominator; the difference in chi-squares between the models being compared serves as the numerator.

39% (South Carolina) to 45% (Dayton) in information over the single-factor model. Because the five-factor oblique model consistently did significantly better than the four-factor model in accounting for the data, we rejected the four-factor specification as the lower order factor solution.

Combined Sites Analyses

The consistency of the results across sites justified a summary combined-sites analysis. Results of chi-square difference tests among the four competing models in the combined-sites analyses are shown in the last panel of Table 4. As would be expected, these tests produced similar results to those obtained in the four sites separately (compare with Panels 1–4 of Table 4). The five-factor oblique model did best in accounting for the MHI data, capturing 49% more information in the correlation matrix than the single-factor model.

A Five-Factor Overlapping Model

Up to this point, our models have assumed that each item measured a single factor. The substantial interfactor correlations found across sites and the number and pattern of residual correlations exceeding .10 even with the best model led us to explore a five-factor model that allowed some items to load on more than one factor. Residuals in the correlation matrix resulting from our best model (i.e., the five-factor oblique solution) were helpful in identifying items that might be appropriately overlapped. In this matrix, 45 residual correlations were greater than or equal to .10, and the largest residual was .26. Although these noteworthy residuals were scattered among many different items, a few items were involved in a disproportionate number. Many residuals involved depression and behavioral/emotional control items. Close examination of the residuals led us to reason that a better

Table 4
Comparisons of Fit Between Models

Site	Model comparison	Chi-square difference[a]	*df*	Δ
Seattle (N = 1,755)	H1–H2C	2665.7	1	.25
	H2C–H4C	2143.7	5	.20
	H4C–H5C	273.0	4	.03
	H1–H5C	5082.4	10	.48
Dayton (N = 1,046)	H1–H2C	1433.9	1	.26
	H2C–H4C	910.6	5	.16
	H4C–H5C	201.6	4	.04
	H1–H5C	2546.1	10	.45
South Carolina (N = 1,247)	H1–H2C	1891.7	1	.29
	H2C–H4C	512.9	5	.08
	H4C–H5C	103.7	4	.02
	H1–H5C	2508.2	10	.39
Massachusetts (N = 1,041)	H1–H2C	1676.2	1	.27
	H2C–H4C	896.8	5	.15
	H4C–H5C	161.5	4	.03
	H1–H5C	2734.5	10	.45
Combined sites (N = 5,089)	H1–H2C	7540.5	1	.30
	H2C–H4C	4105.3	5	.16
	H4C–H5C	705.7	4	.03
	H5C–H5C′	1826.5	4	.07
	H1–H5C′	14178.0	14	.56

Note. H1 is a single-factor model; H2C is a correlated two-factor model; H4C is a correlated four-factor model; H5C is a correlated five-factor model; H5C′ is a correlated five-factor model with secondary loadings for three items: the "strain" Depression item, the "able to relax" Anxiety item, and the "lonely" Emotional Ties item. Recall that the "lonely" item had secondary loadings on the Depression and Behavioral/Emotional Control factors.
[a] All chi-square comparisons were statistically significant ($p < .001$).

solution might be one that allowed three items (starred in Table 1) to have secondary factor loadings. An emotional ties item ("time felt lonely") was permitted a secondary loading on both Depression and Loss of Behavioral Emotional Control; an anxiety item ("relax without difficulty") was permitted a secondary loading on General Positive Affect; and a depression item ("strain, stress, pressure") was permitted a secondary loading on Anxiety.

The five-factor oblique model that allowed estimates of these four additional parameters was compared with the five-factor oblique model in the combined-sites data. This overlapping model reduced the number of residual correlations greater than or equal to .10 from 45 to 21; the size of the maximum residual decreased from .26 to .17.

The third row of the last panel of Table 4 shows the chi-square difference statistic for the simple five-factor oblique model versus the five-factor overlap oblique model. As can be seen, the five-factor overlap model picks up 7% more information in the MHI correlation matrix than the nonoverlapping five-factor oblique model; it represents an improvement over the baseline single-factor model of 56% (Row 5 of the last panel).[11] Estimates of factor loadings for our final five-factor overlap model in the combined sites analysis are shown in Table 5.

Summary of Results

The cross-validational and combined-sites analyses led us to conclude that the best interpretation of the MHI is a hierarchical factor model composed of a general underlying mental health factor, a higher order factor structure, and a lower order factor structure as illustrated in Figure 1. The higher order factor structure is defined by two correlated factors—Psychological Distress, which consists of all the items that describe negative mental health states, and Psychological Well-Being, which consists of all of the items describing positive mental health states. Thus, both factors are unipolar in nature. The lower order factor structure is defined by five correlated factors with three of the 38 items defining more than one factor. Item groupings for this lower order factor structure are shown in the first column of Table 5.

Table 6 presents summary statistics for summated rating scales based on the five-factor overlap model (top panel), the two-factor model (middle panel), and the single-factor model, that is, the MHI (lower panel). The items that we allowed to overlap in the final five-factor model (items with a superscript "c" in Table 1) were not used when estimating scores for the five factors derived from this model. Scores based on the one- and two-factor models were estimated using all 38 items. (See footnote in Table 1 for items comprising the final two-factor solution.) Hereafter, we refer to the factors as "scales."

Reliability, Stability, and Scale Intercorrelations

Table 5 also presents reliability and stability coefficients and intercorrelations among scales based on the final one-, two-, and five-factor models. The reliability (internal consistency) estimates for all scales are more than satisfactory for group comparisons, ranging from .83 to .91 for scales based on the five lower order factors and from .92 to .96 for scales based on the two higher order factors and the MHI. The stability coefficients are in the .56 to .64 range, indicating that a substantial proportion of the reliable variance in these scales is stable over a 1-year interval.

The top panel of Table 5 presents correlations among the five MHI subscales. As can be seen, these range in absolute magnitude from .34 between Emotional Ties and Anxiety to .75 between Anxiety and Depression. All of these intercorrelations are lower than the reliability coefficients for the five subscales, which indicates that the subscales contain a noteworthy amount of unique reliable variance.

Discussion and Conclusions

Our analyses produced a clear and consistent picture of the factor structure of the Men-

[11] Another "null" model that may be of interest is one that specifies 38 uncorrelated factors each measured by one item, that is, all zeros in the off-diagonals of the input correlation matrix. In the combined-sites analysis, our single-factor model picked up 77% more information in the correlation matrix than this "null" model that assumes correlations among the items are zero.

Table 5
Factor Loading Estimates for the Final Model, Combined Sites (N = 5,089)

Factor groupings/item content	A	D	B	G	E
Anxiety (A)					
Very nervous person	.80	—[b]	—	—	—
Bothered by nervousness	.80	—	—	—	—
Felt tense or high-strung	.77	—	—	—	—
Anxious, worried	.74	—	—	—	—
Difficulty trying to calm down	.74	—	—	—	—
Nervous to jumpy	.69	—	—	—	—
Restless, fidgety, impatient	.69	—	—	—	—
Rattled, upset, or flustered	.68	—	—	—	—
Hands shake when doing things	.50	—	—	—	—
[a]Relax without difficulty	−.29[c]	—	—	.46	—
Depression (D)					
Moody, brooded about things	—	.74	—	—	—
Low or very low spirits	—	.83	—	—	—
Downhearted and blue	—	.82	—	—	—
Felt depressed	—	.74	—	—	—
[a]Strain, stress, pressure	.51	.14	—	—	—
Behavioral/Emotional Control (B)					
Control of behavior, thoughts, feelings	—	—	.60	—	—
Concern about losing control of mind	—	—	.58	—	—
Felt emotionally stable	—	—	.55	—	—
Nothing turns out as wanted	—	—	.57	—	—
Felt like crying	—	—	.66	—	—
Better off if dead	—	—	.59	—	—
Down in the dumps	—	—	.75	—	—
Think about taking own life	—	—	.44	—	—
Nothing to look forward to	—	—	.69	—	—
General Positive Affect (G)					
Happy person	—	—	—	.83	—
Happy, satisifed, or pleased	—	—	—	.72	—
Daily life interesting	—	—	—	.67	—
Felt calm and peaceful	—	—	—	.79	—
Felt cheerful, lighthearted	—	—	—	.77	—
Generally enjoyed things	—	—	—	.75	—
Relaxed and free of tension	—	—	—	.74	—
Living a wonderful adventure	—	—	—	.71	—
Expect an interesting day	—	—	—	.67	—
Wake up fresh, rested	—	—	—	.61	—
Future hopeful, promising	—	—	—	.58	—
Emotional Ties (E)					
Felt loved and wanted	—	—	—	—	.85
Love relations full, complete	—	—	—	—	.82
[a]Time felt lonely	—	.18	.29	—	−.26

[a] These items were overlapped in the final five-factor solution.
[b] These loadings were fixed at zero.
[c] It should be noted that although the overlap model substantially reduced the number of noteworthy residuals, the secondary loadings are smaller than our covergent validity criterion of .4.

tal Health Inventory (MHI). This structure is a blueprint for scoring the MHI in future studies. The analyses also clarified some of the theoretical issues debated in the literature on general population mental health surveys.

The MHI Structural Model

Both the higher and lower order factor dimensions identified in the exploratory and confirmatory factor analyses corresponded

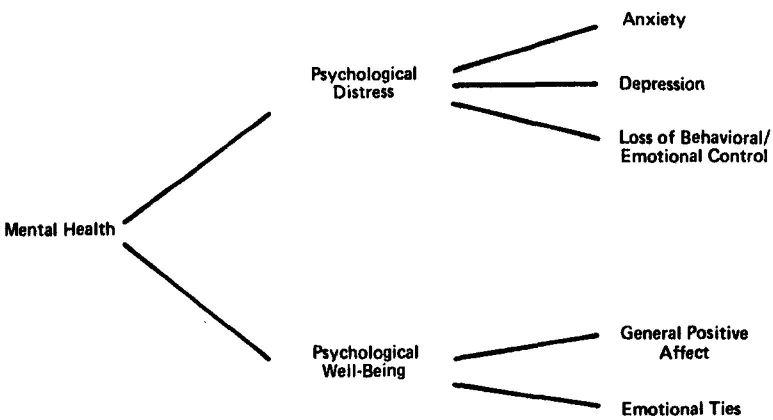

Figure 1. Mental Health Inventory structure.

well with those originally hypothesized. However, some discrepancies between the hypothesized and obtained five-factor structure were observed.

Figure 1 illustrates several of our major conclusions. First, a large mental health factor underlies the MHI. Hence, there is a sound psychometric basis for using the single MHI summary score to define a bipolar psychological distress versus well-being concept. Second, reliance on a single score is associated with significant loss of information. We observed substantial gains with a two-dimensional specification of mental health. As hypothesized, items describing positive states clustered together to define psychological well-being and items describing negative states clustered together to define psychological distress. Of importance here is that these factors are clearly distinct. This result supports the practice of scoring them as separate factors (Bradburn, 1969; Goldberg, 1978; Radloff, 1977; Ware et al., 1979). Third, we consistently identified structures within both psychological well-being

Table 6
Summary Information About Mental Health Index Scales[a]

						Scale intercorrelations						
Scale	k[b]	*M*	*SD*	Reliability[c]	Stability[d]	A	D	B	G	E	PD	PWB
Anxiety (A)	9	19.15	6.85	.90	.63							
Depression (D)	4	8.05	2.97	.86	.56	.76						
Loss of Behavioral Emotional Control (B)	9	15.90	5.57	.83	.58	.71	.77					
General Positive Affect (G)	11	45.64	9.56	.92	.62	−.62	−.70	−.69				
Emotional Ties (E)	2	9.08	2.56	.81	.59	−.39	−.50	−.53	.62			
Psychological Distress (PD)	24	47.54	15.39	.94	.62	.93[e]	.89[e]	.90[e]	−.73	−.52		
Psychological Well-Being (PWB)	14	59.16	12.16	.92	.63	−.63	−.71	−.71	.98[e]	.74[e]	−.75	
Mental Health Index	38	177.56	25.46	.96	.64	−.85[e]	−.86[e]	−.87[e]	.90[e]	.66[e]	−.95[e]	.92[e]

[a] Scale scores are computed by summing scores over all items defining the scale (unlike Table 1, some item scores were reversed, in scoring scales, so that a high score would be consistent with each scale name).
[b] Number of items in scale.
[c] Internal-consistency reliability estimated using Cronbach's alpha (N = 5,089).
[d] Pearson product-moment correlation between scores obtained approximately 1 year apart (N = 3,525).
[e] Correlations inflated due to overlapping definitions.

and psychological distress dimensions. This led us to the hierarchical MHI model shown in Figure 1. These higher order psychological distress and well-being factors were defined by three and two lower order factors, respectively. Fourth, the correlation between distress and well-being was substantial, as were the intercorrelations among the lower order factors. This hierarchical model was confirmed in cross-validation tests using samples of populations with quite different characteristics; this supports the model's generalizability.

Issues Debated in the Literature

Dohrenwend and his colleagues (Dohrenwend et al., 1981; Link & Dohrenwend, 1980) have argued for a single-factor interpretation for the GWB (the basis of the MHI) and other instruments like the MHI. Their position is in disagreement with the results of numerous studies of similar mental health instruments (Bradburn, 1969; Cooke, 1980; Costello & Comrey, 1967; Derogatis et al., 1971; Goldberg, 1978; Radloff, 1977; Wan & Livieratos, 1978; Ware et al., 1979). Our results from the MHI data are quite clear on this issue. Our five-factor overlap model provided a 56 percentage point increase in information gained in the intercorrelation matrix over a single-factor model. This substantial gain in information with a multidimensional specification of mental health was replicated in every site. Further, the observed high reliability and stability coefficients (relative to interscale correlations) leaves little doubt that each of the MHI subscales contains reliable information not contained in the other scales. These findings constitute a strong psychometric basis for a multidimensional specification of mental health as defined by the MHI.

Another important issue is whether mental health factors are correlated or orthogonal. Bradburn (1969) concluded that his Affect Balance Scale (ABS) could best be described by a two-factor model with independent positive and negative factors. We find no support for such a model. Specifically, in our analyses of higher and lower order factor structures, orthogonal models did substantially worse than their oblique counterparts in explaining the variance in interitem correlations in all four sites. Differences in data-analytic strategies and other methodological factors appear to be one reason for the discrepancy. A reanalysis of correlations among ABS items using confirmatory factor analysis led us to reject an orthogonal model of the ABS in favor of one specifying correlated factors (Ware et al., in press). We found a correlation of $-.285$ between ABS positive and negative factors. A two-dimensional correlated model of the ABS accounted for 15% more information in the correlation matrix than a two-dimensional orthogonal model. This difference was highly significant ($p < .001$). These results argue for a correlated rather than an orthogonal two-factor model for the ABS.

The hierarchical MHI model provides interesting scoring options ranging from five distinct mental health constructs to reliance on a single summary index. An intermediate scoring option is two higher order psychological distress and well-being constructs. The tradeoff is between the unique information contained in the subscales versus the simplicity of a single score. Ultimately, the basis for choosing among these options would be which model best predicts mental-health-related behaviors. In a test between our one- and two-factor specification of the MHI, Manning, Newhouse, and Ware (1982) rejected a single- in favor of a two-factor specification (psychological distress and psychological well-being) for predicting ambulatory medical expenditures. Further research is needed to determine whether a five-factor model predicts mental-health-related behaviors significantly better than a single- or two-factor model and whether the gains are worth the additional complexity.

Reference Notes

1. Edwards, D. W., Yarvis, R., Mueller, D. P., Zingale, H., & Wagman, W. J. *Test-taking and the stability of adjustment scales: Can we assess patient deterioration?* University of California at Davis, Department of Psychiatry, 1977.
2. Dupuy, H. J. *Developmental rationale, substantive derivative, and conceptual relevance of the General Well-Being Schedule.* National Center for Health Statistics, Fairfax, Virginia, June, 1973.
3. Dupuy, H. J. *Utility of the National Center for Health Statistics General Well-Being Schedule in the assessment of self-representations of subjective well-being and distress.* National Center for Health Statistics, Fairfax, Virginia, June, 1974.
4. Dupuy, H. J. *Self-representations of general psychoical well-being of American adults.* Paper presented at the annual meeting of the American Public Health Association, Los Angeles, October, 1978.

5. Dupuy, H. J. *A brief description of the research edition of the General Psychological Well-Being Schedule (GWB).* National Center for Health Statistics, Fairfax, Virginia, May, 1979.

References

Beck, A. T. *Depression—Causes and treatment.* University of Pennsylvania Press, 1967.

Bentler, P. M., & Bonett, D. G. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin,* 1980, *88,* 588–606.

Bradburn, N. M. *The structure of psychological well-being.* Chicago: Aldine, 1969.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika,* 1951, *16,* 297–494.

Cleary, P. D., Goldberg, I. D., & Kessler, L. G. Screening for mental disorder among primary care patients. *Archives of General Psychiatry,* 1982, *39,* 837–840.

Cooke, D. J. The structure of depression found in the general population. *Psychological Medicine,* 1980, *10,* 455–463.

Costello C. G., & Comrey A. L. Scales for measuring depression and anxiety. *The Journal of Psychology,* 1967, *66,* 303–313.

Comrey, A. L. *Comrey personality scales.* San Diego, Calif.: Educational and Industrial Testing Service, 1970.

Cronbach, L. J. Coefficient alpha and the internal structure of tests, *Psychometrika,* 1951, *16,* 297–334.

Derogatis, L. R., Lipman, R. S., Covi, L., & Rickels, K. Neurotic symptom dimensions. *Archives of General Psychiatry,* 1971, *24,* 454–464.

Derogatis, L. R., Lipman, R. S., Rickels, K., Chlenhuth, E. H., & Covi, L. The Hopkins symptom checklist (HSCL): A self-report inventory. *Behavioral Science,* 1974, *19,* 1–15.

Dohrenwend, B. P., Oskenberg L., Shrout, P. E., Dohrenwend, B. S., & Cook, D. What brief psychiatric screening scales measure. *Health survey research methods* (DHHS Publication No. PHS 81-3268). Washington, D.C.: U.S. Government Printing Office, 1981.

Dohrenwend, B. P., Shrout, P. E., Egri, G., & Mendelsohn, F. S. Nonspecific psychological distress and other dimensions of psychopathology. *Archives of General Psychiatry,* 1980, *37,* 1229–1236.

Dupuy, H. J. The psychological section of the current health and nutrition examination survey. *Proceedings of the Public Health Conference on Records and Statistics Meeting Jointly with the National Conference on Health Statistics.* Washington, D.C.: National Conference on Health Statistics, 1972.

Goldberg, D. *Manual of the General Health Questionnaire.* Windsor, England: NFER Publishing Co., 1978.

Goldberg, D., Steele, J. J., Johnson, A., & Smith, C. Ability of primary care physicians to make accurate ratings of psychiatric symptoms. *Archives of General Psychiatry,* 1982, *39,* 829–833.

Gurin, G., Veroff, J., & Feld, S. *Americans view their mental health—A nationwide interview survey.* New York: Basic Books, 1960.

Joreskog, K. G., & Sorbom, D. *Advances in factor analysis and structural equation models.* Cambridge, Mass.: Abt Books, 1979.

Joreskog, K. G., & Sorbom, D. LISRL: *Analysis of linear relationships by the method of maximum likelihood.* Uppsala, Sweden: University of Uppsala, 1981.

Langner, T. S. A twenty-two item screening score of psychiatric symptoms indicating impairment. *Journal of Health and Human Behavior,* 1962, *3,* 269–276.

Likert, R. A technique for the measurement of attitudes. *Archives of Psychology,* 1932, *140,* 1–55.

Link, B., & Dohrenwend, B. P. Formulation of hypotheses about the true prevalence of demoralization in the United States. In B. P. Dohrenwend, B. S. Dohrenwend, M. S. Gould, B. Link, R. Neugebauer, & R. Wunsch-Hitzig (Eds.), *Mental illness in the United States: Epidemiologic estimates.* New York: Praeger, 1980.

Macmillan, A. M. The health opinion survey: Technique for estimating prevalence of psychoneurotic and related types of disorder in communities, *Psychological Reports,* 1957, *3,* 325–339.

Manning, W. G., Newhouse, J. P., & Ware, J. E. The status of health in demand estimation; or, beyond excellent, good, fair, and poor. In V. R. Fuchs (Ed.), *Economic aspects of health.* Chicago: University of Chicago Press, 1982.

Newhouse, J. P. A design for a health insurance experiment. *Inquiry,* 1974, *11,* 5–27.

Newhouse, J. P., Manning, W. G., Morris, C. N., et al. Some interim results from a controlled trial of cost sharing in health insurance. *New England Journal of Medicine,* 1981, *305,* 1501–1507.

Nie, N. H., Hull, C. H., & Jenkins, J. G., Steinbrenner, K., & Bent, D. H., *Statistical package for the social sciences.* McGraw-Hill, 1975.

Radloff, L. S. The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement,* 1977, *3,* 385–401.

Stewart, A., Ware, J. E., & Brook, R. H. Advances in the measurement of functional status: Construction of aggregate indexes. *Medical Care,* 1981, *19,* 473–488.

Wan, T. H., & Livieratos, B. Interpreting a general index of subjective well being. *Milbank Memorial Fund Quarterly,* 1978, *56,* 531–556.

Ware, J. E., Johnston, S. A., Davies-Avery, A., & Brook, R. H. *Conceptualization and measurement of health for adults in the Health Insurance Study (Vol. 3, Mental health;* R-1987/3-HEW). Washington, D.C.: U.S. Government Printing Office, 1979.

Ware, J. E., Brook, R. H., Davies-Avery, A., et al. *Conceptualization and measurement of health for adults in the Health Insurance Study* (Vol. 1, *Model of Health and Methodology;* R-1987/1-HEW). Santa Monica, Calif.: The Rand Corporation, 1980. (a)

Ware, J. E., Davies-Avery, A., & Brook, R. H. *Conceptualization and measurement of health for adults in the Health Insurance Study* (Vol. 6, *Analysis of Relationships Among Health Status Measures;* R-1987/6-HEW). Washington, D.C.: U.S. Government Printing Office, 1980. (b)

Ware, J. E., Veit, C. T., & Donald, C. A., *Refinements in the measurement of mental health for adults in the Health Insurance Study* (R-2737-HHS). The Rand Corporation, in press.

Wells, K. B., Manning, W. G., Duan, N., Ware, J. E., & Newhouse, J. P. *Cost Sharing and the Demand for Ambulatory Mental Health Services* (R-2860-HHS). Santa Monica, Calif.: The Rand Corporation, 1982.

Received April 25, 1983 ■