Measuring Health Perceptions in the Health Insurance Experiment

Allyson Ross Davies and John E. Ware, Jr.

Rand

HEALTH INSURANCE EXPERIMENT SERIES

The research reported herein was performed pursuant to Grant No. 016B80 from the U.S. Department of Health and Human Services.

Library of Congress Cataloging in Publication Data

Davies, Allyson Ross, 1947-
Measuring health perceptions in the health insurance experiment.

"R-2711-HHS."
Bibliography: p.
1. Health status indicators. 2. Health surveys. 3. Health attitudes. 4. Self-perception. 5. Psychometrics. 6. Health surveys--United States. I. Ware, John E., 1942- . II. United States. Dept. of Health and Human Services. III. Rand Corporation. IV. Title. [DNLM: 1. Health surveys--United States. 2. Insurance, Health--United States. WA 950 D255m]
RA408.5.D38 362.1'042 81-21086
ISBN 0-8330-0366-6 AACR2

Published by The Rand Corporation

R-2711-HHS

Measuring Health Perceptions in the Health Insurance Experiment

Allyson Ross Davies and John E. Ware, Jr.

October 1981

Prepared under a grant from the
U.S. Department of Health and Human Services

Rand
SANTA MONICA, CA. 90406

PREFACE

This report was written as part of Rand's Health Insurance Experiment, funded under a grant from the U.S. Department of Health and Human Services. A related report, *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. V, General Health Perceptions,* R-1987/5-HEW, September 1978, by John E. Ware, Jr., Allyson Davies-Avery, and Cathy A. Donald, discussed the literature on general health perceptions measures and described those selected for use in the experiment. The present report evaluates their adequacy against standard psychometric criteria using data from enrollment Medical History Questionnaires and annual Health Questionnaires in the six study sites. The information contained here is directed to those who will be using the general health perceptions measures in the Health Insurance Experiment's analyses, and to others who are interested in using or adapting these measures for their own research.

Several other reports and report series issued as part of the Health Insurance Experiment discuss health status measurement. They include:

R-1987-HEW, *Conceptualization and Measurement of Health for Adults in the Health Insurance Study,* Vols. I-VIII, multiple authors, July 1978 through December 1980.

R-2313-HEW, *Conceptualization and Measurement of Health for Children in the Health Insurance Study,* Marvin Eisen, Cathy A. Donald, John E. Ware, Jr., and Robert H. Brook, May 1980.

R-2262-HHS, *Conceptualization and Measurement of Physiologic Health for Adults,* Vols. 1-17, multiple authors, August 1980 and continuing.

A complete list of papers and reports issued through 1978 appears in Rand Paper P-6221, *Overview of Health Insurance Study Publications,* by Joseph P. Newhouse and Rae W. Archibald, November 1978.

SUMMARY

This report presents and discusses the results from psychometric studies of the general health perceptions measures fielded by Rand's Health Insurance Experiment (HIE). Data on general health perceptions will be used in the HIE to test hypotheses about how health care financing mechanisms (different coinsurance and deductible rates, and prepaid versus fee-for-service care) affect health status over a three- to five-year experimental period. General health measures are used in HIE analyses because they provide a comprehensive definition of health status, assessing a construct underlying both physical and mental health, and reflecting both objective information enrollees have about their health and their subjective evaluations of that information. These measures thus provide information beyond that contained in more specific measures of the major health status components, physical and mental.

Our analyses focused on three issues: (1) whether scoring rules developed for the general health perceptions that were adopted by the HIE from other sources could be used in HIE populations and would yield variable and reliable scores; (2) whether we could construct an overall General Health Ratings Index that would aggregate across all general health perceptions constructs and yield a reliable and interpretable score; and (3) further evaluation of the meaning of general health perceptions scores in relation to a wide range of health and health-related variables. Data for the scaling analyses came from questionnaires that were self-administered by almost 5000 adults (aged 14 and older) enrolled in the HIE at all six study sites: Dayton, Ohio; Seattle, Washington; Fitchburg, Massachusetts; Franklin County, Massachusetts; Charleston, South Carolina; and Georgetown County, South Carolina. Data for stability and validity analyses came from some 1200 adults in the largest HIE site; files linking general health measures over time and to other health variables were available only for that site when our analyses began. Missing data rates were very low, with less than 1 percent of possible responses missing.

HIE health questionnaires contain 29 items that assess general health perceptions constructs. Of these, 26 were taken from the Health Perceptions Questionnaire (HPQ) developed by Ware and Karmos.[1] Three single-item measures were adaptations of the commonly used personal ratings of health as excellent, good, fair, or poor (EGFP), of pain, and of health-related worry and concern. We used multitrait scaling to test 27 items (the 26 HPQ items and the single-item worry/concern measure) in six multi-item scales hypothesized to assess the major health perceptions constructs identified in the literature and in empirical tests of the structure of health perceptions. Three reflect time-bound definitions of health: Current Health, Prior Health, and Health Outlook; the other three focus on health-related perceptions of Resistance to Illness, Health Worry/Concern, and Sickness Orientation (a tendency to accept sickness as part of one's life).

Most items met the multitrait criteria for convergent and discriminant validity across sites and in each one of the HIE sites. This replication of results in HIE samples and their close similarity to results reported by Ware and Karmos (1976) in their original studies of the HPQ measures supported separate scoring of the six item groupings in summated ratings scales. The few errors that we observed in multitrait scaling analyses were traced to low convergent validity for the two Concern items in the Worry/Concern item grouping, and relatively lower reliability for the Worry and Concern items than others we studied. Additional analyses supported separate scoring of the Worry (but not the Concern) items as a multi-item scale.

[1]J. E. Ware, Jr., and A. H. Karmos, *Development and Validation of Scales to Measure Perceived Health and Patient Role Propensity: Volume II of a Final Report,* National Technical Information Services, NTIS Publication No. PB 288-331, Springfield, Va., 1976.

We tested a new scoring algorithm for the EGFP item, using the Current Health scale as a criterion measure. The algorithm we developed reflects the unequal intervals between the EGFP response choices and should provide a more precise score than the more commonly used scoring algorithms for this item.

We used principal components analysis to identify the common dimension underlying the general health perceptions items for use in an overall General Health Ratings Index. We chose this approach to defining an overall index in contrast to one that would have selected and/or weighted the various general health constructs according to their importance in predicting some non-health criterion, such as use of medical care services.

All but two of the general health perceptions constructs (health-related concern and sickness orientation) were important to the definition of this common dimension. The current health construct emerged as the most important or "core" construct, capturing much of the information shared in common by the HIE general health ratings. Results of the principal components analyses replicated across HIE sites, supporting generalization of our findings. Scoring studies indicated that the simple summated ratings method would yield the simplest Index score and one that was just as reliable (approximately 0.90) as scores based on either a standardized, unit-weighted summated ratings algorithm or on a factor scoring algorithm.

All the HIE general health perceptions measures yielded scores with adequate variability, despite the age restriction on enrollment into the study (a very small proportion of the sample is older than 60). Studies of reliability based on the internal-consistency approach indicated that the measures were more than reliable enough for group comparisons, their intended use in tests of HIE hypotheses. Reliability estimates for all the multi-item measures equaled, and usually exceeded, the minimum standard of 0.50 for group comparisons. Estimates for Current Health and the General Health Ratings Index approached 0.90. Comparisons of reliability estimates across HIE sites, whose samples vary considerably in sociodemographic characteristics that are related to data quality, and between groups differing in educational attainment, indicated that reliability was somewhat lower in socioeconomically disadvantaged samples. In all cases, however, reliability remained adequate for the intended group comparisons, which must take the observed differences in reliability into account. The single-item measures (EGFP and Pain) were less reliable than the multi-item measures; EGFP, but probably not Pain, seems reliable enough for group comparisons.

Analyses of stability over approximately a one-year interval indicate that the Current Health construct is the least stable over time. Most of the measures had stability coefficients that indicated they reflected changes over time, but were stable enough to warrant their use in repeated-measures designs such as that of the HIE and to enhance precision of hypothesis-testing in such designs. The stability coefficients of both Prior Health and Sickness Orientation were near unity, suggesting their scores reflect very little change over time. This finding was consistent with the content of items in the Prior Health scale, which refer to health states in the relatively unchanging past. Such content could not explain the stability of Sickness Orientation scores; our finding suggested that this measure assessed more of a personal trait than a health perception, which would be expected to change over time.

Validity studies indicated that the HIE health questionnaires contain a comprehensive sample of general health rating items that well represent those fielded in published studies. Empirical evidence based on analyses of correlations between the general health measures and 30 other self-reported and independently measured variables support the use of most HIE general health perceptions measures in tests of hypotheses about group differences in health status and about changes in health over time. We have a strong basis for concluding that the HIE scales, with the exception of Sickness Orientation, contain information about both physical and mental health status. Moreover, they reflect the different operational definitions of *both* physical and mental health constructs. Such findings, particularly notable for Current Health and the Index, support their interpretation as *general* health measures. Current Health stands out as the "most" valid of the subscales, given its prominence in defining the

construct underlying these general health ratings and its more substantial correlations (than the other subscales) with both physical and mental health variables. Because of its importance, we did item-level validity studies and propose a short-form Current Health scale that can be used in place of the longer nine-item HIE scale in studies that do not require as much precision. Our validity analyses indicated that Sickness Orientation should not be used to test hypotheses about health status. Although HIE data provided little empirical evidence about what scores on this measure do mean, we hypothesize on the basis of stability evidence and its content that this scale assesses a health-related personality trait.

We also found, as hypothesized, that the HIE general health perceptions measures had only weak relationships with measures of social contacts and social resources. These findings, consistent with those reported in other HIE analyses and the published literature, indicate that the measures should not be used to assess group differences or changes in social well-being. They also indicate that these aspects of social well-being are not considered part of "health" as it is defined by persons in general, U.S. populations.

Regression analyses with longitudinal data indicated that the general health measures, scored separately, contain information that predicts use of medical care services over and above that predicted by measures of physical and mental health.

Our analyses examined two methodologic issues, namely the effects of response set and of questionnaire placement on general health perceptions scores. Response sets unrelated to item content (acquiescent and opposition response sets) did not affect the scores. Social desirability response set, which is related to item content, affected the interpretation of scores and biased most group comparisons. Placing the general health perceptions battery at the end, rather than in the middle, of a lengthy health survey affected mean scores on the general health ratings but had little systematic effect on reliability. In a randomized-groups experiment, means computed for later questionnaire placement suggested respondents had poorer health, more worry/concern, and greater health-related pain than those who responded to earlier placement of the battery.

We concluded that our analyses of HIE general health perceptions measures supported (with the exception of Sickness Orientation) their use in tests of hypotheses about the effects of health care financing arrangements on personal health status. The measures show considerable variability in general populations, even those with few persons older than 62 (the cutoff for the HIE sample). The multi-item measures (along with EGFP) provide reliable scores for use in group comparisons. The stability of the measures indicates that they will enhance precision in repeated-measures designs such as that of the HIE. Validity studies that went considerably beyond those reported to date support their interpretation as integrative measures of both physical and mental health, and indicate that the measures provide information beyond that contained in specific measures of physical and mental health for predictive analyses. Given these findings, the HIE general health perceptions measures should permit precise tests of experimental hypotheses that examine the health-related effects of differences in coinsurance rate and organization of care.

ACKNOWLEDGMENTS

We would like to thank Lois A. Maiman, University of Rochester, and our Rand colleagues John Winkler and Joseph P. Newhouse for their thoughtful reviews and comments on an earlier version of this report. Our thanks also go to Cathy A. Donald, who constructed several of the analytic files that we used, and to Dorothy E. Boyd, Marilyn Muntz, Marie Knight, and Caren Kamberg, who saw us through several drafts of the manuscript and accompanying tables.

CONTENTS

FIGURES

TABLES

I. INTRODUCTION

BACKGROUND

Self-ratings of general health, commonly used to assess health status in general population surveys, hold a central place in Rand's Health Insurance Experiment's (HIE) measurement strategy. These ratings are considered measures of *general* health because they do not focus on specific components of health, such as physiologic, physical, or mental. In contrast to operational definitions of health constructs based on descriptions of behavior or feeling states that do not use the word "health" (e.g., "ability to walk upstairs" as a measure of physical functioning; "feeling downhearted" as a measure of mental health), general health ratings usually ask the respondent explicitly about personal *health*.

Our previous report[1] reviewed the reasons for including general health ratings, along with measures of physical and mental health and social well-being, in HIE analyses of the effect of coinsurance on individual health status and on the use of medical care services. To recapitulate briefly, the reasons fell into two groups. First, general health ratings provide a comprehensive definition of health status, because they are presumed to assess a construct underlying both physical and mental health. Persons with poor physical or mental health should have less favorable general health ratings than those with better physical or mental health. Moreover, persons with poor physical *and* mental health should score even lower than those with either poor physical *or* mental health. To the extent that changes in such an integrative measure can be interpreted in terms of real changes in both physical and mental health, general health ratings could provide a simple and useful summary indicator in tests of HIE hypotheses about health status outcomes.

Second, general health ratings differ from many of the widely used component-specific measures of health status because they allow respondents to indicate the objective information they have about their own health and how they feel about or evaluate that information. This evaluative aspect also lets them place greater weight on one or the other health component—physical or mental—in rating their general health. Although two persons may have the same profiles across a range of health variables, the emphasis they place on different aspects of their profiles may differ. These different subjective evaluations would be reflected in their general health ratings but not in scores based on more objective health measures. The inclusion of this evaluative component means that general health ratings should make a unique contribution to predictions of medical care consumption and other health-related behavior (beyond that made by measures of physical and mental health), presuming that people seek care in response to a self-assessed need for it.

Highlighting the importance of such measures in forecasting use and expenditures, recent analyses indicate that patients initiate about 60 percent of physician office visits (Rossiter, 1980).[2] Theory suggests that self-assessment of need and voluntary health-related actions may involve, among other things, integrating perceptions of both physical and mental health status along with other health-related perceptions such as worry and concern about health, perceived resistance or susceptibility to illness, and sickness orientation or the tendency to believe that illness is a part of life (Becker, 1974; Becker et al., 1977a).

For these reasons, measures of these constructs—termed general health perceptions—should be important in HIE analyses of how personal health status and use of health services

[1]Ware, Davies-Avery, and Donald (1978). Hereafter cited as WDD (1978).

[2]Patient-initiated visits were defined as "any visit which the patient arranged by calling up for an appointment or walking in without an appointment" (Rossiter, 1980, p. ii).

change in response to different health care financing arrangements. As we described in our earlier report (WDD, 1978), the measures we selected came from the Health Perceptions Questionnaire developed by Ware and Karmos (1976). They fulfilled the requirements that health status measures used in the HIE should reflect contemporary conceptualizations of the constructs identified and represent state-of-the-art measures. In concluding that report, we noted that several other requirements for these measures had to be evaluated with HIE data before they were used in experimental analyses.

FOCUS OF THE REPORT

This report presents the results from our further evaluation of the HIE general health perceptions measures. It documents how these health measures are scored in the HIE and adds considerably to information regarding the appropriate interpretation of scores on these measures. The report should be useful to HIE and other analysts who will be using these measures and interpreting HIE results or the results of other health services research that uses these measures in similar populations.

In particular, the analyses reported and discussed in this volume focus on three sets of goals. First, we wanted to examine whether the scoring rules recommended for these measures by their developers (Ware and Karmos, 1976) could be used in the HIE, and to evaluate the reliability and stability of scores in HIE samples.

Second, we wanted to explore construction of a general health ratings index that would be simple to score, reliable, and valid in relation to a range of health status constructs. Few comprehensive summary indexes were available in the literature, and earlier work with the measures adopted by the HIE had stopped short of developing such an index (Ware and Karmos, 1976).

Third, and perhaps the focal point of our analyses, we wanted to further evaluate the validity of these measures. Our earlier work indicated that their validity—the meaning of scores on general health perceptions measures—required far more study against a wide range of health and health-related variables. The extent of the HIE data base and the analytic importance of understanding the meaning of scores on these measures made further validation studies both possible and necessary. The validity analyses reported in this volume used cross-sectional data and focused on understanding the meaning of self-assessed health status and health-related perceptions in relation to other self-reports of health (e.g., chronic disease status, physical functioning, emotional stability). In addition, two validity variables were based in part on more objective reports of health status from screening examination data. Using already reported results from a longitudinal analysis that included the HIE general health perceptions measures, we were able to discuss their predictive and incremental validity. We also examined possible threats to validity that arise from questionnaire response biases.

Our earlier report (WDD, 1978) identified other analyses of general health perceptions measures that could be done using HIE data to provide further information about their validity. While their importance remains undiminished, we did not address them in the work done for this report. We considered the most important validity issue to be the extent to which these measures reflect an individual's perceptions of his own health. Therefore, the validity of self-assessments of health in relation to assessments made by other means (e.g., physicians' diagnoses, objective reports of the extent of disability) has not yet been evaluated.

Most of the validity studies that require longitudinal data remain to be done. We recognize that information about the meaning of changes in health status scores (on all measures, not only general health perceptions) is critical to their usefulness to the HIE and to other health services research. At the same time, no one set of analyses can cover all important issues.

Moreover, we believe the studies that will provide information about the meaning of changes in perceptions over time are best done in the context of developing longitudinal models of health status, only one facet of which relates to the validity of general health perceptions measures. We have therefore confined our validity studies here to cross-sectional analyses.

ORGANIZATION OF THE REPORT

Section II presents a brief literature review of empirical studies that relate to the three goals of our analyses. Section III contains information about data-gathering procedures, a description of the general health perceptions measures fielded in the HIE, and details of our analysis plan and the analytic methods we used. The final three sections present results pertaining to the three goals outlined earlier. Section IV documents the construction and scoring of scales, reliability and stability estimates, and descriptive statistics for HIE samples on the general health measures adopted from other sources and on the overall General Health Ratings Index developed for HIE use. Section V summarizes the results of analyses to date that provide information about the validity of all the HIE general health perceptions measures. We discuss the implications of these results and draw conclusions about the usefulness of the measures for HIE analyses in Sec. VI. The appendixes reproduce questionnaire batteries, document scoring rules for the measures, summarize the analyses of response bias, and provide supporting tabular material.

II. LITERATURE REVIEW

We organized our literature review in three parts, each pertaining to one of the goals we set for our analyses of the HIE general health perceptions measures. Because our earlier review and discussion of literature on these measures covered much of the same ground in considerable detail (WDD, 1978), we need to provide only enough background from the literature so that our discussion of HIE results can be put in context.

PRIOR RESULTS: HPQ MEASURES

This section is brief because extensively documented original studies of the HPQ are already available in the literature (Ware and Karmos, 1976; Ware, 1976) as well as in our earlier report (WDD, 1978). We will refer again to specific results from these earlier studies when we present and discuss the results of HIE analyses.

On the basis of literature reviews and empirical studies of the dimensions revealed by general health ratings, the Health Perceptions Questionnaire (HPQ) was developed to measure six hypothesized general health perceptions constructs: Current Health, Prior Health, Health Outlook, Resistance to Illness, Health Worry/Concern, and Sickness Orientation.[1] Successful scaling studies, based on factor analytic and multitrait criteria, yielded six summated ratings scales, each measuring one of the constructs mentioned. Ware and Karmos (1976) noted the marked similarity of these findings across five field tests in general population samples that differed widely in socioeconomic characteristics. These replications supported generalization of findings regarding scale development.

Both internal-consistency and test-retest reliability estimates indicated that the six HPQ scales were reliable enough for group comparisons, even in the most disadvantaged site where data quality was poorest. Reliability estimates for all scales ranged from 0.45 to 0.92; those for Health Worry/Concern and Sickness Orientation tended to be lowest but were usually above 0.50, a recommended standard for group comparisons. Reliabilities for Current Health were near or above 0.90, the recommended standard for individual comparisons. The magnitudes of two-year stability coefficients for the four measures (Current Health, Prior Health, Resistance, and Health Worry/Concern) fielded in a follow-up study ranged from 0.31 to 0.62. These results suggested that general health ratings are stable enough over time to increase the precision of hypothesis-testing in repeated-measures designs such as that of the HIE. Results also indicated that measures of general health perceptions may be less reliable at any time and less stable over time for the disadvantaged.

MULTI-ITEM GENERAL HEALTH MEASURES

Our earlier review (WDD, 1978) identified seven published multi-item scales that included one or more general health ratings, as we defined them in the Introduction.[2] Since then, at least one new such scale has been introduced in the literature (Berki and Ashcraft, 1979).

---

[1]Because these are the measures we tested in HIE analyses, we present their operational definitions and item content in Sec. III, Methods.

[2]We selected these measures after content analyses; many measures that are labeled "health indexes" do not use self-ratings of health in general.

In Table 1, we classify the content of these eight scales according to whether the items reflect one of the six general health perceptions constructs (current health, prior health, health outlook, resistance, health worry/concern, sickness orientation), specific components of health (physical and mental), or other constructs. Across measures, the construct most commonly included was current health; Suchman et al. (1958) and Renne (1974) assessed only that construct, although in the Suchman et al. measure only two items were *self-ratings* of current health. Five of the eight (Suchman et al., 1958; Thompson and Streib, 1958; Berkman, 1971; Renne, 1974; Berki and Ashcraft, 1979) included a rating of current health as excellent, good, fair, or poor (EGFP).

Few of the other general health constructs identified by Ware and Karmos (1976) were represented in these measures. Only one measure included items that described specific health components (Berki and Ashcraft, 1979). Their items focused primarily on attitudes toward physical ability (response choices for all items but EGFP asked for ratings of satisfaction with the aspects of physical ability mentioned). Except for Renne's measure, all the multi-item scales published by others included items that assessed constructs other than perceptions about health in general (e.g., use of services, illness or disability episodes, quality of life).

Several scaling methods were used to evaluate these measures. Thompson and Streib (1958) and DiCicco and Apple (1958) tested the unidimensionality and cumulative nature of their indexes using Guttman's Scalogram Analysis, and reported reproducibility coefficients of 0.95 and 0.90, respectively. We inferred from published information that several indexes were scored as summated ratings measures (Suchman et al., 1958; Berkman, 1971; MacPherson, 1972); whether they were evaluated using traditional Likert criteria was not clear, and no internal-consistency estimates were reported. In addition, we found no stability estimates for these indexes.

Berki and Ashcraft (1979) used principal components analysis to study the dimensionality of their nine-item measure. They recommended factor scoring of the nine items using the coefficients (weights) derived for the first rotated principal component, on which the items loaded positively and roughly equally. They did not report reliability estimates.[3]

In sum, we identified eight published multi-item measures that include general health ratings, chiefly of current health. Only one (Renne, 1974) could be considered solely a measure of self-perceived general health. Another (Berki and Ashcraft, 1979) appeared to measure perceived physical health. The confounded content of most of these measures, which combine ratings of general health with measures of several other variables (e.g., use of services, quality of life), calls into question their validity as measures of perceived general health.

VALIDITY OF GENERAL HEALTH PERCEPTIONS MEASURES

In the literature we reviewed earlier (WDD, 1978), as well as in previous studies of the HPQ measures (Ware and Karmos, 1976), we found that the validity (i.e., meaning or interpretation) of general health ratings had most often been studied in relation to measures of

[3]Using the inter-item product-moment correlation matrix they made available to us (Ashcraft, personal communication), we estimated the theta reliability of the factor score for the first unrotated component at 0.88. We also noted that the nine items loaded positively and roughly equally on this first component; the equality of their loadings suggested that the weights may be unnecessary. In fact, the internal-consistency reliability of the simple algebraic sum of the nine items was 0.87, suggesting that neither the weighting nor standardization of items for the factor score contributed much to reliability. The homogeneity (average inter-item correlation) of the nine items equaled 0.43. Supporting interpretation of the measure as perceived physical health status, the homogeneity of the six physical health items was 0.55 (compared with 0.28 for the remaining items) and their internal-consistency reliability was 0.88, just a bit higher than that of all nine items.

Table 1

Content of General Health Ratings Scales Appearing in the Literature

Construct	INVESTIGATOR/MEASURE							
	Suchman et al. (1958) --	Thompson and Streib (1958) --	DiCicco & Apple (1958) Index of Health	Pratt (1971) Level of Health Index	Berkman (1971) Adult Physical Health	MacPherson (1972) --	Renne (1974) --	Berki and Ashcraft (1979) Health Status
GENERAL HEALTH PERCEPTIONS								
Current Health	Rate your health at present Any physical or health problems at present	Rate your health at present Any physical or health problems at present	Health now Present trouble with health Health better or worse than other people your age	Health at present	Health generally excellent, good, fair, or poor	Health these days	Health generally excellent, good, fair, or poor Health better or worse than health of others your age	Compared to others your age, health excellent, good, fair, or poor
Prior Health	--	Health changed during past year	--	--	--	--	--	--
Health Outlook	--	--	--	--	--	--	--	--
Resistance	--	--	--	--	--	--	--	Resistance to illness
Worry/Concern	--	--	--	--	--	--	--	--
Sickness Orientation	--	--	--	--	--	--	--	--
SPECIFIC HEALTH COMPONENTS								
Physical	--	--	--	--	--	--	--	Way you usually feel physically Physical ability to do things need to do Physical fitness Ability to get around outdoors Physical ability to do things you want to do Ability to be as active as others your age
Mental	--	--	--	--	--	--	--	--

Table 1—continued

OTHER VARIABLES								
Illness Episodes	--	--	Illness during past 2 weeks	Illness during past 2 weeks	--	--	--	--
Disability Days	--	--	Bed days in previous year	--	--	--	--	--
Use of Services	--	Seen by doctor in past year	--	--	During past 5 years, number of nights in hospital, sanitarium or nursing home	--	--	--
Social Contacts	--	--	Health keep you from seeing people	--	--	--	--	--
Quality of Life	--	--	--	--	--	All things together, how are things these days How satisfying do you find way of spending these days	--	--
Physician Ratings	Subject's health at present	--	--	--	--	--	--	--
Other	--	--	--	--	--	--	--	Health that runs in your family

physical health. Correlations between general health ratings and measures of mental health and of health and illness behavior (e.g., use of services) were next most often reported.

Few studies offered specific hypotheses, based on theory or previous empirical findings, as to the relationships that should be observed between valid general health perceptions measures and the validity variables used. Virtually all information pertaining to the validity of the measures was based on cross-sectional data, so their predictive validity in the sense of "forecasting" has been little studied. With the exception of preliminary multivariate analyses reported by Ware and Karmos (1976), no analyses took relationships among the health perceptions measures into account when examining their meaning. Our recent study (Ware, Davies-Avery, and Brook, 1980), which used an earlier version of the HIE health status measures, took into account the relationships among the validity variables (e.g., physical and mental health).[4] Estimates of reliability were infrequently reported, which made it difficult to interpret the magnitudes of the observed relationships between general health perceptions measures and other variables.[5]

Despite these drawbacks, which are common to the literature on most health status measures (Ware et al., 1981), we have summarized consistent findings from the literature regarding the meaning of scores on general health perceptions measures in relation to other health and health-related variables. Evidence reported in that literature generally supports the following directional hypotheses about valid (and favorably scored) general health perceptions measures:

- *Function Status:* Favorably defined measures of function status, which reflect variation in physical and mental components of health and, to a lesser extent, in social activities, are positively associated with general health measures.
- *Physical Health:* Favorably defined measures of physical health are positively and usually substantially associated with general health perceptions measures. The physical health variables studied are usually unfavorably defined (so the correlation with general health is negative); they include chronic conditions, disability, number of health problems or symptoms, sensory impairment and immobility, bed days, and role activity limitations. Current Health, among the scales in the HPQ, is most strongly associated with variables in this category.
- *Mental Health:* Mental health measures (e.g., life satisfaction, positive well-being) are positively associated with general health measures; negative definitions of mental health (e.g., depression) are negatively associated.
- *Social Circumstances:* Social interaction and participation are positively associated with general health measures; those of isolation are negatively related to general health. Little has been published about the strength of these relationships.
- *Physicians' Assessments:* A physician's ratings of a patient's functioning and health in general are positively related to the patient's ratings of his own general health status. These relationships are less frequently significant than are those between self-ratings of general health status and self-reported physical and mental health status.
- *Mortality:* Measures of mortality are negatively associated with general health rat-

[4]Relationships among the general health ratings were not taken into account by Ware, Davies-Avery, and Brook (1980) because the earlier HIE health status measures included only one rating for analysis.

[5]Measures cannot be valid unless they are reliable, and in theory no validity coefficient can exceed in magnitude the square root of a variable's reliability (Nunnally, 1978); here, the validity coefficient is the correlation between a measure and a variable with which it should (or should not) be associated, according to theory. Without reliability estimates, therefore, it is usually impossible to know how much of the information contained in a measure is information about a particular construct.

ings at a point in time. These ratings are less favorable for those who die within the next three years than for those who live (in the one longitudinal validity study reported).

- *Health and Illness Behavior:* Measures of illness behavior (e.g., recent use of services) are substantially and negatively related to general health; those of health behavior (e.g., getting a checkup when well) are usually unrelated. Behaviors studied in relation to general health perceptions measures for which these associations held include previous physician and hospital use, number of physician visits, and compliance. Ware and Karmos (1976) found that Current and Prior Health were more strongly and consistently related to illness behaviors than were the other HPQ scales. Only Current Health and Worry/Concern were significantly related to compliance in their study (Current Health negatively and Worry/Concern positively).
- *Rejection of Sick Role:* Measures of the tendency to reject the sick role (RSR) show inconsistent relationships with those of general health perceptions. Current and Prior Health and Health Outlook are not related to RSR; Worry/Concern is positively related and Resistance is negatively related to RSR.
- *Age:* Age is commonly considered an appropriate validity variable, although other sociodemographic variables should not be regarded as providing information about the validity of health measures pending further research (Ware, Davies-Avery, and Brook, 1980). In general, health status is expected to get worse as people age. Age is negatively related to Current and Prior Health and Health Outlook, and positively related to Resistance to Illness. Perhaps contrary to what one would think, previous studies consistently indicate that Health Worry/Concern and Sickness Orientation scores are significantly lower for older persons.

The preceding summary is based entirely on reported bivariate correlations between general health perceptions measures and those of variables in the categories mentioned. The general health perceptions measures are interrelated, as are the other health and health-related variables studied to understand general health. Thus, these bivariate relationships may not reflect adequately the nature of the relationships between these variables. Taking an example from the mental health area, mental health measures tend to be inconsistently related, if not completely unrelated, to age (Ware et al., 1979). When age-related decreases in physical health and changes in social circumstances are controlled statistically, a significant positive relationship emerges: Mental health scores are more favorable for older persons, at least through the age of 62, the upper bound in our study (Ware, Davies-Avery, and Brook, 1980). Thus, multivariate validity studies can be an important adjunct to the more common bivariate validity analyses (more common probably because a comprehensive set of health and health-related variables is infrequently measured in any one study).

Preliminary multivariate analyses reported by Ware and Karmos (1976) indicated that more than one HPQ measure made independent contributions to regressions that predicted different types of illness behavior. They did not emphasize specific findings for the HPQ measures because of the cross-sectional nature of their study design and its reliance on self-report.[6] They did note, however, that similar results might be observed in longitudinal studies. Their finding, together with those from bivariate validity studies, suggests that different interpretations may be appropriate for different measures of general health constructs. Depending on the types of illness behavior studied, separate scoring and interpretation of the different measures would be useful.

---

[6]Scores on the HPQ scales were used to "predict" self-reports of previous use of health care services (Ware and Karmos, 1976, p. 189).

Using earlier versions of the HIE measures of physical, mental, and general health and of social circumstances, along with life stress and age, Ware, Davies-Avery, and Brook (1980) developed a preliminary multivariate model of health status applicable to a nonaged general population.[7] Our model indicates that self-ratings of health in general reflect the independent contributions of self-reported physical and mental health status and age (the latter both directly and indirectly through physical health). Social circumstances (defined in the analyses as work adjustment) was much less related to general health ratings than physical or mental health when the other variables were controlled statistically.

Findings from these bi- and multivariate studies suggest two priorities for further validation of general health perceptions measures: We should not expect measures of the different general health perceptions constructs to have identical patterns of relationships with other health and health-related variables; and valid health perceptions measures, if they adequately reflect an underlying health component, should relate more to physical and mental health status than to measures that reflect social interaction and participation.

---

[7]The general health variable was defined by a three-item general health scale included in the original mental health battery, not by the measures discussed in this report.

III. METHODS

DATA-GATHERING METHODS

Most data reported in this volume came from responses to the first questionnaire fielded in each of the HIE sites that contained all HIE health perceptions measures and for which data files had been archived by early 1980.[1] In all sites except Dayton, all health measures first appeared on the self-administered enrollment Medical History Questionnaire (MHQ), fielded at the beginning of the experiment. The Dayton enrollment MHQ included only two of these measures. Therefore, the data on Dayton reported here came from questionnaires fielded after the experiment began (specifically, the third annual Health Questionnaire (HQ) for the five-year sample and the exit MHQ for the three-year sample).[2] Longitudinal data for stability analyses came from files that linked enrollment MHQ and first annual HQ data from Seattle.

Because of its length, the self-administered MHQ was divided into two booklets (Form A and Form B). All adults completed Form A in their homes before the experimental intervention began and again on exit from the study. Those who did not receive the medical screening examination at enrollment also completed Form B of the MHQ at home. Adults who were screened at enrollment and all adults at exit completed Form B at the screening center in their site. Health Questionnaires, which contained a subset of the MHQ health status batteries and were fielded annually near the anniversary of enrollment date, were also self-administered at home by enrollees aged 14 and older.

Each adult who completed the enrollment MHQ received a $5.00 compensation, up to a maximum per family of $20.00. The family maximum included a $2.00 compensation for each MHQ completed for children 13 and under. A random subsample of families in each site received a medical screening examination during enrollment (Smith et al., 1978). In these families, each family head received a $20.00 compensation and a $5.00 compensation for each dependent, up to a family maximum of $50.00. Families received $5.00 per family head for completing the annual Health Questionnaires.

Questionnaires were carefully checked for missing responses against stringent field-edit specifications. Respondents were contacted if responses were missing on more than six items (out of many hundreds), and the missing information was obtained over the telephone. If the respondent had problems answering the questions, interviewer assistance was provided. When the usually self-administered questionnaire was interviewer-administered, this difference was noted in the data base. Data were processed using standardized coding procedures and "cleaned" by a computer program that checked for possible coding errors and assigned a data status indicator that described data quality for each item in the questionnaire.

---

[1]For details on the design of the Health Insurance Experiment, including site and sample selection, sample eligibility requirements, and a description of the experimental health insurance plans, see Sec. II in Ware et al. (1980). Table 2 provides a brief outline of questionnaire administration by site and sample.

[2]Not all the health perceptions measures appeared on the first annual HQ fielded in Dayton. Although they were administered on the second annual HQ in Dayton, that file had not been archived when these analyses began because of the priority placed on analyzing three-year results and thus on archiving the data from the third annual questionnaires. Across sites, approximately 70 percent of the sample is enrolled for three years and the remainder for five years; differences in length of enrollment will permit analyses of the responsiveness of demand to enrollment period, and extends the analysis of health outcomes to five years for 30 percent of the sample.

Table 2

Administration of Medical History Questionnaire (MHQ) and Health Questionnaire (HQ), by Site, Plan Type, and Enrollment Period

	Dayton, OH		Seattle, WA					Fitchburg & Franklin County, MA		Charleston & Georgetown County, S.C.	
	All-Fee-for-Service		Fee-for-Service		Prepaid Group Practice			All-Fee-for-Service		All-Fee-for-Service	
Time During Experiment	3-Year	5-Year	3-Year	5-Year	3-Year	5-Year	Control	3-Year	5-Year	Pre-experimental	5-Year
Before experiment	-- All samples do Baseline Questionnaire[a] --										
At enrollment	MHQ[a]	MHQ[a]	MHQ[b]	MHQ[b]	MHQ[b]	MHQ[b]	MHQ[b]	MHQ[b]	MHQ[b]	MHQ[b,c]	MHQ[b]
End of year 1	-- All samples do HQ 1 --										
End of year 2	HQ2	HQ2	HQ2	HQ2	HQ2	HQ2	HQ2	HQ2	HQ2	MHQ[d]	HQ2
End of year 3	MHQ[b] exit	HQ3[b]	MHQ exit	HQ3	MHQ exit	HQ2	HQ3	MHQ exit	HQ3	HQ3	HQ3
End of year 4	-	HQ4	-	HQ4	-	-	HQ4	HQ4	-	-	HQ4
End of year 5	-- All remaining samples do exit MHQ --										

[a] All questionnaires but these three included all the health perceptions measures discussed in this report.
[b] Data analyzed in this report came from these questionnaires and samples.
[c] Sample not enrolled for experimental benefits at this point.
[d] Enrolled at this point for next three years.

SAMPLE CHARACTERISTICS

Table 3 presents data on selected characteristics of the 4763 adult respondents (aged 14 and older) at each HIE site and across sites who provided data for general health perceptions measures on the MHQs and HQ described above. As indicated in Table 2, there are six HIE sites: Dayton, Ohio; Seattle, Washington; Fitchburg, Massachusetts; Franklin County, Massachusetts; Charleston, South Carolina; and Georgetown County, South Carolina. Throughout this report, samples at the two Massachusetts and two South Carolina sites have been combined.[3]

As Table 3 shows, the analytic sample had slightly more women than men, and was predominantly white; the South Carolina sample had the largest proportion of nonwhite enrollees. Ages ranged from 14 to 67,[4] with a mean of 33.2 years in the combined-site sample; the Dayton sample was somewhat older and the South Carolina sample somewhat younger than the average. Education of family heads ranged from zero to 27 years of schooling completed, with a combined sample mean of 12.4. Family incomes (1973 dollars for Dayton; 1974 dollars for other sites) ranged from zero to $38,550, with a combined sample average of $12,405; families in Dayton reported the highest average annual income and those in South Carolina, the lowest.

DESCRIPTION OF GENERAL HEALTH PERCEPTIONS MEASURES

The first enrollment MHQ fielded in the HIE in Dayton included two single-item general health perception measures assessing health-related pain and worry. Revisions in the enrollment MHQ[5] before it was fielded in the other HIE sites added a third single-item rating of health as excellent, good, fair, or poor (EGFP) and 26 items from the Health Perceptions Questionnaire (HPQ) to assess the general health construct. These 29 items, which also appeared on annual HQs and on all exit MHQs, constitute the HIE general health perceptions measures and are described in greater detail below.[6]

Complete content for the 29 HIE general health perceptions items appears in Table 4. The pages from the Seattle enrollment MHQ on which these items appeared are reproduced in App. A; directions to the respondent and full response choices for these items also appear on the reproduction. The items are identified in Table 4 by their questionnaire number from the Seattle MHQ. For convenience elsewhere in the text and in all tables, these numbers (simply

---

[3]For these analyses, the Seattle sample excludes the control group that receives care from the prepaid group practice. The HIE randomly assigned some experimental families to receive care from a prepaid group practice in Seattle, and identified as a control group families that met eligibility criteria but had already chosen the group practice. Because of the possible self-selection bias (e.g., healthier families choosing the group practice), control group data were not used in scaling studies. The South Carolina sample includes the pre-experimental group, a sample intended for three-year enrollment, that self-administered the MHQ when the five-year South Carolina sample enrolled. In all sites but South Carolina, the three- and five-year samples enrolled at the same time and exited two years apart. In South Carolina, the five-year sample began participating two years before the three-year group, and the two will exit simultaneously (see Table 2). Because of these differences in sample definition, the sample sizes reported in Table 3 do not equal the enrollment sample sizes, and distributions of some characteristics may not exactly match those of the enrollment samples. Such differences are not likely to have affected the scaling decisions made or conclusions drawn in this report.

[4]A small number of persons older than 61 who were in eligible families were enrolled in the HIE but are not themselves eligible for insurance benefits. Although data are collected from these people and were used in scaling studies, they will not be included in the experimental analyses.

[5]These revisions are documented in the R-1987-HEW series; see Stewart et al. (1978, physical health); Ware, Davies-Avery, and Donald (1978, general health perceptions); and Ware et al. (1979, mental health).

[6]Scaling and validity studies reported by Ware and Karmos (1976) demonstrated that the other six HPQ items dealing with rejection of sick role and attitude toward going to the doctor assessed sick role propensity, as hypothesized, but shared little variance with the 26 general health items. For these reasons, they are not considered measures of general health perceptions and we do not discuss them here.

Table 3

SUMMARY OF ADULT SAMPLE CHARACTERISTICS

Characteristics	All Sites Combined	Dayton	Seattle[a]	Massachusetts[b]	South Carolina[c]
Sample Size	4763	789	1606	1068	1300
Age (years)[d]					
Mean	33.2	34.9	33.1	33.5	31.8
Range	14-67	14-64	14-62	14-67	14-59
Sex (%)					
Male	46.5	46.9	48.1	46.4	44.4
Female	53.5	53.1	51.9	53.6	55.6
Race (%)[e]					
Black	12.4	9.6	1.8	1.2	38.9
White	86.4	89.9	95.5	98.5	60.8
Other	1.2	0.5	2.7	0.3	0.3
Education (years)[f]					
Mean	12.4	12.7	13.1	12.5	11.3
Range	0-27	3-22	4-25	2-22	0-27
Family Income[g]					
Mean	$ 12405	$ 14826	$ 12884	$ 12207	$ 10513
Range	$0-38550	$0-38550	$0-27500	$0-27400	$0-30252

[a]Excludes adults in Group Health Cooperative control group.

[b]Combines Fitchburg and Franklin County samples.

[c]Combines Charleston and Georgetown County samples.

[d]See text for explanation of age restriction in HIS sample.

[e]Obtained for adult heads of household only.

[f]Obtained for respondents ages 18 and older only.

[g]1973 annual family income for Dayton, Seattle, and Massachusetts; 1974 annual family income for South Carolina. See text for explanation of income restriction in HIS sample.

A-FF for 128A-128FF) and the labels EGFP, Pain, and Worry are used to refer to the 26 HPQ items and the other three general health items, respectively.

EGFP asks for a personal evaluation or rating of the respondent's health as excellent, good, fair, or poor; the time frame implicit in the verb tense is the present. The other two non-HPQ items and their tailored response choices inquire about the amount of pain (Pain) and health-related worry or concern (Worry) that the respondent experienced during the past three months. On MHQs and HQs, these three items were among the first questions asked.

Two major reasons lay behind selection of EGFP, Pain, and Worry for use in HIE health questionnaires. First, they represent the type of general health rating items most commonly

Table 4

QUESTIONNAIRE NUMBERS, ABBREVIATED CONTENT, AND HYPOTHESIZED CONSTRUCT FOR GENERAL HEALTH PERCEPTIONS ITEMS FIELDED IN THE HEALTH INSURANCE EXPERIMENT

Item No.	Abbreviated Item Content	Hypothesized Construct
3	EGFP	Current health
5	Pain	Pain
6	Worry or concern about health	Worry/concern
128A	Doctors say health is now excellent	Current health
128C	Get sick easier than others	Resistance
128D	Feel better now than ever	Current health
128E	Probably sick a lot in future	Future health
128F	Never worry about my health	Worry
128G	People get sick easier than I do	Resistance
128I	Somewhat ill	Current health
128J	Expect to have better health than others	Future health
128K	So sick once thought I might die	Past health
128L	Not as healthy now as used to be	Current health
128M	Worry about health more than others	Worry
128O	Body resists illness	Resistance
128P	Getting sick part of life	Sickness orientation
128Q	Healthy as anybody I know	Current health
128R	Health worse in future than now	Future health
128S	Never had a long illness	Past health
128T	Others more concerned about health	Concern
128V	Health is excellent	Current health
128W	Expect a healthy life	Future health
128X	Health is a concern	Concern
128Y	Accept that sometimes I'm sick	Sickness orientation
128Z	Been feeling bad lately	Current health
128BB	Never been seriously ill	Past health
128CC	Something going around, catch it	Resistance
128DD	Doctors say now in poor health	Current health
128FF	Feel as good now as ever	Current health

fielded in general population surveys (e.g., the National Health Interview Survey) and by studies that require some measure of health status. Inclusion of these items therefore allows comparison of HIE data with those from other surveys. Second, although the reliability and validity of these measures in relation to a comprehensive set of health and health-related variables had not been well studied before the HIE (WDD, 1978), they will be useful in HIE analyses of the meaning of multi-item measures constructed from the HPQ items.

Each of the 26 items from the HPQ is a complete statement of opinion, worded favorably or unfavorably, about personal health or health-related perceptions. Five response choices accompany each item: definitely true, mostly true, don't know, mostly false, and definitely false. The entire HPQ battery appeared near the end of Form A in enrollment MHQs (except in Dayton) and near the middle of alternate annual HQs.[7] As reported by Ware and Karmos (1976) for non-HIE general population samples, these 26 items yield multi-item measures for six constructs: perceptions of past, present, and future health (health outlook), resistance to illness, sickness orientation (tendency to accept sickness as a part of life), and health-related worry/concern. The construct that each item was hypothesized to measure also appears in Table 4.

HPQ measures of these variables were added to HIE questionnaires to provide more complete data on the general health construct believed to be the common link between measures of the specific health dimensions, physical and mental. Among the multi-item general health measures available in the literature, the HPQ scales assessed the major general health constructs identified in our literature review. Their scoring had been studied, and results indicated that they would probably be reliable and reflect variability in health status for HIE samples.

ANALYSIS PLAN

Studies of the HIE general health perception measures were designed to achieve the three major goals we identified in Sec. I: (1) to evaluate in HIE samples the previously developed rules for scoring the construct-specific measures; (2) to develop and test an overall General Health Ratings Index; and (3) to increase understanding of how scores on the HIE general health perceptions measures should be interpreted. The analytic steps done to achieve these goals are described below in order of their presentation in the results. The methods we used to construct scales and to study validity are discussed in detail after the analysis plan.

General Health Perceptions Measures: Construct-Specific and Index

1. Examine item variability.
2. Evaluate the six summated ratings scales (Current Health, Prior Health, Health Outlook, Resistance to Illness, Health Worry/Concern, and Sickness Orientation) as recommended by Ware and Karmos (1976).
3. Evaluate a scoring algorithm for EGFP that assumes equal intervals between ratings of health as excellent, good, fair, or poor.
4. Construct a summary index that reflects the general health constructs assessed by the subscales and assesses both physical and mental components of health status.
5. Evaluate scoring options for the General Health Ratings Index to identify one that would be simple to compute and would provide reliable and valid scores.

---

[7]The HPQ items were approximately the 239th through 270th of some 280 items in Form A of the adult enrollment MHQ and the 142nd through 173rd of some 230 in the third annual HQ in Dayton. Precise counts of the number of items answered before HPQ items is difficult because skip patterns were used for many earlier questions.

6. Report descriptive statistics for the construct-specific measures and the Index. Compare HIE sample scores on the construct-specific measures with those on similar measures fielded in other general populations; particularly, with scores reported for a national sample representing adults in the National Health Interview Survey, and with HPQ scale scores reported for general population samples in earlier studies of these measures.
7. Estimate internal-consistency reliability for all measures in each HIE site and for the multi-item measures in groups that differ in sociodemographic characteristics.
8. Estimate the effect of respondent burden on mean scores and on reliability by comparing means and reliability coefficients calculated for early versus late questionnaire placement of the multi-item subscales and the Index; the later the items appear in the questionnaire, the greater the respondent burden in answering those items.
9. Estimate the stability of scores over a one-year interval.
10. Examine the sociodemographic correlates of the measures.

Interpretation of General Health Perceptions Scores

1. Test validity hypotheses for general health perceptions measures using information from correlations between the measures and 30 health and health-related measures (including age). In addition, for the Index, examine the clinical significance of scores in relation to five indicators of substantial physical and emotional impairment.
2. Address possible threats to validity from three types of response styles: acquiescent response set, opposition response set, and socially desirable response set.

METHODS OF ANALYSIS

Item Scoring

Before analyses began, EGFP, Pain, and Worry were each scored so that a higher score indicated more of the construct identified by the item's label: better health, more pain, and more worry or concern, respectively. To accomplish this, item responses appearing in the questionnaire were recoded to reverse scoring. No attempts were made to estimate missing responses for these items. (Additional studies were done to investigate better scoring algorithms for EGFP; these are described and results are presented in Sec. IV.)

HPQ items were also scored so that a higher score indicated more of the construct named by the scale to which the item belonged. Thus, for example, all items in the Current Health scale were scored (reversing precoded questionnaire responses when necessary) to indicate perceptions of better health at present; those in the Health Worry/Concern scale were scored to reflect greater worry and concern. Missing responses were estimated whenever possible during the multitrait scaling analyses described below.

Scale Construction Methods

Multitrait scaling techniques[8] were used to evaluate the six subscales: Current Health, Prior Health, Health Outlook, Resistance to Illness, Health Worry/Concern, Sickness

[8]This terminology has been used by Ware and others (1976, 1980) to refer to tests of item scalability that include multiple traits or constructs. This analytic strategy is not as complete as convergent-discriminant validation with

Orientation. Results of these multitrait scaling analyses and a principal components analysis were used in index construction.

Subscales. Multitrait scaling was used to test both the internal-consistency of the six subscales and the discriminant validity of items in those scales. Item groupings were defined to correspond to those identified in earlier HPQ studies (Ware and Karmos, 1976). The single-item Worry measure was added to the Worry/Concern item grouping and tested as an addition to that scale. Respondents who did not answer at least one item in each hypothesized scale were excluded from multitrait analyses. Otherwise, missing item scores were estimated for respondents included in these analyses by assigning each missing item the respondent's average score for other items in the same scale.[9] The combined-sites sample for these analyses numbered 4717.

We used two criteria to evaluate item-scale correlations, which had been corrected for overlap using the technique recommended by Howard and Forehand (1962). This correction provided more stringent tests of scaling criteria by removing the effect of the item being evaluated from the total scale score; because the scales that we tested were short, each item had a considerable influence on the total scale. Similar criteria were used by Ware and Karmos (1976) in their development of the HPQ measures.

First, to satisfy the Likert-type (internal-consistency) criterion underlying the Method of Summated Ratings (Likert, 1932), correlations between each item and its subscale should be substantial (absolute value ≥ 0.40) and positive. The second criterion, that of item discriminant validity, requires that the corrected correlation between an item and its hypothesized scale be higher than its correlations with other scales. In our analyses, the item discriminant validity criterion was satisfied and a scaling "success" counted each time the corrected correlation between an item and its own scale was more than two standard errors higher than its correlation with another scale. When the correlation between an item and another subscale was within two standard errors of its correlation with its own scale, a "probable" scaling error was counted. In such cases, we had reason to doubt whether correlations between items and their scales would be higher or lower than correlations with other scales upon replication. To take such marginal results into account, we defined and counted "probable" errors. Whenever the correlation between an item and its hypothesized scale was more than two standard errors below its correlation with another scale, we counted a "definite" scaling error.

General Health Ratings Index. After testing and finalizing the general health subscales, we explored construction of an overall General Health Ratings Index. We used the principal components method to extract the first unrotated component from a matrix of correlations among the 26 HPQ items. (Worry, EGFP, and Pain were not included in tests of the Index so the resulting Index would be comparable to indexes that could be constructed by others who have already fielded the HPQ.)

The first unrotated component, by definition, accounted for the largest proportion of item variance. We evaluated the component for equality of loadings and interpretability in each HIE site and across sites, emphasizing replication of findings because the component extraction process capitalizes on chance associations. Items that did not consistently correlate at least 0.30 (absolute value) with the first unrotated component in all sites were dropped from consideration for the General Health Ratings Index.

We compared three scoring options for the Index: (1) the algebraic sum of item responses for a simple summated ratings score; (2) the algebraic sum of standardized and unit-weighted item responses; and (3) factor scoring, weighting each standardized item score by its coefficient on the first unrotated component and then summing weighted scores.

---

the multitrait-multimethod matrix (Campbell and Fiske, 1959) because only one measurement method is represented. Multitrait scaling does extend beyond traditional internal-consistency analyses because it provides discriminant tests of item validity across traits.

[9]These procedures, which were also used by Ware and Karmos (1976) in their HPQ scaling studies, may have increased the internal consistency of the multi-item measures very slightly.

Reliability

We estimated the reliability of multi-item scales and the Index with Cronbach's (1951) Alpha coefficient, which uses the internal-consistency approach. This approach treats common item variance as true score (reliable variance) and both unique item variance and random error as error. Calculation of Alpha uses data collected on a single administration of the items. At each level of item homogeneity (average inter-item correlation), scale length directly affects the magnitude of internal-consistency estimates. Therefore, we also computed homogeneity estimates for each scale to permit direct comparisons between scales that differed in the number of items.

Following Helmstadter's (1964) recommendations, we judged internal-consistency (Alpha) estimates of 0.50 or above to be appropriate for group comparisons. Such comparisons represent the primary use of health status variables in testing HIE hypotheses about the effects of differences in coinsurance level.

The internal-consistency approach is not appropriate for single-item measures, and we considered the year-long interval between their administrations too long for estimating score reliability (but appropriate for examining trait stability). We therefore used multiple correlations (communality estimates) to estimate the point-in-time reliability of EGFP and Pain. These estimates were obtained by regressing EGFP and Pain, separately, on all the other general health perceptions items. As Guertin and Bailey (1970, pp. 215-216) point out, the communality estimate provides a lower-bound estimate of the item's test-retest reliability. Because the general health items used in the regressions represent a comprehensive definition of the general health construct, we expected the multiple correlation to provide a reasonably accurate lower-bound estimate of reliability for EGFP.

We first calculated reliability estimates for each multi-item scale separately in each site and across sites. The sociodemographic characteristics of samples in the HIE sites differ, particularly on education and income, variables known to be related to data quality. Between-site comparisons of reliability estimates thus provided information relevant to whether reliability would be acceptable in the "worst case," when the poorest data quality might be expected.

We then computed reliability estimates for each multi-item scale across sites and separately by site for subgroups that differed in sociodemographic characteristics. For these computations, subgroups were defined on the basis of education (years of school completed, available only for respondents aged 17 and older). We defined groups on the basis of education only, rather than education and income. Multivariate analyses of the relationship between respondent characteristics and the reliability of health-related ratings (Ware and Young, 1976) indicate that income does not make a unique contribution to predictions of reliability when education is in the equation. After examining sample distributions on the education variable, six subgroups were defined: 0-9, 10-11, 12, 13-15, 16, and 17+ years of schooling. Across sites, sample sizes for these subgroups ranged from 233 to 1342; within each site, sample sizes for the subgroups ranged from 22 to 535 and exceeded 100 in the majority of our analyses.

Stability

We examined the stability of the HIE general health perceptions measures with data from respondents in Seattle, the site with the largest enrollee sample. Product-moment correlations were computed between enrollment MHQ scores and scores on the same measures from the first annual HQ, administered an average of 13 months later.[10] (Files linking these

[10]The Seattle sample was enrolled between January and September 1976; the first annual HQ was fielded in June 1977.

questionnaires in other sites and for the combined-sites sample were not available when these analyses were done.) Only respondents who provided complete data on general health and other health measures at both times were included in stability analyses; the sample size for these correlations was approximately 1200.

Interpretation of General Health Perceptions Scores

To understand the meaning of scores on the general health perceptions measures, we relied on a construct validity approach because there are no generally accepted criterion measures in the health status field. Details of these methods and the hypotheses tested during construct validity studies appear below. (The methods used to study the effect of response sets on measurement validity are discussed in App. C.)

The construct validity approach involves examining patterns of relationships among measures of general health and other health and health-related variables to which they should (or should not) be related according to theory. More specifically, the construct validity approach requires empirical standards against which to evaluate the direction and magnitude of the observed relationships. We relied on previous studies of general health perceptions measures, particularly those reported by Ware and Karmos (1976) and those summarized in our other reports (WDD, 1978; Ware, Davies-Avery, and Brook, 1980) in proposing the validity hypotheses for HIE general health perceptions measures.

Most of our hypotheses deal with expected relationships between categories of variables (e.g., general health and physical health), because theory and empirical studies are not complete enough to derive specific hypotheses for different operational definitions of general health variables (e.g., Current Health versus Health Outlook in relation to physical health). In a few cases, the theoretical relationship is intuitive or has been studied with operational definitions much like those used in the HIE, and we have proposed specific hypotheses.

The variables used in our validity studies are identified and defined in Table 5, which also includes reliability estimates that were helpful in interpreting validity coefficients. All variables were scored from HIE data gathered in Seattle on the self-administered MHQ and from the screening examination done at enrollment, and on claims forms for services used during the subsequent year. These variables were not available on enrollment files from other sites when we did our validity studies. The sample size for most validity variables was 1200; for the variable defining subsequent use of services, the sample size was 1557.

We computed product-moment correlations between the HIE general health perceptions measures and these variables to test the validity hypotheses. To be considered valid as measures of general health perceptions, we hypothesized that the HIE measures would be:

- Positively related (if favorably scored) to both physical and mental health measures. The correlations observed with variables in the physical and mental health categories should be the largest observed for general health perceptions measures. We expected Current Health generally to show the most substantial validity coefficients in relation to health. We expected Health Outlook to be most closely associated with measures of chronic problems, and more closely related than the other general health constructs to mental health. We also expected Worry/Concern to be closely related to measures of anxiety, and Resistance to measures of acute symptoms.
- Generally unrelated to measures of social contacts and social resources, a hypothesis based on theory and empirical results discussed by Donald and Ware (forthcoming) as well as results reported in earlier studies of general health perceptions measures.
- Generally unrelated to measures of patient role propensity, with the exception of Worry/Concern and Resistance. We expected these two scales to have positive and

Table 5

DEFINITIONS AND RELIABILITY ESTIMATES FOR VARIABLES USED TO STUDY CONSTRUCT VALIDITY OF GENERAL HEALTH PERCEPTIONS MEASURES

Category/Variable	No. of Items	Reliability Estimate[a]	Operational Definition
PHYSICAL HEALTH			
Chronic Personal Limitations	9	.98	Chronic (>3 months duration) limitations in self-care, mobility and physical activities
Current Personal Limitations	9	.97	Acute or chronic limitations in self-care, mobility, and physical activities
Chronic Role Limitations	3	.93	Chronic (>3 months duration) limitations in major role activity (work, school, housework) and general functioning
Current Role Limitations	3	.93	Acute or chronic limitations in major role activity (work, school, housework) and general functioning
Physical Capacities	12	.99	Performance of or capacity to perform mild, moderate, and strenuous physical activities without difficulty
Strenuous Exercise	2	NA	Number of hours/week spent in strenuous exercise
General Exercise	1	NA	Overall level of physical activity
Acute Symptoms	27	NA	Count of number of physical symptoms, past month, weighted for severity
Chronic Disease (self-report)	23	NA	Count of number of chronic diseases (e.g., diabetes, anemia) present at enrollment weighted for severity
Chronic Disease (screening)	11	NA	
MENTAL HEALTH			
Anxiety	10	.91	Anxious, nervous, tense, worried, past month
Depression	9	.87	Depressed, downhearted, felt blue, past month
Positive Well-Being	8	.91	In good spirits, happy, cheerful, past month
Emotional Ties	2	.83	Felt loved, cared for, wanted by others, past month
Emotional Stability	21	.94	Positively scored, unweighted sum of anxiety, depression, and 2 self-control items
Mental Health Index	38	.96	Mental health summary score (unweighted sum of emotional ties, positive well-being, negative of emotional instability, and 7 other items)
SOCIAL CIRCUMSTANCES			
Close Friends/Relatives	1	NA	Number of close friends, relatives feel close to
Neighborhood Acquaintances	1	NA	Number of people in neighborhood know well enough to visit at home

Table 5—continued

Category/Variable	No. of Items	Reliability Estimate[a]	Operational Definition
Telephone Contacts	1	NA	Number of telephone contacts with close friends, relatives, past month
Getting Along	1	NA	Subjective evaluation of how well get along with others
Attendance at Religious Services	1	NA	Number of times attended religious services, past month
Group Participation	2	.84	Number of formal and informal group memberships maintained, past month or more
Social Contacts	3	.72	Number of visits to and from family, friends, and relatives, past month or more
Social Well-Being Index	9	.69	Social well-being summary score (unweighted sum of preceding measures, except Getting Along)
PATIENT ROLE PROPENSITY			
Attitude Toward Going to the Doctor	2	.76	Favorable attitude toward going to the doctor
Rejection of Sick Role	4	.58	Conscious avoidance of sick/patient role
USE OF SERVICES			
Recent Treatment for Acute Symptoms	1	NA	Whether care was sought for acute physical symptoms, past month (logged variable)
Subsequent Use of Outpatient Care	1	NA	Number of ambulatory care visits (not including mental or dental care) during year after general health ratings were collected (from claims data)
SUBSTANTIAL HEALTH IMPAIRMENT			
Any Physical Impairment	1	NA	Presence of self-reported limitations in self-care, mobility, or physical activities
Any Emotional Impairment	1	NA	Below 20th percentile of score distribution on the Mental Health Index
Any Chronic Disease	1	NA	Presence of one or more chronic diseases (asthma, chronic obstructive pulmonary disease, congestive heart failure, joint problems, anemia, thyroid disease, kidney disease, hay fever, hypertension, diabetes, hearing problems, ulcers)
Any Physical or Emotional Impairment	1	NA	Any substantial physical or emotional impairment (as defined immediately above)
Any Health Impairment	1	NA	Any substantial physical or emotional impairment or chronic disease (as defined immediately above)
OTHER VALIDITY VARIABLES			
Age	1	NA	Age of respondent in years
Life Events	20	NA	Number of life events during past six months, weighted for amount of stress

[a]Reliability of measures in Seattle at enrollment; NA=reliability estimate not available .

low to moderate correlations with the patient role propensity measures, on the basis of empirical findings reported by Ware and Karmos (1976).

- Negatively related (if favorably defined) to measures of use of services. We expected measures of recent use to be more closely related than those of subsequent use to most of the general health measures; "recent" refers to use during the month before the general health ratings were filled out, and "subsequent" to use in the year after the ratings were done.
- Negatively related (if favorably defined) to measures of life stress. Although there is little empirical evidence relevant to this relationship, we hypothesized one because of the known correlation between mental health (which is related to general health perceptions) and life stress.
- Negatively related to age if a time-bound operational definition (i.e., Current Health, Prior Health, and Health Outlook); positively related to age, for Resistance to Illness. On the basis of results reported by Ware and Karmos (1976), we expected Health Worry/Concern and Sickness Orientation to be negatively related to age.

Following an approach to validation of health measures mentioned briefly in our earlier report (WDD, 1978), we examined the clinical significance of scores on the General Health Ratings Index in relation to several dichotomously scored indicators of substantial physical and emotional impairment. These indicators included:

- Any Physical Impairment, on which a score of 1 indicated the presence of self-reported limitations in personal functioning (i.e., limitations in self-care, mobility, or physical activities, as defined by the HIE personal limitations and physical capacities measures; Stewart, Ware, and Brook, 1981) and 0 indicated the absence of any such limitations.
- Any Emotional Impairment, on which a score of 1 indicated that the respondent scored below the 20th percentile on the Mental Health Index (a reasonable cutoff for identifying impairment requiring medical care; Ware et al., 1979) and 0 otherwise.
- Any Chronic Disease, on which a score of 1 indicated presence of one or more chronic diseases (determined using self-report for some diseases and data from a screening examination for others) and 0, no such diseases. (The specific diseases are identified in Table 5.)[11]

We also defined two aggregate indicators of substantial health impairment. Definition of these two indicators allowed us to examine the amount of extra information about Index scores provided by the different definitions of limitations:

- Any Physical or Emotional Impairment, scored 1 if the respondent had a score of 1 on the above measures of physical or psychiatric impairment, and 0 otherwise.
- Any Impairment, scored 1 if the respondent had a score of 1 on the above measures of any physical or psychiatric impairment or chronic disease, and 0 otherwise.

To examine the distribution of health impairments throughout the range, we first divided the Seattle sample into deciles (N = 150 each) on the basis of their Index scores, and plotted the mean score on each dichotomously scored indicator—the probability of the given impair-

[11]Because not all persons were screened, missing data posed a problem in correctly scoring this variable. Estimation of point-prevalence rates was difficult, so anyone with missing data on one or more of the disease variables was presumed to have no chronic diseases, a definition that undoubtedly understates the true prevalence of the selected chronic diseases in this sample and thus underestimates their impact on Index scores.

ment—at the Index mean score for each decile. In addition to examining the relationships visually, we used one-way analyses of variance to test the strength of the relationships and examine their linearity (Nie et al., 1975).

IV. RESULTS: SCALING, SCORING, RELIABILITY, AND STABILITY OF GENERAL HEALTH PERCEPTIONS MEASURES

In this section, we first discuss scaling and scoring of the multi-item subscales (Current Health, Prior Health, Health Outlook, Resistance to Illness, Health Worry/Concern, and Sickness Orientation), the General Health Ratings Index, and the two single-item measures (EGFP and Pain). We also compare HIE scaling results for the multi-item subscales with those reported by Ware and Karmos (1976), who originally developed these measures (see also WDD, 1978). Next, we examine the reliability of all measures at a point in time for the combined-sites sample and of the multi-item scales for groups differing in educational attainment. Following that, we look at the effects of questionnaire placement on measurement reliability and mean scores. We then present one-year stability estimates for all the general health measures. Finally, we report the sociodemographic correlates of the measures. Section V considers the interpretation of scores on the general health perceptions measures.

DESCRIPTIVE STATISTICS: ITEMS

Before scoring and evaluating the multi-item general health perceptions subscales, we examined descriptive statistics for the 27 items used in these scales. Our examination focused particularly on comparability of variances; roughly symmetrical (if not normal) response distributions and standard deviations near 1.00 (when a five-point response continuum is used) are desirable characteristics for items combined in a summated ratings scale. In the results discussed below, all items had been scored to indicate more of the construct assessed by their intended scale (see Table B.1 for item scoring rules).

Table 6 presents means and standard deviations for these items in the combined-sites sample. Frequency distributions for responses and descriptive statistics for each site appear in Tables D.1 and D.2, respectively.

Score distributions for almost all items were skewed with means on the side of the response scale midpoint (2.5 for Worry; 3.0 for A-FF) that indicated more favorable ratings. The only item mean consistently at or below the midpoint across sites was a worry item (M), on which lower scores indicate less worry. Three of the five items with the most extreme mean scores (above 4.00) across sites assess current health (I,Z,DD); one (K) assessed prior health and another (C) reflected perceptions of bodily resistance to illness. This skewness probably reflects the actual perceptions of favorable health in these general, nonaged samples, rather than response bias. To support this, examination of frequency distributions for response choices revealed little tendency to select the most extreme responses (see Table D.1). (The issue of response bias is discussed at greater length in App. C.) Furthermore, the skewness of these items was somewhat less apparent in the South Carolina sites, where health status, as reflected in self-report measures of physical functioning and mental health, is somewhat poorer on average than in other HIE sites (Stewart, Ware, and Brook, 1981; Veit and Ware, forthcoming).

While standard deviations showed a considerable range across items, those for each item were very comparable across sites (see Table D.2). Moreover, standard deviations for 20 of these 27 items were very close to 1.00 (within 0.2 points above or below). Items with the greatest variability across sites generally had standard deviations of 1.30 and above and assessed Prior Health (K,S,BB), Concern (X), and Current Health (L). Because the magnitude

Table 6

MEANS AND STANDARD DEVIATIONS, GENERAL HEALTH RATING ITEMS, ALL SITES COMBINED (N = 4717)

Item		Mean	S.D.
A	Doctors say health is now excellent	3.93	0.97
C	Get sick easier than others*	4.21	1.05
D	Feel better now than ever	3.28	1.16
E	Probably sick a lot in future*	3.85	0.97
F	Never worry about my health*	2.78	1.28
G	People get sick easier than I	3.44	0.92
I	Somewhat ill*	4.28	1.05
J	Expect to have better health than others	3.44	0.88
K	So sick once thought I might die*	4.04	1.42
L	Not as healthy now as used to be*	3.39	1.36
M	Worry about health more than others	2.19	1.09
O	Body resists illness	3.81	0.91
P	Getting sick part of life	3.09	1.29
Q	Healthy as anybody I know	3.78	1.04
R	Health worse in future than now*	3.61	0.96
S	Never had a long illness	3.87	1.40
T	Others more concerned about health*	2.85	1.00
V	Health is excellent	3.74	1.09
W	Expect a healthy life	3.97	0.89
X	Health is a concern	3.52	1.32
Y	Accept that sometimes I'm sick	3.62	1.19
Z	Been feeling bad lately*	4.04	1.15
BB	Never been seriously ill	3.58	1.51
CC	Something going around, catch it*	3.81	1.00
DD	Doctors say now in poor health*	4.38	0.93
FF	Feel as good now as ever	3.77	1.15
6	Worry or concern about health*	1.76	0.91

NOTE: All items were scored as in Table B.1 so a higher score reflected more of the construct assessed by the subscale to which the item belonged. Scoring for those marked with an asterisk (*) has therefore been reversed from that precoded in the questionnaire (see App. A).

of item variances tended to relate to the health perceptions construct measured (particularly notable for Prior Health items), lack of comparable variances was a greater concern during index construction than in scoring the multi-item subscales.

MISSING DATA

Respondents provided virtually complete data on all general health items. The scoring algorithm that estimated missing responses for the multitrait scaling analyses could be used for 4717 (more than 99 percent) of the 4763 adults in the original analytic sample. Site by site, from 95.6 to 99.9 percent of respondents provided data that could be used in our scaling analyses. For respondents included in site-specific analyses, missing responses were extremely rare, ranging from 0.04 to 0.1 percent across all items in each site. Across all sites, only 0.07 percent of all possible responses were missing and had to be estimated (by the respondent's mean score on other items in the scale being tested). Dayton had the highest rate (4.4 percent) of respondents providing data that were too incomplete for use in scaling analyses; South Carolina respondents included in scaling analyses had the highest rate (0.01 percent) of missing item responses. Only one item consistently had notable numbers of missing responses (Q, healthy as anybody I know).

CONSTRUCTION OF THE MULTI-ITEM SUBSCALES

Multitrait Analyses

Table 7 presents the multitrait correlation matrix for the combined-sites sample that we evaluated during scaling analyses of the six multi-item subscales. Site-specific matrices appear in Tables D.3 through D.6. In all these matrices, row entries represent correlations between each item and the sum of items in each scale grouping. Asterisks indicate the hypothesized scale placement of each item and identify item-scale correlations that were corrected for overlap (i.e., correlations between an item and the sum of the remaining items in its scale).

In the combined-sites matrix, items in the Current Health, Prior Health, Health Outlook, and Resistance scales correlated substantially (0.40, after correction for overlap) with their own scales. With only a few exceptions, these item-scale correlations achieved the same standard in the site-specific analyses. Thus, items in these scales satisfied the first multitrait scaling criterion. The item-scale correlations for items in Health Worry/Concern and Sickness Orientation generally did not meet this criterion, although they were most often in the mid to high 0.30s.

The results from our tests of the item discriminant validity criterion for these measures are summarized in Table 8. Entries in Table 8 are ratios of the number of times the criterion was met successfully to the total number of tests for each scale; ratios indicate results in both the combined-sites sample and in site-specific samples. The last row of entries represents the overall success ratio, across scales, for each site and the combined-sites sample.

According to the discriminant validity criterion, items should correlate higher with the scale that defines the construct the item is supposed to measure than with other scales in the matrix. In 148 tests of the item discriminant validity criterion for the combined-sites sample,

Table 7

Item-Scale Correlation Matrix, General Health Perceptions Measures, All Sites Combined (N = 4717)

	Item Grouping/Item	Scale CH	PH	HO	RE	WC	SO
	Current Health (CH)						
V	Health is excellent	.74*	.37	.54	.44	-.41	-.14
FF	Feel as good now as ever	.71*	.28	.42	.32	-.33	-.09
A	Doctors say health is now excellent	.64*	.30	.42	.31	-.31	-.10
Z	Been feeling bad lately	.63*	.29	.37	.38	-.36	-.21
L	Not as healthy now as used to be	.62*	.24	.40	.26	-.33	-.11
Q	Healthy as anybody I know	.62*	.32	.50	.44	-.38	-.08
I	Somewhat ill	.61*	.35	.42	.37	-.33	-.17
D	Feel better now than ever	.57*	.17	.34	.25	-.25	-.05
DD	Doctors say now in poor health	.53*	.28	.34	.30	-.23	-.11
	Prior Health (PH)						
BB	Never been seriously ill	.28	.55*	.18	.19	-.19	-.04
S	Never had a long illness	.32	.45*	.22	.27	-.24	-.10
K	So sick once thought I might die	.32	.40*	.21	.22	-.20	-.08
	Health Outlook (HO)						
R	Health worse in future than now	.40	.16	.58*	.27	-.22	-.17
W	Expect a healthy life	.60	.29	.54*	.39	-.29	-.12
E	Probably sick a lot in future	.42	.21	.52*	.38	-.31	-.21
J	Expect to have better health than others	.30	.13	.43*	.32	-.18	-.09
	Resistance (RE)						
O	Body resists illness	.44	.25	.40	.57*	-.33	-.15
C	Get sick easier than others	.42	.30	.35	.50*	-.30	-.22
CC	Something going around, catch it	.26	.14	.25	.48*	-.18	-.25
G	People get sick easier than I do	.25	.15	.33	.42*	-.25	-.13
	Health Worry/Concern (WC)						
M	Worry about health more than others	-.35	-.22	-.33	-.32	.39*	.13
F	Never worry about my health	-.34	-.18	-.23	-.25	37*	.11
X	Health is a concern	-.17	-.11	-.13	-.12	35*	.15
6	Worry or concern about health	-.57	-.31	-.27	-.30	33*	.13
T	Others more concerned about health	-.10	-.09	-.14	-.22	32*	.05
	Sickness Orientation (SO)						
P	Getting sick part of life	-.18	-.12	-.20	-.28	.16	.36*
Y	Accept that sometimes I'm sick	-.09	-.03	-.13	-.15	.13	.36*

Note: Standard error of the correlation equaled 0.01. '*' denotes hypothesized item groupings and scale correlations that were corrected for overlap.

Table 8

SUMMARY OF MULTITRAIT SCALING RESULTS FOR GENERAL HEALTH PERCEPTIONS MEASURES, ALL SITES COMBINED AND BY SITE[a]

Measure	No. of Items	Site				
		All Sites Combined	Dayton	Seattle	Massachusetts	South Carolina
Current Health	9	45/45	45/45	45/45	45/45	45/45
Prior Health	3	15/15	14/15	15/15	15/15	14/15
Health Outlook	4	23/24	23/24	23/24	23/24	23/24
Resistance	4	24/24	24/24	24/24	23/24	21/24
Health Worry/Concern	5	28/30	17/30	29/30	27/30	24/30
Sickness Orientation	2	10/10	10/10	10/10	10/10	10/10
Overall Success Ratio		144/148	133/148	146/148	143/148	137/148

[a]Entries indicate number of times the item discriminant validity criterion was met successfully over the total number of tests for each scale. The total number of tests equalled the number of items times one less than the number of scales.

we counted 144 successes (97.3 percent).[1] Across sites, results were more than satisfactory. Site-specific overall success rates ranged from 89.9 percent to 98.6 percent. With only one exception (the Health Worry/Concern scale in Dayton), the patterns of scaling successes and errors were strikingly similar in the site-specific and combined-sites analyses. For that reason, we focus here on results from the combined-sites sample.

All Current Health items scaled without error in each site and in the combined-sites analysis. Items in the Prior Health scale successfully met the scaling criterion in the combined-sites sample. Three of the four Health Outlook items also scaled perfectly; a notable finding was the markedly higher correlation of item W (expect a healthy life) with Current Health, which caused definite scaling errors in all analyses. Resistance items successfully passed the scaling tests in the cross-sites analysis and in two of four sites; the four errors noted for these items in site-specific analysis were probable, most resulting from correlations with the Current Health scale. Items in the Sickness Orientation grouping scaled perfectly in the combined-sites analysis and in all sites but one.

Four of the five Health Worry/Concern items scaled perfectly in the combined-sites analyses. An error noted in all analyses was for item no. 6 (worry or concern about health), which correlated highest with Current Health. This item, a hypothesized addition to the four-item scale originally tested by Ware and Karmos (1976), accounted for most of the site-specific errors we observed for this scale. Item no. 6 correlated with all scales but Sickness Orientation, thus displaying a notable lack of discriminant validity. Error patterns for the other Health Worry/Concern items occurred less consistently across sites, and no others were observed in the combined-sites analysis. Both Worry items (M and F), but not the Concern items, accounted for probable errors in South Carolina. As mentioned earlier, errors for this item grouping were particularly noticeable in Dayton. All items but one (T, others more

[1]Definitions of scaling successes and errors appeared in Sec. III.

concerned about health) accounted for errors in Dayton; all Worry items had probable or definite errors on all scales but Sickness Orientation. Health Worry/Concern items also showed less convergent validity than items in the other groupings; their item-scale correlations were notably lower than those for most items in other scales.

Items in the Current Health, Prior Health, Health Outlook, Resistance, and Sickness Orientation scales thus demonstrated considerable discriminant validity. Correlations with Current Health of items hypothesized to assess other constructs accounted for most of the scaling errors observed for the other five multi-item scales. The most noticeable problems in this regard, across sites, were the correlations of item W and item no. 6 with Current Health. The overlap between other items (noted chiefly for Health Worry/Concern and Resistance items in site-specific analyses) and Current Health suggests that Current Health may represent the most integrated (or core) perception of health among those represented in the battery; the principal components analysis that follows confirmed this hypothesis.

In a preliminary set of multitrait analyses (data not reported), we evaluated the possibility of scoring the Health Worry/Concern items in two separate scales, as well as in a combined version. Two Health Worry items (M and F) displayed discriminant validity; although item no. 6 correlated highest with Current Health, it correlated higher with other worry items than with the Health Concern items. In all tests, the two Health Concern items (T and X) consistently failed the discriminant validity criterion; they correlated higher with the Health Worry items than they did with each other. For this reason, they also failed to demonstrate minimal reliability (0.50). Thus, while the Health Worry items might be scored as a separate scale, the two Health Concern items should not be so scored. We therefore dropped this line of evaluation, and continued our analyses with the combined Health Worry/Concern scale.

Comparison of results across sites indicated that the patterns of item scaling errors (with the exception of Health Worry/Concern in Dayton) were generally very similar. Very few scaling errors were observed for these six multi-item scales in Seattle and Massachusetts; apart from problems with Health Worry/Concern, very few errors were observed in Dayton as well. Health Worry/Concern items also accounted for most of the scaling errors we found in South Carolina. We noted more errors for scales in that site than the other sites, and item scale correlations were generally lower in South Carolina, the most socioeconomically disadvantaged HIE site.

Comparison with Earlier Results

One of our goals for the multitrait analyses was to evaluate the multi-item general health perceptions subscales as they had been constructed and scored by their developers (Ware and Karmos, 1976). As part of this evaluation, we compared the item discriminant validity results observed in HIE sites with those reported for the scales by Ware and Karmos.[2] Their results were based on cross-sectional surveys of approximately 2000 adults drawn from five general populations in Illinois and California between 1973 and 1975.

Ware and Karmos used a somewhat more liberal criterion to define scaling successes than we did in HIE analyses, and did not distinguish probable and definite errors.[3] Before comparisons were made, therefore, we applied their criterion to HIE site-specific multitrait tables, and the HIE success/error criteria to their site-specific tables (data not reported).

When the same criteria were applied to HIE and Ware-Karmos data, results were strikingly comparable. When the Ware-Karmos criterion was used, the most notable and consistent

---

[2]As indicated earlier, we added item no. 6 to the originally hypothesized Health Worry/Concern grouping. Results pertaining to this item were therefore not considered in making these comparisons.

[3]Ware and Karmos (1976, p. 137) counted a successful test of the discriminant validity criterion when an item correlated higher with its hypothesized scale (after correction for overlap) than with other scales.

error in both data sets was the overlap of item W from the Health Outlook scale with Current Health. Using the HIE criteria, the Ware-Karmos data showed success and error patterns very similar to those reported above for the HIE scales. In particular, Health Worry/Concern items accounted for many of the errors observed in the Ware-Karmos data. Moreover, most scaling errors were observed in the most disadvantaged site they surveyed (East St. Louis), similar to the HIE finding of somewhat more errors across all scales in South Carolina than in other sites.

Scoring the Multi-Item Subscales

On the strength of these findings and their comparability with results from previous multitrait scaling studies of the items, we decided to score the scales using the item groupings recommended by Ware and Karmos (1976), as shown in Table 7. Item W remained in the Health Outlook scale, despite its error pattern, to maintain consistency with previous scoring rules. Largely because of content similarity, we decided to include item no. 6 in the Health Worry/Concern scale, and to monitor its possibly detrimental effect on reliability and validity.

For all further analyses reported here, we scored the six multi-item general health perceptions scales by using the simple algebraic sum of item responses (after items had been recoded as described earlier). Although summated ratings scales are often scored by taking the average item score, rather than the sum of item scores, we do not recommend this scoring method. We believe it invites inappropriate conclusions based on between-scale score comparisons (e.g., people rate their current health better than their past health); the scaling method does not permit such comparisons. Because item variances and item-scale correlations were generally comparable within each scale, there was no need to standardize or weight items before scoring, thus enabling us to avoid complexity in the scoring rules. The item and scale scoring rules that we used, and recommend to others who might field these items, appear in Table B.1. Briefly, precoded questionnaire responses are reversed where necessary so higher scale scores indicate more of the construct named by the scale. Thus, all scales but Health Worry/Concern and Sickness Orientation are scored positively; these two are scored negatively.

CONSTRUCTION OF A GENERAL HEALTH RATINGS INDEX

Principal Components Analysis

As described in Sec. III, we used a principal components analysis to identify the underlying dimension that accounted for most of the measured variance shared by the 26 items in the Health Perceptions Questionnaire (HPQ).[4] Correlations between these items and the first unrotated component across sites and for each site appear in Table D.7; the inter-item correlation matrix that we used in this analysis is reproduced in Table D.8. The most notable result from this step of the analysis was that four items consistently failed to correlate 0.30 (absolute value) with the first component. We regarded a correlation of that magnitude to be the minimal requirement for considering an item as part of a common general health construct. The four items involved were the two Concern items (X and T) and both Sickness

[4]As we noted earlier, EGFP, Pain, and Worry (item no. 6) were not considered for inclusion so that the resulting Index might be used by others who have fielded only the HPQ.

Orientation items (P and Y). We therefore dropped these items from consideration and extracted the first principal component from the remaining 22 items to re-estimate coefficients for a 22-item index.

Correlations between that unrotated component and the 22 items for the combined-site and individual-site samples appear in Table 9. Items are listed in order of the magnitude of their correlations in the combined-sites analysis. An abbreviation next to each item indicates its scale placement among the construct-specific measures. (For the principal components analyses, item responses were not reversed as shown in App. B; thus, the sign of the correlation indicates the original direction of item wording, favorable or unfavorable.)

Across sites, the first component explained approximately one-third of the variance measured by these items, ranging from 30.6 percent in South Carolina to 38.3 percent in Dayton.

Table 9

CORRELATIONS BETWEEN 22 GENERAL HEALTH PERCEPTIONS ITEMS AND THE FIRST UNROTATED COMPONENT, ALL SITES COMBINED AND BY SITE

Item/Scale	All Sites Combined	Dayton	Seattle	Massachusetts	South Carolina
V CH	.80	.82	.80	.81	.78
FF CH	.70	.76	.70	.71	.67
Q CH	.72	.72	.70	.73	.73
W HO	.70	.75	.70	.68	.70
A CH	.68	.75	.65	.64	.69
I CH	-.68	-.73	-.66	-.67	-.56
Z CH	-.67	-.71	-.65	-.67	-.63
L CH	-.63	-.65	-.63	-.66	-.58
DD CH	-.59	-.66	-.56	-.56	-.57
D CH	.58	.60	.56	.58	.56
O RE	.60	.61	.63	.60	.56
E HO	-.57	-.62	-.57	-.56	-.50
C RE	-.57	-.59	-.58	-.56	-.52
R HO	-.52	-.59	-.51	-.49	-.47
M WC	-.47	-.48	-.49	-.48	-.41
S PH	.44	.50	.36	.47	.47
J HO	.44	.51	.44	.38	.42
F WC	.41	.45	.39	.43	.39
K PH	-.40	-.53	-.32	-.40	-.41
G RE	.40	.45	.44	.33	.39
CC RE	-.40	-.40	-.48	-.37	-.36
BB PH	.38	.47	.34	.37	.37
Percent of Variance Explained	33.1	38.3	32.4	32.4	30.6

NOTE: CH=Current Health; PH=Prior Health; HO=Health Outlook; RE=Resistance; WC=Health Worry/Concern. Item responses were not recoded for the principal component analysis.

Because higher item scores on the favorably worded items indicate better perceptions of health, higher scores on the component thus reflect more favorable general health perceptions. Directions of association were identical in all analyses; favorably worded items correlated positively and unfavorably worded items negatively with the component in each site and across sites.

As required, all items correlated above 0.30 (absolute magnitude) in all analyses. In each analysis, we found that most items correlated 0.50 (absolute value) or higher with the first unrotated component. The magnitude of each item's correlation with the component was markedly similar from site to site; the only notable exceptions were the generally lower correlations for Prior Health items in Seattle than in the other sites. Moreover, the ordering of items on the component was also quite consistent from site to site. We found the consistency of these results encouraging, given our interest in developing a generalizable index.

In terms of magnitude, the Current Health items and one Health Outlook item (W, which had correlated highest with Current Health items in the multitrait analyses) correlated highest with the first unrotated component. Items from the other construct-specific scales were generally mixed in their ordering on the component. The ordering did not reflect the greater-than-average variability of Prior Health items. In fact, Prior Health items tended to correlate lower with the component than did Health Outlook items.

One explanation for the prominence of the Current Health items on the first component might be that they are better represented; 9 (possibly 10 if W is considered) of the 22 items assess Current Health. Thus, a large proportion of the variance in the inter-item correlation matrix was shared with these items. To test this explanation, we randomly selected four Current Health items (two worded favorably—Q and V—and two unfavorably—I and L—to preserve scale balance) so that roughly equal numbers of items assessed each general health construct. Given the substantial correlations among the Current Health items, we judged it very likely that the conclusions from any one random draw would mirror those from multiple replications of the random draw. We then examined the first unrotated component derived from correlations among these 17 items in the combined-sites sample. As the pattern of item-component correlations indicates (see Table D.9), Current Health items still had the highest correlations with the component when we analyzed a similar number of items per construct.

Another explanation for the ordering of items that we observed on the 22-item component could be that of differential item reliability. Items that contain more reliable measured variance (rather than simply more measured variance) would correlate higher with the component, particularly if their reliable variance were also common variance. Were this explanation true, we might expect to find that Current Health items were the most reliable, and Prior Health items among the least reliable, given the ordering of these items on the component. Although we did not estimate HPQ item reliabilities in the HIE, Ware and Karmos (1976, pp. 107-108) reported item reliabilities based on six-week test-retest correlations. Current and Prior Health items had very similar reliabilities (median of high 0.50s for both sets of items). Resistance and Health Worry/Concern items were very slightly less reliable. Differential item reliability therefore does not account for the ordering of the 22 items on the unrotated component.

The best explanation of our findings is that perceptions of Current Health define the core concept underlying these general health ratings. Perceived health in the past contributes little and the other health-related perceptions a middling amount to the dimension's definition. Health-related concern and perceptions that sickness is a part of life contribute negligibly to the common dimension, and thus to an overall index based on it. We consider the interpretation of the dimension (Index) more comprehensively in Sec. V.

Index Scoring

After identifying the items that best defined the common general health construct, we tested three scoring algorithms to identify the simplest method and the most reliable score for the 22-item General Health Ratings Index. The first and simplest scoring algorithm we tested was simply the algebraic sum of unstandardized item responses (Likert-type summated ratings). This method assigns an apparent weight of 1.00 to each raw item score in the sum; items actually contribute to the Index score in amounts proportional to their variances, which are not equal. Because we noted differences in item variability, we took them into account in the second scoring algorithm by standardizing item scores before summing them for the Index score (standardized items, unit weight scoring method). Magnitudes of item-component correlations also differed, so the third scoring algorithm assigned different weights to the standardized items (factor coefficients from the combined-sites component) before summing them for the Index score (factor scoring method).

Reliability estimates for the scores yielded by these three algorithms appear in Table 10. As can be seen, we found virtually no differences in reliability as a function of scoring method. We therefore decided to use the simplest algorithm to score the General Health Ratings Index, namely the simple summated ratings score. To score the Index, all items are first scored so that a high number reflects better health or *less* worry. Item scoring rules for the Index therefore differ—for the Worry items only (F and H)—from those used to score the subscales (see Table B.1). Details of the Index scoring appear in Table B.2; these rules were used in all remaining analyses presented in this report.

Table 10

EFFECTS OF THREE DIFFERENT SCORING METHODS ON RELIABILITY OF GENERAL HEALTH RATINGS INDEX, ALL SITES COMBINED

Scoring Method	Reliability Estimate[a]
Likert-type Summated Ratings	.89
Standardized Items, Unit Weight Scoring	.90
Factor Scoring	.90

[a]The reliability estimate for the summated ratings score is the Alpha coefficient for unstandardized items (Cronbach, 1951); for the standardized scoring, the standardized Alpha; and for factor scoring the theta coefficient (Armor, 1974).

SCORING PAIN AND EGFP

Two of the general health perception items, Pain and EGFP, were not used to construct multi-item scales. Precoded questionnaire responses for each item ranged from one to four, and assumed an equal interval between response choices (see App. A for content of response choices). Because the Current Health scale and EGFP represent different methods of obtain-

ing ratings of personal health at the present, we used the Current Health scale as a criterion to test the equal-interval assumption made in scoring EGFP. We had no such criterion for the Pain item; therefore, its scoring remained as indicated above, after reversing the precoded questionnaire responses so that the higher scores indicated greater pain.

We calculated mean Current Health scores for respondents who chose each of the four EGFP response categories. These means, bracketed by plus and minus one standard deviation, are plotted for each EGFP group in Fig. 1. As can be seen, the middle two-thirds of each response distribution except the "poor" category overlap somewhat; this finding indicates that the coarser EGFP scale does not discriminate between different levels (presumably health levels) as well as does Current Health. Results also indicate that although EGFP scores decreased ordinally from "excellent" to "poor," the intervals were unequal. The largest interval, approximately one and one-half standard deviations, appeared between "poor" and "fair." To reflect these unequal intervals, we recoded each EGFP response choice at the mean Current Health scale score for respondents selecting that choice. (Details of this transformation appear in Table B.1.) EGFP scores were so transformed for all remaining analyses.

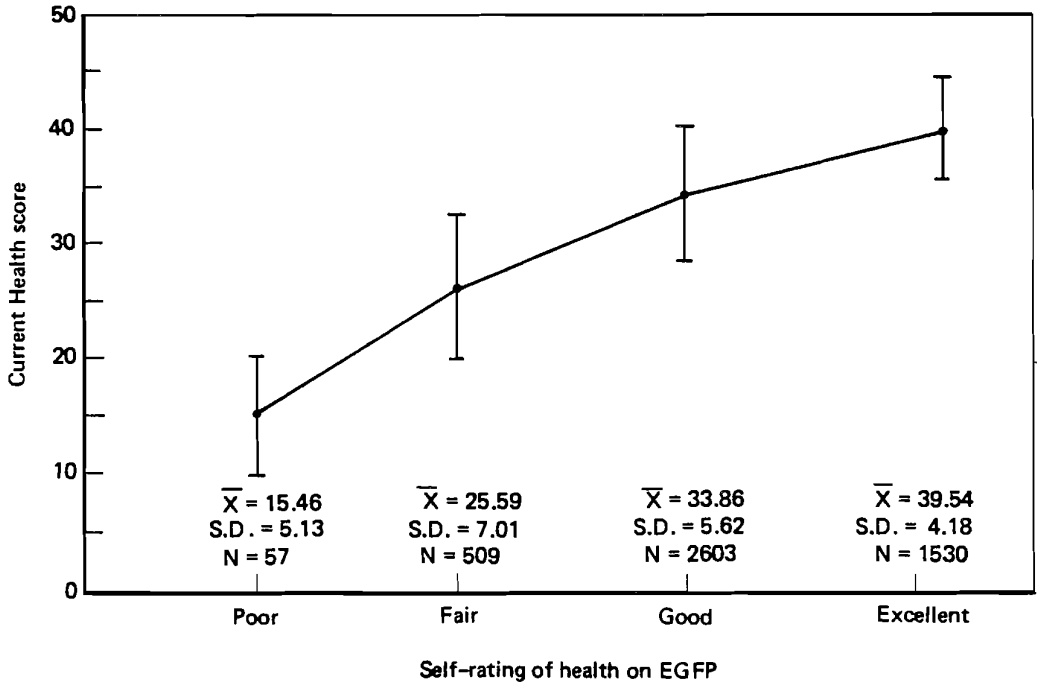

Fig. 1—Current Health mean scores (±1 standard deviation) for categories of health rated as excellent, good, fair, or poor

DESCRIPTIVE STATISTICS

Descriptive statistics for the general health perceptions scales appear in Table 11; site-specific data appear in Table D.10. Entries in the tables indicate the number of items for each

Table 11

DESCRIPTIVE STATISTICS, GENERAL HEALTH PERCEPTIONS MEASURES, ALL SITES COMBINED (N = 4717)

Measure	No. of Items	Highest Possible Score	Midpoint	Mean	S.D.
Current Health	9	45	27	34.58	7.10
EGFP	1	40	27	34.85	4.72
Prior Health	3	15	9	11.50	3.32
Health Outlook	4	20	12	14.87	2.75
Resistance	4	20	12	15.26	2.82
Health Worry/Concern	5	24	14.5	13.06	3.53
Pain	1	4	2.5	1.86	0.76
Sickness Orientation	2	10	6	6.70	2.04
General Health Ratings Index	22	110	66	83.36	13.54

measure (equal to the lowest possible score for all but EGFP), the highest possible score, the midpoint of the score range, and means and standard deviations for the measures.

As with item scores (see Table 6), scores on the multi-item general health perceptions measures were skewed and on the side of the scale midpoint indicating more favorable perceptions (better health, less pain and worry or concern). Sickness Orientation scores were closest to the scale midpoint. Variability was somewhat restricted for the Health Outlook and Resistance scales; their standard deviations were about one-eighth of their scale ranges. Standard deviations for other scales were generally one-fourth to one-sixth of their scale ranges. While means on EGFP and Current Health were very similar, the narrower range of observed EGFP scores reflects the fact that all respondents were assigned the mean of Current Health scores for their respective EGFP categories. Despite the restriction in variance that we observed for Health Outlook and Resistance, variability on all measures was sufficient for the intended correlational analyses.

Respondents in South Carolina generally had less favorable scores, suggesting they may be in poorer health and experience more health-related pain, worry, and concern than respondents in other sites. Variability, particularly on the Current Health scale, was often greater in South Carolina, indicating greater differences in health-related perceptions among respondents in that site, as well.

HIE results for the subscales approximate fairly closely those reported by Ware and Karmos (1976). HIE means were generally higher, undoubtedly because the HIE sample was generally 10 to 20 years younger on average than the samples studied by Ware and Karmos (1976, p. 96). Ware and Karmos also reported lower means for those in their most disadvantaged sample. Variability for subscale scores was similar in the HIE and Ware-Karmos samples.

We also compared HIE results for EGFP with those observed in the National Health Inter-

view Survey (National Center for Health Statistics, 1976, 1977).[5] Because of the way the national data were reported, we could compare only the percentage of respondents choosing each of the categories, rather than means and standard deviations. Data are presented, by HIE site and related census region and for all HIE sites combined and the U.S. total, in Table 12.

Table 12

COMPARISON OF RESPONSE DISTRIBUTIONS TO EGFP RATING ITEM, HIE SAMPLES AND U.S. POPULATIONS

	HIE Site/U.S. Census Region									
Response (Percent)	Dayton	North Central	Seattle	West	Massa-chusetts	North East	South Carolina	South	All Sites	Total U.S.
Excellent	27.9	50.3	35.4	54.1	39.5	48.6	25.9	44.4	32.5	48.6
Good	59.8	38.1	56.5	34.3	52.3	40.3	53.7	39.6	55.3	38.4
Fair	10.7	8.8	7.5	8.6	7.8	8.2	18.0	11.4	11.0	9.7
Poor	1.6	2.3	0.6	2.5	0.4	2.2	2.4	3.9	1.2	2.8

NOTE: Data for individual census regions from 1973 (NCHS, 1976); data for total U.S. population from 1975 (NCHS, 1977); HIE data from 1976-1977. No more recent U.S. data were available in published documents.

Our comparisons indicated that noticeably fewer HIE respondents regarded their health as poor than did respondents in the nationwide sample; approximately half as many in the combined-sites sample as in the national sample. Roughly similar proportions in both samples chose to report their health as "fair." The most notable difference we observed was that respondents in the HIE samples were less likely to choose "excellent." Respondents selecting this most favorable category represented 32.5 percent of the HIE combined-sites sample and 48.6 percent of the U.S. sample. Those selecting "good" represented 55.3 and 38.4 percent of the HIE and U.S. samples, respectively.

The most obvious initial explanation for the difference observed in the percentage selecting "poor" was the difference in age-inclusion criteria for these samples; the age difference would not be expected to explain the smaller proportions of HIE respondents selecting excellent or good. The U.S. sample is composed of noninstitutionalized persons of all ages; the HIE sample includes noninstitutionalized persons between the ages of 14 and 67 only (a very small proportion of whom are over 62). We therefore compared HIE and U.S. results for two groups selected on the basis of the age-specific groups for which national data were reported: 15 to 44 and 45 to 64 years of age.

Several explanations for the difference in the percentages choosing "excellent" and "good," none of which we could test directly, can be offered. First, there are content differences in the two items fielded by these surveys. The HIE asks for an absolute rating of health; the national survey asks for a rating compared to the health of others the person's same age. Adjusting for or considering their age, people may be more likely to view their health as excellent than when rating it in the absolute.

[5]The National Center for Health Statistics does not report data from the single-item health perceptions measures in its regular series on the National Health Interview Survey ("Vital and Health Statistics," Series 10, National Center for Health Statistics, 1963 and continuing); data on EGFP only were available in the Center's annual publication, *Health, United States* (National Center for Health Statistics, 1976, 1977).

Second, administration methods differ for the two measures: the HIE used self-administration while the Health Interview Survey used interviewer-administration. People are more likely to offer socially desirable responses to interviewers than on self-administered questionnaires (Bradburn et al., 1979) and good health may be considered a socially desirable quality. Moreover, socially desirable responding is correlated with measures of both physical health (Stewart, Ware, and Brook, 1981) and mental health (Gove and Geerken, 1977), aspects of health status that are reflected in general health measures. Thus, responses in the national sample may have been biased more by the tendency to rate health favorably.

Another difference that we noted between the two surveys was that of respondent. Although the HIE rarely permits proxy respondents on the adult health measures, the National Health Interview Survey uses adult proxies (usually the female head of household) for all children and other adult respondents. This difference may also have biased the national ratings. Studies of reporting error related to self- vs. proxy-respondent in health surveys have focused on use of services and other events that could be externally verified (e.g., Andersen et al., 1979). To our knowledge, no one has examined the relationship between respondent type and health ratings.

RELIABILITY

Multi-Item Measures

Homogeneity (average inter-item correlations) and reliability estimates for the multi-item measures in the combined sample appear in Table 13. Table D.11 contains these estimates for the site-specific analyses. For all scales in both the combined and site-specific analyses, reliability estimates exceeded the acceptable minimum of 0.50. The one exception was Sickness Orientation, which had an estimated reliability of below 0.50 in one site. Reliabilities for Current Health were all above 0.85; those for Health Outlook were generally above 0.70. Reliability estimates for Prior Health and Resistance ranged from the 0.60s to the low 0.70s, and for Health Worry/Concern and Sickness Orientation from the 0.50s to the mid 0.60s. As expected, the shorter scales were generally somewhat less reliable, yet even the very short Prior Health and Sickness Orientation scales met our reliability standard for group comparisons.

As the homogeneity coefficients indicate, items in the Health Worry/Concern scale were least interrelated. Examination of the inter-item correlation matrices (data not presented) indicated that lower homogeneity for Health Worry/Concern could be traced to lower correlations between the two Concern items than among the Health Worry items, and generally lower correlations for item no. 6 with other items, particularly the Health Concern items. Another reason for the lower homogeneity of the Health Worry/Concern items was traced to their relatively lower reliability in comparison to other HPQ items (Ware and Karmos, 1976, pp. 218-219).

Our examination of index scoring algorithms provided some evidence that the 22-item Index yields a very reliable score. In the combined-sites and site-specific analyses, estimates indicated that the Index would be more than reliable enough for group comparisons, well exceeding the 0.50 standard. In fact, estimates approached, and in Dayton exceeded, the 0.90 standard that is desirable for comparisons of individual respondents.

Table 13

HOMOGENEITY (r_{ii}) AND RELIABILITY (r_{tt}) ESTIMATES, GENERAL HEALTH PERCEPTIONS MEASURES, ALL SITES COMBINED (N = 4717)

Measure	No. of Items	r_{ii}	r_{tt}
Current Health	9	.45	.88
EGFP	1	(a)	.50[b]
Prior Health	3	.38	.65
Health Outlook	4	.40	.73
Resistance	4	.37	.70
Health Worry/Concern	5	.26	.64
Pain	1	(a)	.36[b]
Sickness Orientation	2	.36	.53
General Health Ratings Index	22	.29	.89

[a]Not available for single-item measures.

[b]For the two single-item measures, reliability estimates were based on their communalities with all other general health perceptions items, not on the Alpha coefficient (see Sec. III for details).

Single-Item Measures

Reliability estimates for the two single-item measures, EGFP and Pain, were based on estimates of their communality with all other general health items in the combined-sites analysis. As the estimates indicate, only EGFP was reliable enough to be used in group comparisons.

Reliability for Groups Differing in Educational Attainment

To evaluate reliability in the "worst case," we first compared site-specific results, because the samples differ in sociodemographic characteristics (notably education and income) that are related to data quality. These comparisons indicated that the multi-item scales were somewhat less reliable in South Carolina, the HIE site with the lowest mean educational level and annual family income. The lower reliability was most notable for the Health Outlook, Resistance, and Health Worry/Concern scales in South Carolina. Reliability estimates for the other subscales, particularly Current Health and Sickness Orientation, as well as the Index, were closer or equal to those we observed in the other sites.

Continuing our "worst case" evaluation, we computed subscale and Index reliability estimates separately for groups that differed in educational attainment. Table 14 presents these estimates for the combined-sites sample; those for the individual sites appear in Tables D.12-D.14. In the combined-sites analysis, reliability estimates showed a slight and steady increase, across all scales, with more years of schooling. The median reliability coefficient in the 0-9 years group was 0.64; that for the group with 17+ years of schooling was 0.78. Reliability for the Current Health scale and the Index remained very high regardless of educational attainment; the other multi-item scales showed somewhat lower reliabilities in the one or two least educated groups. Notably, all coefficients except two met or well exceeded the 0.50 standard in all groups examined. The exceptions were coefficients for Sickness Orientation, the shortest scale, in the groups with less than a high school education.

Table 14

RELIABILITY OF GENERAL HEALTH PERCEPTIONS MEASURES IN DIFFERENT EDUCATION GROUPS, ALL SITES COMBINED
(N = 3521)

Measure	Years of Schooling Completed					
	0-9 (N=396)	10-11 (N=486)	12 (N=1342)	13-15 (N=692)	16 (N=372)	17+ (N=233)
Current Health	.89	.89	.88	.88	.88	.84
Prior Health	.61	.64	.66	.65	.71	.73
Health Outlook	.69	.69	.73	.75	.78	.78
Resistance	.56	.72	.71	.77	.76	.78
Health Worry/Concern	.64	.58	.62	.64	.68	.68
Sickness Orientation	.46	.43	.56	.57	.51	.63
General Health Index	.89	.89	.88	.89	.90	.88

NOTE: Samples include only persons ages 17 and older for whom data on education were obtained and who provided complete data on all items in these scales.

While there were occasional exceptions to this general pattern of results in the three site-specific analyses of reliability, trends that indicated improvements in reliability as years of schooling increased were readily apparent in each site (see Tables D.12-D.14). Median reliability estimates across the multi-item scales generally increased with more years of schooling, and reliability estimates for most scales were in the ranges noted above for the majority of respondents (i.e., those who had completed high school or beyond). With only 9 exceptions among 126 coefficients, 6 of which we noted for the shortest scale, estimates indicated that reliability was more than adequate for all respondents, even in the "worst cases."

EFFECT OF QUESTIONNAIRE PLACEMENT ON RELIABILITY AND MEAN SCORES

Table 15 contains the reliability estimates we calculated for the multi-item general health perceptions scales that were placed in the middle of one version of the first annual Health Questionnaire (Form II) in Seattle, Massachusetts, and South Carolina, and at the end of another version (Form I). The two HQ versions were administered to random halves of the HIE enrollee samples in these three sites. Questionnaire placement appears to have had a negligible effect on reliability. Across all scales in the three sites studied, 8 coefficients decreased when the scales appeared last, 12 increased, and one remained unchanged; most differences were very small.

Table 15

EFFECT OF QUESTIONNAIRE PLACEMENT ON RELIABILITY OF GENERAL HEALTH PERCEPTIONS MEASURES, BY SITE[a]

	Seattle		Massachusetts		South Carolina	
	I[a]	II[a]	I	II	I	II
Measure	(N=787)	(N=784)	(N=529)	(N=502)	(N=131)	(N=139)
Current Health	.88	.89	.90	.88	.86	.85
Prior Health	.68	.67	.72	.70	.57	.64
Health Outlook	.74	.76	.74	.76	.60	.67
Resistance	.78	.78	.75	.71	.57	.52
Health Worry/Concern	.63	.62	.68	.66	.52	.54
Sickness Orientation	.63	.61	.43	.52	.57	.48
General Health Index	.89	.90	.90	.89	.89	.86

[a]In Form I, the Health Perceptions battery appeared as the last battery in the questionnaire; in Form II, it was in the middle third.

Mean scores by questionnaire placement and site appear in Table 16. The t-test values in Table 16 test the null hypothesis that questionnaire placement has no biasing effect on mean scores. The data in Table 16 indicate that in virtually all cases this hypothesis was rejected. Only in five tests was the null hypothesis not rejected. When the health perceptions battery appeared last in a lengthy questionnaire, means were significantly lower on all but two measures than when the battery appeared in the middle of the questionnaire; means on Health Worry/Concern and Sickness Orientation were generally significantly higher. Thus, later questionnaire placement tends to bias unfavorably scores on general health perceptions measures. Respondents randomly assigned to answer Form II of the HQ appeared to be in better health and to have less health-related worry and concern and less tendency to accept sickness as a part of life than their counterparts who completed Form I, in which the battery appeared last.

Table 16

T-Tests for Effect of Questionnaire Placement on Mean Scores of General Health Perceptions Measures, by Site

Measure[b]	Seattle			Massachusetts			South Carolina		
	Late[a] (N=787)	Middle[a] (N=784)	t[c]	Late (N=529)	Middle (N=502)	t	Late (N=131)	Middle (N=139)	t
Current Health	34.41 (6.36)	35.66 (6.75)	-3.79**	34.72 (6.56)	36.20 (6.43)	-3.70**	31.53 (7.28)	33.98 (6.54)	-2.92**
Prior Health	11.19 (3.17)	11.67 (3.28)	-3.00**	11.43 (3.14)	11.69 (3.34)	-1.30	10.69 (3.14)	11.55 (3.15)	-2.20**
Health Outlook	14.99 (2.50)	15.39 (2.76)	-3.08**	14.21 (2.56)	14.75 (2.80)	-3.18**	13.81 (2.54)	14.22 (2.52)	-1.28
Resistance	15.13 (2.65)	15.60 (2.76)	-3.36**	14.97 (2.62)	15.33 (2.72)	-2.57**	14.02 (2.71)	15.06 (2.42)	-3.25**
Health Worry/Concern	12.76 (3.12)	12.50 (3.2-)	-1.62	12.97 (3.37)	12.47 (3.52)	2.50*	14.96 (3.21)	14.14 (3.29)	2.05**
Sickness Orientation	6.55 (1.89)	6.34 (2.14)	2.10*	6.88 (1.67)	6.70 (1.98)	1.64	6.76 (1.95)	6.96 (1.95)	-0.83
General Health Ratings Index	82.86 (12.10)	86.55 (13.03)	-5.76**	82.31 (12.51)	85.17 (12.84)	-3.62**	76.16 (13.82)	81.49 (11.89)	-3.39**

[a] In Form I, the Health Perceptions battery appeared at the end of the questionnaire; in Form II, it was in the middle third.

[b] Standard deviations given in parentheses.

[c] t-test value for H_o: $\mu_I - \mu_{II} = 0$

*$p<0.05$, two-tailed test.
**$p<0.01$, two-tailed test.

STABILITY

Stability coefficients for the general health perceptions measures over a one-year interval in Seattle (N = 1200) are presented in Table 17. The table includes observed estimates for both the multi- and single-item measures, along with their internal-consistency estimates (from Table D.11). The magnitudes of the observed coefficients indicates that general health perceptions are generally stable over time. The coefficients can be interpreted as the proportion of true score variance that remained unchanged during the year.[6] Stability coefficients

Table 17

RELIABILITY AND STABILITY COEFFICIENTS, GENERAL HEALTH PERCEPTIONS SCALES
(Seattle, One-Year Interval, N = 1200)

Scale	Enrollment Reliability	Stability Estimates	
		Observed[a]	Adjusted[b]
Current Health	.87	.58	.66
EGFP	.50	.57	~1.00
Prior Health	.64	.67	~1.00
Health Outlook	.73	.59	.80
Resistance	.77	.65	.89
Health Worry/Concern	.62	.50	.79
Pain	.36	.44	~1.00
Sickness Orientation	.56	.55	.96
General Health Ratings Index	.88	.67	.75

[a]From Table D.11; estimates for EGFP and Pain based on combined-sites data from Table 13.

[b]Adjusted estimate equals observed stability estimate divided by square root of product of internal-consistency estimates at enrollment and first anniversary of enrollment (enrollment MHQ and first annual HQ data).

for Prior Health and Sickness Orientation were approximately equal to their internal-consistency estimates, suggesting that these perceptions change least over time of those measured. Comparison of internal-consistency and stability estimates for Current Health showed it to be the least stable over time, as might be expected if the measure reflects perceptions about health status at the time of questionnaire administration.

To compare the stability of the constructs defined by these measures, we had to take differences in measurement reliability into account. We therefore computed adjusted stability esti-

[6]Changes over time that affect all respondents the same (e.g., the effect of one year of age) would not affect stability as defined here.

mates for the measures by dividing the observed coefficient by the square root of the product of the two point-in-time reliability estimates (internal-consistency estimates at enrollment and HQ#1 for the multi-item measures and the communality estimates for the two single-item measures).

The adjusted coefficients illustrate the stability findings even more dramatically.[7] Current Health is clearly the least stable of the constructs, despite being the most reliable measure, indicating its scores are more sensitive to changes than are those of the others. The adjusted estimates also indicate that EGFP scores are less sensitive to change than are Current Health scores. Somewhat more stable than Current Health was the Index (half the items in which are from Current Health), Health Worry/Concern, and Health Outlook. By contrast, the adjusted stability estimates for Sickness Orientation and Prior Health are near or equal to 1.00, suggesting they are more nearly stable traits than states that may change over time. For Prior Health, this finding is not inconsistent with its interpretation as a measure of perceived health, given the focus of its items on the (presumably unchanging) past. For Sickness Orientation, this finding indicates that, if the measure is valid, the tendency to accept illness as a part of life is more of a personal trait than a state that shows much change over a one-year interval.

SOCIODEMOGRAPHIC CORRELATES

Correlations between the general health perceptions measures and four sociodemographic variables (age, sex, race, education, and income) in the combined-sites sample appear in Table 18. Site-specific data are presented in Tables D.15-D.18. Because we considered age to be a validity variable for health status measures, its relationship to the HIE general health perceptions measures is also discussed in Sec. V.

No significant differences were observed between men and women or between whites and nonwhites in perceptions of prior health or the tendency to view sickness as a part of life. The other favorably scored health perceptions measures were significantly and negatively related to both sex (female) and race, indicating that women and nonwhites had poorer perceptions of their health at the present (Current Health and EGFP) and in the future, of resistance to illness, and overall (Index). Health Worry/Concern and Pain, both unfavorably scored, were significantly and positively correlated with sex and race; men and nonwhites had lower scores on these measures than did women and whites. Older persons had less favorable perceptions of current, past, and future health, had greater pain, but were less apt to report health-related worry, sickness orientation, and susceptibility to illness. The site-specific results were very similar with respect to direction and significance of relationships.

Educational attainment was significantly related to all health perceptions measures, positively for those that were favorably scored and negatively for the unfavorably scored measures. Those who had completed more years of schooling had better perceptions of their health at present, in the past, and in the future, viewed themselves as more resistant to illness, exhibited less health-related worry, concern, and pain, and were less likely to consider illness as a part of life. Although the correlations were lower in magnitude, the general health perceptions measures related to income in the same way as they related to education. Again, results across the sites were very similar.

Despite the fact that it is less reliable, EGFP tended to have stronger associations with these measures than did Current Health; both are considered measures of health at the present. This finding raises the possibility that these measures may have different interpre-

[7]To the extent that the communality estimates *under*estimate the point-in-time reliability of the single-item measures, the adjusted coefficients *over*estimate their stability.

Table 18

CORRELATIONS BETWEEN CONSTRUCT-SPECIFIC GENERAL HEALTH PERCEPTIONS MEASURES AND SOCIODEMOGRAPHIC VARIABLES, ALL SITES COMBINED

Measure	Sex	Race	Education	Income
Current Health	-.08	-.09	.18	.10
EGFP	-.12	-.16	.30	.14
Prior Health	-.04	-.03	.12	.06
Health Outlook	-.03	-.10	.17	.05
Resistance	-.13	-.06	.11	.08
Health Worry/Concern	.02	.16	-.14	-.13
Pain	.17	.06	-.13	-.08
Sickness Orientation	.04	.01	-.06	-.08
General Health Ratings Index	-.09	-.11	.20	.11

NOTE: All correlations >.04 significant at p<.001; sex scored 1 = male, 2 = female; race scored 1 = white, 2 = other.

tations, an issue considered further in Sec. V. Relatively lower reliability for the Sickness Orientation scale than the other health perceptions measures undoubtedly reduced the strength of the relationships we observed here.

We explored the curvilinearity of the relationships between general health perceptions and education, income, and age in the combined-sites sample. The five education groups were defined as in the subgroup reliability studies (see above). We defined six income groups that divided the sample into roughly equal subgroups (≤$7,000; $7,001-$10,500; $10,501-$14,500; $14,501-$18,500; $18,501+). Six age categories were also defined to achieve roughly equal groups (14-20, 21-25, 26-30, 31-35, 36-45, and 46+ years). Using analysis of variance, we computed mean scores on each of the general health perceptions measures separately at the mean of each education, income, and age group, and tested for deviations from linearity in group means (using the breakdown procedure in Nie et al., 1975).

Results for these analyses appear in Tables D.19-D.21. Most measures had linear relationships with education. Tests of nonlinearity were significant only for EGFP and Prior Health; in both cases, those in the lowest education groups had notably lower scores. We observed significant departures from linearity in the relationship with income for all measures but Prior Health, Health Outlook, and Sickness Orientation. On Current Health, EGFP, Resistance, Health Worry/Concern, Pain, and the Index, those in the lowest income group generally accounted for the deviations, with scores indicating poorer health perceptions and more health-related worry and pain. We also noted significant nonlinearity in relationships with age for most measures; Resistance and Sickness Orientation showed linear relationships with the age variable. The most notable deviations were for the oldest age group (46+ years) when nonlinearity was significant. Thus, Sickness Orientation and Health Outlook had linear relationships with education, income, and age; Resistance was linear with education and age; and

most general health perceptions had linear relationships with education. When nonlinearity was apparent, those with the least income or in the highest age groups had notably poorer perceptions of their health.

For all measures, comparisons between r^2 and η^2 (Thorndike, 1978) indicated that product-moment correlations represented the relationships between general health perceptions and income and age quite well except at the extreme, usually the group with the least income or oldest age (see Table D.22).[8]

---

[8] η, or eta, makes no assumptions about linearity or level of measurement for the "independent" variable. η^2 is the ratio of the between-groups sum of squares to the total sum of squares, and is equal to r^2 when the assumption of linearity is met.

V. RESULTS: MEANING OF GENERAL HEALTH PERCEPTIONS SCORES

This section presents the results from our studies of the validity of the HIE general health perceptions measures. It documents findings from studies of relationships among the measures and a range of other health and health-related variables that provide information about the interpretation of differences in general health perceptions scores.

RELATIONSHIPS AMONG THE GENERAL HEALTH MEASURES

Before testing our validity hypotheses about the relationships between the general health measures and other variables, we computed correlations among all the general health measures; the inter-scale correlation matrix appears in Table 19. Correlations between the subscales and the Index that are marked with asterisks have been corrected for overlap (recall that Sickness Orientation, EGFP, Pain, and the two Concern items are not included in the Index). Internal-consistency reliability coefficients appear in the matrix diagonal.

Table 19

CORRELATIONS BETWEEN GENERAL HEALTH PERCEPTIONS MEASURES, ALL SITES COMBINED[a]

Measure	CH	EGFP	PH	HO	RE	WC	PN	SO	IN
Current Health (CH)	(.88)								
EGFP	.65	(.71)							
Prior Health (PH)	.40	.34	(.65)						
Health Outlook (HO)	.58	.43	.26	(.73)					
Resistance (RE)	.47	.34	.29	.46	(.70)				
Health Worry/Concern (WC)	-.45	-.34	-.27	-.34	-.36	(.64)			
Pain (PN)	-.47	-.42	-.24	-.25	-.24	.30	(.60)		
Sickness Orientation (SO)	-.16	-.10	-.10	-.20	-.26	.18	.14	(.53)	
General Health Ratings Index (IN)	.66*	.63	.41*	.60*	.53*	.65*	-.44	-.23	(.89)

[a]Diagonal entries are reliability estimates.
*Indicates correlation has been adjusted for overlap; one Worry item (no. 6) and the two Concern items are not included in the Index.

Of the general health constructs, Current Health and Health Outlook correlated highest with the other general health subscales. Sickness Orientation, on the other hand, accounted for the lowest inter-scale correlations; this remained true even after adjusting for reliability, despite the fact that Sickness Orientation was the least reliable of the general health measures. The other subscales generally had low to moderate correlations with each other. The centrality of the current health construct is again apparent in the high scale-Index correlations, which are similar in magnitude to correlations between the subscales and Current Health (see the first column and last row of Table 19).

RELATIONSHIPS BETWEEN GENERAL HEALTH AND THE VALIDITY VARIABLES

The hypotheses that we evaluated in the bivariate validity studies were presented in Sec. III. To reiterate, we expected valid measures of general health perceptions to be significantly related to both mental and physical health, to be generally unrelated to social circumstances and patient role propensity (except for Worry/Concern and Resistance with patient role propensity), and to be significantly related to use of services, life stress, and age. We also expected favorably defined general health perceptions to be positively correlated with favorable definitions of health and social circumstances, and negatively correlated with use of services and stressful life events; we expected the reverse pattern for the unfavorably defined general health perceptions. We expected Current Health, Prior Health, Resistance, and Health Outlook to be negatively correlated with age. Although they are negatively defined, we expected Worry/Concern and Sickness Orientation to be positively correlated with age.

Correlations between each of the HIE general health perceptions measures and 30 validity variables appear in Table 20. The validity variables are classified in six groups: physical health, mental health, social circumstances, patient role propensity, use of services, and other (age and life stress). Definitions and reliability estimates for the validity variables appeared in Table 5. Correlations greater than 0.05 have chance probabilities $p < 0.05$; those equal to or greater than 0.07, $p < 0.01$; and those equal to or greater than 0.09, $p < 0.001$, one-tailed tests (not corrected for multiple tests).

Direction of Association

With only one exception, we observed the hypothesized direction of association for all general health perceptions measures with the validity variables that defined physical and mental health, use of services, and life events. The exception was a negative correlation between the unfavorably defined Sickness Orientation scale and the unfavorably defined chronic disease measure based on screening examination data. Many inconsistencies were noticed in direction of relationships between the health perceptions measures and those defining social circumstances, all of which were favorably scored. The patient role propensity measures, which are negatively intercorrelated, also showed mixed directions of association with the general health perceptions measures. The directions of associations with age were as we had hypothesized.

The number of coefficients in Table 20 makes it difficult to discern patterns of relationships. We therefore summarized the major part of the matrix by computing median validity coefficients for each of the general health perceptions measures across the physical and mental health and social circumstances variables; this summary appears in Table 21. Because correlations between the general health measures and measures in both the physical and mental health categories were always in the hypothesized direction, our calculation of median coefficients could ignore directionality; the calculation therefore assumed that all physical and mental health validity variables were favorably defined. Although all social variables were favorably defined, there were inconsistencies in the direction of correlations; our calculation of the median correlation for social circumstances therefore took directionality into account. In addition, the overall mental health and social indexes were not considered when medians were computed, because they contain the other variables in their respective categories.

Table 20

Correlations Between General Health Perceptions Measures and Validity Variables, Seattle (N = 1200)[a]

Validity Variables[b]	General Health Perceptions Measures								
	Current Health	EGFP	Prior Health	Health Outlook	Resistance	Worry/ Concern	Pain	Sickness Orientation	Index
PHYSICAL HEALTH									
Chronic Personal Limitations	-.42	-.37	-.22	-.23	-.14	.23	.29	.02	-.38
Current Personal Limitations	-.40	-.34	-.21	-.22	-.14	.22	.31	.03	-.36
Chronic Role Limitations	-.40	-.35	-.24	-.23	-.15	.22	.31	.03	-.38
Current Role Limitations	-.40	-.34	-.24	-.22	-.17	.23	.36	.05	-.38
Physical Capacities	-.40	-.38	-.21	-.22	-.14	.15	.28	.00	-.36
Strenuous Exercise	.14	.11	.00	.11	.04	.01	-.07	.03	.10
General Exercise	.22	.14	.04	.17	.12	-.05	-.08	-.01	.20
Acute Symptoms	-.50	-.38	-.30	-.26	-.33	.29	.41	.12	-.51
Chronic Disease (self-report)	-.44	-.41	-.32	-.26	-.22	.24	.34	.08	-.45
Chronic Disease (screening)	-.39	-.41	-.24	-.26	-.19	.16	.20	-.06	-.38
MENTAL HEALTH									
Anxiety	-.40	-.27	-.19	-.18	-.26	.28	.30	.16	-.39
Depression	-.39	-.25	-.15	-.20	-.27	.26	.26	.10	-.38
Positive Well-Being	.41	.24	.06	.27	.23	-.27	-.17	-.10	.38
Emotional Ties	.24	.14	.08	.13	.12	-.16	-.11	-.01	.23
Emotional Stability	.39	.28	.18	.19	.24	-.28	-.29	-.16	.38
Mental Health Index	.45	.28	.15	.25	.29	-.32	-.27	-.13	.44
SOCIAL CIRCUMSTANCES									
Close Friends/Relatives	.08	.04	.02	.05	.02	-.04	.00	-.03	.07
Neighborhood Acquaintances	.08	.05	-.02	.05	.05	-.02	-.04	.00	.07
Telephone Contacts	.08	.03	.00	.08	-.02	-.03	.04	.00	.06
Getting Along	.13	.10	-.03	.12	-.02	-.03	-.07	.03	.09
Attendance at Religious Services	-.02	.00	-.08	.02	.01	-.04	.04	-.01	-.02
Group Participation	.03	.03	-.05	.08	.06	-.05	.02	.00	.04
Social Contacts	.11	.08	.04	.09	.01	-.02	.00	.00	.10
Social Well-Being Index	.11	.08	-.01	.12	.04	-.05	.02	.00	.10
PATIENT ROLE PROPENSITY									
Attitude Toward Going to the Doctor	.04	.00	.02	-.02	-.09	.10	-.02	.04	-.01
Rejection of Sick Role	.09	.00	-.01	.17	.22	-.18	.02	-.08	.15
USE OF SERVICES[c]									
Prior Treatment for Acute Symptoms	-.16	-.14	-.10	-.08	-.13	.09	—	.03	-.17
Subsequent Use of Outpatient Services	-.05	-.05	-.08	-.02	-.02	.07	—	.01	-.06
OTHER									
Age	-.09	-.18	-.08	-.12	.16	-.10	.05	-.19	-.04
Life Events	-.15	-.06	-.10	.00	-.10	.11	.08	.07	-.15

[a] Sample includes persons who had no missing scale scores on the entire set of variables.

[b] Definitions and reliability estimates for the validity variables appear on Table 4.

[c] For these two variables only, the sample included respondents aged 18 and older who were in the HIE for the entire first year of the study in Seattle and Massachusetts and provided complete data on the enrollment MHQ and/or the first annual HQ (N=1557).

Table 21

MEDIAN VALIDITY COEFFICIENTS, BY CATEGORY, FOR GENERAL HEALTH PERCEPTIONS SCALES

Validity Category		Scales								
		Current Health (+)	EGFP (+)	Prior Health (+)	Health Outlook (+)	Resistance (+)	Worry/ Concern (-)	Pain (-)	Sickness Orientation (-)	Index (+)
Physical Health	(+)	.40	.36	.23	.22	.14	-.22	-.31	-.03	.38
Mental Health	(+)	.40	.26	.12	.18	.24	-.26	-.22	-.10	.38
Social Circumstances	(+)	.08	.04	-.02	.07	.01	.03	.00	.00	.07

NOTE: Median assumes that all validity variables in each category represent favorable definitions of health; thus, signs were reversed for the limitations, symptoms, disease, anxiety, and depression measures before computing median. Calculation of medians excluded Mental Health Index and Social Well-Being Index, which contain the other measures in those categories. "+" and "-" indicate direction of scoring; where necessary, variables in the validity categories were reversed so the scores on all variables would indicate better health and social circumstances.

Associations with Physical Health

With one exception, the measures were substantially correlated with most physical health variables. Sickness Orientation was generally unrelated to most measures of physical health (other than occurrence of acute symptoms during the past 30 days). As the median coefficients in Table 21 indicate, Current Health, EGFP, and the Index (which contains Current Health) were more closely associated with the physical health measures than were the other general health scales.

The measures were correlated both with acute and chronic physical health limitations, and with physical ability as well as physical limitations. For most measures, correlations with acute symptoms and self-reported chronic diseases were the largest we observed in the matrix; this finding was notable for Resistance, which correlated considerably higher with Acute Symptoms than with other physical health measures. With one exception, the measures correlated somewhat higher with the weighted count of chronic diseases scored from self-reports than with the one scored chiefly with data from the screening examination. (The self-report score was based primarily on yes/no answers to questions about whether a doctor had ever told the respondent he had a particular disease or problem.) The exception was Health Outlook, which correlated equally with the two operational definitions of chronic diseases.

Associations with Mental Health

As we hypothesized, the general health perceptions measures were generally significantly related to most measures of mental health. Current Health and the Index generally had the highest correlations with mental health variables, and their median correlations with mental health variables were equal (in absolute magnitude at the second decimal point) to their correlations with the physical health variables (see Table 21).

Although Current Health and EGFP had similar correlations with the physical health measures, EGFP was notably less related to mental health than physical health. Pain and Health Outlook were also roughly equally correlated with physical and mental health (on average), although less so than Current Health or the Index. Although Sickness Orientation

was least related to mental health of the general health perceptions measures, its highest health-related correlations were seen with mental health validity variables.

Of the mental health constructs studied, all but Emotional Ties correlated roughly equally with Current Health, EGFP, Resistance, and Health Worry/Concern. Prior Health and Pain were less associated with Positive Well-Being than the others, and Health Outlook was more closely associated with Positive Well-Being.

Associations with Social Circumstances

We observed very few significant relationships between the general health perceptions measures and variables that reflected social circumstances, as we had hypothesized. Across all social circumstances measures, median correlations were significant, although low, only for Current Health, Health Outlook, and the Index (see Table 21). The significant correlations in this category were generally with Getting Along, a subjective assessment of social relationships, and with the Social Well-Being Index, which contains all the other variables in this category but Telephone Contacts.

Associations with Patient Role Propensity

All measures but Worry/Concern and Resistance were unrelated to Attitude Toward Going to the Doctor, a measure of patient role propensity, as we had hypothesized. Resistance and Worry/Concern, as well as Current Health (but not EGFP), Health Outlook, and the Index were significantly related to Rejection of Sick Role.

Associations with Use of Services

The general health perceptions measures were more closely related to use of services during the past month than to use of outpatient services during the year after the measures were administered, as we had hypothesized. Again, Sickness Orientation was an exception, because it had insignificant relationships with both measures of use. Most measures were not significantly related to the measure of subsequent use of outpatient medical services (mental health and dental care services were not included in the latter).

Associations with Age

The favorably defined general health perceptions, with the hypothesized exception of Resistance to Illness, were negatively related to age. Older persons had more favorable perceptions of their resistance than did younger persons, but less favorable perceptions of their health at the present, in the past, or in the future. We noted that EGFP was considerably more closely related to age than was Current Health; scores on the Index were generally unrelated to age in these bivariate analyses.

As hypothesized based on results reported by Ware and Karmos (1976), the unfavorably scored Health Worry/Concern scale was negatively related to age; thus, older persons exhibited less health-related worry and concern than younger persons. Similarly, the Sickness Orientation scale was positively related to age, also as hypothesized on the basis of Ware and Karmos' results. Older persons indicated less tendency to accept illness as part of their lives than younger persons.

Associations with Life Stress

All the general health perceptions measures were significantly related to stressful life events except EGFP and Health Outlook. Although we had no specific validity hypotheses for the measures in regard to this variable, we theorized that a relationship might arise given the known negative relationship between mental health and stressful life events (Williams, Ware, and Donald, in press). Supporting this line of reasoning, the two general health perceptions measures that were most related to mental health measures (Current Health and the Index) were also more closely related to stressful life events than were the other scales.

Summary

To summarize, the pattern of relationships between the general health perceptions measures and validity variables generally confirmed the validity hypotheses we had for the measures. Among the general health perceptions measures, Current Health and the Index generally had the most substantial relationships where significant relationships had been hypothesized.

Such a finding for Current Health and the General Health Ratings Index might have occurred simply because they are more reliable than the other measures. We tested this explanation for our findings by adjusting the correlations in the matrix for imperfect reliability (dividing the correlations by the square root of the reliability coefficient for the relevant general health perceptions measure). Our finding remained unchanged even after these adjustments were made; Current Health and the Index continued to show stronger relationships with most validity variables (with the exception of age) than did the other measures. Thus, the Current Health scale and the General Health Ratings Index appear to contain the most information about health status as it is defined by the validity variables we studied.

SPECIAL STUDIES OF THE INDEX

As mentioned in the methods section, our other bivariate validity studies focused exclusively on the meaning of Index scores; in particular, on the relationship of Index scores to clinically relevant indicators that defined substantial health impairment. To reiterate briefly, we divided the sample into deciles (roughly 150 persons in each) based on their Index scores and plotted the score on each dichotomously scored indicator of health impairment for each decile. These indicators included: probability of having any serious chronic disease; probability of having a serious emotional problem (as defined by the bottom 20th percentile on the Mental Health Index); probability of being limited in physical functioning or capacity; probability of being emotionally *or* physically impaired; and probability of being emotionally *or* physically impaired *or* having any serious chronic disease. These probabilities were plotted at the midpoint of each decile's range from the Index score distribution. We inspected these plots and tested for nonlinearity in the relationship to better understand the meaning of scores on the General Health Ratings Index. The plots and relevant test statistics appear in Figs. 2 through 6; means, standard deviations, and samples sizes for each decile on each of these indicators are presented in Table D.23.

As Fig. 2 indicates, the probability of having a serious chronic disease drops by about half over the Index score range, from 0.59 for persons in the lowest decile of the Index score distribution to 0.29 for those in the highest decile. People with an Index score of about 77 or below have near or greater than average probability of any serious chronic disease (0.39). The test for nonlinearity was not significant. Thus, the probability of having a chronic disease increases linearly with Index scores.

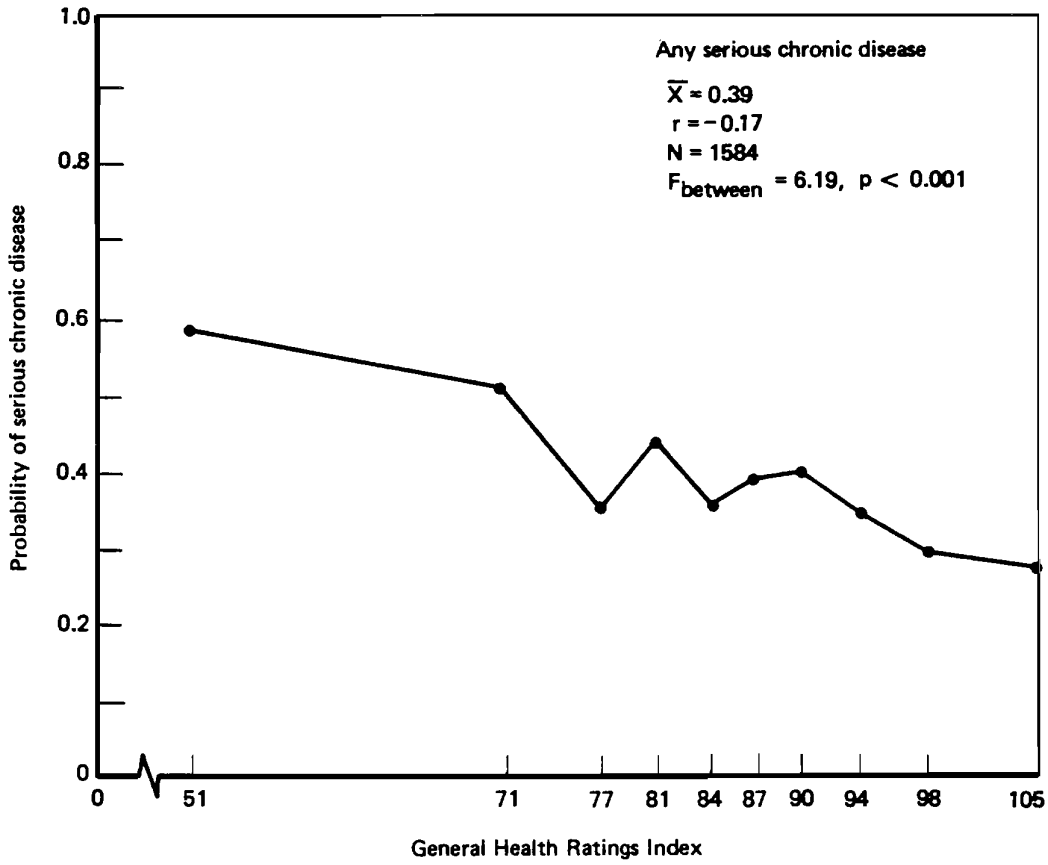

Fig. 2—Probability of any serious chronic disease at mean of each decile of General Health Ratings Index score distribution

Figures 3 and 4 reveal that the probability of having any impairment in physical functioning or of being emotionally impaired decreases markedly throughout the Index score range. The probability of impairment in personal functioning is roughly nine times greater (0.65 versus 0.07) for those in the lowest as opposed to the highest decile of Index scores (see Fig. 3). The test for nonlinearity was not significant at conventional levels ($p < 0.08$); only 7 percent of the variance was related to nonlinearity, $\eta^2 - r^2/r^2 = .07$. Thus, we do not reject the hypothesis that the probability of functional impairment decreases quite steadily throughout the Index score range. Persons with Index scores of about 80 or below had the same or a higher probability of functional impairment than did the sample on average.

Similarly, a tenfold decrease (0.61 to 0.06) in the probability of being emotionally impaired was observed from the lowest to highest deciles (see Fig. 4). The line clearly begins to flatten in the upper deciles, and the test for nonlinearity was significant ($p < 0.002$); about 16 percent of the variance was related to nonlinearity. Thus, changes in General Health Ratings Index scores between the fourth and tenth decile are only slightly related to changes in the probability of emotional impairment (0.13 to 0.06 from fourth to tenth deciles). Differences in this probability are much more dramatic between the fourth and the first deciles (0.13 versus

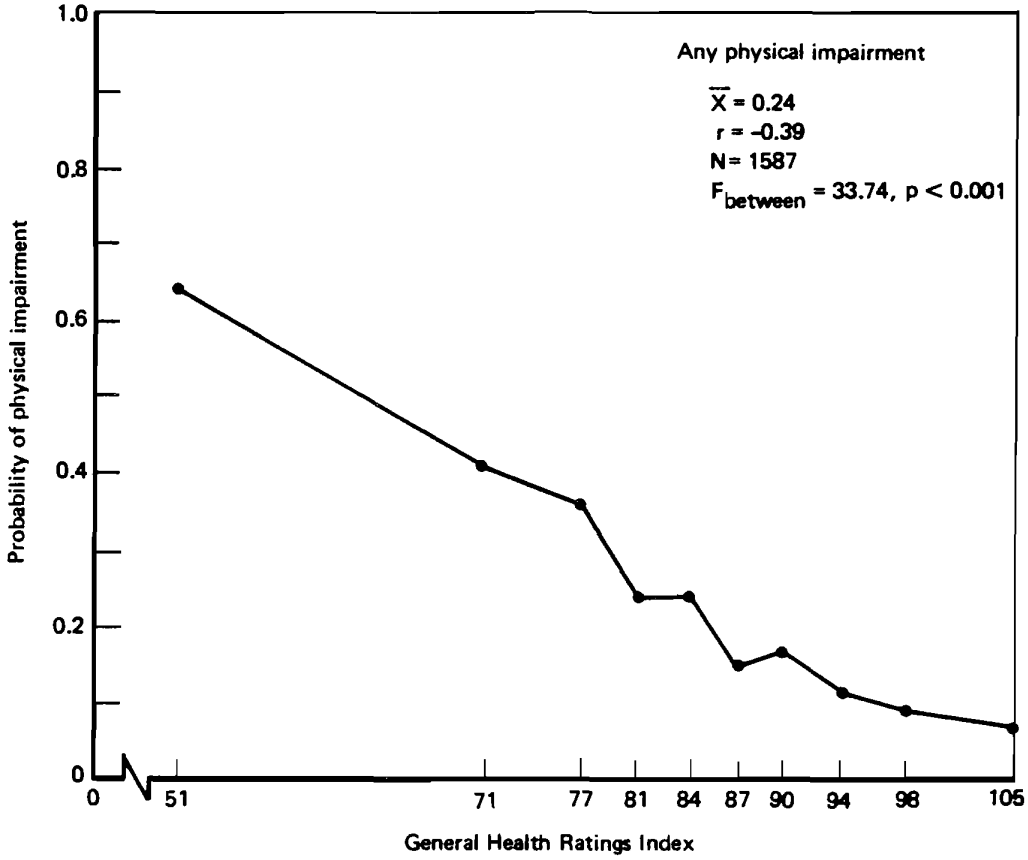

Fig. 3—Probability of serious physical impairment at mean of each decile of General Health Ratings Index score distribution

0.61, respectively). The average probability of emotional impairment was 0.19; thus, persons with a median Index score of roughly 77 or below had a better than average chance of being emotionally impaired. From the data in Figs. 3 and 4, therefore, we can conclude that general health ratings are more strongly related to physical and emotional impairment than to chronic disease (as defined here).

As Fig. 5 illustrates, the probability of emotional and/or functional impairment also decreases dramatically throughout the Index score range, from 0.87 to 0.12 between the first and tenth deciles of the Index score. As was the relationship with emotional impairment (but not impairments in personal functioning), the decrease was significantly nonlinear; the most sizeable changes were observed from the first to the fourth deciles (a change in probability from 0.87 to 0.34).

Finally, Fig. 6 illustrates the relationship between the probability of any serious health problem (problems in personal functioning, emotional impairment, and/or any serious chronic disease) and Index scores. The probability of any serious health limitation showed a nonlinear (and nonmonotonic) decrease with increases in Index scores. The most sizeable shift occurred between the first and second deciles (0.92 to 0.76); we also noted a slight *increase*

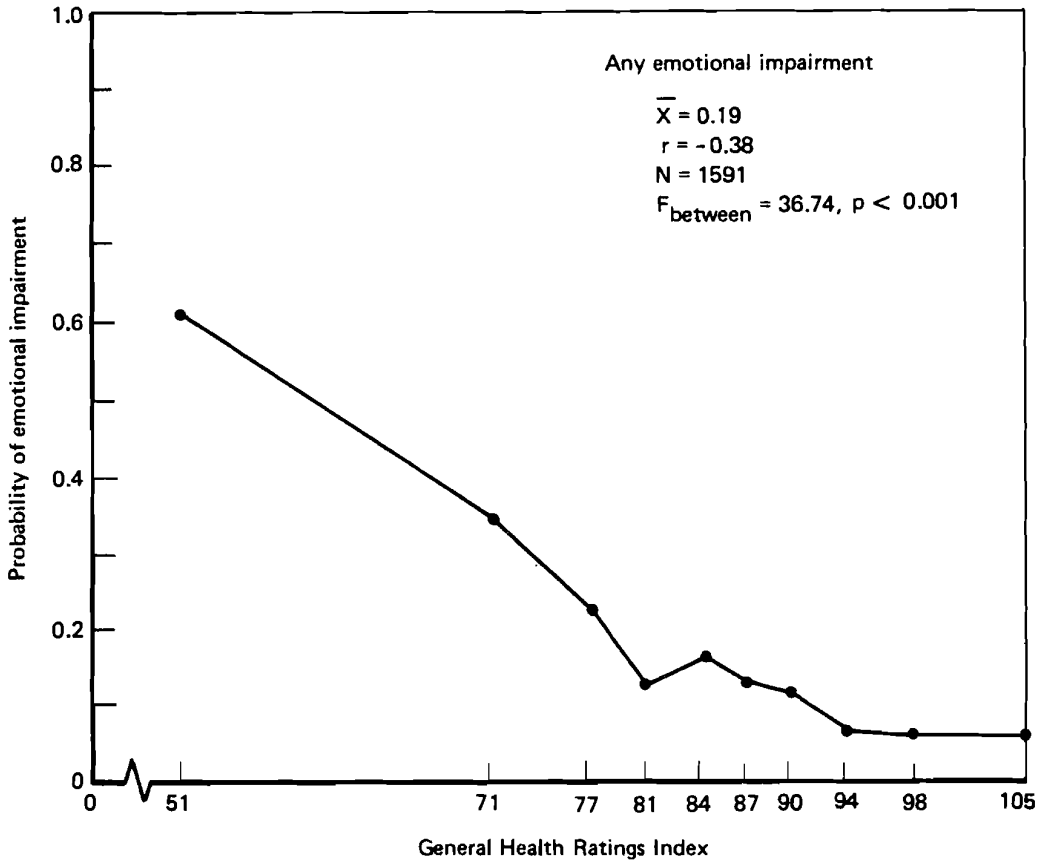

Fig. 4—Probability of emotional impairment at mean of each decile of General Health Ratings Index score distribution

between the sixth and seventh deciles (0.49 to 0.56), counter to the trend of decreasing probabilities as Index scores rose.

To summarize, scores on the General Health Ratings Index throughout its range reflect differences in physical and emotional impairment. The relationship with physical limitations is roughly linear, while that with emotional impairment is notable nonlinear. Index scores do not appear to be as sensitive to presence of the chronic diseases included in our definition of Any Chronic Diseases as to physical and emotional impairments. The latter finding may relate to three factors. First, presence of some of the chronic diseases was scored from the screening examination rather than from the self-report; chronic diseases that were unknown to respondents and had no perceived symptoms could not have been taken into account in the general health ratings. Second, our definition of chronic diseases was conservative and understated somewhat their prevalence in HIE samples, which could account for the low correlation observed. Third, the physical and emotional limitations variables may capture much of the direct effect of chronic disease on general health; alternatively, the presence of chronic diseases to which respondents may have become accustomed may in fact receive less weight in a general health rating than physical and emotional impairments. The latter explanations require further analysis in multivariate models of health status.

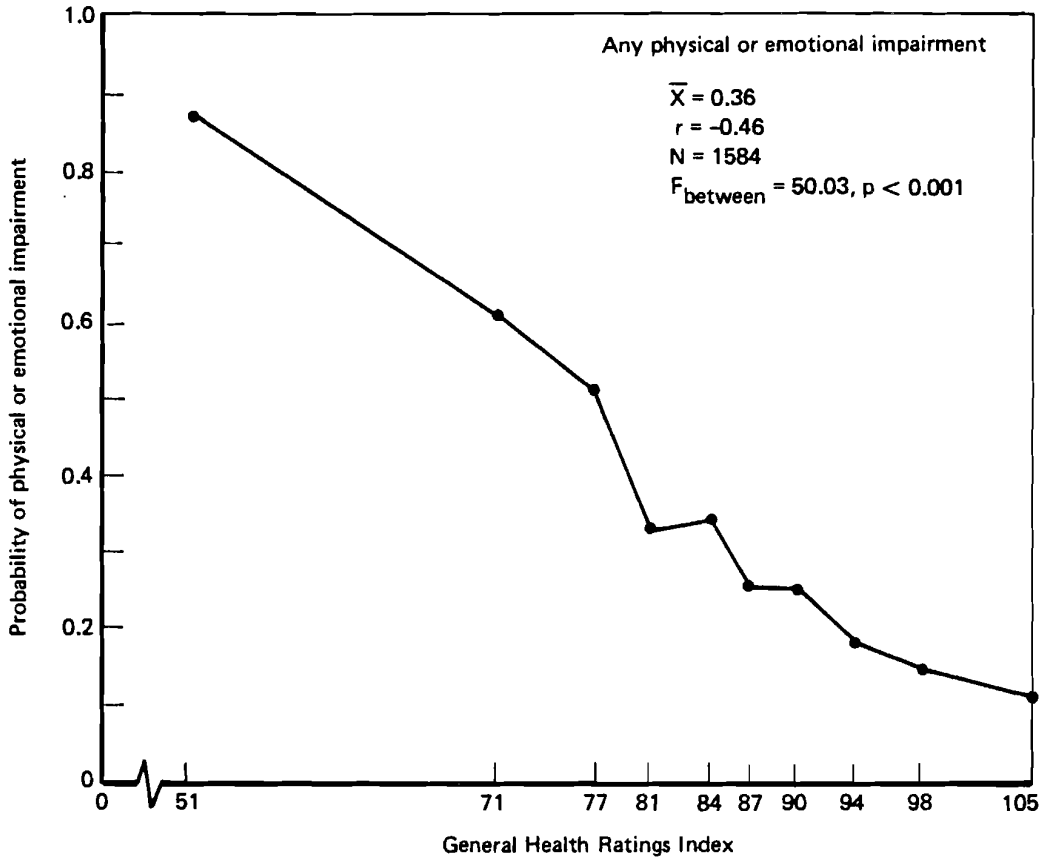

Fig. 5—Probability of serious physical and/or emotional impairment at mean of each decile of General Health Ratings Index score distribution

These findings also suggest that scores of about 80 or less on the General Health Ratings Index are the ones in general, non-aged populations at or below which respondents have an equal or higher than average probability of some *substantial* health impairment. This should not be interpreted as a "critical value"; the linearity of the relationship between Index scores and both chronic diseases and limitations in physical functioning indicate that similar shifts in probabilities are seen with a given increase in Index scores at all points on that score continuum. Given the observed nonlinearity of the Index in relation to probability of emotional limitations, an Index score of 77 to 80 probably does suggest a "critical value"; decreases below that score are associated with much larger increases in the probability of emotional impairment than anywhere else in the Index score range.

ADDITIONAL STUDIES OF CURRENT HEALTH AND EGFP

Given the importance of the Current Health scale as revealed by our scaling and validity studies, and the widespread use of the EGFP item as a general health perceptions measure,

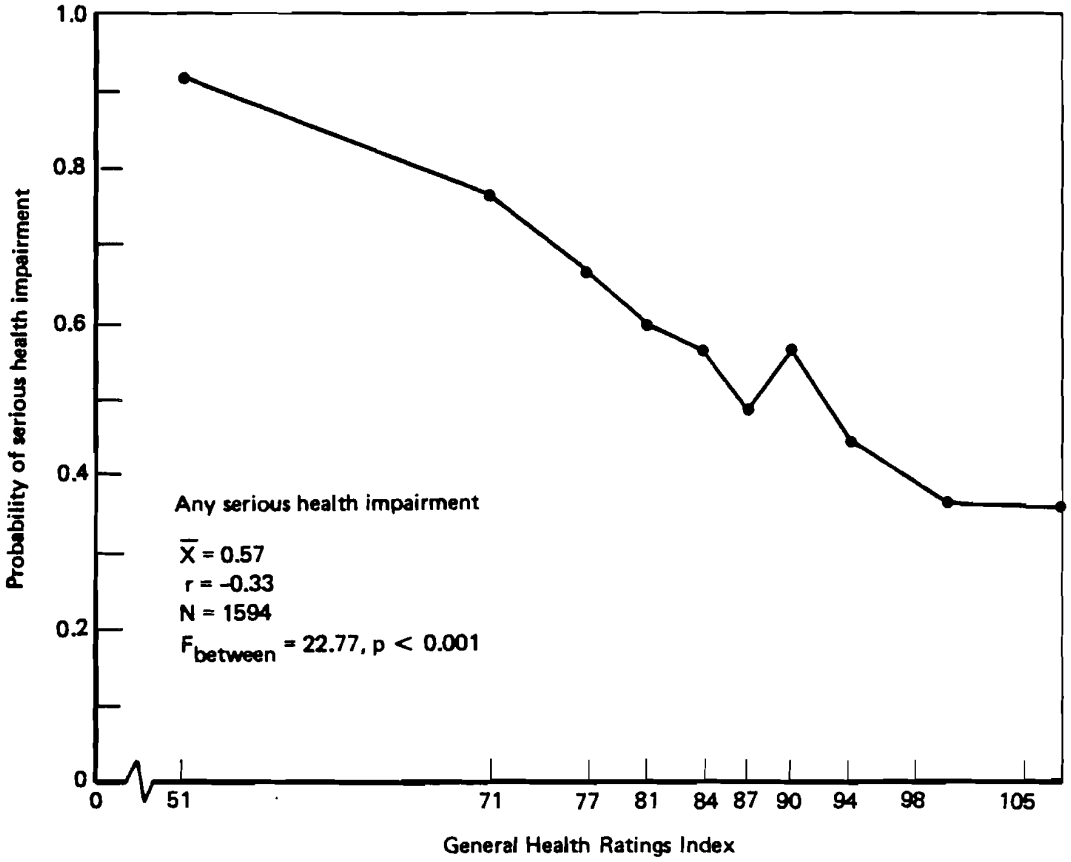

Fig. 6—Probability of any serious health impairment at mean of each decile of General Health Ratings Index score distribution

we decided to examine the validity of the single items in these measures against the physical and mental health validity variables and age. The lowest, median, and highest coefficients between each of the 10 items involved (nine from Current Health and the single-item EGFP measure) and variables in the physical and mental health categories are presented in Table 22, along with correlations between the items and age. The validity variables used in calculations of the median are defined in a table footnote; these calculations assumed that all physical and mental health variables included were positively scored (i.e., indicated better health). The direction of item scoring reflected the favorable or unfavorable perception defined by item content and is indicated by the abbreviated item wording in Table 22. The matrix of correlations between these items and each of the physical and mental health validity variables and age appears in Table D.24.

Several conclusions can be drawn from the data presented in Tables 22 and D.24. First, the favorably worded items were positively correlated with favorable definitions of health and the unfavorably worded items were negatively correlated with such definitions, as would be hypothesized. Second, six of the Current Health items and EGFP correlated somewhat or

Table 22

Lowest, Median, and Highest Correlations for Current Health Items and EGFP with Physical and Mental Health and Age

		Validity Category						
		Physical			Mental			
	Item and Direction of Scoring	L	M[a]	H	L	M[b]	H	Age
A	Doctors say health is now excellent (+)	-.07	-.30	-.33	-.15	-.23	-.26	-.07
D	Feel better now than ever (+)	.10	.24	.31	.16	.28	.39	-.08
I	Somewhat ill (-)	-.09	-.31	-.39	-.15	-.29	-.31	.05
L	Not as healthy now as used to be (-)	-.11	-.29	-.34	-.12	-.25	-.27	.17
Q	Healthy as anybody I know (+)	.09	.28	.32	.18	.25	.26	-.08
V	Health is excellent (+)	.08	.34	.43	.18	.30	.32	-.08
Z	Been feeling bad lately (-)	-.05	-.29	-.46	-.19	-.38	-.40	-.06
DD	Doctors say now in poor health (-)	-.03	-.26	-.30	-.15	-.22	-.23	-.01
FF	Feel as good now as ever (+)	.12	.29	.39	.19	.33	.35	-.02
EGFP	Health rated excellent, good, fair or poor (+)	.08	.38	.42	.14	.27	.28	-.18

[a]Median of correlations with the following validity variables: Chronic Personal Limitations, Current Personal Limitations, Chronic Role Limitations, Current Role Limitations, Physical Capacities, Strenuous Exercise, General Exercise, Acute Symptoms, Chronic Disease (self-report), assuming all validity variables scored positively.

[b]Median of correlations with the following validity variables: Anxiety, Depression, Positive Well-Being, Emotional Ties, Emotional Stability, assuming all validity variables scored positively.

notably higher with physical than mental health variables. Third, the two items that had highest median correlations with mental health validity variables were two of the three that included "feel" or "feeling" in the item stem. Because the mental health validity variables focus on feeling states, this is not a surprising finding. Fourth, two items in the Current Health scale (I and FF) reflect physical and mental health variables approximately equally; one of these includes the "feeling" concept in its item stem (FF). Fifth, EGFP and V (my health is excellent) had the most similar patterns of correlations with the validity variables, not surprising given that Ware and Karmos constructed item V to represent EGFP.

This analysis indicates why the Current Health scale has roughly equal median correlations with physical and mental health (see Table 21). Some of the items reflect primarily the physical health dimension, others the mental health dimension, and two reflect both dimensions roughly equally. With the exception of item V, the median correlations of the Current Health items with physical health were similar; more differences were seen in the extent to which the items reflected mental health.

These data may also suggest one reason that the correlation observed between Current Health and age is lower than that between EGFP and age: Two of the favorably worded Current Health items (Z and DD) correlate negatively with age; the former correlation is significantly different from zero. Thus, when all items are scored positively to score Current Health, these items correlate in the opposite direction from the other seven. One explanation for this finding is apparent from the item validity analyses reported above: of all the Current Health items, Z contains the most mental health-related variance. The zero-order correlation between mental health (as defined by the Mental Health Index) and age in the HIE is positive, indicating that mental health improves with age; physical health, on the other hand, declines with age. Thus, the negative correlation between Z and age, which indicates that people are less likely to report feeling bad lately if they are older, may relate to the mental health nature of the item.

To further explore empirical evidence about this issue, we did secondary analyses of data tapes from the Ware and Karmos (1976) study, correlating each of the Current Health items with age in their four general population samples. The correlation matrix appears in Table D.25. Without exception, the significant correlations between age and the Current Health items in Ware-Karmos data were in the expected direction; the only anomaly was not significantly different from zero. More importantly, these secondary analyses indicated that the items had much more substantial correlations with age in non-HIE samples. The most likely explanation for this finding is the restricted age range in the HIE sample. Thus, final statements about the item- or scale-level relationships between Current Health and age in general populations should not be based solely on HIE findings.

As with the health perceptions scales (see Table 21), the highest correlations in the physical health areas were with the acute symptoms and chronic diseases variables. The lowest physical health correlations were always with the Strenuous Exercise measure. Elimination of the acute symptoms, chronic diseases, and exercise variables did not alter conclusions about the median validity coefficients for physical health, because the items correlated roughly equally with the other five variables in that category, which defined limitations and capacity. The lowest mental health correlations were always with Emotional Ties, similar to what we observed for all health perceptions scales in the bivariate analyses. Highest correlations were generally with Emotional Stability or one of its component scales, Anxiety or Depression.

EFFECTS OF RESPONSE SET ON VALIDITY

Appendix C discusses in detail the results of our analyses of response set effects on measurement validity, a methodological issue that has been little studied for health status mea-

sures. We offer a brief summary of the results here. Those persons who require additional detail about the method used to study response sets and results are referred to App. C.

Acquiescent (ARS) and opposition (ORS) response sets occurred very infrequently and inconsistently in responses to HPQ items; ARS occurred somewhat more frequently than ORS.[1] Response errors of the opposition and acquiescent type were not correlated with other errors of the same type. Because these errors must occur and occur consistently before they can threaten validity, we dropped further investigation of the effects of ARS and ORS on the validity of the HIE general health perceptions measures.

Social desirability response set (SDRS)[2] occurred relatively consistently in responses to the general health perceptions measures. Moreover, SDRS was correlated both with the variables defining groups that would be of interest to compare on the general health measures (age, sex, education, and income) and with all general health measures but EGFP. These findings indicated that SDRS had a clear potential for biasing results of group comparisons. Our multivariate analyses of the existence of this bias supported our hypotheses about the existence and direction of this bias. Differences between men and women and between different education groups were underestimated when SDRS was not controlled statistically. Income group differences were always overestimated without control for SDRS. As hypothesized, age group differences on all but three measures were underestimated without control for SDRS; on Sickness Orientation, Health Worry/Concern, and Resistance, age group differences were overestimated. The effects of SDRS bias were particularly notable for group comparisons that used the General Health Ratings Index.

---

[1]ARS and ORS are tendencies to agree and disagree, respectively, with questionnaire items *regardless* of their content.

[2]SDRS is the tendency to respond to items with the socially desirable answer, and thus is related to item content.

VI. DISCUSSION

In this section, we discuss the results of our analyses in terms of the three major goals we had identified: confirmation of HPQ scaling and scoring rules in HIE samples; construction of an overall General Health Ratings Index; and increased understanding of the meaning of scores on general health perceptions measures. Several other issues that emerged during our analyses are also discussed more fully here.

CONSTRUCTION OF SUBSCALES

Our first goal for the analyses reported in this volume was to confirm, in HIE samples, the scaling and scoring decisions made by Ware and Karmos (1976) for the Health Perceptions Questionnaire (HPQ); most HIE health perceptions measures were drawn from the HPQ. We tested 27 items in six multi-item subscales to measure Current Health, Prior Health, Health Outlook, Resistance to Illness, Health Worry/Concern, and Sickness Orientation. Item groupings for these subscales were those hypothesized and tested by Ware and Karmos (1976); we also tested the addition of a new item to the Health Worry/Concern grouping.

Item descriptive statistics indicated skewed response distributions, with most means consistently on the side of the response scale midpoint indicating more favorable perceptions. Most item standard deviations were very close to 1.00, a desirable characteristic for items with a five-point response continuum that are to be combined in summated ratings scales.

Overall, the scaling studies were very successful. The notable similarity of the multitrait scaling results across the HIE samples, and the replication of results reported by Ware and Karmos (1976), led us to conclude that we had met the first goal of our analyses. These replications were particularly important given the differences in sample characteristics within both studies (HIE and Ware-Karmos) as well as between the two studies.

Almost all items met both criteria used in the multitrait scaling analyses: high (0.40 or over, absolute magnitude) correlations with the sum of other items in the hypothesized scale, and notably higher correlations with the hypothesized scale than with other scales in the battery. According to these criteria, items in the Current Health, Prior Health, Health Outlook, Resistance to Illness, and Sickness Orientation displayed both convergent and discriminant validity. Decisions regarding the appropriate scale placement of most items in these scales were therefore quite straightforward. The two Sickness Orientation items displayed less convergent validity than items in the other scales, correlating slightly below 0.40 with each other. They never failed a discriminant validity test, however, so they were appropriately scored together.

Items in the Health Worry/Concern scale met the two scaling criteria less consistently than items in the other scales, as had been observed by Ware and Karmos (1976). Some problems were caused by our addition of a new item (no. 6, worry or concern about health) to this scale grouping. Our decisions regarding scale placement of the five Health Worry/Concern items therefore relied more heavily on our assessment of the similarity of their content, as well as on the precedent of scoring rules developed by Ware and Karmos (1976).

We traced several of the Health Worry/Concern errors to lack of convergent validity for the two Concern items. They did not correlate highly with each other or with the three other items in the hypothesized scale. The two Worry items from the scale originally defined by Ware and Karmos and our hypothesized addition, correlated more highly among themselves than did the Concern items. Thus, inclusion of the Concern items resulted in a considerably

more heterogeneous measure than if the Worry items had been scored separately. The added Health Worry/Concern item (no. 6), although it discriminated less well than other worry items, scaled more like the Worry than Concern items. Additional scaling studies confirmed that the three Worry items (F,M, no. 6) could appropriately be combined in a single scale score, but that the Concern items could not be scored separately.

We had found that one reason for the lower homogeneity of Worry and Concern items was their relatively lower reliability than other general health perceptions items. We wondered whether another explanation for their lack of convergent validity might relate to the different rating methods used in the items included in these scales: absolute and relative ratings. One item in each set asks for an absolute rating of personal worry (F) or concern (X); another asks for a rating of personal worry (M) or concern (T) relative to that perceived for other persons. Several other scales had such item pairs (an absolute and a relative rating worded in opposite directions); these included items C and G in Resistance and I and Q in Current Health. We therefore examined whether the inter-item correlations for such pairs were consistently lower than other inter-item correlations in their scale groupings. With the exception of the Concern items, the other relative-absolute item pairs (including the pair of Worry items) did not account for the lowest inter-item correlations in their respective item groupings.

We also considered the possibility that the Concern items assessed a construct somewhat different from worry or anxiety about health as a point in time. Some investigators who have studied "concern" in analyses of health-related behavior appear to have considered it related to, but distinct from, worry or anxiety about health. Others have combined worry and concern items in aggregate scales, which suggests they are less distinct as constructs.

Berki and Ashcraft (1979, p. 1168) conceptualized health-related concern as "the salience of the value of health to the individual." Their operational definition referred to the amount of time spent thinking about health now and in the future, and satisfaction with future health, which appears to assess salience of health more than its value to the person. They did not study a worry or anxiety variable in their predictions of enrollment behavior and use of preventive services (Berki et al., 1977a; Berki and Ashcraft, 1979).

Becker et al. (1977b) refer to concern about health matters (broader in scope than concern about health per se) in a category of variables they term "health motivation," to explain health-related behaviors that appear to be motivated by desires for good health rather than only by avoidance of disease. In their expansion of the Health Belief Model, they included general measures of worry about illness in a distinct category, that of perceptions of "health threat." However, empirical analyses of their operational definitions of health-related worry and concern (both on the mother's part about her child's health) led them to combine the measures in an "Overall Health Concern" measure; this was further aggregated with measures of perceived susceptibility in an index of "General Health Threat to the Child" (Maiman et al., 1977; Becker et al., 1977b).[1]

Ware and Young (1976) showed that the Health Worry/Concern subscale correlated as high or higher than the other HPQ general health subscales with both rating and ranking measures of value placed on health. These results suggest that one or more of the Health Worry/Concern items measures the value placed on health, rather than perceived health status. Unfortunately, Ware and Young did not report correlations separately for the two Worry and two Concern items in their HPQ subscale.

Thus, the evidence in the literature is inconsistent with respect to the amount of conceptual and empirical distinction that is made between health-related worry and concern. Few of the studies cited provided explicit guidelines, beyond the conceptual or operational definitions they offered, as to how valid measures of worry and of concern should relate to other variables.

---

[1]The concern and susceptibility variables appear to have considerable overlap in operational definition; one concern item asks about concern about the child's getting (i.e., being susceptible to getting) sick.

In our analyses, the Concern items apparently do assess a construct somewhat different from that assessed by the Worry items, given their relatively poor convergent validity. Moreover, the Concern items probably also reflect a construct or constructs other than the underlying health construct revealed in most of the other general health perceptions items. This conclusion is supported by their failure to correlate substantially (above 0.30) with the first unrotated factor defined by the items. Moreover, the construct they do assess is considerably more stable over time than that assessed by the three Worry items. Further analysis of these items, beyond those possible with the validity variables available to us, will be required to understand better the meaning of the Concern items.

Most of the scaling errors that we observed for the Health Worry/Concern grouping occurred in Dayton, Ohio, and the two South Carolina sites. Data quality was slightly poorer overall in South Carolina, and item-scale correlations were generally lower in that site than in the others; this may have contributed to the scaling problems we encountered for Health Worry/Concern in South Carolina. Apart from the fact that Health Worry/Concern is a more heterogeneous item grouping relative to the others tested, as shown in results reported by Ware and Karmos (1976), it is difficult to explain why we observed relatively more scaling errors for Health Worry/Concern in Dayton, Ohio, than the other sites. Data quality was not a problem in Dayton. Data from other sites came from the first administration of the health perceptions items at the beginning of experimental participation; Dayton data came from the administration of items after three years of experimental participation. Although greater experience with the items or some experimental effect might have resulted in the scaling problems we observed, it is difficult to understand why they would have affected only the Health Worry/Concern scale. Items in the other five scales studied met scaling criteria as well in Dayton as they did in the other sites.

We noted that item W (expect a healthy life) failed most scaling tests of its discriminant validity by correlating higher with Current Health than with Health Outlook, its hypothesized scale grouping. Despite this error pattern, which was also noted by Ware and Karmos (1976), we decided to use item W to score Health Outlook rather than Current Health. We made that decision to maintain consistency with the original scoring rules for these items. Our decision could well be questioned; empirically, item W is much more a Current Health than Health Outlook item. Moreover, its inclusion in the Health Outlook scale is a major factor in the notably high correlation between the two scales. The zero-order correlation between Health Outlook and Current Health was 0.58; when item W is removed from the Health Outlook scale, this correlation drops to 0.49. Thus, just less than one-third[2] of the overlap between Health Outlook (scored with item W) and Current Health is due to the inclusion of item W. The lower correlation when item W is removed is not due in any great part to reduced reliability; the internal-consistency reliability of Health Outlook with item W is 0.73, and without item W, 0.69.

The scaling studies indicated that the 27 items were appropriately scored, in the six hypothesized groupings, as summated ratings scales. The scoring algorithms we adopted for HIE use, and recommend to other investigators who field these items, require recoding of item responses (from those printed in the questionnaire) so that a high score on each item indicates more of the construct measured by its scale. After recoding, the simple algebraic sum of relevant item scores yields the scale score. As a result, high scores on Current Health, Prior Health, Health Outlook, and Resistance to Illness reflect favorable perceptions related to these constructs. High scores on Health Worry/Concern and Sickness Orientation indicate more worry and greater tendency to accept illness as a part of one's life, respectively.

To summarize, items in five of the six general health perceptions subscales—Current

[2]Correlations must be squared to compare their relative magnitudes; $0.58^2 - 0.49^2 = 0.34 - 0.24 = 0.10$, which is about one-third of 0.58^2.

Health, Prior Health, Health Outlook, Resistance to Illness, and Sickness Orientation—successfully passed most scaling tests, confirming the scaling rules originally tested by Ware and Karmos (1976). Items in the Health Worry/Concern scale were somewhat less successful in meeting the assumptions of the scaling studies, as had been the case in the original scaling studies (Ware and Karmos, 1976). For use in the Health Insurance Experiment, the original scaling rules were therefore adopted for all the subscales and a new item was added to Health Worry/Concern.

If further work is done to improve these scales, we recommend that special attention be placed on improving both the convergent and discriminant validity of the Worry items. Although we believe the concern construct might be an important one, these items do not relate substantially to others in this general health perceptions battery. If additional work is done, item W should be tested as an addition to the Current Health scale, and a new item might be added to Health Outlook to replace it.

CONSTRUCTION OF A GENERAL HEALTH RATINGS INDEX

As a second major goal for our analyses, we wanted to develop a General Health Ratings Index that would combine all the general health perceptions constructs in an easily scored and reliable overall measure. Although several multi-item "general health" scales had been reported in the literature, few contained items from all the general health constructs identified in the literature. Moreover, most included items that assessed constructs other than perceptions toward health in general. The multi-item general health scales discussed in the literature did have in common a focus on current health, usually measured with self-ratings of health as excellent, good, fair, or poor and of current health in relation to that of others. Ware and Karmos (1976) had not developed an index during their studies of the HPQ measures.

In our earlier report (WDD, 1978), we proposed two different strategies for developing an overall General Health Ratings Index. The first approach, relying on principal components analyses, would define an index based on the major dimension underlying the general health items (i.e., the dimension accounting for most of the variance shared by the items). The second approach, relying on both theory and empirical results from predictive studies, would define an index (or indexes) using the general health construct(s) that best predicted some outcome of interest. As Ware and Karmos (1976) had indicated, this second approach would result in several indexes, each particular to the outcome being analyzed. In their studies of four different outcomes, both theory and empirical results indicated that the subscales that contributed most to the predictions varied depending on the outcome variable considered. The principal components approach would result in one index, considerably more general but with poorer predictive validity for any one outcome. We opted for the more general approach in the analyses reported here, considering it important to define and study the meaning of the major dimension underlying these general health perceptions items. If this dimension were defined by all (or most) of the construct-specific scales, these scales could still be scored separately and combined in appropriate subsets to yield the best aggregate measure for a particular predictive analysis.

Content of the Index

Because the statistical process of identifying the common dimension (principal components analysis) can capitalize on chance associations among items and conclusions may vary across populations, we were concerned that the definition of this dimension would be generalizable. We therefore placed considerable emphasis on replicating the common dimension across the

HIE sites. The more nearly similar the common factor (i.e., first unrotated component), the more confidence we could have that the dimension thus identified was actually the general construct underlying all the items.

Although we did not test statistically the similarity of components across the HIE sites, careful visual examination of the direction and magnitude of item-component correlations revealed a high degree of correspondence in results. The direction of item-component correlations was always identical; all favorably worded items correlated positively, and unfavorably worded items correlated negatively with the component. (For the principal components analysis, items were scored as they appeared in the questionnaire.) Thus, the component represented a favorable definition of health in general. Although the magnitude of individual item-component correlations differed somewhat across the sites, largely reflecting site differences in item variability, the correlations and the ordering of items on the component were quite comparable from site to site. Thus, we achieved our goal of replicability across sites for the Index definition. For this reason, we chose to define the HIE General Health Ratings Index on the basis of the first unrotated component from the combined-sites analysis.

The principal components analyses indicated that all but two of the general health perceptions constructs are required to define the general dimension underlying the 26 items studied.[3] None of the Concern or Sickness Orientation items correlated above 0.30 (i.e., shared more than 9 percent of their variance) with the general dimension. We considered this cutoff to be a minimum requirement for including items in the definition of the common factor. Items from all the other general health constructs—current health, prior health, health outlook, health-related worry, and resistance to illness—had substantial to very high correlations with the component and thus were important to its definition. Given these results, we judged that another of our goals for the Index—that it represent all of the major general health constructs—had been achieved. We raised questions above about the extent to which the Concern items represent a general health construct; similar issues arose regarding the Sickness Orientation items, which we discuss at greater length in a later section.

Scoring the Index: Issues of Simplicity Versus Weights

Other important goals related to index construction were those of identifying a simple scoring algorithm and achieving a reliable score. We believe simple scoring algorithms are desirable because they are easy to document, understand, and use, thus making data preparation and analysis less complicated and the study variables easier to describe. To the extent that these characteristics can be achieved by a scoring algorithm that yields as reliable a score as would a more complicated algorithm, there is no reason to adopt the more complicated scheme. Moreover, empirical evidence suggests that when the item-component correlations do not differ markedly in magnitude, the simple summated ratings score captures as much reliable variance as the more complicated factor scoring method (Armor, 1974).

We tested three scoring algorithms for the 22-item General Health Ratings Index: a simple summated ratings score, the sum of item responses; the sum of standardized and unit-weighted item responses; and a factor score, the sum of standardized item responses weighted by their item-component correlations. The simplest algorithm, the summated ratings score, actually weights each item proportionately to its variance. Because item variances differed, although most not markedly from 1.00, we examined the standardized item scoring algorithm. Finally, because item-component correlations varied in magnitude across the items used to define the Index, we weighted each standardized item score by its component coefficient for the factor score.

[3]Recall that we did not include the Worry, EGFP, or Pain items in these analyses. We wanted the resulting Index to be useful to those who have fielded only the Health Perceptions Questionnaire.

Reliability estimates for each of the scores were very high, and we found virtually no difference in reliability as a function of scoring method. The summated ratings method yielded a score with an internal-consistency reliability of 0.89; the other two methods yielded scores with reliabilities of 0.90. We did not consider that this difference at the second decimal place warranted the use of a more complicated scoring algorithm. The gains in reliable variance captured by standardizing and/or weighting item scores were trivial beyond that reflected in the summated ratings score.

This finding was particularly important because many investigators apparently assume that factor scoring should be used for an index based on a principal components analysis. Our findings suggest that even though the weights are derived so that they maximize the reliable variance of the first component, and capture as much variance among the original items as possible, the simple summated ratings score does equally well in this regard. As we noted before, this similarity did not result because items with large variances also had high item-component correlations, and thus were weighted more heavily in both the summated ratings and factor scoring algorithms. The scoring rules we have adopted for use in the HIE therefore recode item responses (from those printed in questionnaire) so that high scores on *all* items indicate more favorable health perceptions. Item scores are then summed to achieve a score on the General Health Ratings Index.

The issue of weights relates not only to score reliability but also to the predictive validity of the Index, as we have indicated already. The Index includes the items that best predict their common source of variance, rather than some criterion or outcome variable of interest. As Manning, Newhouse, and Ware's (1980) comparisons of health definitions indicated, the General Health Ratings Index does not predict subsequent consumption of medical services as well as do the subscales scored separately. This finding indicates that the subscales have different weights in the prediction of consumption than they do in the present Index. For example, Current Health and Prior Health had the largest b-weights in the prediction of subsequent use when the HPQ scales were scored separately and used as the definition of health status; examination of the item-component correlations for the Index indicates that while Current Health items generally had the highest weights, Prior Health items usually had the smallest weights.

RELIABILITY OF GENERAL HEALTH RATINGS

Before the studies of general health perceptions measures reported by Ware and Karmos (1976), the reliability of subjective self-reports of health had been severely criticized (see their literature review). These criticisms were based in part on fact, because the most commonly used general health ratings were single-item measures, known to be less reliable than their multi-item counterparts. Dichotomous scoring of many of the single-item measures, which had the effect of decreasing the information (variability) provided by the resulting score, also contributed to their relatively poor reliability. Ware and Karmos' (1976) developmental work on the HPQ general health measures demonstrated that multi-item measures of general health constructs, even short two- and three-item scales, met and usually exceeded minimum reliability standards for group comparisons in all populations studied. Similar findings of adequate reliability were noted by investigators who reported such information for the multi-item health indexes we found in our literature review, although few were studied in diverse population samples.

An important reason for HIE selection of the HPQ measures was the expectation that they would prove reliable in HIE samples. Our internal-consistency reliability studies indicated that the measures met this expectation. Reliability estimates for all the measures equaled, and usually exceeded, our minimum standard of 0.50 for group comparisons. Estimates cal-

culated for the Current Health scale and the General Health Ratings Index were very close to or equaled 0.90, indicating they could be used to compare individuals. Only for the very shortest measure, the two-item Sickness Orientation scale, did we observe a reliability estimate (0.46) below the 0.50 standard, and then only in one site.

A concern in our reliability studies was whether the measures would prove reliable even in the "worst case," in HIE samples with the lowest levels of education and income. Both variables are directly related to data quality; if reliability proved problematic in such samples, the intended group comparisons would be compromised. Ware and Karmos (1976) had studied this issue in their samples and reported favorable results. Our comparisons of reliability estimates across the HIE sites, whose samples differ considerably in sociodemographic characteristics, reached similar conclusions. Although reliability estimates were somewhat lower in the South Carolina samples, which are the most disadvantaged of HIE samples with respect to education and income, they remained adequate for the intended group comparisons.

Ware and Karmos had also studied "worst-case" reliability by computing reliability estimates for subgroups differing in educational attainment. Their results indicated that reliability remained acceptable, but because of sample size problems they did not examine subgroup reliability in very disparate groups. The larger HIE sample sizes, particularly that of the combined-sites sample, permitted us to study reliability in groups that differed much more in educational attainment than those studied by Ware and Karmos. In HIE analyses, reliability estimates for most scales showed a slight and steady increase with more years of schooling. Estimates for the longer scales, Current Health and the Index, remained high regardless of educational attainment. The most marked improvements in reliability were noted at the high school graduation level; reliability improved only slightly after that educational level. For those with less than a high school education, and even for persons with less than nine years of schooling, reliability remained adequate for group comparisons. Because the mean educational level in most HIE samples is near 12 years, and because the most notable improvement in reliability with educational level occurs at this point, we suggest that differences in reliability be kept in mind when comparing groups above and below the mean educational level.

To summarize, the reliability of the HIE general health perceptions measures was adequate for their use in HIE hypothesis-testing. Despite the relatively greater heterogeneity of the Health Worry/Concern items, which was directly related to the problems we observed in convergent and discriminant validity for several of these items, reliability estimates for this scale always exceeded the 0.50 minimum. The only reliability estimates below 0.50 that we observed, in the site-by-site analyses as well as in the educational subgroup analyses, were for Sickness Orientation. These low coefficients reflect both the relatively lower convergent validity of the items (below 0.40) and the shortness of this two-item measure. These findings, which generally replicate those reported by Ware and Karmos (1976), should lay to rest challenges to the use of such general health ratings on the grounds that they do not provide reliable information.

VALIDITY OF GENERAL HEALTH RATINGS

Our validity analyses were designed to provide information about the meaning of general health rating scores. We wanted to know how differences in rating scores should be interpreted in terms of personal health status, and specifically, what they tell us about both physical and mental health.

Studies of Validity Before HIE Analyses

Before the HIE analyses were done, three types of findings had been reported (and were summarized in our literature review in WDD, 1978). All provided some support for the use of

general health ratings in tests of hypotheses about health status. First, studies had shown that general health rating scores vary greatly in general populations. Thus, assuming that their measured variance is reliable, ratings have the potential to discriminate between persons. As noted above, findings about variability in these ratings scales were confirmed in the "nonaged" HIE samples, which underrepresent persons older than 60 years. Second, earlier studies (particularly those reported by Ware and Karmos, 1976) indicated that general health ratings are reliable, both in an internal-consistency sense and in terms of stability across administrations repeated over short intervals (e.g., a month). Third, published analyses had begun to document the meaning of general health ratings scores in relation to measures of health and illness constructs. The major problem with these validity studies (reported before HIE analyses were done) was that few had available a wide range of health-related validity variables for the same population. In particular, while most studies included one or more physical health measures, only a few included any measures of mental health, the other important health construct that valid general health measures should assess.

Content Validity

HIE analyses provided two types of evidence relevant to measurement validity. First, we showed that the HIE health questionnaires include a comprehensive sample of general health rating items that represent well those fielded in published studies. Included are items that ask about health in the past, present, and future, as well as about perceived resistance to illness and health-related worry/concern. Thus, from the perspective of what is described in HIE items, the general health measures have content validity.

Empirical Evidence of Validity

Second, we evaluated empirical evidence pertinent to the validity of the measures in relation to a wide range of self-reported and independently measured variables that should be related to valid measures of general health perceptions. Findings from these empirical studies support the use of most HIE general health perceptions measures in tests of hypotheses about group differences in health status and changes in health over time. Important features of our empirical analyses, and questions that remain to be addressed, are discussed below.

Dimensionality of Health Perceptions. The factor analyses reported by Ware and Karmos (1976) and the multitrait tests of the validity of general health items done with HIE data clearly demonstrate the structure of the health perceptions battery. These analyses consistently yielded distinct dimensions of health perceptions, which have been labeled Current Health, Prior Health, Health Outlook, Resistance to Illness, Health Worry/Concern, and Sickness Orientation. Thus, there is an empirical basis for distinguishing these dimensions and for scoring and interpreting them separately.

Validity as "General" Health Measures. HIE analyses contributed substantially to our understanding of the interpretation of the general health subscales and Index in terms of the specific physical and mental health constructs the measures were intended to assess. Correlations between the general health measures and physical and mental health variables were consistently statistically significant and in the hypothesized direction. Furthermore, many of these associations were of substantial magnitude. With one exception, we have a strong basis for concluding that the HPQ subscales contain information about both physical and mental health status.

The exception is the Sickness Orientation scale, which had the weakest correlations with physical and mental health measures among the general health measures. Among the physical and mental health measures, Sickness Orientation correlated significantly only with the

self-report of chronic diseases and with Anxiety; the correlation with Anxiety probably explains its significant correlation with the overall Mental Health Index, as well. These findings, along with those regarding the virtually perfect stability of the Sickness Orientation construct, suggest that the Sickness Orientation scale should *not* be used to test hypotheses about group differences in health status or about changes in health over time. The validity variables that were included in HIE analyses provide very little information about the meaning of scores on the Sickness Orientation measure. At this time and considering its stability coefficient, our best guess is that this scale measures a health-related personality trait.

We were also very interested in whether the general health measures contain different amounts of information about physical and mental health status, and about different operational definitions within each of these two major health components. All measures (except Sickness Orientation) correlated substantially with all the physical health measures we studied (except the exercise variables, which may reflect personal interest and disposition as much as physical health). Thus, we have an empirical basis for using the measures to test hypotheses about physical health status, including chronic disease status, acute symptomatology, and limitations in self-care, mobility, and physical functioning (personal functioning). With two exceptions, the measures did not differ remarkably in the amount of information shared with the different operational definitions of physical health. The exceptions were Resistance to Illness and Pain, both of which are notably more sensitive than the other measures to differences in acute physical symptoms (e.g., colds, flu symptoms). This exception was hypothesized for Resistance on the basis of earlier findings; we had not hypothesized this result for the Pain measure.

Also as hypothesized, the measures (again with the exception of Sickness Orientation) were substantially related to each of the mental health measures we studied. The Current Health scale overlapped most with mental health; the other measures had fairly similar amounts of overlap with mental health. Finally, the strength of associations with physical and mental health were roughly equal for most of the scales; this was particularly notable for Current Health and the Index. Thus, the general health ratings generally reflect a balanced mix of these two major health status components. This pattern of results supports their use as *general* health measures.

We have attempted to summarize the results from the bivariate validity studies in tabular form (see Table 23), so that readers would have an easily referenced guide to the interpretation of scores on each of the HIE general health perceptions measures. Rows in Table 23 identify the general health measures; columns indicate the category of validity variable (for Use of Services, we have considered only the true predictive validity of the measures by focusing on consumption of medical care services in the year after the ratings were obtained). A "yes" indicates that the general health measure can be used to predict a given validity variable, and a "no," that it should not be used in such a prediction. We have indicated the direction of the prediction for the age variable; in the physical and mental categories, direction depends on whether the validity variable represents a favorable or unfavorable definition of health (directions of these relationships can be checked in Table 20). We have also highlighted findings related to the relative balance between mental and physical health variance and notable correlations with specific operational definitions of physical and mental health constructs.

Incremental Validity. HIE analyses demonstrate that most of the general health measures (except Sickness Orientation) are valid summary measures of both physical and mental health. Two questions remain unanswered by our analyses; both refer to their incremental validity (Sechrest, 1967). The first is whether each of the general health perceptions measures contributes unique information about health beyond that contained in the other measures. Evidence about incremental validity at this level is important, for example, to decisions about whether to use the Index, the subscales, or even to rely on one of the subscales. The second is whether the ratings contain useful information about health status *beyond* that found in measures that focus specifically on physical and mental health.

Table 23

SUMMARY OF VALIDITY STUDIES INDICATING WHETHER EACH GENERAL HEALTH PERCEPTIONS MEASURE CONTAINS INFORMATION ABOUT PHYSICAL AND MENTAL HEALTH, USE OF SERVICES, AND AGE

General Health Perceptions Measure	Physical Health	Mental Health	Subsequent Use of Services	Age[a]
Current Health	Yes Best measure of physical health among HIE general health measures; most related to acute symptoms. Roughly equally related to physical and mental, on average.	Yes Best measure of mental health among HIE general health measures.	Yes	Yes (-)
EGFP	Yes More than mental.	Yes Less than physical.	Yes	Yes (-) More than Current Health.
Prior Health	Yes More than mental.	Yes Less than physical.	Yes	Yes (-)
Health Outlook	Yes Roughly equally related to physical and mental, on average.	Yes	No	Yes (-)
Resistance to Illness	Yes Particularly to acute symptoms.	Yes More than physical.	No	Yes (+)
Health Worry/Concern	Yes Roughly equally related to physical and mental, on average.	Yes	Yes	Yes (-)
Pain	Yes Best measure of acute symptoms of all HIE general health measures.	Yes	Not available	Yes (-)
Sickness Orientation	Yes Only to acute symptoms and self-reported chronic disease.	Yes Smallest correlations of all HIE general health measures.	No	Yes (-) Highest correlation of all HIE general health measures.
General Health Ratings Index	Yes Next best measure (to Current Health) of physical health. Roughly equally related to physical and mental, on average.	Yes Next best measure (to Current Health) of mental health.	Yes	No

[a]Sign in parentheses indicates direction of relationship with age.

Analyses reported by Manning, Newhouse, and Ware (1980) provide information relevant to both questions.[4] In addressing the question of the gain in explanatory power achieved by increasingly comprehensive definitions of health status, they compared the performance of five different definitions in their prediction of medical care consumption: (1) EGFP; (2) Current Health; (3) the General Health Ratings Index; (4) the multi-item general health subscales, scored separately (i.e., Current Health, Prior Health, Health Outlook, Resistance to Illness, Health Worry/Concern, and Sickness Orientation); and (5) the six general health subscales together with other measures of physical and mental health and social circumstances.

Their results indicated that Current Health provided a significantly better definition of health status for the prediction of medical services consumed than did EGFP, but that the Index did no better than Current Health in this longitudinal prediction. The subscales, scored separately, did predict consumption significantly better than did the Index. This finding suggests that the separate scales do display incremental validity, at least in relation to consumption of medical care services. Whether this finding would hold in predicting other outcomes of interest cannot be concluded from this evidence; that provided by Ware and Karmos (1976) in cross-sectional "predictions" suggests that different sets of the subscales demonstrate incremental validity depending on the variable being predicted.

Regression analyses reported by Manning, Newhouse, and Ware (1980) also indicate clearly the value of the general health measures over and above those of physical and mental health. Thus, the general health measures contain "additional" information that predicts use of medical care services. This finding fits our hypothesis that the general health measures provide more information about health status than if they simply summarized some of the information contained in more traditional physical and mental health measures. As noted in Sec. I, our best guess is that general health perceptions measures improve prediction over component-specific health status measures because they reflect both individual differences in the weights placed on those components when evaluating health and differences in personal concerns about different levels of health within each component. Further research is necessary to test these hypotheses.

Results reported by Manning, Newhouse, and Ware (1980) can be interpreted from another point of view. The physical and mental health status measures still made unique and significant contributions to the prediction of medical care expenditures when the general health ratings were included in the prediction (although their coefficients were smaller than when the ratings were not included). Thus, general health ratings do not contain all of the information that is important to the prediction. Such findings, particularly if they are replicated in predictions of other health-related outcomes, suggest that the appropriate measurement strategy for the health field would be to develop a core set of measures that assess the major health components (physical and mental) and that capture subjective evaluations of health in general.

The possibility remains that not all of the additional information in general health ratings is about *health*. Some may, as we have argued, reflect the subjective evaluation of the objective information a person has about his own health and his relative preference for different aspects of health status (i.e., physical and mental). In addition, the general health measures may reflect one or more constructs, other than health, that also predict medical care consumption.

A plausible candidate for this nonhealth-related variance, attitudes toward doctors and

---

[4]Their definitions of these variables were the same as those presented in Table 5; they used data from 1557 respondents aged 18 and older who participated in the HIE during the entire first year of the experiment in Seattle and Massachusetts and provided complete data on both the enrollment MHQ and first annual HQ in those sites (Manning, Newhouse, and Ware, 1980).

medical care services, can be rejected. Certainly no items among the HIE general health items appear to measure attitudes toward care. In addition, these general health measures relate only weakly to measures of health care attitudes. Again, our best data source is the study reported by Manning, Newhouse, and Ware (1980); attitudes toward care were measured and included in the predictive model along with the general health ratings and other health measures. The attitudinal measures made unique and significant contributions to the prediction of medical care expenditures. We would not have expected this result had the remaining variance in the general health ratings been attributable to attitudes toward medical care. Thus, this explanation for the interpretation of the unexplained variance in general health ratings is not supported.

Similarity of Meaning for Current Health and the Index. The empirical validity analyses indicated that among the HIE general health measures, Current Health and the Index yielded the "most" valid scores. Each one correlated with both the physical and mental health measures, and their relationships with these measures were the most substantial we observed (even after adjusting for differences in reliability). In fact, the direction and magnitude of validity coefficients for these two measures were virtually identical, indicating that their scores have very similar interpretations. We observed this similarity again in the multivariate analyses, when Current Health and the Index provided virtually identical and equally predictive definitions of health in the prediction of medical care expenditures.

This similarity of meaning reflects several features of the Index and its parts. Although it aggregates most of the general health subscales, and we have shown that those subscales have different validities, all correlate in the same direction as Current Health with the physical and mental health variables when they are scored to indicate favorable health ratings. Therefore, we would expect Current Health and the Index to correlate in the same direction with the health status variables. The similarity in magnitude of their validity coefficients reflects their high reliabilities, and the fact that by definition, much of the variance in Index scores reflects variance in current health, the core concept among the HIE general health perceptions measures. Most differences in magnitude of validity coefficients, such as the age correlations, can be explained by noting that the subscales included in the Index correlate in *opposite* directions with age when they all are scored to indicate more favorable health perceptions.

The similarity is also related to our decision not to weight subscales in the Index according to their relative contributions in predicting some external (i.e., non-health) criterion. We wanted the Index to maximally predict "general health," which the principal components and validity analyses indicate it does. Moreover, those analyses revealed that the current health concept captures much of the information that the HIE general health perceptions measures have in common. Thus, the similarity in their interpretation is not unexpected.

The Meaning of Score Units on the Index. Several of our validity analyses focused on the meaning and importance of score units throughout the General Health Ratings Index score range in terms of substantial health impairment. We defined three dichotomous variables for these analyses, indicating impaired physical and emotional health and presence of one or more chronic diseases.

Our analyses indicated that Index scores throughout the range are sensitive to physical (i.e., limitations in bodily functioning) and emotional impairments. In the HIE sample, Index scores below 77 to 80 indicated that respondents had a better than average probability of severe physical and/or emotional impairment.[5] The relationship between limited physical functioning and Index scores was not significantly different from linear, indicating that the probability of functional impairment increases quite steadily as Index scores get lower. The relationship between severe emotional impairment and the Index scores, by contrast, was

[5]This score may well differ in general population samples that include a more representative proportion of those aged 62 and older than does the HIE sample.

significantly nonlinear; Index scores for respondents in the bottom three or four deciles of the distribution showed a more marked relationship to the probability of emotional impairment than did those in the upper deciles. Thus, Index scores in the lower 30 to 40 percent of the distribution (in this general population sample) are considerably more sensitive to serious emotional than physical impairments.

Although we were also interested in examining Index scores in relation to an indicator of acute disease (i.e., symptoms experienced during the preceding two weeks, as defined by the Acute Symptoms variable), we could not construct a dichotomous indicator of acute symptoms. Perhaps coincidentally, all persons included in the validity analyses (N = 1200 in Seattle only) reported experiencing at least one symptom during the preceding two weeks from among those included on the list. Thus, the extent to which extreme Index scores are sensitive to acute symptomatology remains to be studied.

Moreover, we regard our operational definition of the presence of serious chronic disease to be a preliminary one. We were able to consider data from only a subset of the chronic diseases whose possible presence and severity the HIE measures at enrollment (see Brook et al., 1980 and continuing), so the definition is a limited one. In addition, we did not distinguish between diseases whose presence was determined solely on the basis of objective screening examination data and those whose presence was scored using (alone or in addition) self-report data. This approach may well have resulted in our underestimating the relationship between chronic disease and Index scores throughout the range, particularly if diseases scored using the objective data were only those that respondents were not aware of, and thus did not consider in their rating of general health status. Given these drawbacks to our definition of the chronic disease indicator, further analyses of the meaning of Index scores in relation to such a definition of health impairment would be useful.

Social Well-Being and Health. We observed very weak relationships between ratings of health in general and measures of social contacts (e.g., visits with friends and relatives) and of social resources (e.g., number of close friends). We drew several conclusions from these trends. First, because general health ratings contain little or no information about social well-being (as defined in the HIE), they should not be used to test hypotheses about group differences in social well-being or changes in social well-being over time. Our results in this regard are consistent both with previously reported HIE results (Ware, Davies-Avery, and Brook, 1980) and with the published literature (reviewed in Donald et al., 1978; Donald and Ware, forthcoming). Second, quantitative differences in social contacts and more qualitative differences in social resources are apparently not considered part of "health" as defined by non-aged persons in general, U.S. populations. If they were, we would expect to see much stronger relationships between general health ratings and measures of social contacts and resources. We continually fail to detect such relationships, despite considerable effort aimed at improving methods for scoring HIE measures of social contacts and resources (Donald and Ware, forthcoming). These relationships are not observed in the published literature, either (as reviewed in the preceding citations). Given that most of the reported analyses of this relationship focus specifically on elderly populations, the lack of an observed relationship in HIE analyses probably cannot be attributed to restricted age range in the HIE sample.

CURRENT HEALTH AS A CORE GENERAL HEALTH CONSTRUCT

The current health construct emerged from our analyses as the "core" concept, reflecting the common variance in perceptions of health in general as defined by the HIE battery. The inclusion of current health items in all of the self-styled general health indexes we identified in the literature had indicated the theoretical importance of this construct. Ware and Karmos (1976) had also assumed that a personal rating of health at the present would be the most

useful general health rating measure, and had overrepresented current health items among those in the HPQ. HIE analyses supported this assumption and the importance of the current health construct.

The first supporting evidence came from the scaling analyses. Most of the scaling errors we observed occurred because items correlated higher with the Current Health scale than with their hypothesized scales, yet no convergent or discriminant validity errors were noted for Current Health items. This pattern of errors suggested that Current Health might define a common source of variance for many of the HPQ items.

The principal components analyses provided further support for concluding that the Current Health construct is central to perceptions of general health at a point in time. Current Health items consistently had the highest correlations with the component in all analyses. We tested and rejected two alternate explanations for these results. The prominence of the Current Health items was not explained by their greater numbers in the HPQ battery (one-third of the items assessed Current Health). When we did a principal components analysis with roughly the same number of items per general health construct, the Current Health items still had the highest item-component correlations. As far as we could determine (using non-HIE data), the Current Health items were no more reliable than those assessing other general health constructs. Greater reliability could have accounted for their dominance of the common construct, particularly if their reliable variance was the variance they shared with the other items.

Evidence from the validity analyses provided even more support for considering Current Health the most important general health construct. Current Health is as good as the other subscales in representing differences in physical health status, and appears to summarize information about mental health status better than the others. That Current Health represents perceptions of health at a point in time was supported by analyses of its stability. Although it was one of the more reliable measures, it was notably less stable over time than the other measures we studied. This instability in its variance suggests that it assesses perceptions that change over time, which is what we would expect of perceptions about health at any one point in time. Given the combined weight of this evidence, we conclude that personal perceptions of health at the present best represent the information shared by the general health perceptions constructs defined in the HIE battery.

TWO MEASURES OF CURRENT HEALTH

Items in the Current Health scale and the single-item EGFP measure refer, explicitly or implicitly, to the respondent's health at the present. Given the importance of the current health construct as revealed in the scaling studies and the similarity of focus in the two measures, an important aspect of our analyses is the ability to compare and contrast the performance and interpretation of the two measures. Such a comparison is particularly interesting in light of the fact that, although the EGFP rating is perhaps the most commonly fielded general health measure (alone or in published indexes), interpretation of EGFP scores has not been systematically studied.

Our findings support the conclusion that EGFP scores are more stable than scores on Current Health. Thus, EGFP scores are less sensitive to change than are scores on Current Health. One feature of EGFP that explains why it is less sensitive to change than Current Health is the much greater coarseness of its response scale. Only four response categories are offered for EGFP, and thus EGFP divides respondents into only four groups. The five response categories offered for each of nine Current Health items allows 37 possible groups of respondents. Much finer distinctions can be made using Current Health than EGFP. When small changes in health occur, they are therefore much more likely to be reflected in different Current Health scores than in different EGFP scores.

Another reason why EGFP has greater stability than Current Health relates more to EGFP's interpretation than to its measurement scale. During analyses of validity, we noted that EGFP correlated much more substantially with age than did Current Health. Because measures of age are highly stable over time, the age-related variance in EGFP accounts for some of its stability.[6]

Comparison of validity coefficients for EGFP and Current Health, keeping in mind the poorer reliability of EGFP, indicates that EGFP reflects less information about mental health; both contain substantial physical health variance and little information about social participation and activities. This comparison, together with the stability data, suggest that Current Health reflects overall health in the present somewhat better than does EGFP.

We did observe in Manning, Newhouse, and Ware's (1980) study that Current Health was a better predictor of subsequent medical expenditures than EGFP. In part, their finding reflects the greater coarseness of EGFP. Differences between EGFP and Current Health in their relationships to mental health variables may also have affected this conclusion. Although the expenditures variable did not include visits to mental health professionals, many outpatient visits to non-psychiatrist physicians are for problems of everyday living that are related to mental health, such as tension or depression (Regier, Goldberg, and Taube, 1978). Thus, we would expect the measure that better reflected mental health variance to better predict outpatient expenditures (so defined).

Although EGFP is less reliable and somewhat less valid as a general measure of health in the present than Current Health, we believe it is still a very adequate general health measure. EGFP scores reflect substantial amounts of variance in the two major health components—physical and mental. Depending on the purpose of the study, the greater prominence that should be given physical health over mental health when interpreting its scores may not be a significant drawback to its use.

A somewhat greater drawback will probably be its coarseness and relatively lower reliability. These features could be improved by using a new response continuum that defined more points on the range, particularly at the positive end of the range, where most respondents place themselves. Although other investigators have occasionally used response choices in addition to the traditional "excellent," "good," "fair," and "poor," most added choices at the negative end of the continuum. Moreover, they have not made the gains in discrimination and precision that were possible because most continued to use a dichotomous scoring algorithm (see the literature review in Ware and Karmos, 1976). The tradeoffs in dichotomous versus continuous scoring of EGFP (with or without added response categories) are discussed below.

SCORING EGFP

EGFP, the most commonly fielded general health rating, is frequently scored either as an interval measure by assuming that the intervals between each response choice are equal, or as a nominal measure generally by collapsing excellent and good in one category and fair and poor in the other (Ware and Karmos, 1976). In the health economics literature, responses to EGFP are often treated as a series of dummy variables.

We examined the appropriateness of the interval-scoring assumption by using the multi-item Current Health scale, a more comprehensive self-rating of health at the present, as a

[6]The stability coefficient for an age measure, adjusted for point-in-time reliability, should be 1.00; if age were measured at time one and again exactly a year later, each respondent would be a year older but still in the same place in the distribution. The intertemporal correlation (stability coefficient) is not affected by additive changes in scores, only by changes in the ordering of respondents within the score distribution.

criterion. The four EGFP groups had a significantly nonlinear relationship with Current Health ($p < 0.00001$). The distance between "poor" and "fair" was half again as large as that between "good" and "excellent" (in terms of standard deviations on the Current Health scale). Despite this, the deviation from linearity accounted for barely 3 percent of the variance in the relationship.

If this is considered an important amount of nonlinearity, in terms of its effect on the precision of measurement, our finding suggests a transformation of EGFP scores to take the unequal intervals into account. For HIE purposes, we considered such a transformation important. Each of the four EGFP response categories was therefore scored at its combined-sample mean score on the Current Health scale (see App. B).

If maintaining the maximum precision were of less concern, or if the simplest possible scoring algorithm were sought, our finding could also be taken to suggest that simple equal-interval scoring (e.g., 4 = excellent, 3 = good, 2 = fair, 1 = poor) captures the lion's share of the information provided by EGFP in samples like that enrolled in the HIE, and for that reason would be an appropriate scoring method.

These findings suggest that, whether the transformation is made or not, dichotomous scoring of EGFP responses results in the loss of considerable information. As Ware and Karmos (1976) noted in their literature review, dichotomous scoring of EGFP was the most common scoring algorithm used by investigators studying this item. The place at which the dichotomy was made differed; some scored "poor" versus other groups, and others "poor" and "fair" versus the other categories. "Poor" and "fair" are clearly more distinct than "good" and "excellent." However, in HIE samples, fewer people scored in the "poor" and "fair" categories (57 and 509 in the combined-sites HIE sample, respectively) than "good" (N = 2603) and "excellent" (N = 1503). Thus, collapsing "poor" and "fair" combines more dissimilar categories yet loses information about far fewer respondents than does combining "good" and "excellent."

Whether this loss of information can be tolerated depends on why EGFP is being used in the analysis. For example, if the analysis attempts to predict use of hospital services, focusing on the "poor" and "fair" groups may be appropriate; one would expect far more variability in hospital use among these subgroups than for those scoring "good" or "excellent."

In general, we recommend that EGFP be scored as a four-level ordinal or interval scale (whether equal-interval or a transformation to account for unequal intervals). Such scoring preserves useful information and thus improves reliability and precision. It also facilitates use of more powerful statistical techniques than would dichotomous scoring. Continuous scoring also maintains the option of further aggregation if combining response categories is warranted by the analysis.

STATES OR TRAITS?

In measuring variables, the distinction is often made between "states" and "traits." Variables that assess states refer generically to some aspect of "existence at a particular time and place"; those that measure traits generally refer to some "distinguishing quality (as of a personal character)," usually some characteristic that is not time- and/or place-bound (*Webster's New Collegiate Dictionary,* 1975). These definitions have clear implications for the measurement properties—particularly the stability—of variables that are states or traits and for their usefulness in testing hypotheses about changes over time. Specifically, we would expect the stability—the amount of measured variance that does not change over time—to be notably higher for measures of traits than those of states.

In comparison to many variables that describe individuals, health status can be considered more a state than an unchanging, or little changing, personal trait. Thus, good measures of health status should reflect such changes over time. In measurement terms, their stability

coefficients—correlations between scores on the measure at two different times—should be substantially less than unity.[7]

While health status constructs may be more states than traits, the state-trait distinction may be useful if viewed as two ends of a continuum rather than a clear dichotomy. This health state/trait continuum is a particularly useful way to examine health status measures that are intentionally time-bound; for example, the Current Health and Prior Health scales. Some aspects of personal health status, such as one's health history (e.g., Prior Health), can be expected to change relatively little over time, as contrasted with one's health at a particular time (e.g., Current Health), which could be expected to show more change over time. Following this logic, we might hypothesize Prior Health to reflect more a trait and Current Health to assess a state, and therefore Prior Health to have a much more substantial stability coefficient than Current Health. In fact, the stability estimates were consistent with this reasoning; the Prior Health estimate (adjusted for point-in-time reliabilities) was approximately 1.00, and the Current Health (adjusted) estimate of 0.66 indicated that it was the least stable of the general health perceptions over time.

The estimate of almost perfect stability for the Prior Health scale is not inconsistent with its interpretation as a measure of a general health construct, given the focus of its items on the past. Similarly, the lower stability of the Current Health scale, relative to the other general health measures, suggests that it assesses the health state we would most expect to change over time. (The stability coefficient provides information only about its "state-ness"; that this is a health state can be concluded from the validity studies.)

Specific hypotheses about whether the other general health measures are more like states or traits are less readily constructed, because they have no specific time-referent or because the theory about the constructs they measure is sparse. Very generally, we would hypothesize that to the extent that they are health measures they are more states than traits, and thus should have stability coefficients notably less than unity. With the exception of Sickness Orientation, the coefficients for the other multi-item general health measures were consistent with this hypothesis; all were between the 1.00 observed for Prior Health and the 0.66 observed for Current Health. Adjusted stability coefficients for the two single-item measures, EGFP and Pain, approximated unity; because the communality estimates used to evaluate their reliability are lower-bound estimates, the adjusted stability coefficients undoubtedly overstate their true stability. Before firm conclusions about the stability of the health perceptions constructs are reached, the three- and five-year intertemporal correlations that will be available in HIE data should be examined.

The adjusted coefficient for Sickness Orientation (0.96) was almost equal to the Prior Health estimate, suggesting that this measure assesses a construct that is more nearly a trait than a state. Information from the validity analyses indicates that some of the stability of this measure derives from its notable relationship with age. As we discussed above, the validity analyses provided little additional information about the meaning of scores on this measure, and further studies will be required for that purpose.

CHOOSING MEASURES OF GENERAL HEALTH PERCEPTIONS

One of the most common requests we receive from readers when reports such as this one are published is to recommend particular measures for their studies. While appropriate selection of measures depends heavily on a given study's objectives, design, and available resources, we can make some general recommendations here about selection of measures from the HIE general health perceptions battery for use in other studies.

---

[7]Imperfect reliability at either point in time will cause the estimated stability to be less than unity. This can be taken into account by dividing the estimate by the square root of the two reliability estimates multiplied together, which better approximates the true intertemporal stability.

Assuming measurement resources are not too constrained and the respondent will not be too heavily burdened by other questions in the survey, we recommend that the entire 22-item General Health Ratings Index be fielded. This option offers the greatest flexibility to the researcher and analyst, because it provides a valid overall general health rating score (the Index) as well as the possibility of disaggregating and scoring each of the subscales separately. As our previous discussion indicated, separate scoring of the subscales may well prove useful in certain predictive studies, or when the specific constructs they assess are of particular interest to the study.

Another option, which would still permit disaggregation, would be to select items from each of the subscales and to construct a shorter version of the Index. Apart from the item validity studies we did on the Current Health scale, we have not done the studies necessary to identify such a subset of items. The inter-item correlation matrix (Table D.8) is available for those who wish to pursue this option.

In the HIE, the Index will be used to examine whether different health care financing mechanisms have any measurable effect on general health status. Given its reliability and ability to make fine distinctions between levels of health, the Index should prove the most precise of any single HIE general health measures for such hypothesis tests. If measurable effects are found, the separately scored subscales can be used to study where and why those effects occur.

In studies that require a measure of the "core" general health construct, the one that captures much of the variance common to the HIE general health ratings, either the full Index or the Current Health subscale could be fielded. When disaggregation is not required for analyses and/or measurement resources dictate a short measure of this "core" construct, the nine-item Current Health scale, or the shorter version we propose below, may well be preferable.

For analyses that require a very short scale to assess the "core" construct, or simply want fewer items that focus on Current Health so more can be devoted to other constructs, the investigators may prefer to use a short version of the Current Health scale. To select a shorter set of items, the table that includes correlations between the Current Health items and validity variables (see Table 22) should prove useful. A major criterion in selecting such items should be that they reflect physical and mental health substantially and roughly equally, as does the complete nine-item scale. Another way to achieve this balance is to select one or two items that assess primarily physical health constructs and a similar number that reflect chiefly mental health, thus possibly minimizing redundancy. The items selected should also maintain the balance in favorable and unfavorable wording of the original scale, so the final score is not confounded with acquiescent or opposition response sets.

With these approaches in mind, we suggest that a valid four-item Current Health scale could be constructed using items I and V, which have substantial and roughly equal median correlations with the physical and mental health validity variables we studied; item Q, which has one of the highest median correlations with physical health variables after I and V;[8] and item Z, which has the most substantial correlation with the mental health variables. Items Q and V are favorably worded and I and Z are unfavorably worded, which maintains scale balance. Preliminary analyses using HIE data indicated that this subset of Current Health items will yield a very reliable (0.80) summated ratings score.

To assist other investigators in choosing and analyzing general health perceptions measures, empirical results and general population norms for the Health Perceptions Question-

[8]Item A actually has the next highest correlation with physical and mental health variables, but some might object to the similarity between its content (Doctors say my health is excellent) and that of item V (My health is excellent). For this reason, we suggest item Q as a first choice for this short-form scale; items A,I,V, and Z yield a short-form scale of similar reliability (0.81).

naire are now available from the studies on some 2000 respondents reported by Ware and Karmos (1976), as well as the nearly 5000 respondents studied in these HIE analyses. Data for international comparisons of general health status will be available from an ongoing Dutch study of medical care expenditures, which is fielding the HPQ as well as other HIE measures of physical and social health (Pederson, personal communication).[9] The questionnaire is available in Spanish (Ware, Wright, and Snyder, 1976) and Dutch (Pederson, personal communication) translations.

METHODOLOGICAL ISSUES

Response Set Effects on General Health Measures

We found no evidence that the HIE general health perceptions measures are affected by response biases that are unrelated to items content, namely, acquiescent and opposition response sets. A major response bias that is related to item content, that of socially desirable responding, does correlate with scores on the measures, which means a significant amount of non-health variance must be taken into account when their scores are interpreted. Moreover, social desirability response set scores differed significantly across most of the groups that would be studied to understand differential responses to health care financing mechanisms. The paucity of information in the literature about the effects of response sets on health status measures other than mental health leads us to suggest that thorough investigations of this issue could be important to the health status measurement field.

Effects of Questionnaire Placement on General Health Scores

As part of our analyses, we examined the effects of questionnaire placement on the reliability and mean scores of the HIE general health perceptions measures. To our knowledge, this issue has not been studied systematically for health status measures. The first annual Health Questionnaire administered in the Seattle, Washington, Massachusetts, and South Carolina sites varied placement of the health and attitudinal batteries, so random halves of these respondent samples completed the general health battery near the middle or at the end of an approximately 300-item questionnaire. When the HPQ appeared in the middle, it was preceded by physical and mental health and social circumstances batteries; when it appeared last, the health-related batteries came after batteries assessing attitudes toward medical and dental care and the effects of medical care. Late questionnaire placement generally resulted in significantly different mean scores on the general health measures; the directions of the differences indicated that health perceptions were less favorable (better health, less worry, less pain) when the measures appeared late in the questionnaire. Late placement did not, however, affect reliability of the general health perceptions measures.

Our analyses of the effect of questionnaire placement on one of the attitudinal batteries (assessing satisfaction with dental care) had indicated that score means were higher and reliability estimates somewhat lower when the battery appeared later in the questionnaire (Davies and Ware, 1981). Several explanations for the differential effect of questionnaire placement on reliability of health and attitudinal measures come to mind. Attitudinal measures may be more reactive than health status measures; the experience of completing ques-

[9]Many other investigators are using all or part of the HIE health perceptions battery in their studies, to judge from the number of requests we have received for norms and scoring rules; we have not attempted to keep a record of these studies.

tionnaire items themselves, particularly if they refer to experiences toward which the respondent's evaluation are obtained later in the questionnaire, may introduce error into the attitude measurement. Health status measures may also be more salient to respondents than questions requesting an expression of sentiment, and thus be less subject to fatigue that may introduce error into a person's responses. This issue deserves far more extensive study to determine whether our findings generalize to other health status and attitudinal measures, and then to explain the results. These results do, however, suggest that it is important to hold the placement of health status batteries constant over different administrations of a given questionnaire (whether to different samples or at different times in the same sample) so comparisons will not be adversely affected by biased means.

CONCLUSIONS ABOUT THE USEFULNESS OF HIE GENERAL HEALTH RATINGS

The analyses of HIE general health perceptions measures that we have reported and discussed in this volume support our conclusion that these measures will prove useful in tests of HIE hypotheses about the effects of health care financing on health status. The measures are simple to score. They demonstrate considerable variability in general population samples and in non-aged samples like those in the HIE. This evidence, together with that of their reliability (even in the "worst" case) and moderate intertemporal stability, indicates they will be very precise in measuring real changes over time in a repeated-measures design such as that of the HIE.

The crucial issues, then, are whether the general health perceptions measures can be used to test hypotheses about health status and, if so, how differences in general health perceptions should be interpreted. We have expanded considerably our understanding of these issues. Validity evidence supports their interpretation (with the exception of Sickness Orientation) as indicators of both physical and mental health status and their use in testing hypotheses about changes in health status over time. Because they reflect both physical and mental health variance, their validity as "general" health measures was supported. If generous health insurance improves both physical and mental health status, this positive effect may best be detected by the general health perceptions measures because they measure both of these components well. Thus, these measures may provide a solution to the problem of aggregating physical and mental health effects in a single health index. However, because the general health scores cannot be disaggregated with respect to physical and mental health effects, they would not be appropriate measures to test hypotheses about interventions that have substantially different effects on mental and physical health.

We identified several areas for additional research on general health measures. Some involve improvements in scale content (e.g., Worry/Concern and Health Outlook), and others in scoring EGFP. One of the most important areas includes studies to provide further information about the variance in general health ratings that is unrelated to physical and mental health as they are defined by HIE measures. Our in-depth studies of the meaning of score units on the General Health Ratings Index provided some information about effect sizes; using indicators that are very sensitive to severe health impairments, we found that the changes of single scale units on the Index are associated with noteworthy shifts in the probability of having severe emotional and physical impairments. Further analyses should pursue the study of effect sizes for general health perceptions measures. Analyses relevant to these issues will undoubtedly result as a by-product of the experimental analyses that are now under way to study the effects of health care financing mechanisms on personal health status.

Appendix A

GENERAL HEALTH PERCEPTIONS MEASURES AS THEY APPEAR IN HEALTH INSURANCE EXPERIMENT QUESTIONNAIRES

Single-Item Measures

GENERAL HEALTH

3. IN GENERAL, WOULD YOU SAY YOUR HEALTH IS EXCELLENT, GOOD, FAIR, OR POOR?

(Circle one)

Excellent	1	19/
Good	2	
Fair	3	
Poor	4	

---

5. DURING THE PAST 3 MONTHS, HOW MUCH PAIN HAVE YOU HAD?

(Circle one)

A great deal of pain	1	21/
Some pain	2	
A little pain	3	
No pain at all	4	

---

6. DURING THE PAST 3 MONTHS, HOW MUCH HAS YOUR HEALTH WORRIED OR CONCERNED YOU?

(Circle one)

A great deal	1	22/
Somewhat	2	
A little	3	
Not at all	4	

HEALTH PERCEPTIONS

DO NOT WRITE IN THIS SPACE

128. PLEASE READ EACH OF THE FOLLOWING STATEMENTS, AND THEN CIRCLE ONE OF THE NUMBERS ON EACH LINE TO INDICATE WHETHER THE STATEMENT IS TRUE OR FALSE FOR YOU.

THERE ARE NO RIGHT OR WRONG ANSWERS.

If a statement is definitely true for you, circle 5.
If it is mostly true for you, circle 4.
If you don't know whether it is true or false, circle 3.
If it is mostly false for you, circle 2.
If it is definitely false for you, circle 1.

SOME OF THE STATEMENTS MAY LOOK OR SEEM LIKE OTHERS. BUT EACH STATEMENT IS DIFFERENT, AND SHOULD BE RATED BY ITSELF.

	Definitely true	Mostly true	Don't know	Mostly false	Definitely false	
A. According to the doctors I've seen, my health is now excellent	5	4	3	2	1	26/
B. I try to avoid letting illness interfere with my life	5	4	3	2	1	27/
C. I seem to get sick a little easier than other people	5	4	3	2	1	28/
D. I feel better now than I ever have before	5	4	3	2	1	29/
E. I will probably be sick a lot in the future	5	4	3	2	1	30/
F. I never worry about my health	5	4	3	2	1	31/
G. Most people get sick a little easier than I do	5	4	3	2	1	32/
H. I don't like to go to the doctor	5	4	3	2	1	33/
I. I am somewhat ill	5	4	3	2	1	34/
J. In the future, I expect to have better health than other people I know	5	4	3	2	1	35/
K. I was so sick once I thought I might die	5	4	3	2	1	36/
L. I'm not as healthy now as I used to be	5	4	3	2	1	37/
M. I worry about my health more than other people worry about their health	5	4	3	2	1	38/

CARD 06

	Definitely true	Mostly true	Don't know	Mostly false	Definitely false	DO NOT WRITE IN THIS SPACE
N. When I'm sick, I try to just keep going as usual	5	4	3	2	1	39/
O. My body seems to resist illness very well	5	4	3	2	1	40/
P. Getting sick once in a while is a part of my life	5	4	3	2	1	41/
Q. I'm as healthy as anybody I know	5	4	3	2	1	42/
R. I think my health will be worse in the future than it is now	5	4	3	2	1	43/
S. I've never had an illness that lasted a long period of time	5	4	3	2	1	44/
T. Others seem more concerned about their health than I am about mine	5	4	3	2	1	45/
U. When I'm sick, I try to keep it to myself	5	4	3	2	1	46/
V. My health is excellent	5	4	3	2	1	47/
W. I expect to have a very healthy life	5	4	3	2	1	48/
X. My health is a concern in my life	5	4	3	2	1	49/
Y. I accept that sometimes I'm just going to be sick	5	4	3	2	1	50/
Z. I have been feeling bad lately	5	4	3	2	1	51/
AA. It doesn't bother me to go to a doctor	5	4	3	2	1	52/
BB. I have never been seriously ill	5	4	3	2	1	53/
CC. When there is something going around, I usually catch it	5	4	3	2	1	54/
DD. Doctors say that I am now in poor health	5	4	3	2	1	55/
EE. When I think I am getting sick, I fight it	5	4	3	2	1	56/
FF. I feel about as good now as I ever have	5	4	3	2	1	57/

CARD 06

Appendix B

SCORING RULES FOR GENERAL HEALTH PERCEPTIONS ITEMS AND SCALES

Table B.1

SCORING RULES FOR HEALTH PERCEPTIONS SUBSCALES

ITEM SCORING

1 = Definitely false 2 = Mostly false 3 = Don't know 4 = Mostly true 5 = Definitely true	A, D, G, J, M, O, P, Q, S, V, W, X, Y, BB, FF
1 = Definitely true 2 = Mostly true 3 = Don't know 4 = Mostly false 5 = Definitely false	C, E, F, I, K, L, R, T, Z, CC, DD
1 = Not at all 2 = A little 3 = Somewhat 4 = A great deal	I6 (worry item)
1 = No pain 2 = A little pain 3 = Some pain 4 = Great deal of pain	I5 (pain)
15 = Poor 26 = Fair 34 = Good 40 = Excellent	I3 (EGFP)

SCALE SCORING

	(Sum these items after scoring as above)
Current Health	A + D + I + L + Q + V + Z + DD + FF
Prior Health	K + S + BB
Health Outlook	E + J + R + W
Resistance to Illness	C + G + O + CC
Health Worry/Concern	I6 + F + M + T + X
Sickness Orientation	P + Y

Note: Questionnaire responses for Health Perceptions battery are printed 1=Definitely false, etc., so only items in second set are recoded, as are I6, EGFP, and Pain, to score construct-specific general health perceptions measures.

Table B.2

Scoring Rules, General Health Rating Index

ITEM SCORING

1 = Definitely true 2 = Mostly true 3 = Don't know 4 = Mostly false 5 = Definitely false	C, E, I, K, L, M, R, Z, CC, DD
1 = Definitely false 2 = Mostly false 3 = Don't know 4 = Mostly true 5 = Definitely true	A, D, F, G, J, O, Q, S, V, W, BB, FF

INDEX SCORING

(Sum these items after scoring as above)

A + C + D + E +
F + G + I + J +
K + L + M + O +
Q + R + S + V +
W + Z + BB + CC +
DD + FF

NOTE: Questionnaire responses are printed 1=Definitely false, etc., so only items in first set are recoded to score General Health Rating Index.

Appendix C

RESPONSE SETS AND THEIR EFFECTS ON GENERAL HEALTH RATINGS

THE LITERATURE

A potential threat to the validity of general health ratings can arise from response sets,[1] which can introduce bias or error into studies based on questionnaire data. The literature characterizes at least three types of response sets: acquiescence, the tendency to agree with statements regardless of content; opposition, the tendency to disagree with statements regardless of content; and social desirability, the tendency to respond in a socially desirable manner. Unlike the other two, socially desirable responding *is* related to item content, particularly content that refers to sensitive or embarrassing subjects or things that have social value. We have abbreviated the terms referring to these sets as ARS, ORS, and SDRS, respectively.

If response set is present, serious problems may result when questionnaire data are analyzed. Response set may be a source of "correlated errors" between items. Such errors can artificially inflate reliability estimates when inter-item correlations are in the same direction, and in such instances will threaten measurement validity by biasing scale scores. This bias can lead to erroneous conclusions about the extent and nature of group differences (Gove and Geerken, 1977, p. 1294). For example, Ware (1978) demonstrated that acquiescent response set so biased mean scores on single-item and on unbalanced[2] multi-item measures of patient satisfaction that differences between groups at several educational levels were overestimated by unbalanced scales containing only favorably worded items, and underestimated or missed entirely by scales with only unfavorably worded items. Thus, when groups differ in response set tendency but not in perceived health, response set bias in the scores may lead to the erroneous conclusion that there are group differences in perceived health. When groups differ in both set tendency and perceived health, estimates of group differences in perceived health will be biased by the presence of set. Response set can therefore threaten both the validity of measures and of conclusions based on these measures.

Biases related to acquiescent and opposition sets and their effects on attitudinal measures have been documented (cf. Messick, 1961; Ware, Snyder, and Wright, 1976; Ware, 1978; Bradburn et al., 1979). Social desirability response set has been studied for a range of measures and questionnaire administration methods (e.g., measures of mental health and of behaviors that are socially undesirable such as illegal activity; items that may be sensitive or threatening; self- versus interviewer-administration of such measures).

Previous studies suggest that scores on measures of health-related attitudes (i.e., patient satisfaction with medical care) can be significantly biased by acquiescent, and to a lesser extent, socially desirable response sets (Ware, Snyder, and Wright, 1976). Although the effects of response set have not been systematically studied for measures of physical health,

[1]By definition, response style represents consistent respondent behavior across questionnaires, such as a tendency to respond desirably or to agree with difficult items. Response set refers to response styles that are observed for a specific questionnaire battery. In this discussion, response set is used synonymously with response style to connote both the behavioral tendency and the actual set score for a respondent in a given questionnaire (Ware, Snyder, and Wright, 1976).

[2]Scale balance refers to the proportion of items that are favorably and unfavorably worded. Unbalanced scales contain only or a sizeable majority of items worded in the same direction.

psychometric studies of mental health measures demonstrate the considerable effects of socially desirable responding on such measures (e.g., Edwards, 1970; Phillips and Clancy, 1970; Gove and Geerken, 1977; Bradburn et al., 1979).

Our earlier literature review (Ware, Davies-Avery, and Donald, 1978) revealed no studies of response set effects on general health ratings other than preliminary studies reported by Ware and Karmos (1976) during development of the Health Perceptions Questionnaire (HPQ). They found little evidence of acquiescent or opposition set in most study sites, although there were some signs of set bias in the most disadvantaged sample they studied. The finding of no ARS and no ORS held only when the scales were scored as they have been in the Health Insurance Experiment (HIE)—predominantly balanced between favorably and unfavorably worded items. When single-item and unbalanced multi-item measures of the six general health perceptions constructs (Current Health, Prior Health, Health Outlook, Resistance to Illness, Health Worry/Concern, and Sickness Orientation) were computed by Ware and Karmos, they correlated with both ARS and ORS.

Ware and Karmos (1976) did not address the issue of SDRS in relation to general health ratings included in the HPQ. Findings that indicate SDRS affects mental health measures may be instructive, because some of the general survey measures of mental health include items similar in content to the general health ratings in the HPQ (cf. Ware et al., 1979). As Edwards (1970) has demonstrated, socially desirable responding increases with items of increasing sensitivity (e.g., those with very personal or embarrassing content). We would also expect to see SDRS effects on general health measures because this tendency is more pronounced when items refer to things that are perceived to have social value, such as health. Because most general health ratings are less sensitive than mental health measures, they should be less affected by socially desirable responding than are mental health measures.

ANALYSIS PLAN AND METHODS FOR STUDYING RESPONSE SET

Acquiescent and Opposition Response Sets

We used the matched-pairs formula (Ware, 1978), which was based on Messick's formula (1961), to calculate ARS and ORS scores. Because the HPQ was balanced between favorably and unfavorably worded items, this method did not require use of additional items to measure ARS and ORS. Item wording was balanced for four of the six construct-specific subscales, and we were able to identify seven matched-item pairs (see Table C.1). Each pair included one favorably and one unfavorably worded item pertaining to the same construct. These item pairs were identified on the basis of scale placement and high negative inter-item correlations.

The logic of the matched-pairs method argues that each instance of incompatible responses to items in the pair represents an occurrence of acquiescence or opposition. For example, one matched-pair contains two items from the Resistance scale: "I seem to get sick a little easier than other people," and "Most people get sick a little easier than I do." Respondents who used affirmative responses ("definitely true" or "mostly true") for both items, thus contradicting themselves, manifested an acquiescent response pattern. By contrast, respondents who used negative responses ("mostly false" or "definitely false") for both items manifested an opposition response pattern. Occasional inconsistent responses to items in the matched pairs could occur as a result of respondent error in marking the questionnaire, random responding, or keypunch errors, and would not by themselves constitute proof of response set. The intent of our analysis was to identify those individuals who *consistently* demonstrated acquiescent or opposition set tendencies.

Table C.1

MATCHED PAIRS USED TO SCORE ACQUIESCENT AND OPPOSITION RESPONSE SET[a]

Pair Number	Pair Number	Item Content
1	1	According to the doctors I've seen, my health is now excellent
	30	Doctors say that I am now in poor health
2	3	I seem to get sick a little easier than other people
	7	Most people get sick a little easier than I do
3	12	I'm not as healthy now as I used to be
	32	I feel about as good now as I ever have
4	9	I am somewhat ill
	22	My health is excellent
5	11	I was so sick once I thought I might die
	28	I have never been seriously ill
6	5	I will probably be sick a lot in the future
	23	I expect to have a very healthy life
7	8	I don't like to go to the doctor
	27	It doesn't bother me to go to the doctor

[a]Attitude toward going to the doctor is the object of measurement for items in pair no. 7. Although this construct is not considered in this volume, theory would suggest that these items are appropriately used to score response set because if set is manifested in the HPQ battery, it would be manifested regardless of specific item content.

For ARS and ORS studies, data were available from samples in all HIE sites. We counted the number of times ARS and ORS response patterns occurred independently for each respondent in each HIE site. Scores for each measure had the potential to range from zero (no set) to seven (maximum set tendency). After counting occurrences of ARS, we computed phi correlations between the scores on each pair of items and those on all other pairs to examine the consistency with which the set occurred; the same procedure was followed to examine ORS. Only if acquiescent or opposition patterns were consistent (rather than randomly occurring) phenomena for an individual respondent would they bias analyses based on the general health measures. If either set (or both) did occur consistently, we planned to correlate set scores with respondent characteristics and with scores on the general health measures. If sets were correlated with general health measures and variables defining population subgroups (e.g., different income groups), their effect would be to bias estimates of group differences. We planned to examine the extent of the bias by regressing each general health score on measures of sociodemographic characteristics, with and without statistical control for ARS or for ORS. If neither set occurred consistently, these correlational investigations of threats to validity from ARS and ORS would be unnecessary.

Social Desirability Response Set

To estimate socially desirable responding, eight items adapted from the Comrey Personality Scales (Comrey, 1970) were interspersed in the mental health battery fielded on Mental Health Questionnaires and Health Questionnaires following the enrollment MHQ in Dayton. (Table C.2 presents item wording and scoring details for the SDRS measure.) Scores had the potential to range from 0 (minimal SDRS tendency) to 6 (maximum SDRS tendency). Scoring studies of SDRS (Veit and Ware, forthcoming) indicated that this set occurred relatively consistently; the internal-consistency reliability estimate was 0.50, indicating the score was just reliable enough to permit group comparisons. The SDRS studies used only data from respondents in Seattle, the site for which our files included both the general health perceptions measures and the SDRS score; sample sizes for these studies ranged from 1940 to 1999.

RESULTS

ARS and ORS Effects on General Health Ratings

Frequency distributions, means, and average inter-item correlations (an estimate of consistency) for our measures of ARS, ORS, and SDRS appear in Table C.3. Data for ARS and ORS are from the combined-sites sample; those for SDRS, from Seattle only. Site-specific data for ARS and ORS are presented in Table C.4.

These results show that both ARS and ORS occurred very infrequently and inconsistently in responses to HPQ items. ARS did occur somewhat more often that did ORS, in line with results reported for the HPQ measures by Ware and Karmos (1976). Results in each site followed the same pattern. We did note that relatively more respondents in South Carolina demonstrated ARS than in the other sites, although not consistently enough to be a problem. These data indicate that there were errors in responses to items in the seven matched pairs, but that the errors associated with any given pair were not related to those associated with any other pair. Because response errors must occur and must be intercorrelated before they can present threats to validity, these findings indicate that neither ARS nor ORS threaten the validity of the HIE general health perceptions measures. We therefore dropped further analyses of the effects of ARS and ORS.

Table C.2

ITEMS AND SCORING RULES FOR SOCIAL DESIRABILITY RESPONSE SET MEASURE[a]

Item Content	Response Categories	Item Scoring
How often do you eat too much?	Very often, fairly often, sometimes, almost never	0
	Never	1
In general, would you say your morals have been above reproach?	Yes, definitely	1
	Yes, probably; I don't know; probably not, definately not	0
How often have there been times in your life when you felt you acted like a coward?	Very often, fairly often, sometimes, almost never	0
	Never	1
Would you say that you give every penny you can to charity?	Yes, definitely; yes, for the most part	1
	Yes, I try; no	0
In choosing your friends, how important to you are things like their race, their religion, or their political beliefs?	Always very important, almost always important, usually important, not too important, hardly ever important	0
	Not important at all	1
If it is more convenient for you to do so, how often will you tell a lie?	Very often tell a lie, fairly often, sometimes tell a lie, almost never	0
	Never tell a lie	1
How often have you done anything of a sexual nature that society does not approve of?	Very often, fairly often, sometimes, almost never	0
	Never	1
Are your table manners at home just as good as they are when you are invited out to dinner?	Yes, always as good	1
	Yes, with rare exceptions; yes, usually just as good; no, usually worse at home; no, quite a bit worse at home; no, very bad at home	0

[a]Socially Desirable Response Set (SDRS) was scored in two steps. First, item scores were summed after scoring as indicated above; scores for SDRS ranged from 0 to 8. Second, SDRS scores were recoded as follows: 8=6, 7=5, 6=4, 5=3, 4=2, 2=0, and 1=0. See Veit and Ware (forthcoming) for further details on the logic of scoring SDRS.

Table C.3

MEANS, FREQUENCY DISTRIBUTIONS, AND AVERAGE INTER-ITEM CORRELATIONS FOR ACQUIESCENT, OPPOSITION, AND SOCIAL DESIRABILITY RESPONSE SETS, ALL SITES COMBINED

	Response Set					
	ARS[a]		ORS[a]		SDRS[b]	
Occurrences	No.	Percent	No.	Percent	No.	Percent
0	2765	58.6	3315	70.3	1194	60.0
1	1472	31.2	1130	23.7	391	19.6
2	405	8.6	225	4.8	265	13.3
3	53	1.1	29	0.6	97	4.9
4	14	0.3	10	0.2	33	1.7
5	3	0.1	5	0.1	8	0.4
6	2	0.0	0	0.0	2	0.1
7	1	0.0	1	0.0	(c)	(c)
Mean	0.537		0.369		0.702	
Average Inter-Item Correlation	0.06		0.06		0.14	

[a]Scores could range from one to seven.

[b]Only studied in Seattle sample.

[c]Scores could range from one to six only.

SDRS Effects on General Health Ratings

The data in Table C.3 show that SDRS occurred somewhat more often and with considerably more consistency than did ARS and ORS. Because of the relatively greater consistency of its occurrence, we pursued further analyses designed to investigate the impact of SDRS on validity.

Correlations between SDRS and the general health perceptions measures and sociodemographic variables appear in Table C.5. As the data in Table C.5 show, SDRS correlated significantly with all the general health perceptions measures but EGFP, and with all sociodemographic variables but race. Thus, the potential exists for socially desirable response set to bias results of group comparisons: SDRS occurred relatively consistently, correlated with the variables that define the groups, and correlated with the health perceptions measures on which the groups are to be compared. The direction of the expected bias in group differences is a function of the direction of SDRS's correlation with both the group variable and the particular health perception measure being considered. Our hypotheses for the direction of expected bias follow:

Table C.4

FREQUENCY DISTRIBUTIONS AND MEANS FOR ACQUIESCENT AND OPPOSITION RESPONSE SETS BY SITES

	Site/Set							
	Dayton		Seattle		Massachusetts		South Carolina	
Occurrences	ARS	ORS	ARS	ORS	ARS	ORS	ARS	ORS
0	459	535	980	1147	693	783	633	850
1	231	179	484	381	292	236	465	334
2	63	40	128	62	63	35	151	88
3	3	3	9	12	10	7	31	7
4	1	2	1	0	3	2	9	6
5	1	0	0	0	0	0	2	5
6	0	0	0	0	2	0	0	0
7	1	0	0	0	0	0	0	1
Mean	0.503	0.364	0.481	0.338	0.444	0.315	0.702	0.455

- Underestimates of education group differences on all health perceptions measures.
- Underestimates of differences between men and women on all health perceptions measures.
- Overestimates of income group differences on all health perceptions measures.
- Underestimates of age group differences on all general health perceptions measures but Resistance to Illness, Health Worry/Concern, and Sickness Orientation; on the latter three measures, age group differences will be overestimated.
- No bias in differences between racial groups, because SDRS was not correlated with race.

To test these hypotheses about the effects of SDRS bias, we regressed each of the general health ratings on the sociodemographic variables (age, income, education, sex, and race) with and without statistical control for SDRS. Results of these regressions appear in Table C.6. Entries in the columns are the standardized regression coefficients from two different regression equations, the first without and the second with statistical control for SDRS. We present standardized coefficients because they can be more readily compared.

The regression results indicate that in most cases SDRS had the hypothesized biasing effects on estimates of the magnitude of group differences (i.e., the magnitude of the regression coefficient). With few exceptions, however, the magnitude of the coefficient changed very slightly when SDRS was controlled statistically. The most notable changes in magnitude were observed for the Index in relation to education and income. Only one changed coefficient was observed for group differences in Prior Health, probably because Prior Health had a low (although statistically significant) correlation with SDRS.

Group differences in education and between men and women tended to be underestimated when SDRS was not controlled, as hypothesized (compare col. I entries with those in col. II for age, education, and sex). Age group differences on Current Health, EGFP, Prior Health,

Table C.5

CORRELATIONS BETWEEN SOCIAL DESIRABILITY RESPONSE SET AND GENERAL HEALTH PERCEPTIONS AND SOCIODEMOGRAPHIC VARIABLES, SEATTLE

Measure	SDRS
GENERAL HEALTH PERCEPTIONS	
Current Health	.12**
EGFP	.01
Prior Health	.04*
Health Outlook	.13**
Resistance to Illness	.12**
Health Worry/Concern	-.12**
Pain	-.07**
Sickness Orientation	-.10**
General Health Ratings Index	.15**
SOCIODEMOGRAPHIC VARIABLES	
Age	.14**
Family Income	.12**
Education	-.08**
Sex	.07**
Race	.00

*Significant at $p<0.05$.
**Significant at $p\leq 0.001$.

Health Outlook, Pain, and the Index were underestimated; those for Resistance, Worry/Concern, and Sickness Orientation were overestimated, as we hypothesized. Only the direction of bias of differences between men and women on the Worry/Concern scale was in the opposite direction from that hypothesized. This may have occurred because the zero-order correlation between sex and Worry/Concern was insignificant. As hypothesized, income group differences were always overestimated when SDRS was not controlled, and the magnitude of differences between whites and nonwhites was unaffected by SDRS.

To summarize, estimates of the magnitude of group differences in general health perceptions (with the probable exception of Prior Health) will be slightly biased if SDRS is not taken into account. The bias will be particularly apparent in studies of differences using the Index, which aggregates most of the construct-specific measures.

Table C.6

REGRESSION COEFFICIENTS FOR SOCIODEMOGRAPHIC VARIABLES IN PREDICTIONS OF GENERAL HEALTH PERCEPTIONS SCALES, WITHOUT (I) AND WITH (II) STATISTICAL CONTROL FOR SOCIAL DESIRABILITY RESPONSE SET

	Age		Sex[a]		Education		Income		Race[b]	
Measure	I[c]	II[d]	I	II	I	II	I	II	I	II
Current Health	-.14	-.16	-.04	-.05	.11	.13	.10	.08	-.06	-.06
EGFP	-.24	-.25	-.10	-.11	.20	.20	.16	.15	-.07	-.07
Prior Health	-.12	-.12	-.03	-.03	.08	.08	.08	.07	.04	.04
Health Outlook	-.15	-.17	.00	-.01	.12	.13	.02	.01	-.06	-.07
Resistance	.11	.10	-.11	-.12	.03	.04	.04	.03	-.07	-.07
Worry/Concern	-.06	-.05	-.04	-.03	-.10	-.10	-.04	-.04	.06	.06
Pain	.09	.10	.13	.14	-.05	-.06	-.08	-.07	.00	.00
Sickness Orientation	-.16	-.15	.07	.07	.04	.04	-.03	-.02	.02	.02
General Health Ratings Index	-.10	-.12	-.05	-.06	.09	.14	.12	.07	-.06	-.06

[a]Scored male = 1, female = 2.

[b]Scored white = 1, nonwhite = 2.

[c]Column I contains standardized coefficients from regression of measure on age, sex, education, income,and race.

[d]Column II contains standardized coefficients from regression of measure on age, sex, education, income, race,and SDRS.

NOTE: Coefficients $\geq$.06 are significant at $p<0.05$, two-tailed test.

DISCUSSION

Studies of the effects of response styles, biases, or sets have been uncommon for health status measures. With regard to attitudinal measures, data indicate that the consistency with which ARS and ORS occur can be minimized considerably if not entirely by relying on balanced measures (Cloud and Vaughn, 1970; Ware, 1978).

Because the HPQ scales are balanced, containing roughly equal numbers of favorably and unfavorably worded items for each general health construct, we did not expect response sets that occur independently of item content—ARS and ORS—to affect reliability estimates. With unbalanced scales, in which all items are worded and scored in the same direction, bias unrelated to content would tend to inflate inter-item correlations. When estimating reliability with an internal-consistency approach, this consistent bias could not be differentiated from common true score variance. Thus, unbalanced scales would appear to be more reliable than was actually the case. With balanced scales, responses to all favorably or all unfavorably worded items are reversed (depending on whether the precoded values indicated favorable or unfavorable perceptions) before scoring the scale. Response set thus correlates in opposite

directions with favorably and unfavorably worded items in a balanced scale. When the items are summed to score the scale, response set effects tend to cancel out and reliability estimates are generally little affected.

Balanced scales cannot, on the other hand, minimize the effects of content-related response sets such as SDRS. Respondents who tend to present themselves favorably will agree with favorably worded items and disagree with unfavorably worded items. After item scores are recoded, this set will correlate in the *same* direction with items in balanced scales and tend to inflate reliability estimates. Therefore, reliability estimates for the multi-item general health subscales may have been overestimated slightly due to the occurrence of SDRS among the HIE respondents.

If general health ratings were correlated with response set, their validity would be affected because their scores would reflect variance in a construct that was not the object of measurement. Response set could also have notable effects on conclusions regarding group differences if two conditions were met: The groups compared differed in response set tendency, and response set measures correlated with the health status measures on which the groups were to be compared. Because of the importance of group comparisons in HIE hypothesis-testing, our analyses of the effects of response set on the general health perceptions measures provided valuable information about their usefulness to HIE analyses.

Replicating results reported by Ware and Karmos (1976) for the HPQ scales, we found minimal evidence of acquiescent or opposition response set in HIE samples. Occurrences of acquiescent and opposition response patterns to matched-item pairs (one item worded favorably and one worded unfavorably) were rare, and did not occur with any consistency from pair to pair. We noted this finding in all HIE sites, including South Carolina, the least socioeconomically advantaged sample in the HIE.

File availability permitted us to study the effects of SDRS only in one HIE site (Seattle, Washington). The large sample size (N of roughly 2000) and our replication of Ware and Karmos' results suggest that our findings regarding the effects of SDRS on the general health measures will generalize to other HIE (and non-HIE) samples. SDRS correlated significantly with all but one general health measure (EGFP), and we found differences in SDRS tendency among most groups of interest: between men and women, across age groups, among income groups, and between persons differing in educational attainment. We found no differences in SDRS tendency for racial groups (nonwhite and white).

The effects of this bias were that group differences in education and differences between men and women on all the general health measures (but EGFP, which did not correlate with SDRS) were underestimated when SDRS was not controlled. Income group differences on the general health measures were always overestimated without this statistical control. Age group differences were overestimated for Worry/Concern, Resistance to Illness, and Sickness Orientation, and underestimated for the other general health measures. (The difference in direction of the bias for age groups with the various general health measures related to their differing correlations with SDRS and with age.) The effect of this bias was particularly apparent in estimates of group differences using the General Health Ratings Index, which aggregates most of the multi-item subscales. We recommend therefore that SDRS be measured and its effect controlled statistically in analyses of group differences on general health variables. Such controls will improve the precision of estimates of group differences using the HIE general health measures.

Appendix D

SUPPORTING TABLES

Table D.1

FREQUENCY DISTRIBUTIONS OF RESPONSES TO GENERAL HEALTH RATING ITEMS, BY SITE AND ALL SITES COMBINED

Item	Site[a]	Responses[b] 1	2	3	4	5	Missing
A	D	17	42	173	335	192	0
	S	35	36	335	636	542	7
	FI	8	21	107	159	188	1
	FR	4	13	107	217	243	0
	C	25	36	153	232	184	1
	G	37	30	212	232	151	2
	ALL	126	178	1106	1809	1498	11
C	D	16	45	134	235	329	1
	S	36	85	144	434	905	12
	FI	18	22	43	117	283	1
	FR	21	26	50	138	349	0
	C	29	43	71	179	308	2
	G	23	49	108	167	315	4
	ALL	143	270	549	1268	2487	20
D	D	63	157	164	306	69	5
	S	135	307	382	620	160	8
	FI	39	64	96	205	79	3
	FR	41	97	128	254	64	0
	C	76	125	94	236	99	2
	G	74	124	90	282	92	2
	ALL	428	873	952	1902	562	20
E	D	8	17	355	155	224	3
	S	7	18	535	339	705	6
	FI	6	12	206	71	188	1
	FR	1	5	237	97	244	1
	C	13	16	324	109	168	0
	G	17	30	383	91	141	4
	ALL	52	98	2038	860	1669	15
F	D	91	354	43	219	52	0
	S	226	726	78	434	140	9
	FI	92	181	16	134	60	2
	FR	82	270	25	156	51	1
	C	87	226	30	187	100	4
	G	89	257	30	185	101	6
	ALL	666	2012	223	1315	501	22

Table D.1—continued

Item	Site[a]	Responses[b] 1	2	3	4	5	Missing
G	D	11	48	418	218	64	3
	S	43	92	701	528	240	5
	FI	16	32	244	123	68	0
	FR	29	32	280	152	91	3
	C	30	37	300	178	85	2
	G	33	38	306	195	90	5
	ALL	162	278	2245	1394	638	18
I	D	16	71	58	223	391	1
	S	23	101	147	326	1007	10
	FI	6	29	40	93	315	2
	FR	13	28	38	105	400	1
	C	21	58	70	138	343	3
	G	27	61	94	156	324	5
	ALL	106	348	447	1040	2776	22
J	D	11	33	452	174	89	1
	S	18	54	820	440	272	5
	FI	17	19	284	90	73	3
	FR	12	22	337	128	85	2
	C	29	25	341	137	98	2
	G	34	27	388	129	84	7
	ALL	121	180	2618	1098	700	20
K	D	66	74	36	128	455	3
	S	170	144	78	227	985	8
	FI	55	32	23	72	301	3
	FR	60	44	31	78	371	1
	C	87	61	29	97	356	3
	G	85	72	27	95	383	3
	ALL	522	427	224	696	2848	21
L	D	63	187	134	206	169	0
	S	145	344	266	390	459	7
	FI	43	75	70	114	181	0
	FR	40	107	85	156	196	0
	C	78	155	95	130	172	3
	G	105	149	111	137	160	2
	ALL	474	1016	761	1132	1334	12
M	D	234	201	269	43	12	2
	S	148	344	266	390	459	7
	FI	189	111	120	45	18	0
	FR	222	147	157	41	17	0
	C	192	148	179	70	41	1
	G	199	147	209	69	38	0
	ALL	1641	1165	1404	357	147	7

Table D.1—continued

Item	Site[a]	Responses[b] 1	2	3	4	5	Missing
O	D	13	61	139	437	109	8
	S	28	116	243	896	321	2
	FI	10	43	74	260	96	0
	FR	9	33	93	321	128	1
	C	25	64	99	316	126	0
	G	24	58	93	363	124	0
	ALL	109	375	738	2592	903	11
P	D	100	191	93	310	65	5
	S	28	116	243	896	321	2
	FI	59	107	59	185	73	3
	FR	95	136	63	217	73	1
	C	88	123	76	231	112	4
	G	93	137	92	235	105	2
	ALL	670	1133	595	1746	573	20
Q	D	33	52	181	339	154	1
	S	43	99	323	686	453	9
	FI	13	31	95	170	174	6
	FR	17	34	122	243	168	5
	C	40	60	157	216	157	5
	G	50	54	185	236	137	5
	ALL	196	330	1061	1889	1241	31
R	D	7	42	397	154	5	2
	S	12	75	700	355	462	3
	FI	8	14	258	78	125	3
	FR	6	28	285	92	173	4
	C	16	29	329	107	147	0
	G	21	34	405	81	121	4
	ALL	70	222	2372	866	1187	16
S	D	70	96	25	256	312	1
	S	193	180	46	410	775	3
	FI	46	48	26	126	237	0
	FR	63	58	13	143	307	2
	C	74	85	23	172	276	1
	G	86	63	28	174	311	2
	ALL	537	530	161	1278	2216	9
T	D	48	181	421	84	25	0
	S	157	380	819	181	67	6
	FI	51	116	218	68	30	0
	FR	63	142	285	61	33	1
	C	68	113	260	104	85	0
	G	60	143	274	85	100	3
	ALL	448	1074	2277	578	340	10

Table D.1—continued

Item	Site[a]	Responses[b] 1	2	3	4	5	Missing
V	D	50	77	132	347	153	0
	S	64	109	246	774	411	5
	FI	29	32	68	199	155	2
	FR	20	33	82	278	171	4
	C	52	61	131	242	144	4
	G	57	53	191	232	129	5
	ALL	272	365	849	2069	1162	20
W	D	15	19	179	375	171	0
	S	9	40	332	698	525	2
	FI	7	10	137	164	165	2
	FR	9	8	148	230	189	0
	C	14	19	146	247	204	4
	G	23	20	182	252	185	4
	ALL	77	116	1122	1964	1438	12
X	D	80	155	56	342	126	0
	S	177	333	121	652	321	4
	FI	65	96	27	164	131	0
	FR	75	103	28	245	133	3
	C	41	90	33	240	226	4
	G	49	82	36	246	249	3
	ALL	486	855	302	1888	1186	14
Y	D	56	105	90	394	114	2
	S	148	224	172	734	326	4
	FI	29	45	60	226	123	1
	FR	39	69	64	283	129	3
	C	64	62	75	261	168	3
	G	54	70	79	281	178	5
	ALL	389	575	537	2178	1038	18
Z	D	20	116	18	338	267	1
	S	47	154	55	602	746	5
	FI	16	39	15	157	256	2
	FR	13	38	18	183	332	1
	C	45	113	17	236	219	3
	G	45	123	19	242	233	7
	ALL	186	582	141	1757	2051	19
BB	D	96	126	30	223	284	2
	S	270	237	93	383	621	6
	FI	59	63	26	119	216	2
	FR	100	72	21	130	261	1
	C	117	90	27	157	239	3
	G	112	86	45	156	263	13
	ALL	754	674	242	1166	1881	27

Table D.1—continued

Item	Site[a]	Response[b] 1	2	3	4	5	Missing
CC	D	11	82	103	428	135	1
	S	33	202	153	868	348	1
	FI	18	66	42	259	98	1
	FR	13	69	56	332	114	1
	C	23	78	57	280	192	1
	G	20	74	54	327	187	1
	ALL	118	571	464	2492	1072	6
DD	D	10	20	104	177	448	1
	S	11	30	259	274	1030	4
	FI	5	13	81	70	314	2
	FR	8	8	74	87	407	2
	C	8	17	112	112	381	2
	G	26	15	121	128	372	4
	ALL	68	103	750	848	2948	15
FF	D	37	115	57	382	168	0
	S	63	209	142	786	404	2
	FI	20	58	26	188	191	0
	FR	25	83	41	256	179	1
	C	50	96	43	271	170	1
	G	50	94	40	296	182	1
	ALL	245	655	347	2178	1292	5
3	D	12	81	454	212	--	1
	S	10	121	906	567	--	6
	M	4	83	556	420	--	1
	SC	31	233	694	335	--	6
	ALL	57	517	2609	1534	--	14
5	D	238	401	98	22	--	0
	S	576	787	209	32	--	3
	M	403	500	132	29	--	0
	SC	31	233	694	335	--	7
	ALL	1620	2292	666	139	--	10
6	D	366	296	68	29	--	2
	S	742	648	173	41	--	5
	M	517	369	111	34	--	1
	SC	542	531	134	85	--	7
	ALL	2184	1851	490	192	--	15

[a]D=Dayton; S=Seattle experimental; FI=Fitchburg, Mass.; FR= Franklin County, Mass.; C=Charleston, South Carolina, Georgetown County, South Carolina; M=Massachusetts sites combined; SC=South Carolina sites combined; ALL=all sites combined.

[b]Responses have been recoded as shown in the item scoring table (B.1); only 4 response choices were provided for item nos. 3, 5, and 6. (See App. A).

Table D.2

Means and Standard Deviations (in Parentheses) for General Health Rating Items, by Site

	Abbreviated Item Content	Sites[a] D	S	FI	FR	C	G
A	According to doctors, my health is excellent	3.85 (0.94)	4.01 (0.92)	4.03 (0.96)	4.17 (0.85)	3.82 (1.04)	3.65 (1.05)
C	Get sick easier than others	4.08 (1.02)	4.30 (0.99)	4.29 (1.05)	4.32 (1.04)	4.10 (1.13)	4.06 (1.12)
D	Feel better now than ever	3.21 (1.12)	3.23 (1.12)	3.46 (1.15)	3.35 (1.10)	3.25 (1.27)	3.30 (1.24)
E	Probably sick at lot in future	3.75 (0.94)	4.07 (0.92)	3.88 (1.01)	3.99 (0.93)	3.64 (0.97)	3.47 (0.96)
F	Never worry about my health	2.72 (1.20)	2.71 (1.25)	2.77 (1.36)	2.70 (1.35)	2.98 (1.36)	2.95 (1.35)
G	People get sick easier than I	3.36 (0.78)	3.52 (0.91)	3.40 (0.92)	3.42 (0.98)	3.40 (0.95)	3.41 (0.96)
I	I am somewhat ill	4.19 (1.06)	4.37 (0.98)	4.41 (0.96)	4.46 (0.97)	4.15 (1.14)	4.04 (1.17)
J	Better health in future than others	3.39 (0.80)	3.56 (0.85)	3.38 (0.91)	3.43 (0.86)	3.40 (0.95)	3.31 (0.93)
K	So sick once thought I'd die	4.10 (1.35)	4.07 (1.40)	4.10 (1.40)	4.12 (1.38)	3.91 (1.49)	3.94 (1.49)
L	Not as healthy new as before	3.30 (1.28)	3.42 (1.34)	3.65 (1.35)	3.62 (1.30)	3.26 (1.40)	3.15 (1.42)
M	Worry about health more than others	2.21 (1.00)	2.07 (1.01)	2.16 (1.15)	2.12 (1.09)	2.40 (1.21)	2.40 (1.18)
O	My body resists illness well	3.75 (0.85)	3.85 (0.88)	3.81 (0.93)	3.90 (0.86)	3.72 (1.02)	3.76 (0.97)
P	Getting sick part of my life	3.06 (1.24)	2.97 (1.26)	3.22 (1.28)	3.06 (1.32)	3.25 (1.33)	3.18 (1.31)
Q	As healthy as anybody I know	3.75 (1.01)	3.88 (0.97)	3.96 (1.02)	3.88 (0.98)	3.63 (1.14)	3.54 (1.13)
R	Health worse in future than now	3.55 (0.91)	3.74 (0.95)	3.62 (0.95)	3.68 (0.98)	3.55 (0.98)	3.37 (0.94)

Table D.2—continued

	Abbreviated Item Content	Sites[a] D	S	FI	FR	C	G
S	Never had a long illness	3.85 (1.33)	3.87 (1.42)	3.95 (1.34)	3.98 (1.38)	3.78 (1.42)	3.85 (1.43)
T	Others more concerned about health than I	2.81 (0.84)	2.76 (0.92)	2.81 (1.01)	2.76 (0.97)	3.04 (1.15)	3.03 (1.14)
V	My health is excellent	3.63 (1.11)	3.85 (1.01)	3.87 (1.11)	3.94 (0 98)	3.58 (1.18)	3.49 (1.14)
W	Expect a healthy life	3.88 (0.85)	4.05 (0.83)	3.97 (0.92)	4.00 (0.88)	3.97 (0.93)	3.84 (0.98)
X	My health a concern	3.37 (1.27)	3.38 (1.31)	3.41 (1.41)	3.44 (1.35)	3.83 (1.24)	3.85 (1.25)
Y	Accept that sometimes sick	3.53 (1.27)	3.54 (1.22)	3.76 (1.11)	3.67 (1.14)	3.65 (1.25)	3.69 (1.20)
Z	Been feeling bad lately	3.94 (1.10)	4.15 (1.06)	4.24 (1.06)	4.34 (0.97)	3.75 (1.29)	3.75 (1.29)
BB	Never been seriously ill	3.62 (1.44)	3.53 (1.53)	3.77 (1.44)	3.65 (1.55)	3.49 (1.55)	3.57 (1.52)
CC	Usually catch what is going around	3.78 (0.91)	3.81 (0.98)	3.73 (1.05)	3.80 (0.96)	3.86 (1.09)	3.89 (1.03)
DD	Doctors say I have poor health now	4.36 (0.91)	4.42 (0.87)	4.40 (0.93)	4.50 (0.87)	4.33 (0.94)	4.22 (1.07)
FF	Feel as good now as ever have	3.70 (1.12)	3.79 (1.08)	3.98 (1.14)	3.82 (1.14)	3.66 (1.24)	3.70 (1.22)
3	EGFP	34.54 (4.72)	35.42 (4.20)	35.67[b] (4.22)		33.66[c] (4.22)	
5	PAIN	1.87 (0.74)	1.82 (0.75)	1.80[b] (0.76)		1.97[c] (0.86)	
6	Worry or concern about health	1.87 (1.18)	1.71 (0.80)	1.67[b] (0.80)		1.84[c] (0.92)	

[a]D=Dayton; S=Seattle experimentals; FI=Fitchburg, Mass.; FR=Franklin County, Mass.; C=Charleston, South Carolina; G=Georgetown County, South Carolina.

[b]Means and standard deviations for item nos. 3, 5, and 6 for combined Massachusetts samples.

[c]Means and standard deviations for item nos. 3, 5, and 6 for combined South Carolina samples.

Table D.3

Item-Scale Correlation Matrix, General Health Perceptions Measures, Dayton (N = 759)

Item Grouping/Item		Scale CH	PH	HO	RE	WC	SO
	Current Health (CH)						
V	Health is excellent	78*	46	60	46	-43	-19
FF	Feel as good now as ever	77*	35	50	37	-42	-15
A	Doctors say health is now excellent	71*	40	51	35	-54	-15
Z	Been feeling bad lately	67*	36	40	43	-38	-21
I	Somewhat ill	66*	41	48	41	-39	-20
L	Not as healthy now as used to be	65*	27	48	26	-39	-15
D	Feel better now than ever	62*	22	40	29	-29	-08
Q	Healthy as anybody I know	62*	38	54	45	-39	-12
DD	Doctors say now in poor health	59*	36	44	32	-29	-08
	Prior Health (PH)						
BB	Never been seriously ill	35	63*	26	27	-23	-05
S	Never had a long illness	37	55*	27	29	-30	-11
K	So sick once thought I might die	41	47*	31	29	-26	-07
	Health Outlook (HO)						
R	Health worse in future than now	47	22	65*	31	-30	-17
E	Probably sick a lot in future	48	25	59*	42	-35	-25
W	Expect a healthy life	66	37	58*	38	-34	-20
J	Expect to have better health than others	38	24	50*	36	-23	-19
	Resistance (RE)						
O	Body resists illness	48	30	38	59*	-35	-18
C	Get sick easier than others	44	31	40	55*	-30	-24
G	People get sick easier than I do	29	24	39	47*	-28	-19
CC	Something going around, catch it	25	19	24	47*	-18	-25
	Health Worry/Concern (WC)						
I6	Worry or concern about health	-61	-40	-36	-32	36*	14
X	Health is a concern	-20	-11	-19	-12	34*	16
F	Never worry about my health	-39	-26	-28	-25	31*	13
T	Others more concerned about health	-11	-09	-15	-27	24*	12
M	Worry about health more than others	-36	-24	-35	-30	23*	15
	Sickness Orientation (SO)						
P	Getting sick part of life	-20	-12	-20	-27	18	43*
Y	Accept that sometimes I'm sick	-13	-03	-21	-23	18	43*

Note: Standard error of the correlation equaled 0.04. '*' denotes hypothesized item groupings and correlations that were corrected for overlap.

Table D.4

ITEM-SCALE CORRELATION MATRIX, GENERAL HEALTH PERCEPTIONS MEASURES, SEATTLE (N = 1604)

	Item Grouping/Item	Scale CH	PH	HO	RE	WC	SO
	Current Health (CH)						
V	Health is excellent	74*	31	52	46	-38	-13
FF	Feel as good now as ever	71*	24	41	33	-35	-09
A	Doctors say health is now excellent	62*	25	40	28	-28	-08
L	Not as healthy now as used to be	62*	18	41	27	-32	-07
Q	Healthy as anybody I know	61*	27	47	42	-34	-06
I	Somewhat ill	60*	29	40	38	-31	-15
Z	Been feeling bad lately	60*	24	36	39	-34	-17
D	Feel better now than ever	57*	12	34	24	-26	-06
DD	Doctors say now in poor health	51*	24	32	28	-21	-08
	Prior Health (PH)						
BB	Never been seriously ill	25	55*	17	18	-18	-03
S	Never had a long illness	25	44*	19	23	-20	-10
K	So sick once thought I might die	25	39*	14	18	-16	-06
	Health Outlook (HO)						
R	Health worse in future than now	40	10	58*	27	-20	-13
W	Expect a healthy life	60	26	54*	38	-30	-09
E	Probably sick a lot in future	41	18	52*	41	-31	-16
J	Expect to have better health than others	29	11	46*	35	-17	-10
	Resistance (RE)						
O	Body resists illness	45	23	43	65*	-33	-20
CC	Something going around, catch it	31	13	32	57*	-28	-32
C	Get sick easier than others	42	28	37	56*	-28	-22
G	People get sick easier than I do	27	14	34	51*	-26	-20
	Health Worry/Concern (WC)						
M	Worry about health more than others	-37	-17	-36	-33	44*	17
F	Never worry about my health	-32	-15	-20	-23	38*	13
X	Health is a concern	-14	-11	-10	-14	34*	15
T	Others more concerned about health	-09	-08	-14	-22	32*	12
I6	Worry or concern about health	-55	-26	-27	-31	32*	11
	Sickness Orientation (SO)						
P	Getting sick part of life	-17	-10	-18	-33	17	39*
Y	Accept that sometimes I'm sick	-06	-04	-09	-19	18	39*

Note: Standard error of the correlation equaled 0.02. '*' denotes hypothesized item groupings and correlations that were corrected for overlap.

Table D.5

Item-Scale Correlation Matrix, General Health Perceptions Measures, Massachusetts (N = 1067)

	Item Grouping/Item	Scale CH	PH	HO	RE	WC	SO
	Current Health (CH)						
V	Health is excellent	73*	38	51	44	-40	-10
FF	Feel as good now as ever	71*	28	42	30	-33	-09
L	Not as healthy now as used to be	65*	29	37	36	-35	-11
Z	Been feeling bad lately	65*	32	31	35	-36	-18
I	Somewhat ill	63*	36	34	32	-31	-16
A	Doctors say health is now excellent	62*	27	35	26	-24	-06
Q	Healthy as anybody I know	61*	32	48	45	-40	-08
D	Feel better now than ever	57*	18	35	24	-29	-10
DD	Doctors say now in poor health	53*	25	27	28	-21	-08
	Prior Health (PH)						
BB	Never been seriously ill	27	54*	18	16	-19	-06
S	Never had a long illness	36	43*	24	28	-27	-12
K	So sick once thought I might die	31	42*	18	19	-22	-08
	Health Outlook (HO)						
R	Health worse in future than now	35	18	61*	23	-23	-12
W	Expect a healthy life	57	27	58*	37	-29	-09
E	Probably sick a lot in future	41	22	54*	39	-26	-18
J	Expect to have better health than others	25	11	40*	28	-16	-02
	Resistance (RE)						
O	Body resists illness	43	24	41	58*	-33	-19
CC	Something going around, catch it	23	13	23	50*	-17	-24
C	Get sick easier than others	42	29	31	48*	-29	-19
G	People get sick easier than I do	19	12	28	35*	-24	-10
	Health Worry/Concern (WC)						
M	Worry about health more than others	-35	-26	-32	-30	50*	08
F	Never worry about my health	-35	-19	-23	-26	44*	10
X	Health is a concern	-19	-14	-11	-13	38*	09
T	Others more concerned about health	-06	-10	-16	-23	35*	-02
I6	Worry or concern about health	-58	-31	-23	-27	33*	13
	Sickness Orientation (SO)						
P	Getting sick part of life	-17	-14	-15	-29	13	29*
Y	Accept that sometimes I'm sick	-06	-03	-07	-10	04	29*

Note: Standard error of the correlation equaled 0.03. '*' denotes hypothesized item groupings and correlations that were corrected for overlap.

Table D.6

Item-Scale Correlation Matrix, General Health Perceptions Measures, South Carolina (N = 1292)

	Item Grouping/Item	Scale CH	PH	HO	RE	WC	SO
	Current Health (CH)						
V	Health is excellent	71*	36	51	42	-40	-14
FF	Feel as good now as ever	67*	28	37	30	-27	-06
A	Doctors say health is now excellent	62*	32	42	35	-32	-11
Q	Healthy as anybody I know	62*	35	50	43	-35	-07
Z	Been feeling bad lately	59*	27	37	35	-31	-26
I	Somewhat ill	58*	37	42	37	-31	-18
L	Not as healthy now as used to be	56*	25	36	23	-28	-13
D	Feel better now than ever	55*	18	32	26	-19	-01
DD	Doctors say now in poor health	50*	28	35	33	-21	-16
	Prior Health (PH)						
BB	Never been seriously ill	27	50*	15	19	-18	-05
S	Never had a long illness	35	41*	23	29	-25	-08
K	So sick once thought I might die	33	36*	23	23	-17	-12
	Health Outlook (HO)						
R	Health worse in future than now	36	16	51*	26	-15	-22
W	Expect a healthy life	58	31	47*	41	-24	-11
E	Probably sick a lot in future	38	20	45*	29	-24	-25
J	Expect to have better health than others	31	09	38*	30	-17	-07
	Resistance (RE)						
O	Body resists illness	40	26	36	46*	-32	-06
C	Get sick easier than others	39	30	32	42*	-28	-24
CC	Something going around, catch it	26	15	22	38*	-12	-20
G	People get sick easier than I	25	14	32	33*	-24	-02
	Health Worry/Concern (WC)						
M	Worry about health more than others	-29	-22	-26	-31	36*	10
F	Never worry about my health	-31	-16	-22	-24	34*	09
T	Others more concerned about health	-10	-08	-08	-18	33*	-01
I6	Worry or concern about health	-58	-29	-26	-29	31*	14
X	Health is a concern	-11	-09	-07	-10	29*	17
	Sickness Orientation (SO)						
P	Getting sick part of life	-17	-14	-21	-23	14	35*
Y	Accept that sometimes I'm sick	-12	-04	-17	-09	11	35*

Note: Standard error of the correlation equaled 0.03. '*' denotes item-total correlations that were corrected for overlap.

Table D.7

CORRELATIONS BETWEEN 26 GENERAL HEALTH PERCEPTIONS ITEMS AND THE FIRST UNROTATED COMPONENT, ALL SITES COMBINED AND BY SITE

Item/Scale	All Sites Combined	Dayton	Seattle	Massachusetts	South Carolina
V CH	80	82	79	80	78
Q CH	72	72	69	72	72
W HO	70	74	70	68	68
FF CH	69	76	70	70	66
I CH	-68	-73	-66	-67	-66
A CH	67	74	64	62	68
Z CH	-67	-70	-64	-67	-64
L CH	-62	-65	-62	-65	-58
O HO	60	62	64	60	55
DD CH	-58	-65	-55	-56	-57
E HO	-58	-62	-58	-56	-51
C RE	-57	-60	-59	-56	-53
D CH	56	60	55	58	55
R HO	-52	-59	-51	-49	-47
M WC	-49	-49	-51	-50	-42
S PH	44	49	36	47	47
J HO	44	52	45	37	42
F WC	42	46	40	45	40
G RE	41	46	46	34	39
K PH	-40	-52	-32	-40	-41
CC RE	-40	-40	-49	-38	-37
BB PH	37	46	34	37	37
P SO	-28	-29	-29	-27	-27
X WC	-24	-25	-22	-27	-17
T WC	22	24	22	20	19
Y SO	-15	-19	-16	-09	-17
Percent of Variance Explained	28.6	33.2	28.1	28.0	26.4

NOTE: Decimals have been omitted from correlations; CH=Current Health; PH=Poor Health; HO=Health Outlook; RE=Resistance; WC=Health Worry/Concern; SO=Sickness Orientation.

Table D.8

Correlations Between 29 General Health Perceptions Items, All Sites Combined

	A	D	I	L	Q	V	Z	DD	FF	K	S	BB	E	J	R	W	C	G	O	CC	F	M	6[a]	T	X	P	Y	5[a]	3[a]
A																													
D	.39																												
I	-.44	-.34																											
L	-.41	-.46	.42																										
Q	.47	.39	-.42	-.43																									
W	.65	.44	-.52	-.50	.62																								
Z	-.40	-.40	.54	.43	-.40	-.50																							
DD	-.51	-.27	.42	.35	-.36	-.45	.40																						
FF	.45	.57	-.45	-.54	.51	.57	-.54	-.36																					
K	-.22	-.12	.30	.23	-.22	-.27	.26	.24	-.20																				
S	.26	.16	-.28	-.16	.29	.31	-.22	-.21	.24	-.28																			
BB	.22	.11	-.23	-.16	.24	.27	-.18	-.19	.20	-.41	.46																		
E	-.29	-.20	.37	.30	-.36	-.36	.32	.29	-.28	.18	-.17	-.14																	
J	.26	.23	-.17	-.16	.31	.32	-.15	-.14	.22	-.06	.14	.09	-.28																
R	-.25	-.23	.29	.35	-.32	-.32	.25	.24	-.28	.14	-.10	-.11	.54	-.34															
W	.47	.35	-.39	-.38	.50	.60	-.36	-.36	.45	-.21	.26	.21	-.39	.43	-.43														
C	-.27	-.17	.39	.27	-.36	-.37	.36	.29	-.27	.26	-.24	-.19	.40	-.14	.23	-.27													
G	.19	.17	-.16	-.10	.28	.26	-.15	-.12	.18	-.08	.16	.10	-.19	.38	-.16	.27	-.31												
O	.33	.25	-.31	-.21	.43	.44	-.31	-.25	.32	-.15	.26	.18	-.27	.31	-.21	.41	-.43	.41											
CC	-.14	-.14	.20	.14	-.21	-.23	.26	.22	-.17	.12	-.13	-.08	.22	-.13	.18	-.20	.40	-.27	-.42										
F	.24	.25	-.20	-.23	.28	.31	-.25	-.15	.27	-.09	.19	.14	-.16	.16	-.13	.24	-.16	.19	.26	-.12									
M	-.22	-.16	.28	.28	-.30	-.28	.27	.22	-.23	.18	-.18	-.14	.34	-.12	.26	-.26	.33	-.16	-.26	.17	-.32								
6[a]	.40	.34	-.45	-.42	.39	.48	-.46	-.32	.43	-.26	.24	.21	-.22	.12	-.17	.30	-.28	.13	.27	-.17	.32	-.28							
T	.09	.05	-.05	-.03	.15	.14	-.07	.00	.08	-.02	.13	.06	-.10	.20	-.03	.11	-.09	.28	.22	-.07	.25	-.30	.13						
X	-.10	-.03	.14	.15	-.14	-.15	.15	.10	-.11	.10	-.08	-.07	.18	-.02	.11	-.05	.13	-.07	-.10	.07	-.23	.31	-.18	-.25					
P	-.11	-.06	.19	.11	-.10	-.15	.21	.12	-.10	.11	-.12	-.06	.22	-.08	.16	-.12	.26	-.12	-.17	.25	-.09	.15	-.13	-.03	.12				
Y	-.05	-.03	.08	.06	-.03	-.07	.13	.05	-.04	.03	-.04	-.01	.13	-.07	.11	-.06	.10	-.08	-.08	.16	-.09	.07	-.08	-.04	.13	.36			
5[a]	.32	.28	-.36	-.33	.29	.40	-.41	-.24	.35	-.20	.20	.18	-.20	.12	-.16	.26	-.24	.10	.21	-.14	.19	-.16	.55	.06	-.09	-.12	-.10		
3[a]	-.55	-.33	.48	.44	-.48	-.64	.43	.40	-.44	.28	-.25	-.24	.33	-.23	.26	-.45	.33	-.18	-.31	.16	-.22	.23	-.48	-.09	.13	.12	.04	-.42	

[a]Correlations with item nos. 3, 5, and 6 were not included in the matrix we subjected to principal components analysis; see Table 4 or Appendix A for item content.

Table D.9

Correlations Between 17 General Health Perceptions Items (Roughly Equating Items per Construct) and the First Unrotated Component, All Sites Combined

Item/Scale	Correlation
V CH	77
Q CH	72
W HO	71
I CH	-65
O RE	64
C RE	-61
E HO	-61
L CH	-58
R HO	-55
M WC	-51
J HO	48
S PH	46
G RE	-43
K PH	-42
F WC	42
BB PH	40

NOTE: Decimals have been omitted; CH = Current Health; PH = Prior Health; HO = Health Outlook; RE = Resistance; WC = Health Worry/Concern.

Table D.10

Descriptive Statistics, General Health Perceptions Scales, by Site

Measure	No. of Items	Highest Possible Score	Midpoint	Dayton Mean	Dayton S.D.	Seattle Mean	Seattle S.D.	Massachusetts Mean	Massachusetts S.D.	South Carolina Mean	South Carolina S.D.
Current Health	9	45	27	33.88	7.26	35.11	6.66	36.04	6.80	33.14	7.47
EGFP	1	45	27	34.54	4.72	35.42	4.20	35.67	4.22	33.66	5.42
Prior Health	3	15	9	11.57	3.31	11.47	3.33	11.78	3.27	11.26	3.35
Health Outlook	4	20	12	14.57	2.72	15.42	2.65	14.98	2.78	14.28	2.72
Resistance	4	20	12	14.98	2.65	15.48	2.89	15.35	2.81	15.10	2.83
Health Worry/Concern	3	15	9	12.78	3.32	12.61	3.33	12.74	3.65	14.03	3.60
Pain	1	4	2.5	1.87	0.74	1.81	0.73	1.80	0.76	1.95	0.81
Sickness Orientation	2	10	6	6.60	2.00	6.51	2.06	6.85	1.97	6.79	2.23
General Health Ratings Index	22	110	66	82.07	13,79	84.71	12.76	85.29	13.06	80.43	13.84

[a]For all but EGFP, the lowest possible score equals the number of items in the measure because EGFP was transformed, its lowest possible score was 15 (see App. B).

Table D.11

Homogeneity (r_{ii}) and Reliability (r_{tt}) Estimates for General Health Perceptions Measures, by Site

Scale	No. of Items	Dayton		Seattle		Massachusetts		South Carolina	
		r_{ii}	r_{tt}	r_{ii}	r_{tt}	r_{ii}	r_{tt}	r_{ii}	r_{tt}
Current Health	9	.51	.90	.44	.87	.45	.88	.42	.86
Prior Health	3	.47	.73	.38	.64	.38	.65	.35	.61
Health Outlook	4	.46	.78	.41	.73	.41	.74	.34	.67
Resistance	4	.40	.72	.45	.77	.36	.69	.29	.62
Health Worry/Concern	5	.27	.65	.25	.62	.27	.65	.22	.58
Sickness Orientation	2	.43	.60	.39	.56	.28	.44	.35	.52
General Health Ratings Index	22	.33	.91	.28	.88	.28	.89	.27	.88

Table D.12

Reliability of General Health Perceptions Measures in Different Education Groups, Seattle (N = 1394)

Scale	Years of Schooling Completed					
	0-9 (N=57)	10-11 (N=167)	12 (N=535)	13-15 (N=343)	16 (N=177)	17+ (N=115)
Current Health	.91	.89	.88	.87	.86	.82
Prior Health	.46	.65	.67	.61	.72	.73
Health Outlook	.79	.75	.72	.74	.72	.72
Resistance	.47	.77	.77	.80	.80	.79
Health Worry/Concern	.59	.55	.61	.64	.58	.66
Sickness Orientation	.66	.43	.58	.55	.61	.59
General Health Index	.90	.88	.89	.88	.89	.87

NOTE: Samples include only persons ages 17 and older for whom data on education were obtained and who provided complete data on all items in these scales.

Table D.13

Reliability of General Health Perceptions Measures in Different Education Groups, Massachusetts (N = 871)

Scale	Years of Schooling Completed					
	0-9 (N=89)	10-11 (N=111)	12 (N=360)	13-15 (N=169)	16 (N=92)	17+ (N=50)
Current Health	.91	.91	.88	.88	.87	.85
Prior Health	.67	.62	.65	.74	.69	.73
Health Outlook	.66	.66	.78	.78	.77	.79
Resistance	.43	.75	.70	.73	.66	.71
Health Worry/Concern	.67	.65	.62	.68	.70	.74
Sickness Orientation	.50	.42	.48	.54	.29	.53
General Health Index	.87	.90	.88	.90	.85	.87

NOTE: Samples include only persons ages 17 and older for whom data on education were obtained and who provided complete data on all items in these scales.

Table D.14

Reliability of General Health Perceptions Measures in Different Education Groups, South Carolina (N = 981)

Scale	Years of Schooling Completed					
	0-9 (N=225)	10-11 (N=173)	12 (N=335)	13-15 (N=136)	16 (N=66)	17+ (N=46)
Current Health	.87	.86	.86	.87	.90	.87
Prior Health	.62	.62	.62	.62	.64	.70
Health Outlook	.62	.63	.66	.74	.84	.80
Resistance	.62	.67	.60	.71	.82	.82
Health Worry/Concern	.60	.52	.61	.64	.75	.61
Sickness Orientation	.40	.35	.53	.63	.53	.66
General Health Index	.89	.87	.87	.88	.93	.90

NOTE: Samples include only persons ages 17 and older for whom data on education were obtained and who provided complete data on all items in these scales.

Table D.15

Correlations Between General Health Measures and Sociodemographic Variables, Dayton

Measure	Sex	Race	Education	Income
Current Health	-.08	-.05	.23	.13
EGFP	-.07	-.10	.29	.16
Prior Health	-.04	-.01	.11	.07
Health Outlook	-.01	-.05	.24	.04
Resistance	-.12	-.06	.14	.05
Health Worry/Concern	-.02	.06	-.06	-.10
Pain	.13	.02	-.09	-.10
Sickness Orientation	.02	-.04	.03	.06
General Health Index	-.08	-.05	.23	.11

Table D.16

Correlations Between General Health Measures and Sociodemographic Variables, Seattle

Measure	Sex	Race	Education	Income
Current Health	-.07	-.06	.10	.11
EGFP	-.15	-.08	.17	.13
Prior Health	-.03	.05	.04	.05
Health Outlook	-.02	-.06	.10	.02
Resistance	-.12	-.06	.07	.10
Health Worry/Concern	-.02	.04	-.10	-.09
Pain	.17	-.02	-.07	-.08
Sickness Orientation	.05	.02	-.01	-.09
General Health Index	-.08	-.04	.11	.10

Table D.17

CORRELATIONS BETWEEN GENERAL HEALTH MEASURES AND SOCIODEMOGRAPHIC VARIABLES, MASSACHUSETTS

Measure	Sex	Race	Education	Income
Current Health	-.04	-.04	.09	.12
EGFP	-.09	.00	.27	.14
Prior Health	-.02	-.05	.08	.08
Health Outlook	-.04	-.05	.12	.04
Resistance	-.11	-.05	.08	.09
Health Worry/Concern	-.02	.04	-.08	-.07
Pain	.13	.07	-.11	-.10
Sickness Orientation	.06	-.01	-.01	-.06
General Health Index	-.05	-.06	.12	.12

Table D.18

CORRELATIONS BETWEEN GENERAL HEALTH MEASURES AND SOCIODEMOGRAPHIC VARIABLES, SOUTH CAROLINA

Measure	Sex	Race	Education	Income
Current Health	-.12	-.04	.24	.06
EGFP	-.14	-.16	.35	.13
Prior Health	-.07	-.10	.22	.05
Health Outlook	-.04	-.09	.17	.06
Resistance	-.16	-.03	.13	.08
Health Worry/Concern	.11	.26	-.14	-.17
Pain	.20	.06	-.15	-.04
Sickness Orientation	.03	.03	-.11	-.07
General Health Index	-.13	-.10	.26	.09

Table D.19

MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) FOR GENERAL HEALTH PERCEPTIONS SCALES IN SIX EDUCATIONAL GROUPS, ALL SITES COMBINED

Scales	Educational Groups						F-ratios	
	<9 years (N=536)	9-11 years (N=585)	12 years (N=1455)	13-15 years (N=732)	16 years (N=389)	17+ years (N=246)	Group Differences	Deviation from Linearity
Current Health	31.72 (8.22)	33.78 (7.45)	34.68 (6.79)	35.56 (6.62)	35.96 (6.54)	36.36 (5.74)	28.56**	2.26
EGFP	31.88 (6.28)	34.39 (4.64)	34.82 (4.26)	35.76 (3.92)	36.38 (3.97)	37.21 (3.29)	76.01**	5.04**
Prior Health	10.58 (3.34)	11.03 (3.38)	11.64 (3.30)	11.69 (3.30)	12.26 (3.21)	11.67 (3.44)	15.78**	4.50**
Health Outlook	13.95 (2.79)	14.65 (2.75)	14.84 (2.65)	15.34 (2.71)	15.49 (2.79)	15.65 (2.72)	25.38**	1.31
Resistance	14.84 (2.78)	14.96 (3.00)	15.39 (2.76)	15.55 (2.81)	15.28 (2.76)	15.95 (2.65)	9.05**	1.14
Worry/Concern	14.02 (3.72)	13.60 (3.48)	12.98 (3.43)	12.63 (3.46)	12.64 (3.43)	11.99 (3.53)	19.05**	1.43
Pain	2.08 (0.88)	1.89 (0.80)	1.84 (0.75)	1.83 (0.74)	1.73 (0.67)	1.75 (0.64)	13.22**	2.36
Sickness Orientation	7.05 (2.02)	6.77 (2.04)	6.57 (2.06)	6.70 (2.08)	6.57 (1.95)	6.44 (2.14)	5.38**	2.28
Index	77.75 (14.54)	81.10 (13.72)	83.65 (12.67)	85.43 (12.84)	86.45 (12.85)	87.08 (12.02)	35.21**	2.26

*p<0.05, two tailed test

**p<0.01, two-tailed test

Table D.20

Means and Standard Deviations (in Parentheses) for General Health Perceptions Scales in Five Income Groups, All Sites Combined

Scales	Income Groups					F-ratios	
	≤$7000 (N=1090)	$7001-10500 (N=708)	$10501-14500 (N=921)	$14501-18500 (N=855)	$18501+ (N=843)	Group Differences	Deviation from Linearity
Current Health	33.32 (7.52)	35.08 (6.81)	34.44 (7.19)	35.33 (6.68)	35.67 (6.59)	17.12**	5.12**
EGFP	33.84 (5.11)	35.13 (4.60)	34.77 (4.61)	35.35 (4.36)	36.00 (3.95)	29.29**	4.83**
Prior Health	11.16 (3.29)	11.56 (3.36)	11.50 (3.33)	11.86 (3.23)	11.76 (3.30)	6.52**	2.16
Health Outlook	14.64 (2.75)	14.96 (2.73)	14.88 (2.71)	14.98 (2.74)	15.17 (2.77)	4.65**	0.80
Resistance	14.81 (3.05)	15.45 (2.76)	15.28 (2.79)	15.42 (2.60)	15.56 (2.76)	10.80**	4.12**
Worry/Concern	13.84 (3.65)	12.83 (3.64)	13.16 (3.45)	12.62 (3.33)	12.38 (3.29)	25.78**	6.23**
Pain	1.94 (0.79)	1.80 (0.76)	1.90 (0.78)	1.80 (0.72)	1.76 (0.70)	9.89**	4.51**
Sickness Orientation	6.93 (2.03)	6.74 (1.99)	6.75 (1.97)	6.56 (2.06)	6.44 (2.11)	8.30**	0.53
Index	80.64 (13.96)	84.14 (12.99)	83.13 (13.23)	84.82 (12.85)	85.41 (12.97)	19.76**	5.44**

** $p<0.01$, two-tailed test

Table D.21

MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) FOR GENERAL HEALTH PERCEPTIONS SCALES IN SIX AGE GROUPS, ALL SITES COMBINED

Scales	Age Groups (in Years)						F-ratio	
	14-20 (N=705)	21-25 (N=599)	26-30 (N=666)	31-35 (N=532)	36-45 (N=649)	46+ (N=745)	Group Difference	Deviation from Linearity
Current Health	36.67 (6.31)	35.02 (6.74)	35.25 (6.64)	35.59 (7.00)	34.44 (7.30)	32.86 (7.67)	16.07**	3.67**
EGFP	35.62 (4.22)	35.39 (4.05)	35.68 (4.22)	35.52 (4.51)	34.74 (4.71)	32.96 (5.54)	37.25**	8.82**
Prior Health	11.70 (2.98)	11.79 (3.07)	11.56 (3.39)	11.83 (3.26)	11.42 (3.46)	10.80 (3.58)	9.14**	2.81*
Health Outlook	14.96 (2.77)	15.25 (2.69)	15.34 (2.73)	15.27 (2.84)	14.84 (2.72)	14.17 (2.60)	18.07**	9.06**
Resistance	14.86 (3.02)	14.88 (3.18)	15.24 (3.02)	15.47 (2.74)	15.64 (2.61)	15.80 (2.46)	12.69**	1.27
Worry/Concern	13.63 (3.56)	13.45 (3.56)	13.00 (3.66)	12.80 (3.44)	12.78 (3.45)	12.87 (3.53)	6.83**	3.19*
Pain	1.73 (0.76)	1.81 (0.73)	1.89 (0.76)	1.80 (0.75)	1.87 (0.74)	1.99 (0.81)	9.85**	2.50*
Sickness Orientation	7.14 (1.94)	6.91 (1.95)	6.79 (1.99)	6.72 (2.07)	6.53 (2.06)	6.27 (2.16)	15.61**	0.26
Index	83.85 (12.61)	83.82 (12.76)	84.51 (13.18)	85.37 (13.35)	83.45 (13.78)	80.82 (13.75)	9.12**	5.50**

*$p<0.05$, two-tailed test
**$p<0.01$, two-tailed test

Table D.22

Product Moment (r, r^2) and Eta (η, η^2) Estimates of Association Between General Health Perceptions Scales and Education, Income, and Age

Scale	Education[a]				Income[b]				Age[c]			
	r	r^2	η	η^2	r	r^2	η	η^2	r	r^2	η	η^2
Current Health	.18	.03	.19	.04	.11	.01	.12	.02	-.13	.02	.14	.02
EGFP	.28	.08	.30	.09	.15	.02	.16	.02	-.19	.04	.21	.05
Prior Health	.12	.01	.14	.02	.07	.00	.08	.00	-.09	.01	.11	.01
Health Outlook	.17	.03	.18	.03	.06	.00	.06	.00	-.12	.01	.15	.02
Resistance	.10	.01	.11	.01	.08	.01	.10	.01	.12	.01	.13	.02
Worry/Concern	-.15	.02	.15	.02	-.14	.02	.15	.02	-.07	.00	.09	.01
Pain	-.11	.01	.13	.02	-.07	.00	.09	.01	.10	.01	.11	.01
Sickness Orientation	-.06	.00	.08	.01	-.08	.01	.09	.01	-.14	.02	.14	.02
General Health Ratings Index	.20	.04	.21	.04	.12	.01	.13	.02	-.08	.00	.11	.01

[a]Six education groups defined: ≤12 years, 12-15 years, 16 years, 17-19 years, 20 years, 21+ years.

[b]Five income groups defined: ≤$7000, $7001-10500, $10501-14500, $14501-18500, $10501+.

[c]Six age groups defined: 14-20, 21-25, 26-30, 31-35, 36-45, 46+.

Table D.23

Means and Standard Deviations for Indicators of Health Impairment, by Decile of Score Distribution on General Health Ratings Index

Decile	Index Mean	N	Any Chronic Disease		Any Physical Impairment		Any Emotional Impairment		Any Physical or Emotional Impairment		Any Impairment	
			Mean[a]	S.D.	Mean[a]	S.D.	Mean[a]	S.D.	Mean[a]	S.D.	Mean[a]	S.D.
1	51	160[b]	.59	.49	.65	.48	.61	.49	.87	.34	.92	.27
2	71	145[b]	.51	.50	.41	.49	.35	.48	.62	.49	.76	.42
3	77	172[b]	.35	.48	.36	.48	.23	.42	.51	.50	.67	.47
4	81	142[b]	.44	.50	.24	.43	.13	.34	.34	.48	.60	.49
5	84	145[b]	.36	.48	.24	.43	.17	.37	.35	.48	.56	.50
6	87	177[b]	.39	.49	.15	.36	.13	.34	.26	.44	.49	.50
7	90	155	.40	.49	.17	.37	.12	.33	.26	.44	.56	.50
8	94	166[b]	.35	.48	.11	.31	.07	.26	.18	.38	.45	.50
9	98	150	.30	.46	.09	.28	.06	.24	.15	.36	.37	.48
10	105	182[b]	.28	.45	.07	.26	.06	.23	.12	.32	.36	.48

[a]Mean can be interpreted as probability of given impairment for particular decile.

[b]Sample sizes in these deciles were one or two persons smaller for the Any Physical Impairment and Any Physical or Emotional Impairment variables because of missing data.

NOTE: Figs. 2 through 5 were drawn using the data in this table.

Table D.24

Correlations for Current Health and EGFP Items[a] with Physical and Mental Health Validity Variables and Age

Validity Variable	Items and Original Direction of Scoring									
	A (+)	D (+)	I (-)	L (-)	Q (+)	V (+)	Z (-)	DD (-)	FF (+)	EGFP (+)
PHYSICAL HEALTH										
Chronic Personal Limitations	-.30	-.24	.31	.30	-.29	-.34	.28	.26	-.30	-.38
Current Personal Limitations	-.29	-.24	.31	.28	-.28	-.34	.29	.26	-.29	-.37
Chronic Role Limitations	-.30	-.24	.31	.30	-.29	-.34	.28	.26	-.30	-.39
Current Role Limitations	-.29	-.24	.31	.28	-.28	-.34	.29	.15	-.29	-.37
Physical Capacities	.30	.24	-.27	-.29	.28	.33	-.24	-.27	.29	.39
Strenuous Exercise	.07	.10	-.09	-.11	.09	.08	-.05	-.03	.12	.08
General Exercise	.15	.14	-.14	-.16	.14	.16	-.10	-.11	.18	.11
Acute Symptoms	-.34	-.31	.39	.32	-.30	-.43	.46	.30	-.39	-.39
Chronic Disease (self-report)	-.33	-.26	.36	.34	-.32	-.39	.30	.26	-.34	-.42
MENTAL HEALTH										
Anxiety	-.23	-.28	.31	.26	-.25	-.32	.40	.23	-.34	-.27
Depression	-.21	-.26	.29	.20	-.26	-.30	.40	.23	-.32	-.28
Positive Well-Being	.26	.39	-.26	-.25	.26	.30	-.34	-.21	.35	.25
Emotional Ties	.15	.16	-.15	-.12	.18	.18	-.19	-.15	.19	.14
Emotional Stability	.23	.28	-.30	-.27	.24	.32	-.38	-.22	.33	.27
Mental Health Index	.28	.25	-.32	-.27	.29	.35	-.43	-.25	.38	.30
Age	-.07	-.08	.05	.17	-.08	-.08	-.06	-.01	-.03	-.18

[a]See Table 2 in text for abbreviated item content.

Table D.25

CORRELATIONS FOR CURRENT HEALTH ITEMS, EGFP, AND WORRY/CONCERN ITEMS WITH AGE IN NON-HIE AND HIE SAMPLES

Item and Direction of Scoring		Sample[a] SAC	ESC	LAC	FPC	HIE
Current Health						
A	(+)	-.26	-.11	-.17	-.22	-.07
D	(+)	-.24	-.23	-.16	-.11	-.08
I	(-)	.31	.22	.04	.17	.05
L	(-)	.34	.22	.09	.20	.17
Q	(+)	-.21	-.17	-.20	-.20	-.08
V	(+)	-.27	-.16	-.20	-.19	-.08
Z	(-)	.14	.14	-.02	.08	-.06
DD	(-)	.21	.17	.00	.20	-.01
FF	(+)	-.23	-.22	-.18	-.12	-.03
EGFP	(+)	(b)	(b)	(b)	-.26	-.18
Worry/Concern						
F	(+)	.02	.19	-.04	.03	.07
M	(-)	.04	-.01	-.10	-.11	-.07
T	(-)	.13	-.04	-.02	.12	.14
X	(+)	-.03	-.05	-.08	.00	-.08
I6	(+)	(b)	(b)	(b)	(b)	.05

[a]First four columns are samples studied by Ware and Karmos (1976). SAC=Sangamon County, median age 45, range 17-84; ESL=East St. Louis, median age 43, range 17-88; LAC=Los Angeles County, median age 43, range 18-92; FPC=Family Practice Center, median age 32, range 17-84; HIE=Health Insurance Experiment, median age 30.78, range 14-62.

[b]Not studied in that sample.

REFERENCES

Andersen, R., J. Kasper, M.R. Frankel, and Associates, *Total Survey Error,* Jossey-Bass, San Francisco, 1979.

Armor, D.J., "Theta Reliability and Factor Scaling," in H.L. Costner (ed.), *Sociological Methodology, 1973-1974,* Jossey-Bass, San Francisco, 1974, pp. 17-50.

Ashcraft, M.L., personal communication, 1980.

Becker, M.H. (ed.), "The Health Belief Model and Personal Health Behavior," *Health Education Monographs* 2:326-473, 1974.

Becker, M.H., D.P. Haefner, S.U. Kasl, J.P. Kirscht, L.A. Maiman, and I.M. Rosenstock, "Selected Psychosocial Models and Correlates of Individual Health-Related Behaviors," *Medical Care* 15 (Supplement to No. 5):27-46, 1977a.

Becker, M.H., L.A. Maiman, J.P. Kirscht, D.P. Haefner, and R.H. Drachman, "The Health Belief Model and Prediction of Dietary Compliance: A Field Experiment," *Journal of Health and Social Behavior* 18:348-366, 1977b.

Berki, S.E., and M.L. Ashcraft, "On the Analysis of Ambulatory Utilization: An Investigation of the Roles of Need, Access and Price as Predictors of Illness and Preventive Visits," *Medical Care* 17:1163-1181, 1979.

Berki, S.E., M. Ashcraft, R. Penchansky, and R. Fortus, "Health Concern, HMO Enrollment, and Preventive Care Use," *Journal of Community Health* 3:3-31, 1977a.

Berki, S.E., R. Penchansky, R.S. Fortus, et al., "Enrollment Choice in Different Types of HMOs: A Multivariate Analysis," *Medical Care* 16:682-697, 1977b.

Berkman, P., "Life Stress and Psychological Well-Being: A Reapplication of Langner's Analysis in the Midtown Manhattan Study," *Journal of Health and Social Behavior* 12:35-45, 1971.

Bradburn, N.M., S. Sudman and Associates, *Improving Interview Method and Questionnaire Design,* Jossey-Bass, San Francisco, 1979.

Brook, R.H., and multiple authors, *Conceptualization and Measurement of Physiologic Health for Adults,* The Rand Corporation, R-2262/1-17-HHS, 1980 and continuing.

Campbell, D.T., and D.W. Fiske, "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin* 56:81-105, 1959.

Cattell, R.B. (ed.), *Handbook of Multivariate Experimental Psychology,* Rand McNally and Company, Chicago, 1966.

Cloud, J., and G.M. Vaughn, "Using Balanced Scales to Control Acquiescence," *Sociometry* 33:193-202, 1970.

Comrey, A.L., *Comrey Personality Scales,* Testing Services, San Diego, 1970.

Cronbach, L.J., "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika* 16:-297-334, 1951.

Davies, A.R., and J.E. Ware, Jr., "Measuring Patient Satisfaction with Dental Care," *Social Science and Medicine,* 1981, in press.

DiCicco, L., and D. Apple, "Health Needs and Opinions of Older Adults," *Public Health Reports* 73:479-487, 1958.

Donald, C.A., and J.E. Ware, Jr., *Quantifying Social Activities and Resources to Measure Health and Social Well-Being,* The Rand Corporation, R-2738-HHS, forthcoming.

Donald, C.A., J.E. Ware, Jr., R.H. Brook, et al., *Conceptualization and Measurement of Health*

for Adults in the Health Insurance Study: Vol. IV, Social Health, The Rand Corporation, R-1987/4-HEW, 1978.

Edwards, A.L., *The Measurement of Personality Traits by Scales and Inventories,* Holt, Rinehart and Winston, Inc., New York, 1970.

Gove, W.R., and M.R. Geerken, "Response Bias in Surveys of Mental Health: An Empirical Investigation," *American Journal of Sociology* 82:1289-1316, 1977.

Guertin, W.H., and J.P. Bailey, Jr., *Introduction to Modern Factor Analysis,* Edwards Bothers, Inc., Ann Arbor, 1970.

Helmstadter, G.C., *Principles of Psychological Measurement,* Appleton-Century-Crofts, New York, 1964.

Howard, K.I., and G.G. Forehand, "A Method for Correcting Item-total Correlations for the Effect of Relevant Item Inclusion," *Educational and Psychological Measurement* 22:731-735, 1962.

Likert, R., "A Technique for the Measurement of Attitudes," *Archives of Psychology* 140:1-55, 1932.

MacPherson, A.S., "The Measurement of Mental Health in a General Population," Thesis, Master of Science in Medicine, McMaster University, Hamilton, Ontario, 1972.

Maiman, L.A., M.H. Becker, J.P. Kirscht, D.P. Haefner, and R.H. Drachman, "Scales for Measuring Health Belief Model Dimensions: A Test of Predictive Value, Internal Consistency, and Relationships Among Beliefs," *Health Education Monographs* 5:215-230, 1977.

Manning, W.G., Jr., J.P. Newhouse, and J.E. Ware, Jr., *The Status of Health in Demand Estimation: Beyond Excellent, Good, Fair, and Poor,* The Rand Corporation, R-2696-HHS, 1980.

Messick, S., "Separate Set and Content Scores for Personality and Attitude Scales," *Educational and Psychological Measurement* 21:915-923, 1961.

National Center for Health Statistics, *Health, United States, 1975,* DHEW Publ. No. (HRA) 76-1232, Health Resources Administration, Rockville, Md., 1976.

National Center for Health Statistics, *Health, 1976-1977,* DHEW Publ. No. (HRA) 77-1233, Health Resources Administration, Rockville, Md., 1977

National Center for Health Statistics, "Vital and Health Statistics," Series 10, Health Resources Administration, Rockville, Md., 1963 and continuing.

Nie, H.E., C.H. Hull, J.G. Jenkins, et al., *SPSS: Statistical Package for the Social Sciences,* 2d ed., McGraw-Hill Book Company, New York, 1975.

Nunnally, J.C., *Psychometric Theory,* 2d ed., McGraw-Hill Book Company, New York, 1978.

Pederson, K., personal communication, 1981.

Phillips, D.L., and K.J. Clancy, "Response Biases in Field Studies of Mental Illness," *American Sociological Review* 35:503-515, 1970.

Pratt, L., "The Relationship of Socioeconomic Status to Health," *American Journal of Public Health* 61:281, 1971.

Regier, D.A., I.D. Goldberg, and C.A. Taube, "The defacto U.S. Mental Health Services System: A Public Health Perspective," *Archives of General Psychiatry* 35:685-693, 1978.

Renne, K.S., "Measurement of Social Health in a General Population Survey," *Social Science Research* 3:25, 1974.

Rossiter, L.F., "Who Initiates Visits to a Physician," Data Preview 3, National Health Care Expenditures Study, DHHS Publication No. (PHS) 80-3278, National Center for Health Services Research, Hyattsville, Md., 1980.

Sechrest, I., "Incremental Validity," in D.N. Jackson and S. Messick (eds.), *Problems in Human Assessment,* McGraw-Hill Book Company, New York, 1967, pp. 368-371.

Smith, L.H., G.A. Goldberg, R.H. Brook, et al., *The Health Insurance Study Screening Examination Procedures Manual,* The Rand Corporation, R-2101-HEW, September 1978.

Stewart, A.L., J.E. Ware, Jr., and R.H. Brook, *Construction and Scoring of Aggregate Status Indexes: Volume I,* The Rand Corporation, R-2551-HHS, 1981.

Stewart, A.L., J.E. Ware, Jr., R.H. Brook, et al., *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. II, Physical Health in Terms of Functioning,* The Rand Corporation, R-1987/2-HEW, 1978.

Suchman, E.A., B.S. Phillips, and G.E. Streib, "An Analysis of the Validity of Health Questionnaires," *Social Forces* 26:223, 1958.

Thompson, W.E., and G.F. Streib, "Situational Determinants: Health and Economic Deprivation in Retirement," *Journal of Social Issues* 14:18-45, 1958.

Thorndike, R.M., *Correlational Procedures for Research,* Gardner Press, Inc., New York, 1978.

Veit, C., and J.E. Ware, Jr., *Refinements in the Measurement of Mental Health for Adults in the Health Insurance Study,* The Rand Corporation, R-2737-HHS, forthcoming.

Ware, J.E., Jr., "Scales for Measuring General Health Perceptions," *Health Services Research* 11:396-415, 1976.

Ware, J.E., Jr., "Effects of Acquiescent Response Set on Patient Satisfaction Ratings," *Medical Care* 16:327-336, 1978.

Ware, J.E., Jr., R.H. Brook, A.R. Davies, et al., "Choosing Measures of Health Status for Individuals in General Populations," *American Journal of Public Health* 71:620-625, 1981.

Ware, J.E., Jr., R.H. Brook, A. Davies-Avery, et al., *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. I, Model of Health and Methodology,* The Rand Corporation, R-1987/1-HEW, 1980.

Ware, J.E., Jr., A. Davies-Avery, and C.A. Donald, *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. V, General Health Perceptions,* The Rand Corporation, R-1987/5-HEW, 1978.

Ware, J.E., Jr., A. Davies-Avery, and R.H. Brook, *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. VI, Analyses of Relationships Among Health Status Measures,* The Rand Corporation, R-1987/6-HEW, 1980.

Ware, J.E., Jr., S.A. Johnston, A. Davies-Avery, et al., *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. III, Mental Health,* The Rand Corporation, R-1987/3-HEW, 1979.

Ware, J.E., Jr., and A.H. Karmos, *Development and Validation of Scales to Measure Perceived Health and Patient Role Propensity: Volume II of a Final Report,* National Technical Information Services, NTIS Publ. No. PB 288-331. Springfield, Va., 1976.

Ware, J.E., Jr., M.K. Snyder, and W.R. Wright, *Development and Validation of Scales to Measure Patient Satisfaction with Health Care Services: Volume I of a Final Report,* National Technical Information Services, NTIS Publ. Nos. PB 288-329/30, Springfield, Va., 1976.

Ware, J.E., Jr., W.R. Wright, and M.K. Snyder, *The Development and Validation of Scales to Measure Key Health Concepts, Methodological Appendix: Vol. IV of a Final Report,* Southern Illinois University School of Medicine, Carbondale, Ill., 1976.

Ware, J.E., Jr., and J. Young, *The Conceptualization and Measurement of Health as a Value: Volume III of a Final Report,* National Technical Information Services, NTIS Publ. No. PB 288-332, Springfield, Va., 1976.

Ware, J.E., Jr., and J. Young, "Issues in the Conceptualization and Measurement of Value Placed on Health," in S.J. Mushkin and D.W. Dunlop (eds.), *Health: What Is It Worth? Measures of Health Benefits,* Pergamon Press, New York, 1979, pp. 141-166.

Webster's New Collegiate Dictionary, G. & C. Merriam Company, Springfield, Mass., 1975.

Williams, A.W., J.E. Ware, Jr., and C.A. Donald, "A Model of Mental Health, Life Events and Social Supports Applicable to General Populations," *Journal of Health and Social Behavior,* 1981 (in press).