

The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection

Author(s): John E. Ware, Jr. and Cathy Donald Sherbourne

Reviewed work(s):

Source: *Medical Care*, Vol. 30, No. 6 (Jun., 1992), pp. 473-483

Published by: [Lippincott Williams & Wilkins](#)

Stable URL: <http://www.jstor.org/stable/3765916>

Accessed: 13/12/2012 06:27

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Lippincott Williams & Wilkins is collaborating with JSTOR to digitize, preserve and extend access to *Medical Care*.

<http://www.jstor.org>

The MOS 36-Item Short-Form Health Survey (SF-36)

I. Conceptual Framework and Item Selection

JOHN E. WARE, JR., PHD,* AND CATHY DONALD SHERBOURNE, PHD†

A 36-item short-form (SF-36) was constructed to survey health status in the Medical Outcomes Study. The SF-36 was designed for use in clinical practice and research, health policy evaluations, and general population surveys. The SF-36 includes one multi-item scale that assesses eight health concepts: 1) limitations in physical activities because of health problems; 2) limitations in social activities because of physical or emotional problems; 3) limitations in usual role activities because of physical health problems; 4) bodily pain; 5) general mental health (psychological distress and well-being); 6) limitations in usual role activities because of emotional problems; 7) vitality (energy and fatigue); and 8) general health perceptions. The survey was constructed for self-administration by persons 14 years of age and older, and for administration by a trained interviewer in person or by telephone. The history of the development of the SF-36, the origin of specific items, and the logic underlying their selection are summarized. The content and features of the SF-36 are compared with the 20-item Medical Outcomes Study short-form. Key words: SF-36; Medical Outcomes Study. (Med Care 1992; 30:473-483)

Among the most important health care developments made during the past 10 years is an increasing consensus regarding the cen-

trality of the patient's point of view in monitoring medical care outcomes.¹ Indeed, the goal of medical care for most patients today is to obtain a more "effective" life² and to preserve functioning and well-being.³⁻⁶ Although the patient is usually the best judge of whether these goals have been achieved, data concerning a patient's experiences of disease and treatment are not routinely collected. One reason for the lack of information is the lack of valid methods of data collection that are easy to use.

* From The Health Institute and International Resource Center for Health Care Assessment, New England Medical Center Hospitals, Boston, Massachusetts.

† From the Social Policy Department, RAND Corporation, Santa Monica, California.

Address correspondence to: John E. Ware, Jr., PhD, The Health Institute, NEMC-Box 345, 750 Washington Street, Boston, MA 02111.

Development of the SF-36 Short-Form Health Survey and preparation of this paper were supported by grants from the Henry J. Kaiser Family Foundation, Menlo Park, CA; National Institute on Aging, Bethesda, MD; and The Hartford Foundation of New York. The Medical Outcomes Study (MOS) is also sponsored by The Robert Wood Johnson Foundation, Princeton, NJ; The Pew Charitable Trusts, Philadelphia, PA; the Agency for Health Care Policy and Research, Rockville, MD; and the National Institute of Mental Health.

Scoring standardized responses to standardized questions is an efficient way to measure health status. Carefully constructed sets of survey questions have greatly assisted research efforts for the past 10 years.⁷⁻¹⁰ Among the surveys most useful with diverse groups and treatments are sur-

veys that address general health concepts not specific to any age, disease, or treatment group. These scales measure such basic human values as functioning and emotional well-being.¹¹ General health measures can be used in ways not possible with disease- or treatment-specific measures, including comparisons of the relative burden of different diseases and the relative benefits of different treatments. However, using general health measures on a large scale has not been practical because of their length.

One solution to this practical constraint is a standardized health status survey that is comprehensive, psychometrically sound, and brief. Such a survey can help fill the gap between lengthy health surveys used successfully in research projects¹²⁻¹⁴ and the relatively coarse single-item health measures used in national surveys and numerous clinical investigations.¹⁵⁻¹⁸

This study describes an improved 36-item short-form survey (SF-36) constructed for use in the Medical Outcomes Study (MOS). We summarize and briefly discuss: 1) background information, including the research that led to the development of the SF-36; 2) the conceptual framework underlying the health concepts represented in the SF-36; and 3) the logic and evidence for the selection of specific questionnaire items. Another article presents the results of preliminary psychometric tests of the validity of SF-36 scales as measures of physical and mental health.¹⁹

Selection and Origin of Items

A survey can be shortened by excluding some health concepts. However, minimum standards of comprehensiveness, i.e., content validity in relation to accepted definitions of health, require the representation of numerous health concepts. From these standards, the authors chose to represent the health concepts most frequently included in widely-used health surveys (physical, social and role functioning, mental health, and

general health perceptions) along with two additional concepts that are strongly supported by empirical study (i.e., bodily pain and vitality).¹⁴

During the 7-year period since the 18-item and 20-item MOS short-forms^{20,21} were first used, we have accumulated considerable experience with the tradeoffs involved in the construction of more efficient scales for measuring a core set of general health concepts worthy of inclusion in a short-form survey.¹¹ We also have identified strategies for evaluating and improving the precision of short-form scales used to measure these concepts.¹⁴ The result is the SF-36 survey described in this study.

As summarized in Table 1, SF-36 includes one multi-item scale measuring each of eight health concepts: 1) physical functioning; 2) role limitations because of physical health problems; 3) bodily pain; 4) social functioning; 5) general mental health (psychological distress and psychological well-being); 6) role limitations because of emotional problems; 7) vitality (energy/fatigue); and 8) general health perceptions. The content of SF-36 items selected to measure these concepts will be familiar to those who follow the health status assessment literature (See Appendix). Most of these items have been adapted from instruments that have been used for 20 to 40 years or longer. We reviewed the content of various source instruments used to measure limitations in physical, social, and role functioning^{14,22,23}; general mental health^{14,24}; and general health perceptions.^{14,25} In fact, it has been the accumulation of experience with these full-length scales that made it feasible to construct useful short-form health scales.

A major problem in the field is the absence of criteria for the construction and validation of health scales. In selecting items for each SF-36 scale, we used the corresponding full-length MOS scale as the criterion. Items in each SF-36 scale were selected to reproduce the "parent" scale as much as possible. Other psychometric standards also were con-

TABLE 1. Information About SF-36 Health Status Scales and the Interpretation of Low and High Scores

Concepts	No. of Items	No. of Levels	Meaning of Scores	
			Low	High
Physical functioning	10	21	Limited a lot in performing all physical activities including bathing or dressing	Performs all types of physical activities including the most vigorous without limitations due to health
Role limitations due to physical problems	4	5	Problems with work or other daily activities as a result of physical health	No problems with work or other daily activities as a result of physical health, past 4 weeks
Social Functioning	2	9	Extreme and frequent interference with normal social activities due to physical and emotional problems	Performs normal social activities without interference due to physical or emotional problems, past 4 weeks
Bodily pain	2	11	Very severe and extremely limiting pain	No pain or limitations due to pain, past 4 weeks
General mental health	5	26	Feelings of nervousness and depression all of the time	Feels peaceful, happy, and calm all of the time, past 4 weeks
Role limitations due to emotional problems	3	4	Problems with work or other daily activities as a result of emotional problems	No problems with work or other daily activities as a result of emotional problems, past 4 weeks
Vitality	4	21	Feels tired and worn out all of the time	Feels full of pep and energy all of the time, past 4 weeks
General health perceptions	5	21	Believes personal health is poor and likely to get worse	Believes personal health is excellent

sidered. Significantly more data were available for applying these strategies to construct the SF-36 than in the SF-20. The SF-36 and SF-20 forms are compared in Table 2 in terms of the numbers of items and scale levels for each concept. Specific strategies for constructing SF-36 scales, which varied across concepts, are summarized below.

Physical Functioning

Because of the importance of distinct aspects of physical functioning and the necessity of sampling a range of severe and minor physical limitations, the full-length (10-item) MOS Physical Functioning Scale

was adopted without modification. This scale reflects two important improvements over the SF-20. First, items were added to better represent levels and types of limitations between the extremes, including lifting and carrying groceries, climbing stairs, bending, kneeling, and walking moderate distances. Second, standardized response choices were revised to estimate the severity of each limitation, and thereby to increase score precision. This substantial departure from the SF-20 form was based on methodologic comparisons that showed gains in precision due to the distinction between those able to perform physical activities with and without difficulty.²⁶ SF-36 items cap-

TABLE 2. Comparison of Number of Items and Scale Levels for each Concept in the MOS SF-20 and SF-36 Health Surveys

Concept	SF-20		SF-36	
	No. Items	No. Levels	No. Items	No. Levels
Physical Functioning	6	7	10	21
Role Functioning	2	3	—	—
Role—Physical	—	—	4	5
Role—Emotional	—	—	3	4
Pain	1	6	2	12
Social Functioning	1	6	2	9
Mental Health	5	26	5	26
Energy/Fatigue	—	—	4	21
Current Health Perceptions	5	21	—	—
General Health Perceptions	—	—	5	21
Change in Health	—	—	1	5

ture both the presence and extent of physical limitations using a three-level response continuum. Thus, the number of scale levels defined was tripled relative to the number achieved by the SF-20 form (Table 2).

Role Functioning

The SF-20 role functioning scale was constructed from two widely used questions about health-related limitations in the type or amount of work.²⁰ The result was a rather coarse, three-level scale. The SF-36 includes a subset of the 11 role functioning items from MOS long-forms. They differ from the SF-20 and other widely used surveys in two important respects.²⁷ First, the SF-36 items cover a greater array of role limitations. Thus, in addition to defining more levels of role limitations because of health problems, the two SF-36 role functioning scales are, on the face of it, more applicable to retired individuals and those with more than one usual role. Second, SF-36 items define two scales that distinguish between role limitations because of physical health and mental problems. Role limitations because of mental problems are sometimes missed by the SF-20 scale and similar surveys that do not ask explicitly about limitations because of personal and emotional problems.^{19,27}

Bodily Pain

The SF-36 retains the SF-20 question concerning the frequency of bodily pain or discomfort and adds a second item measuring the extent of interference with normal activities because of pain. The latter item was chosen because it is the best predictor ($r = 0.84$) of the total score for the Behavioral Effects of Pain scale used in the MOS.²⁸ The result is an increase in scale reliability and improved precision (i.e., 10 scale levels as opposed to 6 scale levels), relative to the SF-20 version.¹⁹

Social Functioning

The SF-20^{20,21} included only one social functioning item. The SF-36 retains an improved form of that item and adds a second item. These two items, a subset of the long-form social functioning items developed for the MOS, assess health-related effects on social activities.²⁹ Most measures of social activity require respondents to report the number of contacts and activities or frequency of participation in different activities.²¹ To measure *health* outcomes, SF-36 items ask specifically about the impact of physical health or emotional problems on social activities. The resulting two-item SF-36 scale defines more levels of social functioning and

achieves a higher level of validity and precision.¹⁹

Mental Health

The five-item mental health scale (MHI-5) used in SF-20 has been retained without modification. The MHI-5 has been in use for nearly 8 years.^{14,21,30-35} The MHI-5 was constructed from the 5 items that best predicted the summary score for the 38-item Mental Health Inventory. It includes one or more items from each of the four major mental health dimensions (anxiety, depression, loss of behavioral or emotional control, and psychological well-being) confirmed in factor-analytic studies of the full-length MHI.²⁴ The simple sum of the five short-form items (without weights) correlated 0.95 with the full-length 38-item MHI. This correlation was 0.93 on cross-validation using data from the Health Insurance Experiment (HIE).

Vitality.

A four-item measure of vitality (energy level and fatigue), not included in SF-20, was added to better capture differences in subjective well-being. The items selected have an impressive record of empirical validity and balance between favorably and unfavorably worded items to control for response set effects. These items were adapted from the MHI fielded in the HIE, which was derived from the 1976 HANES survey by the National Center for Health Statistics.²⁹ These studies all yielded thorough evaluations of the scale's psychometric properties and documented item-discriminant validity and scale reliability. The scale's sensitivity to the impact of disease and treatment has been demonstrated in recent clinical trials involving patients with hypertension,³³ prostate disease,³¹ and those differing in severity of AIDS.^{34,35}

General Health Perceptions

The SF-20 combined the widely used single-item rating of health (ranging from ex-

cellent to poor) and four items from the Current Health scale constructed from the Health Perceptions Questionnaire (HPQ).²⁵ Although this 5-item scale has performed well in studies to date, a number of potential improvements have been achieved with the SF-36 5-item version. The SF-36 scale: 1) achieves a more comprehensive sample of the content of the HPQ; 2) correlates highly ($r = 0.96$) with the 22-item General Health Rating Index (GHRI), constructed from the HPQ; and (3) is more acceptable to respondents because its items appear to be less redundant. The new scale also balances between favorably and unfavorably worded items to control for response set effects. We chose to reproduce the GHRI summary score to represent health perceptions in SF-36, rather than only the Current Health subscale used in SF-20, because of the substantial empirical evidence of validity accumulated for the GHRI.^{14,25}

A sixth item, which asks respondents to rate the amount of change in their general health status over a 1-year period, is included in SF-36, although it is not used to score any of the 8 multi-item scales. Pending results from analyses of the accuracy and meaning of self-reported transitions in response to this item, it should be analyzed as a categorical variable or as an ordinal level scale.

Using the SF-36 Short-Form

The SF-36 permits scoring of a set of eight scales displayed as a profile of health status concepts, as in the case of six-scale SF-20 profiles.^{20,21} The scores are easy to compute, and considerable information regarding their interpretation is rapidly being accumulated. The advantages and disadvantages of other scoring and display options warrant further study.

Table 1, which defines the meaning of high and low scores, is offered as a guide to the interpretation of the eight SF-36 scales. The content of the 36 questionnaire items

and response choices used are documented in the Appendix. These items have been standardized and are available for use in a variety of proven questionnaire formats suitable for self-administration. Both key-punch data entry and optical scanning forms have been used successfully. Permission to use these forms or to reproduce them in an approved format is granted royalty free upon completion of a user agreement form available from the senior author.

The SF-36 items and scales were constructed for scoring using the Likert method of summated ratings.³⁶ Analysis and interpretation of the resulting linear scales assumes that item scores, on average, linearly related to the underlying health concept being measured. Research to date offers positive support for this assumption for SF-36 items.¹⁴ All scales are favorably scored to facilitate display and interpretation of a health profile. Rules for scoring items and scales are documented in the *SF-36 Scoring Manual*, which is available from the senior author.³⁷

Discussion

A number of tradeoffs are involved in the construction of a short-form health survey. A major tradeoff exists between breadth and depth of measurement. Breadth is an issue of comprehensiveness; depth relates to precision in measuring each concept. To achieve breadth of measurement, we included measures of the most frequently studied functional status and well-being concepts described in accepted definitions of health status.¹¹ To achieve depth of measurement for each health concept, we constructed a short, multi-item scale from a subset of items shown to best reproduce a full-length measurement scale of proven validity. The comprehensiveness of SF-36 was improved by adding concepts not represented in the first short-form tested in the MOS, the SF-20. Our goal for SF-36 was to enhance content validity and construct scales that were likely to more precisely de-

tect medically and socially relevant differences in health status and changes in health over time.

To represent a broad array of health concepts in a 5- to 10-minute survey, it is necessary to restrict the number of items within each conceptual domain. Some investigators have taken this strategy to the extreme of relying on one questionnaire item per concept.^{15-18,38} We also adopted this strategy for some concepts in the SF-20.^{20,21} However, we have rejected this strategy for the SF-36 because the coarseness of single-item measures appears to limit their usefulness in detecting small to moderate differences between groups and even large differences for *individual* patients.¹⁹

In constructing the SF-36, we also placed a high value on comparability with the SF-20. Direct comparability between one or more items in SF-20 and SF-36 was maintained for three of the six SF-20 concepts (General Mental Health, Pain, and General Health Perceptions). Thus, it is possible to compare some results across studies using either of the two forms.

Like the SF-20, the SF-36 form is designed for self-administration, telephone administration, or administration during a personal interview. All three administration methods have been used successfully. However, different forms and instructions are required. Items were selected or constructed so that response choices would be identical within each scale, with few exceptions. Such standardization makes it possible to print questions and responses in less space, and greatly facilitates oral administration by phone or in person.

Further research is necessary to better understand the tradeoffs involved in using short- versus long-form versions of health scales. Most reports of empirical tests compared instruments varying in a number of ways, including length of the measures, concepts represented, and methods of scale construction and enumeration.^{30,39} The HIE comparisons between short- and long-form

measures in terms of predictive validity³⁴ clearly indicate that multi-item scales, even short ones, are more valid than single-item measures, and that longer and more comprehensive questionnaires are the most valid. However, such analyses do not answer the critical issue of whether a short-form measure is acceptable and, if so, for what types of studies. Recent comparisons between 18-item, 5-item, and single-item scales from the Mental Health Inventory (MHI) exemplify the kinds of studies that are useful to address this issue.^{32,40}

Of the eight short-form scales in SF-36, the five-item MHI has been evaluated most extensively outside the MOS. It has been used successfully in a large evaluation of an inpatient primary care delivery program,⁴¹ empirical validity studies,^{19,30,32} and in clinical trials comparing quality-of-life outcomes for patients with benign prostatic hyper trophy³¹ and AIDS.^{34,35} The five-item MHI also has been compared favorably with the emotional reactions score from the Nottingham Health Profile⁴¹ and has been shown to correlate substantially with the much longer summary SIP Psychosocial Scale.³⁰ Perhaps the most impressive results of the MHI-5 pertain to its validity in discriminating psychiatric patients from those with other medical conditions. Berwick et al.,³² who studied acute and chronic primary care patients, systematically compared MHI-5 with the 18-item MHI version and other widely used mental health measures including the General Health Questionnaire.⁴¹ MHI-5 discriminated between patients with major depression, severe affective disorders, and anxiety disorders identified using the Diagnostic Interview Schedule as well or better than the three longer scales.

Although the SF-36 includes eight distinct health status concepts and one item measuring self-reported health transition, important health concepts are not represented. Among those concepts omitted are concepts currently being evaluated in the MOS and

other studies: health distress, family functioning, sexual functioning, cognitive functioning, and sleep disorders.¹⁴ Testing whether these and other omitted concepts add enough information to warrant lengthening the SF-36 beyond its current brief form is underway in the MOS. Other instruments noteworthy for their comprehensiveness include: 1) the Sickness Impact Profile,¹³ covering 12 health status concepts; 2) the full-length MOS health survey, covering 20 concepts;¹⁴ and 3) the HIE survey, covering 15 concepts.¹² However, these instruments have roughly a four-fold greater respondent burden than the SF-36.

Any health survey designed for young and old patients, and for use among sick and well populations, represents a substantial challenge. Measuring a comprehensive set of health concepts and the full range of levels for each concept does not allow for a great level of detail. Therefore, short-form measures are likely to have at least two types of problems: 1) ceiling effects, which entail substantial numbers of people getting the highest possible scores; and 2) floor effects, which include substantial numbers of people receiving the lowest possible score; in a given population. Floor effects were demonstrated in some SF-20 scales in severely ill hospital patients.⁴⁴ Subsequent studies demonstrate that floor effects are rare for SF-36 in the MOS, even among patients with serious chronic disease.¹⁹ Adding items and improving response choices in SF-36, relative to SF-20, appear to be worthwhile in this regard.

However, for studies of severely ill populations, it may be desirable to add a supplemental battery of items to represent the extreme low end of the continuum defined by some health scales. Most noteworthy is the physical functioning scale, which includes only one item focusing on daily self-care activities. When a large proportion of the sample scores appear at the scale's low end, it may be necessary to supplement the SF-36 with additional items that measure basic ac-

tivities of daily living.⁴⁵ A two-stage measurement strategy may enable the normative comparisons possible with SF-36 and the in-depth measurements required for precision in testing hypotheses involving severely ill patients.

During the past few years, a developmental version of SF-36 has been tested in numerous projects. The SF-36 has now been standardized in final form as documented here. Use of the developmental version is no longer recommended. The popularity of SF-36 appears to be largely driven by its brevity and comprehensiveness. These two competing measurement goals were achieved using very short multi-item scales. Whether this tradeoff results in an unacceptable loss of measurement precision requires further study. Preliminary results support the use of SF-36 scales in studies based on *group*-level analyses.^{19,36} Additional cross-sectional and longitudinal tests are needed to test the generalizability of these results, and to address the appropriateness of using SF-36 in monitoring outcomes for individual patients. The authors of this study hope that the standardization and publication of the form at this time will facilitate such studies.

Acknowledgments

The authors acknowledge the contributions of their MOS colleagues and consultants, including: Cameron Cushing, MA, Allyson R. Davies, PhD, J.F. Rachael Lu, MA, and Colleen A. McHorney, PhD, for assistance in analyzing data; Sandra R. Berry, MA; Audrey Burnam, PhD; Karen Camberg, MA; Paul Ellwood, MD; Sheldon Greenfield, MD; Ronald D. Hays, PhD; Elizabeth McGlynn, PhD; Eugene Nelson, ScD; William Rogers, PhD; Anita L. Stewart, PhD; Alvin R. Tarlov, MD; and Kenneth Wells, MD.

References

1. Geigle R, Jones SB. Outcomes measurement: a report from the front. *Inquiry* 1990;27:7.
2. McDermott W. Absence of indicators of the influence of physicians on a society's health. *Am J Med* 1981;70:833.
3. Cluff LE. Chronic Disease, Function and the Quality of Care. *Journal of Chronic Disease* 1981;34:299.
4. Tarlov AR. Shattuck lecture: the increasing supply of physicians, the changing structure of the health-ser-

vices system, and the future practice of medicine. *N Engl J Med* 1983;308:1235.

5. Schroeder SA. Outcome assessment 70 years later: are we ready? *N Engl J Med* 1987; 216:160.

6. Ellwood PM. Outcomes management: a technology of patient experience. *N Engl J Med* 1988; 318:1549.

7. American Cancer Society. Proceedings of the Working Conference on Methodology in Behavioral and Psychological Cancer Research—1983. *Cancer* 1984;53(suppl):2217.

8. Wenger NK, Mattson ME, Furberg CD, et al. Assessment of quality of life in clinical trials of cardiovascular therapies. New York: Le Jacq Publishing Company, 1984.

9. Lohr KN, Ware JE. Advances in health assessment. *Journal of Chronic Disease* 1987;40(suppl):15

10. Lohr KN. Advances in health status assessment: overview of the conference. *Medical Care* 1989;27:3.

11. Ware JE. Standards for validating health measures: definition and content. *Journal of Chronic Disease* 1987;40(6):473.

12. Brook RH, Ware JE, Davies-Avery A, et al. Overview of adult health status measures fielded in RAND's health insurance study. *Medical Care* 1979;17(suppl):1

13. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The sickness impact profile: development and final revision of a health status measure. *Medical Care* 1981;19:787.

14. Stewart AL, Ware JE, eds. Measuring functioning and well-being: the medical outcomes study approach. Duke University Press (in press).

15. Campbell A, Converse PE, Rodgers WL. The quality of american life: perceptions, evaluations, and satisfactions. :Russell Sage Foundation, 1976.

16. Spitzer WO, Dobson AJ, Hall J, et al. Measuring the quality of life of cancer patients: a concise QL-index for use by physicians. *Journal of Chronic Diseases* 1981;34:585.

17. Coates A, Gebiski V, Stat M, et al. Improving the quality of life during chemotherapy for advanced breast cancer. *N Engl J Med* 1987;10:1490.

18. NCHS:Health, United States. U.S. Department of Health and Human Services, 1989.

19. McHorney CA, Ware JE, Rogers W, et al. The validity and relative precision of MOS short- and long-form health status scales and Dartmouth COOP charts: results from the Medical Outcomes Study. *Med Care* (in press).

20. Ware JE, Sherbourne C, Davies A. A general population application: developing and testing the MOS 20-item short-form health survey. In: Stewart AL, Ware JE, eds. Measuring functioning and well-being: the medical outcomes study approach. Duke University Press (in press).

21. Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey: reliability and validity in a patient population. *Med Care* 1988;26:724.

22. Stewart AL, Ware JE, Brook RH, Davies-Avery A. conceptualization and measurement of health for

adults in the health insurance study, volume II, physical health in terms of functioning. Santa Monica, CA: The RAND Corporation. publication no. R-1987/2-HEW, 1978.

23. Donald CA, Ware JE. The measurement of social support. In: Greenley JR, ed. *Research in community and mental health*. Greenwich, CT: JAI Press, 1984:325.
24. Veit CT, Ware JE. The structure of psychological distress and well-being in general populations. *J Consult Clin Psychol* 1983;51: 730.
25. Davies AR, Ware JE. Measuring health perceptions in the health insurance experiment. Santa Monica, CA: The RAND Corporation. publication no. R-2711-HHS, 1981.
26. Stewart AL, Kamberg CJ. Physical functioning measures. In: Stewart AL, Ware JE, eds. *Measuring functioning and well-being: The Medical Outcomes Study Approach*. Durham, NC: Duke University Press, 1992:86.
27. Sherbourne CD, Stewart AL, Wells KB. Role functioning measures. In: Stewart AL, Ware JE, eds. *Measuring functioning and well-being: The Medical Outcomes Study Approach*. Durham, NC: Duke University Press, 1992:205.
28. Sherbourne CD. Pain measures. In: Stewart AL, Ware JE, eds. *Measuring functioning and well-being: The Medical Outcomes Study Approach*. Durham, NC: Duke University Press, 1992:220.
29. Sherbourne CD. Social functioning: social activity limitations measures. In: Stewart AL, Ware JE, eds. *Measuring functioning and well-being: The Medical Outcomes Study Approach*. Durham, NC: Duke University Press, 1992:173.
30. Read LJ, Quinn RJ, Hofer MA. Measuring overall health: an evaluation of three important approaches. *Journal of Chronic Diseases* 1987;40(suppl):75.
31. Fowler FJ, Wennberg JE, Timothy RP, et al. Symptom status and quality of life following prostatectomy. *JAMA* 1988;259(20):3018.
32. Berwick DM, Murphy JM, Goldman PA, et al. Performance of a five-item mental health screening test. *Med Care* 1991;29:169-175.
33. Croog SH, Levine S, Testa MA, et al. The effects of antihypertensive therapy on quality of life. *N Engl J of Med* 1986; 314:1657.
34. Wu AW, et al. A health status questionnaire using 30 items from the medical outcomes study: preliminary validation in persons with early HIV infection. *Med Care* 1991;29:786.
35. Wachtel T, Piette J, Mor V, et al. Quality of life in persons with AIDS, as measured by the Medical Outcome Study's instruments. *Ann Int Med* (in press).
36. Likert R. A technique for the measurement of attitudes. *Archives of Psychology* 1932;140:1-55.
37. International Resource Center for Health Care Assessment. *How to score the SF-36 short-form health survey*. Boston: The Health Institute, 1992.
38. Nelson E, Wasson J, Kirk A, et al. Assessment of function in routine clinical practice: description of the COOP chart method and preliminary findings. *Journal of Chronic Diseases* 1987, 40(suppl): 55S.
39. Liang MH, Fossel AH, Larson MG. Comparison of five health status instruments for orthopedic evaluation. *Med Care* 1990;28:632.
40. Weinstein MC, Berwick DM, et al. A comparison of three psychiatric screening tests using receiver operating characteristic (ROC) analysis. *Med Care* 1989;27:593.
41. Brook RH, Fink A, Kosecoff J, et al. Educating physicians and treating patients in the ambulatory setting: where are we going and how will we know when we arrive? *Ann Int Med* 1987;107:392.
42. Anderson JSC, Sullivan F, Usherwood TP. The Medical Outcomes Study Instrument (MOSI)—use of a new health status measure in Britain. *J Fam Pract* (in press).
43. Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. *Psychol Med* 1979; 9:139.
44. Bindman AB, Keane D, Lurie N. Measuring health changes among severely ill patients: the floor phenomenon. *Med Care* 1990;28:1142.
45. Katz MM, Lyerly SB. Methods for measuring adjustment and social behavior in the community. I. rationale, description, discriminative validity and scale development. *Psychological Reports*, 1963;13:503.

Appendix. SF-36 Questions^a

1. In general, would you say your health is:
2. *Compared to one year ago*, how would you rate your health in general *now*?
3. The following items are about activities you might do during a typical day. Does *your health now limit you* in these activities? If so, how much?
 - a. *Vigorous activities*, such as running, lifting heavy objects, participating in strenuous sports
 - b. *Moderate activities*, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf
 - c. Lifting or carrying groceries
 - d. Climbing *several* flights of stairs
 - e. Climbing *one* flight of stairs
 - f. Bending, kneeling, or stooping
 - g. Walking *more than a mile*
 - h. Walking *several blocks*
 - i. Walking *one block*
 - j. Bathing or dressing yourself
4. During the *past 4 weeks*, have you had any of the following problems with your work or other regular daily activities *as a result of your physical health*?
 - a. Cut down the *amount of time* you spent on work or other activities.
 - b. *Accomplished less* than you would like
 - c. Were limited in the *kind* of work or other activities
 - d. Had *difficulty* performing the work or other activities (for example, it took extra effort)
5. During the *past 4 weeks*, have you had any of the following problems with your work or other regular daily activities *as a result of any emotional problems* (such as feeling depressed or anxious)?
 - a. Cut down the *amount of time* you spent on work or other activities
 - b. *Accomplished less* than you would like
 - c. Didn't do work or other activities as *carefully* as usual
6. During the *past 4 weeks*, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?
7. How much *bodily pain* have you had during the *past 4 weeks*?
8. During the *past 4 weeks*, how much did *pain* interfere with your normal work (including both work outside the home and housework)?
9. These questions are about how you feel and how things have been with you *during the past 4 weeks*. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the *past 4 weeks*

- a. Did you feel full of pep?
 - b. Have you been a very nervous person?
 - c. Have you felt so down in the dumps that nothing could cheer you up?
 - d. Have you felt calm and peaceful?
 - e. Did you have a lot of energy?
 - f. Have you felt downhearted and blue?
 - g. Did you feel worn out?
 - h. Have you been a happy person?
 - i. Did you feel tired?
10. During the *past 4 weeks*, how much of the time has your *physical health or emotional problems* interfered with your social activities (like visiting with friends, relatives, etc.)?
11. How TRUE or FALSE is *each* of the following statements for you?
- a. I seem to get sick a little easier than other people
 - b. I am as healthy as anybody I know
 - c. I expect my health to get worse
 - d. My health is excellent

SF-36 Response Choices^a

- 1. Excellent, Very Good, Good, Fair, Poor
- 2. Much better now than one year ago, Somewhat better now than one year ago, About the same as one year ago, Somewhat worse now than one year ago, Much worse than one year ago
- 3. Yes, Limited a lot; Yes, Limited a little; No, Not limited at all
- 4a-d. Yes, No
- 5a-c. Yes, No
- 6. Not at all, Slightly, Moderately, Quite a bit, Extremely
- 7. None, Very mild, Mild, Moderate, Severe, Very severe
- 8. Not at all, A little bit, Moderately, Quite a bit, Extremely
- 9. All of the time, Most of the time, A good bit of the time, Some of the time, A little of the time, None of the time
- 10. All of the time, Most of the time, Some of the time, A little of the time, None of the time
- 11. Definitely true, Mostly true, Don't know, Mostly false, Definitely false

^aCopyright© The MOS Trust, Inc.; 1990. All rights reserved.