

THE QUALITY OF WELL-BEING SCALE: COMPARISON OF THE INTERVIEWER- ADMINISTERED VERSION WITH A SELF- ADMINISTERED QUESTIONNAIRE

ROBERT M. KAPLAN*, WILLIAM J. SIEBER and THEODORE G. GANIATS

*Health Outcomes Assessment Program (HOAP), Department of Family
and Preventive Medicine, University of California, San Diego, USA*

The Quality of Well-being Scale (QWB) is a general measure of health-related quality of life that has been used in a wide variety of population and clinical studies. One of the major disadvantages of the QWB is that it requires a trained interviewer. Recently, a self-administered version of the QWB was developed. The purpose of this study was to compare the self-administered QWB with the established interviewer-administered form. The respondents were 218 English speaking adults who attended primary care clinics. Each respondent was evaluated twice with an interval of one month in-between. At each session respondents were randomly assigned to complete either the interviewer-administered or self-administered QWB, resulting in a 2×2 factorial design. Data from the study demonstrated that the self-administered QWB yields scores equivalent to the interviewer-administered form. Further, QWB scores remain stable over the course of a one month interval. The results suggest that an inexpensive self-administered QWB may produce data comparable to the more difficult and expensive interviewer-administered version. Disadvantages of self-administered forms are also discussed.

KEY WORDS: Quality of Well-being Scale (QWB), Quality of Life Measurement, assessment method, interviewer.

INTRODUCTION

The assessment of quality of life has developed dramatically over the past two decades. Among the many instruments now available, only a few have been validated in different populations. Further, there are a limited number of measures that can be used for cost-utility comparisons. One of these instruments is the Quality of Well-being Scale (QWB). The Quality of Well-being scale is a preference-weighted measure combining three scales of functioning with a measure of symptoms and problems to produce a point-in-time expression of well-being that runs from 0 (for death) to 1.0 (for asymptomatic full function). The QWB is used as one component in a General Health Policy Model (GHPM) (Kaplan, 1993a, 1993b; Kaplan and Anderson, 1996; Kaplan, Anderson and Ganiats, 1993). This model includes several components. The major aspects of the model include mortality (death) and morbidity (health-related quality of life). In several papers, we have suggested that diseases and disabilities are important for two reasons. First, illness may cause the life expectancy to be shortened. Second, illness may make life less desirable at times prior to death (health-related quality of life) (Kaplan and Anderson, 1996; Kaplan *et al.*, 1993).

* Corresponding author. E-mail: RKaplan@ucsd.edu.

Central to the General Health Policy Model is a general conceptualization of quality of life. The model separates aspects of health status and life quality into distinct components. These are life expectancy (mortality), functioning and symptoms (morbidity), preference for observed functional states (utility) and duration of stay in health states (prognosis).

Using the QWB patients are classified according to objective levels of functioning. These levels are represented by scales of mobility, physical activity, and social activity. In addition to classification into these observable levels of function, individuals are also classified by the symptom or problem that each patient found most undesirable. On any particular day, nearly 80% of the general population is optimally functional. However, fewer than half of the population experience no symptoms. Symptoms or problems may be severe, such as serious joint pain, or minor such as taking medication or following a prescribed diet for health reasons.

Human value studies have been conducted to place the observable states of health and functioning onto a preference continuum for the desirability of various conditions, giving a "quality" rating between 0 for death and 1.0 for completely well. A quality-adjusted life year is defined as the equivalent of a completely well year of life, or a year of life free of any symptoms, problems, or health-related disabilities. The well-life expectancy is the current life expectancy adjusted for diminished quality of life associated with dysfunctional states and the durations of stay in each state. The model quantifies the health activity or treatment program in terms of the quality adjusted life years that it produces or saves.

One important aspect of the QWB is that it provides a metric for quality adjusted survival analysis. This analysis adjusts survival time by diminished quality of life. The output is usually called a Quality Adjusted Life Year (QALY). Consider, for example, a person who has a set of symptoms and is in an objective state of functioning that is rated by community peers as 0.5 on a 0.0 to 1.0 scale. If the person remains in that state for one year, he or she would have lost the equivalent of 1/2 of one year of life. Thus, a person limited in activities who requires a cane or walker to get around the community would be hypothetically rated at 0.50. If he or she remained in that state for an entire year, the individual would lose the equivalent of one-half year of life. However, a person who has the flu may also be rated as 0.50. In this case, the illness might only last three days and the total loss in QALYs might be $3/365 \times 0.50$ which is equal to 0.004 QALYs. This may not appear as significant an outcome as the disabled person. But suppose that 5,000 people in a community get the flu. The well years lost would then be $5,000 \times 0.004$ which is equal to 20 years.

THE QUALITY OF WELL-BEING SCALE - SELF-ADMINISTERED (QWB-SA)

One of the major criticisms of the QWB is that it is more expensive and difficult to administer than measures such as the SF-36 (McHorney, Ware, Lu and Sherbourne, 1994). The QWB requires a trained interviewer. Further, the QWB questionnaire is long and complex because it employs branching and probe questions. The reason that the current QWB uses a complex interviewer-administered questionnaire is described in several publications by Anderson, Bush, and Berry (1986, 1988). In the original QWB validation study, a simple self-administered form was compared with a more complex interviewer-based questionnaire. Results from these studies suggested that the self-administered form was associated with several errors, ultimately resulting in underestimates of population

health status. We believe that these problems are not unique to the QWB, but are associated with all self-administered questionnaires (Anderson *et al.*, 1986, 1988; Anderson, Kaplan and DeBon, 1989). Refinements in the questionnaire may circumvent these problems. There would be widespread application of a utility-based, self-administered instrument if such refinements could be incorporated into a measure that produces QALYs. The demand for rapid health status assessment is exemplified by the current widespread use of the SF-36. In response to this demand, we are in the process of developing a self-administered version of the QWB, known as the QWB-SA. The current version of the QWB-SA can be printed on two sides of a single page and takes about 10 minutes to complete.

The development of new forms of the QWB has gone through several stages. First, a new list of symptoms and problems was developed. The current version of the QWB uses a list of 26 symptoms or problems. The QWB-SA has 58 symptoms. The improved symptoms assessment not only better reflects health status, it more closely resembles a clinical review of symptoms, thus increasing the clinical utility of the QWB-SA.

The format for the QWB-SA includes five sections. The first section asks about acute and chronic symptoms. Respondents are asked, in a yes-no format if they have each of 19 chronic symptoms or problems. Examples are blindness or severely impaired vision in both eyes (a separate item asks about one eye), speech problems, and so on. Part 1 also asks about 25 acute physical symptoms (i.e. headache, coughing or wheezing, shortness of breath or difficulty breathing) and 14 mental health symptoms (trouble falling asleep or staying asleep, spells of feeling upset, downhearted or blue, feelings of being lonely or isolated). The format for these items requests respondents to think back over the last three days and indicate if the symptom was present on no days, yesterday, two days ago, or three days ago (multiple responses are allowed). Part II uses a similar format and asks about self care. It includes two items asking if the respondent had been in an institution (hospital) and whether they needed help with self care such as eating, dressing, bathing, or getting around the home. Part III asks about mobility, use of public transportation or driving or confinement to a home or hospital. Part IV considers physical activity and asks about performance of physical functioning, such as walking, confinement to a bed or chair or movement of a wheelchair. Performance of usual activity, such as work, school, or housework is evaluated in Part V.

The purpose of this study was to evaluate the reliability and validity of the questionnaire as well as compare its psychometric properties to the interviewer-administered version of the QWB.

METHOD

Sample

A total of 375 English-speaking adults were approached at primary care clinics within the University of California, San Diego (UCSD) Healthcare Network in San Diego, California. Of these, 94 people refused to participate or did not provide enough information to be contacted for interview. Of the remaining 281 respondents, 63 were dropped from the analyses because they did not complete an assessment at either baseline or one month. Thus, a total of 218 respondents completed assessments at both timepoints. The sample included 123 females (56%) and 95 males (46%) between the ages of 18 and 85 (mean age = 48 years).

Design

The study used a 2×2 (Mode \times Time) factorial design with one between-subjects factor and one within-subjects factor. The within-subjects factor was time of administration. All respondents were evaluated two times, with a one-month interval in between. The between-subjects factor was mode of administration. There were two types: (1) interviewer-administered, and (2) self-administered. These factors were repeated at the second evaluation.

The features of the experimental design allow for two types of comparison. First, since respondents were randomly assigned to group, we would anticipate that the mean QWB score at baseline should be the same independent of mode of administration. Thus, we expect to retain the null hypothesis of no differences between groups at baseline. Differences by mode of administration may suggest that different formats produce different results. A second feature of the design is that there is a built-in replication at one month. For example, if there are differences by mode of administration, we would expect that these differences would be replicated at the one month evaluation.

A second feature of the design is that it allows for within subject comparisons. Respondents were selected because they were in stable health states. When they were recruited from medical clinics, they did not complete the QWB assessment until 10 days after their consent was obtained, thus minimizing the possibility that respondents were presenting with health complaints only at baseline. Although variation from month-to-month is expected, this variation is presumed to be random and not systematically associated with time of assessment. Thus, the mean scores at each follow-up should be equivalent. One half of these respondents completed the same questionnaire (either QWB-SA or interviewer administered QWB) at the one month retest. The remaining half were evaluated using the other form of the QWB. Differences by mode of administration that are detected in the between-subjects component of the experimental design can be verified using this within-subjects component. Finally, the half of respondents receiving the same mode of administration allows test-retest reliability to be ascertained for both forms of administration. Order effects can also be assessed. The design allows for the evaluation of several alternate hypotheses. For example, if there are practice effects, QWB scores may change systematically over the course of time, independent of mode of administration. The between and within-subjects components of the design allow cross-comparisons.

RESULTS

The data were analyzed by comparing QWB scores at both the baseline and one month evaluations. At each evaluation individuals were tested using either the QWB-SA or the interviewer-administered form. This design yielded four independent groups representing combinations of measurement at the two consecutive evaluations: (1) interviewer-interviewer, (2) interviewer-SA, (3) SA-SA and (4) SA-interviewer. The Analysis of Variance (ANOVA) was calculated separately for first administration and second administration. The means and standard deviations for those completing the QWB-SA and the interviewer-administered QWB at baseline and one month evaluations are summarized in Table 1. As the table suggests, the two measures yielded nearly identical QWB scores at both the baseline and the one month evaluations. Since respondents had been randomly assigned to measures and we would expect differences between groups to be represent only sampling error, this finding implies that the methods produce equivalent results.

Table 1 Means and standard deviations for QWB by mode of administration

Mode	Baseline		One month	
	Mean	SD	Mean	SD
Interviewer	0.704	0.105	0.718	0.106
SA	0.706	0.108	0.709	0.120

Table 2 Summary of 2 × 2 ANOVA outcome of QWB scores at baseline and 1 month

Factor	Baseline			One month		
	df	Ms	F	df	Ms	F
Method (baseline)	1	0.008	0.68	1	0.001	0.082
Method (One Month one month)	1	0.004	0.361	1	0.004	0.322
Interaction	1	0.001	0.100	1	0.0004	0.039
Residual	241	0.012		240	0.012	

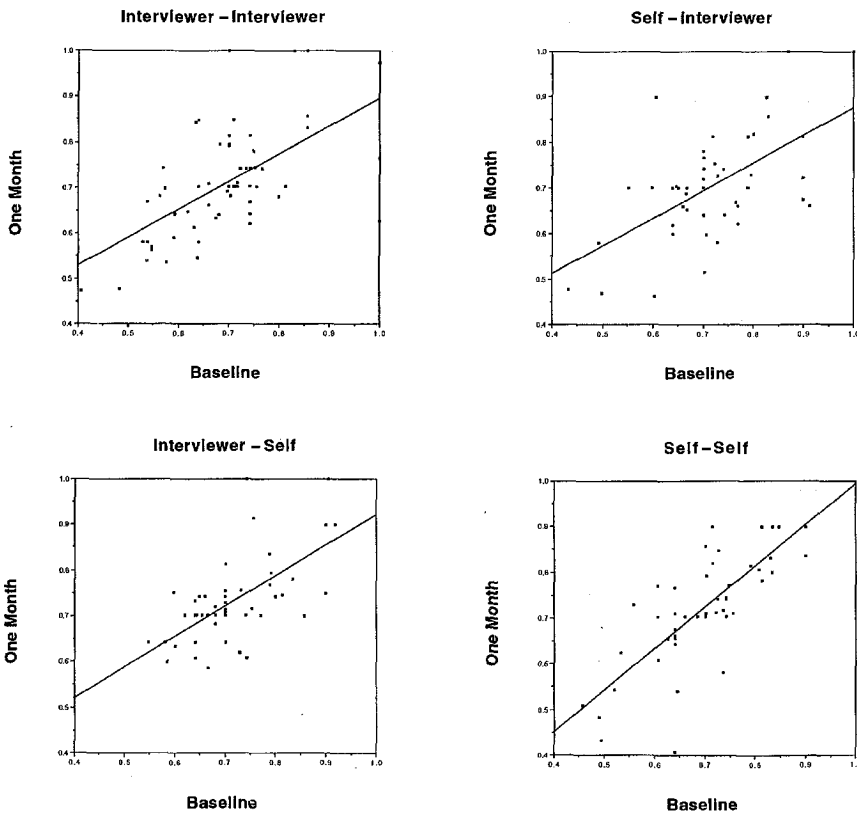


Figure 1 Scatterplots of baseline versus one month scores for interviewer-interviewer, $r = 0.60$ (upper left), self-interviewer, $r = 0.61$ (upper right), interviewer-self, $r = 0.65$ (lower left) and self-self, $r = 0.77$ (lower right) conditions.

The four groups were also evaluated using a 2×2 ANOVA separately at baseline and one month evaluations. All main effects and interactions in the model were non significant (see Table 2).

Figure 1 summarizes the relationship between baseline and one month follow-up separately for each combination of QWB-SA and Interviewer administration. Each quadrant in the figure represents a unique combination. All four quadrants show a strong linear relationship. The correlations ranged from 0.60 (Interviewer-Interviewer) to 0.77 (Self-Self).

DISCUSSION

The results from these analyses suggest that the QWB-SA is comparable to the full QWB interviewer-administered questionnaire. These methods differ in more ways than form of administration. The two questionnaires actually ask different questions. For example, the QWB-SA has no branching questions and records more symptoms than the interviewer-administered form. Data from the initial evaluation indicated that, on a population basis, mean scores for the two forms of administration were nearly identical. This finding was replicated at one month follow-up. All the patients were selected to be in stable health states. The test-retest evaluation suggests that QWB scores remain stable over a one month time period for relatively healthy adults and are not influenced by mode of administration.

One important component of the QWB is the preference weighting system. In this study, the standardized weights for the interviewer-administered QWB were applied to responses on the self-administered form. This allowed a pure test of the effect of administration. We are currently developing a new preference weighting system that will be specific to the QWB-SA. The new system is desirable because the QWB-SA has more symptom-problem complexes than does the interviewer-administered form. It is expected that the application of a different weighting system will be associated with greater discrepancies between interviewer and self-administration than were observed in this study.

Despite the advantages of the self-administered questionnaire, we also urge caution in the use of self-administered methods. Our previous work has shown that interviewer-administered questionnaires may detect higher rates of dysfunction than self-administered versions (Anderson *et al.*, 1988; Kaplan, Anderson and Erikson, 1989). Another concern is that the high correlations observed in this study may be specific to the relatively well population that was studied. Future studies, assessing persons with acute medical conditions are expected to show fluctuations in QWB scores.

In other studies we have found that non-completion rates are significantly higher for self-administered questionnaires. Thus, the use of self-administered questionnaires carries the risk of more missing data. Despite these disadvantages, we feel further investigation of the self-administered forms is warranted.

In the present study, QWB scores were relatively stable over time. Future research must address responsiveness to clinical change. Several studies have demonstrated that the QWB is responsive to clinical change in a variety of patient populations.

The consistency of the QWB upon repeated administrations might argue that the measure is so stable that it is not responsive to clinical change. However, several studies have demonstrated that the QWB is responsive to clinical change in a variety of patient populations. For example, studies have shown that antibiotic treatment is associated with improvements in QWB scores for patients with cystic fibrosis (Orenstein, Nixon, Ross

and Kaplan, 1989) and the method has been responsive to a variety of different surgeries including surgery for chronic sinusitis (Hodgkin, 1994) and cochlear implant (Harris, Anderson and Novak, 1995). In addition, the QWB is responsive to medications expected to have a minor effect, such as oral gold treatment for patients with arthritis (Bombardier, Ware, Russell, Larson, Chalmers and Read, 1986) or medications that have a larger effect such as AZT for patients with HIV infection (Kaplan, Anderson, Wu, Mathews, Kozin and Orenstein 1989).

It might be argued that the failure to find differences between the two forms of administration is the result of low statistical power. In order to consider this issue, we conducted a post-hoc power analysis assuming a power of 0.90 with the probability of a type one error of 0.95. Given the variability seen in these samples, we should have been able to detect a difference of 0.02 (on the 0 to 1.0 scale). Differences less than 0.02 are usually regarded as so small that they are not clinically meaningful.

The validity of the QWB-SA remains unestablished. However, these findings are encouraging because they suggest that the QWB-SA may serve as a simple and inexpensive proxy for the traditional interviewer-administered QWB. The QWB has now been used in a wide variety of different studies. The measure has been used in both population and clinical studies. For example, estimates from the U.S. National Health Interview Survey compared three problems: sinusitis, diabetes, and chronic lung disease. In each of three age groups, sinus disease is shown to be a less serious problem than diabetes which, in turn, has less impact than emphysema (Erickson *et al.*, 1989).

The QWB has also been compared to measures of cognitive impairment for patients with Alzheimer's Disease. Those with disease are significantly lower on the QWB and the degree of illness is systematically related to QWB score (Kerner, Patterson and Kaplan, 1997). A variety of studies have evaluated the relationship between QWB and HIV disease. These studies, which were conducted at the UCSD HIV Neurobehavioral Research Center (HNRC), evaluated the relationship between neurocognitive impairment, disability, and the QWB. Those with less than 200 CD4+ cells (full AIDS) had significantly lower QWB scores than those with 200–500 and those with over 500 cells. The differences between those with less than 200 CD4+ cells and those with greater than 500 cells was about 0.13 units of well being, suggesting that individuals lose 13/100 QALYs for each year they are in the AIDS category in comparison to the asymptomatic groups. In comparison to the uninfected controls, this would equal a one year of life loss for each 7.69 infected individuals. The QWB was shown to be significantly associated with Beta-2 Micro-globulin quartile ($p < 0.03$), neurologists' ratings of dysfunction ($p < 0.001$), clinician ratings of neuropsychological impairment ($p < 0.04$), future vital status ($p < 0.05$) and several psychiatric variables including Profile of Mood States (POMS) scores for vigor ($p < 0.001$) and dejection ($p < 0.001$). Multivariate models demonstrated high covariation between predictors of QWB. These results suggest that the QWB is a significant correlate of biological, neuropsychological, neurological psychiatric, and mortality outcomes for male HIV-infected patients (Kaplan, Anderson, Patterson *et al.*, 1995).

Other applications of the QWB include chronic obstructive pulmonary disease (Kaplan, Atkins and Timms, 1984), AIDS (Kaplan *et al.*, 1989), cystic fibrosis (Orenstein *et al.*, 1989), diabetes mellitus (Kaplan, Hartwell, Wilson and Wallace, 1987), atrial fibrillation (Ganiats, Palinkas and Kaplan, 1992), lung transplantation (Squier, Ries, Kaplan *et al.*, 1995), arthritis (Kaplan, Kozin and Anderson, 1988), cancer (Kaplan, 1993c) depression (Kaplan, 1997) schizophrenia (Patterson, Kaplan, Grant *et al.*, 1996) and several other conditions (Kaplan, 1993a). Further, the method has been used for

health resource allocation modeling and has served as the basis for an innovative experiment on rationing of health care by the U.S. State of Oregon (Kaplan, 1993a, 1993b).

In summary, the QWB is a general measure of health-related quality of life with some accumulated evidence for validity in a variety of different specific diseases. One of the major drawbacks for those interested in using the QWB is that it requires a trained interviewer. Data presented here suggest that a new self-administered form may correspond well to the validated interviewer-administered method. The self-administered version is expanded to better assess psychological states and self-perceptions of health. Studies are currently underway to determine the utility weights for these additional items as well as recalibrating the weighted scores for this new mode of administration of the QWB. We urge additional studies to validate the self-administered version.

Acknowledgement

Supported by Grant PO 1-AR-40423 from the National Institute of Arthritis, Musculoskeletal, and Skin Disorders of the National Institutes of Health, and Grant R01 HS 09170 from the agency for Health Care Policy and Research.

References

- Anderson, J.P., Bush, J.W. and Berry, C.C. (1986) Classifying function for health outcome and quality of life evaluation. *Medical Care*, **24**, 454-469.
- Anderson, J.P., Bush, J.W. and Berry, C.C. (1988) Internal Consistency Analysis: a method for studying the accuracy of function assessment for health outcome and quality of life evaluation. *Journal of Clinical Epidemiology*, **41**, 127-37.
- Anderson, J.P., Kaplan, R.M. and DeBon, M. (1989) Comparison of responses to similar questions in health surveys. In F. Fowler (Ed.), *Health Survey Research Methods*. (pp. 13-21) Washington, D.C.: National Center For Health Statistics.
- Bombardier, C., Ware, J., Russell, I.J., Larson, M., Chalmers, A. and Read, J.L. (1986). Auranoftin therapy and quality of life in patients with rheumatoid arthritis. *American Journal of Medicine*, **81**, 565-578.
- Erickson, P., Kendall, E.A., Anderson, J.P. and Kaplan, R.M. (1989) Using composite health status measures to assess the nation's health. *Medical Care*, **27**(Suppl 3), S66-S76.
- Gannias, T.G., Palinkas, L.A. and Kaplan, R.M. (1992) Comparison of Quality of Well-being scale and functional status index in patients with atrial fibrillation. *Medical Care*, **30**, 958-964.
- Harris, J.P., Anderson, J.P. and Novak, R. (1995) An outcomes study of cochlear implants in deaf patients. Audiologic, economic, and quality-of-life changes. *Archives of Otolaryngology - Head and Neck Surgery*, **121**, 398-404.
- Hodgkin, P.S. (1994) Health impact of endoscopic sinus surgery assessed by the Quality of Well-being (QWB) Scale. Unpublished paper, University of California, San Diego.
- Kaplan, R.M. (1993a) Application of a general health policy model in the American health care crisis. *Journal of the Royal Society of Medicine*, **86**, 277-281.
- Kaplan, R.M. (1993b). Hippocratic Predicament: *Affordability, Access, and Accountability in Health Care*. San Diego: Academic Press.
- Kaplan, R.M. (1993c) Quality of Life Assessment for Cost/Utility Studies in Cancer. *Cancer Treatment Reviews*, **19**(Suppl A) 85-96.
- Kaplan, R.M. (1997) Health-Related Quality of Life in Mental Health Services Evaluation. In N.E. Miller and K.M. Magruder (Eds.), *The cost effectiveness of psychotherapy: A guide for practitioners, researchers, and policymakers*. New York: John Wiley and Sons, in press.
- Kaplan, R.M. and Anderson, J.P. (1996) The general health policy model: An integrated approach. In B. Spilker (Ed.), *Quality of Life and Pharmacoeconomics in Clinical Trials*. (pp. 309-322), New York: Raven.
- Kaplan, R.M., Anderson, J.P. and Erickson, P. (1989) Estimating well-years of life for a new public health indicator. *Proceedings of the 1989 Public Health Conference on Records and Statistics* (pp. 298-303). HDDS, National Center for Health Statistics, DHHS Publication Number (PHS) 90-1214.
- Kaplan, R.M., Anderson, J.P. and Gannias, T.G. (1993) The Quality of Well-being Scale: rationale for a single quality of life index. In S.R. Walker and R.M. Rosser (Eds.), *Quality of life assessment: Key issues in the 1990s* (pp. 65-94) London: Kluwer Academic Publishers.

- Kaplan, R.M., Anderson, J.P., Patterson, T.L., McCutchan, J.A., Weinrich, J.D., Heaton, R.H., Atkinson, J.H., Thal, L., Chandler, J. and Grant, I. (1995) Validity of the Quality of Well-being Scale For Persons with HIV Infection. *Psychosomatic Medicine*, **57**, 138–157.
- Kaplan, R.M., Anderson, J.P., Wu, A.W., Mathews, W.C., Kozin, F. and Orenstein, D. (1989). The quality of well-being scale: Applications in AIDS, cystic fibrosis, and arthritis. *Medical Care*, **27**(Suppl 3), S27–S43.
- Kaplan, R.M., Atkins, C.J. and Timms, R. (1984) Validity of a quality of well-being scale as an outcome measure in chronic obstructive pulmonary disease. *Journal of Chronic Diseases*, **37**, 85–95.
- Kaplan, R.M., Hartwell, S.L., Wilson, D.K. and Wallace, J.P. (1987) Effects of diet and exercise interventions on control and quality of life in non-insulin-dependent diabetes mellitus. *Journal of General Internal Medicine*, **2**, 220–228. (republished in *Diabetes Spectrum*, 1989, **2**, 20–25.)
- Kaplan, R.M., Kozin, F. and Anderson, J.P. (1988) Measuring quality of life in arthritis patients (including discussion of a general health-decision model). *Quality of Life and Cardiovascular Care*, **4**, 131–139.
- Kerner, D., Patterson, T.L. and Kaplan, R.M. (1997) Validity of the Quality of Well-being Scale in Alzheimer's disease (submitted).
- McHorney, C.A., Ware, J.E. Jr, Lu, J.F. and Sherbourne, C.D. (1994) The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care*, **32**, 40–66.
- Orenstein, D.M., Nixon, P.A., Ross, E.A. and Kaplan, R.M. (1989) The quality of well-being in cystic fibrosis. *Chest*, **95**, 344–347.
- Patterson, T.L., Kaplan, R.M., Grant, I., Dr, Semple, S.J., Moscona, S., Koch, W.L., Harris, M.J. and Jeste, D.V. (1996) Quality of Well-being in late-life psychosis. *Psychiatry Research*, **63**, 169–181.
- Squier, H., Ries, A.L., Kaplan, R.M., Prewitt, L.M., Smith, C.M., Kriett, J.M. and Jamieson, S.W. (1995). Quality of Well-being predicts survival in lung transplantation candidates. *American Journal of Respiratory and Critical Care Medicine*, **152**, 2032–2036.