

CHAPTER 4

Measures of Health Outcome in Social Support Research

Robert M. Kaplan

A wide variety of papers link social support to health outcomes (see Berkman, 1984; Broadhead et al., 1983; Wallston, Alagna, DeVellis, & DeVellis, 1983). Stressful life events in the personal, social, occupational, or marital realms may have important consequences, and social support may soften the impact of these events. Wallston et al. (1983) suggest that social support is a crucial factor in coping with physical disability and illness. Family, friends, and other social contacts aid in the reduction of emotional distress and problems resulting from illness or injury (Davidson, Bowden, & Tholen, 1979; Porrit, 1979).

The notion that social support enhances health outcomes is widely embraced in the medical, public health, and psychological literatures (Cohen & Syme, 1985). Systematic investigation of this problem, however, is hampered for several reasons. First, measures of social support have varied from study to study; even the definition of social support has been quite inconsistent (Heitzmann & Kaplan, 1988). A second and perhaps more disturbing problem is that few studies relating social support to health have considered the complex issues in assessing health status. This chapter reviews some of the measurement issues that complicate studies relating social support to health outcome. First, I consider issues in the measurement of social support, and then review the measures that social support scales have been validated against. Consideration is also given to

ROBERT M. KAPLAN • Division of Health Care Sciences, Department of Family and Preventive Medicine, University of California, San Diego, La Jolla, California 92093-0622.

Social Support and Cardiovascular Disease, edited by Sally A. Shumaker and Susan M. Czajkowski. Plenum Press, New York, 1994.

the conceptualization and measurement of health outcomes, and problems in relating health status measures to social support measures will be simulated. Finally, directions for future research are suggested.

The Conceptualization and Measurement of Social Support

Although definitions vary, most measures of social support include tangible components (e.g., financial assistance or physical aid) and intangible components (e.g., encouragement and guidance). As noted above, social support has been implicated in the mediation of stressful life events, recovery from illness, and increased program adherence. Some measures emphasize the instrumental function of social support, whereas others focus on its stress-buffering function.

Heitzmann and Kaplan (1988) reviewed the literature on the assessment of social support and identified at least 23 different measurement techniques. Most of the measures had suitable reliability; however, only about half of the measures had any evidence of validity, defined as the correlation between the measures of social support and well-defined criterion measures. This was particularly problematic for studies concerning the relationship of social support to health, because there are few well-validated measures of health status.

Chapters 2 and 3 in this volume consider the conceptualization and measurement of social support. In concert with Heitzmann and Kaplan (1988), these writings suggest that problems in the conceptualization and measurement of social support still remain. Few studies, however, have seriously considered problems in the conceptualization and measurement of the other side of the equation—health status. Some would question why health measures should be used as validity criteria for measures of social support. The rationale is that social support interventions are justified on the basis of their presumed relationships to health outcomes. Authors repeatedly evoke the social support–health outcome connection in discussions of either direct effects or buffering models. It is the evidence for these support–health relationships that I examine here.

Table 1 provides a summary of scales used as validity criteria for social support measures. The left-hand column of the table lists the social support scale, the next column gives the measure the scale was validated against, and the remaining columns describe the nature of the criterion measure and the association. In the Heitzmann and Kaplan (1988) review, 11 of 23 measures were validated against some external criterion. In three of the studies, measures were validated against other social support measures. In another three studies, they were validated against symptom

Table 1. Summary of Scales Used as Validity Criteria for Social Support Measures

<i>Scale</i>	<i>Validity criterion</i>	<i>Nature of criterion measure</i>	<i>Correlation</i>
Norbeck Social Support Questionnaire Norbeck (1981)	SSQ Schaefer et al., (1981)	Social support scale	-.03 to .56
Personal Resource Questionnaire (PRQ; Brandt & Weinert, 1981)	Family integration measure	Social support scale	.21 to .44
Arizona Social Support Interview Schedule (ASSIS; Berrera 1981a)	Inventory of Socially Supportive Behaviors (ISSB; Berrera, 1981b)	Social support measure	.42 (network size)
Interpersonal Support Evaluation Sched- ule (ISES; Cohen et al., 1985)	Psychiatric and physical symptoms	Symptom checklists	-.60 with measures of psychiatric symptoms, -.39 with measures of physical symptomatology
Social Relationship Scale (SRS; McFarlane, et al., 1981)	Clinician reports	Clinical judgment	Specific correlations not reported
Interview Schedule for Social Interaction (ISSI; Henderson, et al., 1980)	Eysenck Personality Inventory (EPI)	Personality test	Modest correlation
Social Support Questionnaire (SSQ; Sarason et al., 1983)	Multiple Affect Adjective Checklist (MAACL) and EPI	Adjective checklist and personality test	-.43 between SSQ and MAACL, -.37 between SSQ and ERI (for women)
Social Support Scale (SSS; Lin et al., 1979)	Psychiatric symptoms	Symptom checklist	.36
Perceived Social Support from Friends (PSS-Fr; Procidano & Heller, 1983)	Psychiatric symptomatology	Symptom checklist	Modest negative correlation with psychiatric symptoms
Work Relationship Index (WRI; Billings & Moos, 1982)	Personal functioning	Personality measure	-.33 for men -.15 for women
Diabetes Family Behavior Checklist (DFBC; Schafer et al., 1984)	Adherence with diabetes regimen	Adherence behavior	Significant negative correlations with changes in 3 categories of adherence to diabetic regimen

checklists. In three further studies, the measures were validated against personality tests. One study validated the social support scales against clinical judgments, and in one case social support was validated against self-reported measures of behaviors.

Inspection of Table 1 suggests that social support measures have rarely been validated against widely accepted measures of health status. Most often, when validity data are presented, mental health measures are used as the outcome. For example, McFarlane, Neale, Norman, Rox, and Streiner (1981) validated their social relationship scale against clinical judgments (the specific correlations were not reported). Henderson, Duncan-Jones, Byrne, and Scott (1980) found modest correlations between their interview schedule for social interaction and the Eysenck Personality Inventory. Sarason, Levine, Bashom, and Sarason (1983) found substantial correlations between their social support questionnaire and the Multiple Affect Adjective Check List. Some of the studies used psychiatric symptoms as validity criteria (Cohen, Mermelstein, Kamarck, & Hoberman, 1985; Lin, Simeone, Ensel, & Kuo, 1979; Procidano & Heller, 1983). A few studies used adherence behaviors as an outcome (Schafer, McCaul, and Glasgow, 1984).

As Table 1 suggests, the "health" variables have been inconsistent across studies of social support and health. In order to understand clearly the relationship between social support and health, we need a definition of health status. The next section of this chapter considers health status in more detail.

Measurement of Health Status

The conceptualization and measurement of health status has been of interest to scholars for many decades. Following the Eisenhower administration, a President's Commission on National Goals identified health status measurement as an important objective. Shortly after, John Kenneth Galbraith, in his noted book *The Affluent Society*, described the need to measure the effect of the health care system upon quality of life. In recent years, there have been many attempts to define and measure health status (see Walker & Rosser, 1988; Wenger, Mattson, Furberg, & Elinson, 1984; Bergner, 1985).

The terms *health status*, *quality of life*, and *health-related quality of life* are often used interchangeably. The term *health status* is often used to describe indicators of health outcome, including mortality rates, disability days, and years of potential life lost. I reserve the term *quality of life* here for

indicators that assume some valuation of states of being (see below). I use the term *health-related quality of life* to refer to the impact of health conditions upon the values associated with function, excluding those quality dimensions associated with work, housing, air pollution, and so forth. Before considering any specific approach, it is worth noting that traditional indicators of "health" have well-identified problems.

Mortality

Mortality remains the major outcome measure in most epidemiological studies and clinical trials. Typically, mortality is expressed in the form of a rate, that is the proportion of deaths from a particular cause occurring in some defined time interval (usually per year). Usually, mortality rates are age adjusted. Case fatality rates express the proportion of persons who died of a particular disease divided by the total number with the disease (including those who die and those who live). There are many advantages to reporting mortality rates: They are "hard" data (despite some misclassification bias) and the meaning of the outcome is not difficult to comprehend. But despite these advantages, there are also some obvious limitations. Mortality rates consider only the dead and ignore the living; many important health variables, including social support, might have little or no impact on mortality rates. Some very important illnesses, such as arthritis, are clearly major public health concerns, yet these conditions have relatively little impact upon mortality. Nevertheless, we would not want to conclude that they are unimportant.

Two chapters in this volume consider the relationship between social support and mortality; Chapter 5 reviews the international evidence, whereas Chapter 6 reviews the U.S. evidence. To date, the epidemiological investigations provide the best evidence of the relationship between social support and health outcomes. There is little disagreement that mortality is an important health indicator.

Morbidity

The most common approach to health status assessment is to measure morbidity in terms of function or role performance (e.g., workdays missed or bed disability days). Most approaches to health status assessment are essentially morbidity indicators. The RAND health status measures (Stewart, Ware, Brook, & Davies-Avery, 1978) include separate categories for the effects of disease or health states upon physical function, social function, and mental function. These measures do not integrate morbidity and

mortality, although as each birth cohort ages, there is accrual of mortality cases.

Death is a health outcome, and it is important that this outcome not be excluded from any expression of health status. For example, suppose we were evaluating the effect of a program of integrated support and treatment, as opposed to no support or treatment, for randomly assigned groups of very ill, elderly nursing home residents. Let us suppose that the program maintained them all at a very low level of function throughout the year, while in the comparison group, the sickest 10% died. Looking just at the living in the follow-up, one finds the comparison group to be healthier, because the sickest have been removed by mortality. By this standard, the program of no supportive treatment might appear to be the better alternative. With a measure that combined morbidity and mortality, however, the story would be very different, with mortality effects dragging the overall health of the comparison group to a very low level.

Some authors believe that the idea of integrating morbidity and mortality into a single measure is problematic because death can be viewed not as a health outcome, but rather as the absence of life and of health. According to this line of reasoning, death is not a level of health but a qualitatively different outcome altogether. We assert that mortality is a very important end point; in fact, many health services are directed toward preventing premature mortality. Also, it has been suggested that mortality and morbidity should be separated because many treatments cause side effects, and those refusing treatment may die earlier but experience fewer side effects and a better quality of life before their deaths. According to this argument, if health status includes mortality, these latter individuals would be seen as having a lower quality of life despite its improvement during the period that they avoided the toxic medications. A comprehensive system that includes morbidity and mortality avoids this sort of problem. If the system includes duration of stay at different states, it may or may not suggest that the treatment is worthwhile. For example, a treatment that extends life by only one month but makes people very disabled prior to their deaths would accumulate a loss of well years of life (discussed later in this chapter) that may exceed the benefit in well years. Separating morbidity from mortality confuses rather than clarifies this issue.

I do not mean to imply that mortality should never be analyzed separately. In fact, in many studies there are separate comprehensive analyses for morbidity and mortality. Yet separating morbidity and mortality essentially forbids an analysis that compares treatment with different objectives. For example, it is sometimes of interest to compare programs that prevent early mortality for a few people versus those that reduce

morbidity for a large number of people. A comprehensive system allows for these types of trade-offs.

The Value Dimension

Scholars have debated about components of "health" for many centuries. Sullivan (1966), synthesizing literature from a variety of different fields, notes that most concepts of morbidity involve three types of evidence: clinical, subjective, and behavioral. Most studies in social support focus on either clinical or subjective outcomes. Clinical outcomes might include clinical judgment as well as the results of tests obtained during physical examination or invasive procedures; subjective evidence might include symptoms and complaints. Clinical evidence is valuable only if it is clearly related to well-defined behavioral health outcomes. For example, significant abnormalities in certain blood proteins are only of concern if these deviations correlate with dysfunction and early mortality. The burden of proof is on the scientist to demonstrate these associations.

Subjective symptoms are also very important in health care, because symptoms are a major correlate of health care utilization. Not all symptoms, however, should be given equal weight. It is not obvious that the number of symptoms depicts the severity of health status. For example, an adult with an acute 24-hour flu may have an enormous number of symptoms, including nausea, headache, aches and pains, vomiting and diarrhea. Yet it is not clear that this condition is more severe than the single symptom of a very severe headache. Several factors need to be taken into consideration. First, we must determine the degree to which the symptoms limit function. One individual may have five symptoms—an itchy eye, a runny nose, coughing, fatigue, and headache. Yet he or she may still feel well enough to work and to perform all usual activities. Another person with the single symptom of a severe headache may be limited to bed and unable to move around. Would we want to call the person with five symptoms less well? Another dimension is the duration of the symptoms; a year in pain is certainly worse than a day in pain. Finally, and perhaps the most often neglected, is the value or preference associated with different types of dysfunction.

Biomedical investigators often avoid reference to values or preferences because these constructs are not considered scientific; however, the value dimension in health status is inescapable. Fishburn (1964) defined value as the quantification of the concept of worth, importance, or desirability. Ultimately, our judgment of the value of health states depend upon subjective evaluations. The judgment that one level of functioning is better than another level of functioning is ultimately tied to this appraisal. If

we advise individuals to change their diet in order to avoid heart disease, we inherently assume that the reduced probability of heart disease later in life is valued more than the immediate but enduring mild displeasure of dietary change. The term *quality of life* presumes a qualitative judgment.

Behavioral Dysfunction

When Sullivan (1966) reviewed the literature on health measurement, he emphasized the importance of behavioral outcomes. Bolstered by the proud accomplishments of behavioral scientists, a convincing argument was developed suggesting that such behavioral indicators as absenteeism, bed disability days, and institutional confinement would be the most important consequences of disease and disability. Ability to perform activities at different ages could be compared to societal standards for these behaviors; restrictions in usual activity were seen as *prima facie* evidence of deviation from well-being. Many other investigators focus on point-in-time measures of dysfunction as measures of health (Bergner, 1985; Katz, Ford, Moskowitz, Jackson, & Jaffe, 1963; Stewart et al., 1978). Clearly point-in-time dysfunction is crucial in our quantification of health, but it is important not to neglect what will happen in the future. The spectrum of medical care ranges from public health, preventive medicine, and environmental control, through diagnosis and medical care, to convalescence and rehabilitation. Many programs affect the probability of occurrence of dysfunction in the future, rather than altering present functional status. For example, a socially supportive family that instills proper health habits in its children may also promote better health in the future, yet it may be years until this benefit is realized. A positive future orientation might lead to the exercise of better health habits or better planning for future health care.

The concept of health must consider not only the ability to function now, but also the probability of future changes in function. A person who is very functional and asymptomatic today may harbor a disease with a poor prognosis. Thus, many individuals are at high risk for mortality attributable to heart disease even though they are perfectly functional today. The term *severity of illness* should take into consideration both dysfunction and prognosis, as many medical treatments may cause near-term dysfunction in order to prevent future dysfunction. For example, coronary artery bypass surgery causes severe dysfunction for a short period of time, but it is presumed that the surgery will enhance function or decrease mortality at a later point in time. Patients may be incapacitated following myocardial infarction and restricted to coronary care units, yet

the treatment is designed to help them achieve better future outcomes. Pap smears are performed and hysterectomies are executed in order to decrease the probability of future deaths caused by cancer. Much of health care involves looking into the future in order to enhance outcomes over the life span, and therefore it is essential to separate out the current and future components of health. I prefer the term *prognosis* to describe the probability of transition among health states over the course of time (Fanshel & Bush, 1970).

Health-Related Quality of Life

There is a growing sentiment that the objectives of health care are twofold. First, health care and health policy should be designed to increase the life expectancy. Second, the health care system should improve the quality of life during the years that people are alive. It is instructive to consider various measures in health care in light of these two objectives. Traditional biomedical indicators and diagnoses are important to us because they may be related to mortality or to quality of life. I prefer the term *health-related quality of life* to refer to the impact of health conditions upon function. Thus, health-related quality of life may be independent of quality of life relevant to work setting, housing, air pollution, and so forth (Rice, 1984).

Numerous new quality-of-life measurement systems have evolved since 1965, representing various traditions in measurement. In the late 1960s and early 1970s, the National Center for Health Services Research funded several major projects to develop general measures of health status. Those projects resulted in the Sickness Impact Profile (Bergner, Bobbitt, Carter, & Gilson, 1981), the Quality of Well-Being Scale (Kaplan & Bush, 1982), and the RAND health status measures (Stewart, et al., 1978). A variety of other measures resulted from this work. Most of these efforts involved extensive multidisciplinary collaboration between behavioral scientists and physicians and focused on the impact of disease and disability upon function and observable behaviors. For example, many of these measures examined the role of disease or disability upon performance of social roles, ability to get around the community, and physical functioning. Some of the systems include separate components for the measurement of social and mental health. All of the systems were guided by the World Health Organization definition of health status: "Health is a complete state of physical, mental, and social well-being and not merely absence of disease" (WHO, 1948). Three of the more commonly used methods include the Sickness Impact Profile (SIP), the Index of Activities of Daily Living scales, and the RAND measures.

Sickness Impact Profile (SIP)

The Sickness Impact Profile (SIP) is one of the best-known and widely used quality-of-life measures. It is a general measure applicable to any disease or disability group, and it has been successfully used with a variety of different cultural subgroups. The SIP includes 136 items describing the effect of sickness upon behavioral function. These items are divided into 12 categories, which are further clustered into three groups: independent, physical, and psychosocial. The independent categories include sleep and rest, eating, work, home management, and recreation/pastimes. Physical categories include ambulation, mobility, and body care and movement. The psychosocial categories are social interaction, alertness behavior, emotional behavior, and communication. Examples of SIP items include "I sleep or nap during the day" (sleep and rest), "I am not doing heavy work around the house" (home management), and "I have difficulty reasoning and solving problems—for example, making plans, making decisions, learning new things" (alertness behavior).

Each SIP item has been evaluated by an independent group of judges on a 15-point scale of dysfunction. Using these independent weights, a respondent taking the SIP endorses or does not endorse each of the 136 items. The overall SIP percentage score is obtained by separating the items endorsed by the respondent, summing their scale values, and dividing by the sum of all values for all items on the SIP. This proportion is then multiplied by 100; scores are obtained similarly for each category. Percentage scores for each category can be plotted on a graphic display that looks similar to an MMPI profile. A variety of studies attest to the substantial reliability and validity of the SIP (see Bergner et al., 1981).

Two minor issues are relevant to general use of the SIP. First, it does not integrate morbidity and mortality, and thus it is less appropriate than some other measures for policy analysis. The second problem is that the SIP is sometimes cumbersome to administer. With 136 items, it can be time-consuming, and it requires alertness and attention by the respondent. The SIP, however, is an example of a measurement system that has undergone systematic methodological refinements over many years. It has been widely used, widely tested, and well evaluated.

Index of Activities of Daily Living

Perhaps the oldest general quality-of-life measure is the Index of Activities of Daily Living (ADL) most commonly used in studies of the elderly. Katz was very early to argue that the major importance of disease

and disability was upon function and ability to perform role activities (Katz et al., 1963).

The system includes six subscales: bathing, dressing, toileting, transfer, continence, and feeding. For each category, a judgment is made as to whether the person is independent or dependent. For the category of bathing, people are judged to be independent if they need assistance only in bathing a single part of the body or can bathe themselves; they are judged to be dependent if they need assistance in bathing more than one part of the body. Once a judgement of dependence or independence is obtained for each of the six categories, an overall grade is assigned. To receive the top grade of A, the person must be independent in all six categories. A grade of B is assigned to those who are independent in all but one of these functions. The bottom grade, G, is assigned to those who are dependent in all six functions. Several reliability and validity studies for the ADL have been reported (Katz, Downs, Cash, & Grotz, 1970).

Despite its many important applications in studies of aging, the ADL has been criticized because it does not make distinctions toward the well end of the quality-of-life continuum. Stewart and colleagues (1978) suggest that nearly 80% of the noninstitutionalized population have no functional limitations and would obtain the top score in the ADL system. Other population surveys, however, demonstrate that more than 50% of the population experience one or more symptoms on a given day (Kaplan, Bush, & Berry, 1976). Given the wide array of challenges for a quality-of-life measure, the ADL has some significant limitations. Other measures are required in order to distinguish between those individuals who are toward the healthy end of the functioning continuum.

RAND Health Status Measures

Perhaps the most thorough review in the conceptualization of health-related quality-of-life measures yet available has been accumulated by the RAND Corporation. The RAND group adapted questionnaires developed by Bush, Kaplan, and others to describe physical activity, social activity, and mobility. The social activity category was subdivided to include social activity, role activity, household activity, and leisure activity. The RAND group also adapted Dupey's (1969) General Well-Being Index. In addition, they have added a General Health Perceptions Questionnaire (Ware & Karmos, 1976) and a variety of other measures. Finally, the RAND group uses self-report questionnaires to assess the clinical status associated with a wide variety of medical conditions (Brook et al., 1979). Through considerable testing and evaluation the RAND group revised the measure for the

Medical Outcomes Study (MOS) and eventually shortened it to 36 items. The measure is now known as the 36 item short form, or SF-36.

The RAND approach has the advantage of being very comprehensive. Perhaps the major disadvantage is that it is sometimes difficult to aggregate the measures in order to provide a comprehensive expression of quality of life. For example, the approach may demonstrate that patients with cardiovascular disease have shown minor improvements in certain aspects of mobility and role performance, but are experiencing side effects such as mental confusion and headaches. Unlike some other systems, the RAND approach does not allow a comprehensive statement about whether the patients are getting better or worse on a composite index.

In addition to these approaches that focus on health status, other authors refer to quality of life as something that is independent of health status. Although many investigators believe that symptoms and mortality represent quality of life (see Bush, 1984), Croog et al. (1986) used a wide variety of outcome measures to define the term. Some investigators now use traditional psychological measures and call them quality-of-life outcomes; for instance, Follick et al. (1988) include the patient's subjective evaluation of well-being, physical symptoms, sexual function, work performance and satisfaction, emotional status, cognitive function, social participation, and life satisfaction. Other investigators, including Hunt and McEwen (1983) regard quality of life as subjective appraisals of life satisfaction. In summary, there is no agreement on which dimensions should be considered the standard for assessing quality of life in research studies, yet consideration of recurrent themes in the methodological literature can assist in the evaluation of existent instruments.

Measurement Issues

Unidimensional versus Multidimensional

There is essentially no disagreement that quality of life is a multidimensional construct, yet there is considerable debate about whether outcome measures must necessarily represent this multidimensional structure. There are essentially two major approaches to quality of life assessment: a psychometric approach, and a decision theory approach. The psychometric approach attempts to provide separate measures for the many different dimensions of quality of life. Perhaps the best-known example of the psychometric tradition is the Sickness Impact Profile (above).

The alternative is the decision theory approach, which attempts to weight the different dimensions of health in order to provide a single unitary expression of health status. Supporters of this approach argue that psychometric approaches fail to consider that different health problems are not of equal concern; 100 runny noses are not the same as 100 severe abdominal pains (Bush, 1984). In an experimental trial using the psychometric approach, it is not uncommon to find that some aspects of quality of life improve while others get worse. For example, a medication might reduce high blood pressure but also be associated with headaches and impotence. The decision theory approach attempts to place an overall value on health status by weighting the different dimensions and combining them into an aggregate quality score. It is argued that "quality" is the subjective evaluation of observable or objective health states. The decision theory approach attempts to provide an overall summary of quality of life that integrates subjective function states, preferences for these states, morbidity, and mortality.

Ware et al. (1981) argue that the psychometric approach has greater validity for studies in quality of life. Citing studies on factor analysis, they suggest that different components of health (including mental, physical, and social aspects) might be statistically independent dimensions, and thus any aggregate measure of health status might be considered the same as adding apples to oranges. In rebuttal, Bush (1984) argued that different components of quality-of-life measures indeed are different from one another and might be considered analogous to different pieces of fruit; however, it is the overall evaluation of the basket of fruit that is important. A fruit peddler who regularly delivers a full basket of fruit is preferred over one who delivers a half-empty basket. A basket of fruit in which all pieces are fresh and none are rotten is preferred over one in which some pieces are either missing or decayed. Baskets of fruit thus are associated with preferences or levels of desirability. Even though the contents of baskets may differ, some baskets are preferred over others, and there is a differential willingness to pay for different baskets. Bush argued that the psychometric approach was analogous to comparing one full bowl of fruit to a second bowl of fruit with a banana rotted and a pear missing. Both are bowls of fruit, but they may have different values. Health status often represents combinations of function, symptoms, and disabilities in different systems. Ultimately, our concern is with the overall desirability of the aggregate.

Many investigators prefer the profile approach for assessing side effects of medications. For example, measures such as the SIP allow the investigators to determine if some dimensions of health are getting better

while others are getting worse. In other cases, focus on the aggregate may be more desirable. The aggregate approach allows the investigator to state comprehensibly whether the treatment makes people better or worse. There may be instances in which individual preferences rather than societal preferences are used for these decision processes. Knowing the aggregate may be important for some purposes, but the ability to disaggregate may be important for other purposes. Some investigators may want to know whether or not lowering blood pressure makes people dizzy; here a profile approach would be preferable. Others want to know if, considering all of the benefits and all the side effects, blood pressure treatment improves health status; in this case, the aggregate approach may be more desirable.

Disease-Specific versus General Approaches

Most health-related quality-of-life measures are designed for use with any population. Some investigators, however, feel it is necessary to develop quality-of-life measures for specific diseases. For example, the RAND Corporation has produced a series of booklets describing the conceptualization and measurement of "physiologic health." Each booklet describes the problems in conceptualization and measurement of a specific condition, such as coronary heart disease. The rationale underlying these measures is largely clinical, based on the idea that medical conditions have very specific outcomes: Heart patients are evaluated according to ejection fractions, blood gases, etc. Clearly there are advantages to the clinician in considering outcomes relative to specific diseases. In addition to general physiological indicators, there are also quality-of-life measures designed specifically for particular disease groups, best represented in the arthritis literature (Liang, Cullen, & Larson, 1982).

In contrast to those using disease-specific approaches, many investigators believe that all diseases and disabilities have a general effect upon quality of life. In fact, the purpose of quality-of-life measurement is not to identify clinical information relevant to the disease; instead, it seeks to determine the impact of the disease on general function. For example, a lower ejection fraction may be associated with shortness of breath, weakness, and increased risk of mortality, and medications used to control cardiovascular diseases might cause headaches, irritability, and general confusion. By focusing too specifically on clinical correlates of disease, it is argued that the general impact is overlooked. It has also been argued that the general quality-of-life measures adequately capture a wide variety of dysfunctions associated with cardiovascular diseases. These dysfunctions might be in many different systems and recognized in symptoms such as

confusion, tiredness, sexual impotence, and depression. These outcomes may not be specific to disease condition.

There is considerable debate over generalized versus disease-specific measures. Although my colleagues and I have argued for the more generalized approach, we recognize the value of disease-specific measures in some clinical studies (R. Kaplan & Anderson, 1988). We urge investigators who choose disease-specific approaches, however, not to limit their measures to symptoms or clinical indicators of a specific disease.

Risk Factors and Outcomes

Epidemiological studies identify a variety of risk factors for coronary heart disease. (N. Kaplan & Stamler, 1983). Among the most important of these is blood pressure. Studies consistently show that elevated blood pressure is a predictor of mortality, nonfatal heart attack, and stroke; thus, important interventions have been developed to lower blood pressure. Many studies use blood pressure reduction as the outcome, and interventions that lower blood pressure are deemed successful. Yet blood pressure is a risk factor for bad health outcomes, but not an outcome itself.

One example that might illustrate this point concerns cigarette smoking. The evidence that cigarette smoking is detrimental to health is overwhelming (Holbrook, 1986; Surgeon General of the United States, 1979). In addition, cigarette smoking interacts with other risk factors such as hypercholesterolemia and hypertension to enhance the risk of coronary heart disease (Gotto, 1986). Nevertheless, the effect of cigarette smoking upon blood pressure is difficult to evaluate. Cigarette smoking may cause an acute rise in blood pressure (Benowitz, Kuyt, & Jacobs, 1984). Epidemiological studies, however, consistently find that smokers have lower blood pressure than nonsmokers. In addition, ex-smokers have blood pressures similar to nonsmokers, even after adjustments for the confounding effects of age and weight. Toshima (1987) recently reviewed this literature and found that across a remarkably diverse set of studies, the effects of cigarette smoking on blood pressure are consistent. Yet they go in the unexpected direction: Cigarette smoking may reduce rather than increase blood pressure.

Toshima also evaluated prospective changes in blood pressure across a variety of epidemiologic studies. Again, in several prospective studies, relationships between cigarette smoking and blood pressure were in the unexpected direction. For example, the Normative Aging Study (Seltzer, 1974) found that systolic blood pressure increases when smokers discontinue cigarette use. The Framingham Heart Study (Gordon et al., 1975) also observed slight increases in systolic blood pressure in ex-smokers in

comparison to continuing smokers. Dietary changes were not capable of explaining these changes. Other studies have not observed these relationships, Greene, Aavedel, Tyroler, Davis, & Hames, 1977; Paffenbarger, Thorne, & Wing, 1968); however, most studies simply do not demonstrate that quitting smoking reduces blood pressure.

What can we make of these results? If our outcome measure is blood pressure, we might come to the conclusion that cigarette smoking is good. After all, it appears that habitual cigarette use lowers blood pressure. In addition, we might advise cigarette smokers to continue to smoke; again, the studies consistently demonstrate that smoking cessation is associated with increased blood pressure. But advising smokers to continue would clearly be the wrong conclusion, because the evidence that cigarette smoking has detrimental effects upon health status is overwhelming (Holbrook, 1986). Blood pressure is a risk factor, but not an outcome. Focus of attention on a risk factor may misdirect the purpose of a health care intervention.

Finally, consider the case of insulin-dependent diabetes mellitus. Several studies, including the Diabetes Control and Complications Trial (DCCT, 1993), have suggested that degree of hyperglycemia is associated with the long-term risk of diabetic complications (Tchobroutsky, 1978). A general quality-of-life scale may have substantial advantages for estimating the treatment benefits in diabetes care. In addition to mortality, diabetes may be associated with poor outcomes in a variety of organ systems; for example, poor control may lead to differential rates of retinopathy, kidney failure, and foot infection. The difficulty is in finding one common expression of these outcomes, when some patients may have foot infections that result in amputations while others have eye problems that result in blindness. One purpose of a general quality-of-life measurement system is that it can aggregate these outcomes with death to provide a single expression of the impact of poor control.

In addition to the benefits of the tight management of diabetes, we must also consider the consequences or side effects. Some data suggest that as many as one third of patients who are aggressively managed experience nausea and weakness associated with hypoglycemia on as many as half of the days. A comprehensive view of the benefits of tight control in diabetes must trade the expected benefits against the consequences of tight control. If poor outcomes can be established, they must be represented as probabilities. Retinopathy, for example, may occur in about 50% of insulin-dependent diabetic cases; tight control may reduce this rate to 30%. The real question in diabetes care is how to exchange minor symptoms that occur over an extended period of time with major symptoms that occur later in the life cycle.

Decision Theory Approaches

Within the last few years there has been growing interest in using quality-of-life data to help evaluate the cost/utility or cost-effectiveness of health care programs. Cost-effectiveness analysis typically quantifies the benefits of health care intervention in terms of years of life or *quality-adjusted life years* (QALYs). Cost/utility is a special use of cost-effectiveness that takes expressed preference for health status into consideration (Kaplan & Bush, 1982). In cost/utility analysis, the benefits of medical care, behavioral interventions, or social support are expressed in terms of *well years*; others have chosen to describe the same outcome as QALYs (Weinstein & Stason, 1976) or healthy years of life (Russell, 1986). Because the term *quality-adjusted life years* has become most popular, I will use it in this presentation. QALYs integrate mortality and morbidity to express health status in terms of equivalents of well years of life. If a man dies of heart disease at age 50 and we would have expected him to live to age 75, it might be concluded that the disease was associated with 25 lost life years. If 100 such men died at age 50, we might conclude that 2,500 (100 men times 25 years) life years had been lost.

Yet death is not the only outcome of concern in heart disease. Many adults suffer myocardial infarctions that leave them somewhat disabled over longer periods of time; although they are still alive, the quality of their lives has diminished. QALYs take into consideration the quality-of-life consequences of these illnesses. For example, a disease that reduces quality of life by half will take away 0.5 QALYs over the course of 1 year. If it affects two people, it will take away a total of 1.0 years over the same period. A medical treatment that improves quality of life by 0.2 for each of five individuals will result in the equivalent of 1.0 QALY if the benefit is maintained over a 1-year period. This system has the advantage of considering both benefits and side effects of programs in terms of the common QALY units.

The need to integrate mortality and quality-of-life information is clearly apparent in studies of heart disease. Consider the case of hypertension. People with high blood pressure may live shorter lives if they are untreated; thus, one benefit of treatment is to add years to life. For most patients, however, high blood pressure is not associated with symptoms for many years. Conversely, the treatment for high blood pressure may cause a variety of symptoms. In other words, in the short run, patients taking medication may experience more symptoms than those who avoid it. If a treatment is evaluated in terms of changes in life expectancy, the benefits of the program will be overestimated, because side effects are not taken into consideration. In contrast, considering only current quality of

life will underestimate the treatment benefits, because information on mortality is excluded. A comprehensive measurement may take into consideration side effects and benefits and provide an overall estimate of the net effectiveness of treatment (Russell, 1986).

Although there are several different approaches for obtaining quality-adjusted life years, most of them are similar (R. Kaplan, 1985b). The approach that I prefer involves several steps. First, patients are classified according to objective levels of functioning represented by scales of mobility, physical activity, and social activity. The dimensions and steps for these levels of functioning are shown in Table 2. (The reader is cautioned that these steps are not actually the scale, only listings of labels representing the scale steps.) Standardized questionnaires have been developed to classify individuals into one of each of these scale steps (Anderson, Bush, & Berry, 1986). In addition to classification into these observable levels of function, individuals are also classified by the symptom or problem that bothered them most (see Table 3).

Many measures include separate dimensions for emotional and sexual functioning. In this system, these problems are captured in the list of symptoms and problems (i.e., problems in sexual interest or performance, spells of feeling upset, depressed, or crying, etc.). Systems vary in the attention they give to these symptoms, although most agree on the importance of the problems. Depression is a symptom, though, just as is a cough. It has a duration, and its severity might be judged by the degree to which it inhibits role performance. Although it could be regarded as its own dimension, depression disrupts function just as other symptoms do. There are also advantages in keeping the number of dimensions to a minimum. On any particular day, nearly 80% of the general population is optimally functional; however, fewer than 15% of the population experience no symptoms. Symptoms may be severe (e.g., serious chest pain) or minor (e.g., taking medication or a prescribed diet for health reasons). The functional classification and the accompanying list of symptoms or problems was created after extensive reviews of the medical and public health literature (Kaplan et al., 1976).

Once levels of functioning for observable behavior have been classified, the observable health states are weighted by ratings for the desirability of these conditions on a preference continuum with an anchor of 0 for death and 1.0 for completely well. In several studies, random samples of citizens from a metropolitan community evaluated the desirability of more than 400 case descriptions. Using these ratings, a preference structure that assigned the weights to each combination of an observable state and a symptom or problem was developed (R. Kaplan et al., 1976). Cross-validation studies show that the model can be used to assign weights to all

Table 2. Quality of Well-being/General Health Policy Model: Elements and Calculating Formulas (Function Scales, with Step Definitions and Calculating Weights)

<i>Step number</i>	<i>Step definition</i>	<i>Weight</i>
<i>Mobility Scale (MOB)</i>		
5	No limitations for health reasons	-.000
4	Did not drive a car, health related; did not ride in a car as usual for age (younger than 15 yr), health related, and/or did not use public transportation, health related; or had or would have used more help than usual for age to use public transportation, health related	-.062
2	In hospital, health related	-.090
<i>Physical Activity Scale (PAC)</i>		
4	No limitations for health reasons	-.000
3	In wheelchair, moved or controlled movement of wheelchair without help from someone else; or had trouble or did not try to lift, stoop, bend over, or use stairs or inclines, health related; and/or limped, used a cane, crutches, or walker, health related; and/or had any other physical limitation in walking, or did not try to walk as far as or as fast as other the same age are able, health related	-.060
1	In wheelchair, did not move or control the movement of wheelchair without help from someone else, or in bed, chair, or couch for most or all of the day, health related	-.077
<i>Social Activity Scale (SAC)</i>		
5	No limitations for health reasons	-.000
4	Limited in other (e.g., recreational) role activity, health related	-.061
3	Limited in major (primary) role activity, health related	-.061
2	Performed no major role activity, health related, but did perform self-care activities	-.061
1	Performed no major role activity, health related, and did not perform or had more help than usual in performance of one or more self-care activities, health related	-.106
<i>Calculating formulas</i>		
Formula 1. Point-in-time well-being score for an individual (W):		
$W = 1 + (CPXwt) + (MOBwt) + (PACwt) + (SACwt)$		
where "wt" is the preference-weighted measure for each factor and CPX is Symptom/Problem complex. For example, the W score for a person with the following description profile may be calculated for one day as:		
CPX-11	Cough, wheezing or shortness of breath, with or without fever, chills, or aching all over	-.257
MOB-5	No limitations	-.000
PAC-1	In bed, chair, or couch for most or all of the day, health related	-.077
SAC-2	Performed no major role activity, health related, but did perform self-care	
$W = 1 + (-.257) + (-.000) + (-.077) + (-.061) = .605$		
Formula 2. Well-years (WY) as an output measure:		
$WY = [No. \text{ of persons} \times (CPXwt + MOBwt + PACwt + SACwt) \times \text{Time}]$		

*Table 3. Quality of Well-being/General Health Policy Model:
Symptom/Problem Complexes (CPX) with Calculating Weights*

CPX number	CPX description	Weight
1	Death (not on respondent's card)	-.727
2	Loss of consciousness such as seizure (fits), fainting, or coma (out cold or knocked out)	-.407
3	Burn over large areas of face, body, arms, or legs	-.387
4	Pain, bleeding, itching, or discharge (drainage) from sexual organs—does not include normal menstrual (monthly) bleeding	-.349
5	Trouble learning, remembering, or thinking clearly	-.340
6	Any combination of one or more hands, feet, arms, or legs either missing, deformed (crooked), paralyzed (unable to move), or broken—includes wearing artificial limbs or braces	-.333
7	Pain, stiffness, weakness, numbness, or other discomfort in chest, stomach (including hernia or rupture), side, neck, back, hips, or any joints or hands, feet, arms, or legs	-.299
8	Pain, burning, bleeding, itching, or other difficulty with rectum, bowel movements, or urination (passing water)	-.292
9	Sick or upset stomach, vomiting or loose bowel movement, with or without chills, or aching all over	-.290
10	General tiredness, weakness, or weight loss	-.259
11	Cough, wheezing, or shortness of breath, <i>with</i> or <i>without</i> fever, chills, or aching all over	-.257
12	Spells of feeling upset, being depressed, or of crying	-.257
13	Headache, or dizziness, or ringing in ears, or spells of feeling hot, nervous or shaky	-.244
14	Burning or itching rash on large areas of face, body, arms, or legs	-.240
15	Trouble talking, such as lisp, stuttering, hoarseness, or being unable to speak	-.237
16	Pain or discomfort in one or both eyes (such as burning or itching) or any trouble seeing after correction	-.230
17	Overweight for age and height or skin defect of face, body, arms, or legs, such as scars, pimples, warts, bruises or changes in color	-.188
18	Pain in ear, tooth, jaw throat, lips, tongue; several missing or crooked permanent teeth—includes wearing bridges or false teeth; stuffy, runny nose; or any trouble hearing—includes wearing a hearing aid	-.170
19	Taking medication or staying on a prescribed diet for health reasons	-.144
20	Wore eyeglasses or contact lenses	-.101
21	Breathing smog or unpleasant air	-.101
22	No symptoms or problems (not on respondent's card)	-.000
23	Standard symptom/problem	-.257
X24	Trouble sleeping	-.257
X25	Intoxication	-.257
X26	Problems with sexual interest or performance	-.257
X27	Excessive worry or anxiety	-.257

X—Specific weight not available.

possible states of functioning with a high degree of accuracy ($R^2 = .96$); the regression weights obtained in these studies are given in Tables 2 and 3. Finally, it is necessary to consider the duration of stay in various health states. For example, 1 year in a state that has been assigned the weight of .5 is equivalent to 0.5 QALYs. Table 2 provides an illustrative example of such a calculation.

The *well life expectancy* is the current life expectancy adjusted for diminished quality of life associated with dysfunctional states and duration of stay in each state. Using the system, it is possible to consider simultaneously mortality, morbidity, and the preference weights for these behavioral states of function. When the proper steps are followed, the model quantifies the health activity or treatment program in terms of the quality-adjusted life years that it produces or saves. A quality-adjusted life year is defined conceptually as the equivalent of a completely well year of life, or a year of life free of any symptoms, problems, or health-related disabilities. More detailed descriptions of this system are available in other publications (R. Kaplan, 1985a; R. Kaplan & Bush, 1982).

There are other approaches to integrating values into a quality-of-life measure. DuPuis (1989) argues for a subjective approach to quality-of-life measurement. According to this method, preferences are completely unstandardized; quality of life becomes an individual's preference for states. Many investigators favor this approach because it allows the individual to estimate how treatments affect him or her. The disadvantage is that these approaches are completely unstandardized, do not allow for comparisons between different treatment approaches, and forbid policy analysis because the outcomes are not in the same unit. These subjective approaches are valuable for learning how patients react to their treatments, may give guidance with regard to compliance decisions, and may be held as subjective reactions to treatment.

Reliability and Validity

Measures of health status are often evaluated using common psychometric methods. Despite the attractiveness of applying commonly used psychometric methods to health measures, there are several inherently difficult problems. Some of these problems concern validity, and others concern reliability.

The reliability of a health status measure is difficult to assess, particularly if we consider test-retest reliability (Kaplan & Saccuzzo, 1993). Conceptually, test-retest methods were developed to measure traits—characteristics of individuals that are stable over the course of time. Thus, variation in test scores over the course of time is attributable to measure-

ment error. Health, in contrast, is not assumed to be constant; indeed, it is the variability in health status that is of major interest. Traditional methods of reliability assessment will overestimate measurement error for measures of health status because true changes in health status will be counted as errors. Internal consistency methods are commonly used, but they are limited to assessing error associated with item sampling.

Reliability refers to the portion of variance in a measure that is "true score," or free of measurement error. There are difference sources of unreliability in scores. For example, some methods for evaluating reliability consider the internal consistency of the measure. Using this model, it is assumed that all items in the measure tap the same construct, and that these items are independent samples of characteristic under study. Other methods for evaluating reliability assume that such characteristics as social support or health are stable over the course of time; thus, different scores obtained at different points in time are attributable to measurement error. Reliability is a problem in research because it reduces the chances of finding significant relationships between measures.

Problems in the assessment of the validity of health status measures have been outlined by R. Kaplan et al. (1976). Validity, a frequently misunderstood concept in health status measurement, describes the range of inferences that are appropriate when interpreting a measurement, a score, or the result of a test. In other words, the validity of a measure defines the meaning of a score. Validity is not absolute: It is relative to the domain about which statements are made. If we want to measure what society means by health, then an indicator or index is a valid measure of total health status only to the extent that it expresses or correlates with that construct.

Criterion validity is the correspondence of a proposed measure with some other observation that accurately measures the phenomenon of interest. By definition, the criterion must be a superior, more accurate measure of the phenomenon if it is to serve as a verifying norm. If a criterion exists, only greater practicalities or less expense justify the use of concurrent measures as proxies. If the criterion is not a superior measure, then failure of correspondence by any new measure may be a defect of the criterion itself, making it insufficient as a reference for validity.

Most exercises in validation of health status measures involve construct validity. Construct validation is a process required when "no criterion or universe of content is accepted as entirely adequate to define the quality to be measured" (Cronbach & Meehl, 1955). Construct validation involves assembling empirical evidence to support the inference at a particular level that has meaning. It is an ongoing process, akin to amassing support for a complex scientific theory for which no single set of observa-

tions provides crucial or critical evidence. It is difficult to define a point at which an investigator can declare that his or her measure is valid; instead, the meaning of the measure is established by its empirical connections to other defined measures.

Simulations

The purpose of this chapter is to identify some of the psychometric issues in research linking social support and health outcomes. Most of the focus has been on the definition and measurement of health. I will now return to the measurement issues in identifying relationships between imperfect measures of health and social support.

Some of the psychometric problems associated with establishing the relationship between social support and health can be understood through simulation. The effect of low reliability on correlations has been well documented in the psychometric literature. Observed correlations between two variables are attenuated when either or both variables are measured with error; formulas that describe this relationship are available (R. Kaplan & Saccuzzo, 1993).

In our simulations, my colleagues and I made the following assumptions. First, we assumed that the maximum true correlation between social support and various outcome measures would not exceed .5. This seems reasonable, because most health outcome variables are affected by multiple sources of variability. For example, one of the variables in our simulation is blood pressure; we would not expect the true correlations between social support and blood pressure to exceed .5, because blood pressure is affected by hereditary factors, age, weight, diet, and so forth. The outcome variables selected for the simulation were chosen because they were used in a variety of studies and may reasonably be expected to correlate with social support.

Five outcome variables were chosen somewhat arbitrarily to represent different observed levels of reliability. In a similar fashion, four social support questionnaires were chosen to represent different levels of reliability. Because the simulation is for illustrative purposes only, the outcome variables chosen may or may not actually bear significant relationships to social support. But because social support has been shown to affect physical (Cohen & Syme, 1985; DiMatteo & Hays, 1981) and psychological health (Dean, Lin, & Ensel, 1981; Sarason & Sarason, 1985) and may serve as a buffer against life stress (Cohen & Wills, 1985) it seems reasonable to select outcome variables associated with health concerns.

The first outcome variable used in the simulation was the Mental Health Questionnaire of the Older American Research and Service Center

instrument (OARS). The OARS is a multipurpose assessment questionnaire for evaluating the elderly; it was developed at Duke University and has been used in a wide variety of studies. Although the psychometric data for the OARS are generally good, the test-retest coefficient for the 15-item mental health screening tool was only .32. This measure was chosen for use in the simulation because this relatively low coefficient (Fillenbaum, 1978).

Social scientists often attempt to correlate their measures with ratings by trained medical practitioners. Clinical ratings are known to be fallible, however, and are often measured with considerable error. To demonstrate this point, we chose clinical ratings of dysfunction provided by active medical practitioners for use in the simulation. The reliability obtained in a careful study by Bergner et al. (1981; see their Table 5) of these ratings of dysfunction was .41.

Our next range of reliability was taken from a measure of life stress. Although the Schedule of Recent Events (SRE) has been used successfully in many studies, early reports suggested that it had a test-retest reliability of only .55 (Holmes & Rahe, 1967). A measure of life change was chosen because many investigators hope to demonstrate a relationship between life events and social support.

Some investigators use risk factors as criteria against which to evaluate social support variables. To simulate one physical risk factor, we chose blood pressure. A variety of different studies demonstrate that blood pressure, even when measured under the most rigorous criteria, has a reported reliability of approximately .65 (Hypertension Detection and Follow-Up Program Cooperative Group, 1979). Finally, we used the Sickness Impact Profile (SIP), a widely used general health outcome measure that has a reported reliability for an interviewer-administered form of .97 (Bergner et al., 1981).

The first social support measure was the Dean et al. (1981) Social Support Scale (SSS), which was chosen for its low reliability level of .28. Various studies show a range of reliability coefficients for this measure; the .28 value was chosen because it was the lowest observed reliability coefficient. The second social support measure used for the simulation was the support need measure from Barrera's (1981a) Arizona Social Support Interview Schedule (ASSIS). The reliability coefficient for that measure was .52. The third social support measure, the Social Support Satisfaction Scale (SSSS; Blaik & Genser, 1980) has an internal consistency reliability of .69 for the short form; it was also used in this simulation. Finally, a portion of the Sarason et al. (1983) Social Support Questionnaire (SSQ), which enumerates the number of people in one's social network, was chosen because of its very high (.97) level of reliability.

The expected observed correlations between each social support measure and each outcome measure was estimated using the formula $R = .5\sqrt{r_{11}r_{22}}$, in which R is the expected observed correlations, r_{11} is the reliability of the social support questionnaire, r_{22} is the reliability of the outcome measure, and .5 is the expected true correlation. The simulation is summarized in Table 4. The entries in the table represent the expected observed correlation between each social support and outcome measure pair if the true correlation is .50. Asterisks are also used to identify those that would be statistically significant at the 0.5 level in a study with 50 respondents.

As the table demonstrates, the expected observed correlations between measures is affected by their reliability. If the true correlations between variables were .5, the SSQ would still find a statistically significant correlation when used with all outcome measures except the OARS Mental Health Questionnaire. Conversely, the SSS, with a reliability of .28, would not be able to detect a .5 correlation with any of the chosen outcome measures. In other words, it would be a frustrating effort to employ the SSS and to expect to obtain any significant correlations under our assumptions. The other social support questionnaires represent intermediate capabilities to detect correlations; for example, the ASSIS might be expected to detect a correlation with blood pressure or the SIP, but not with the SRE, the OARS measure, or clinical judgments. Again, this simulation assumes that the true correlations would be .5. Many of the relationships

Table 4. Correlations between Selected Social Support and Criterion Measures

Scale	OARS Mental Health (.32)	Clinical ratings (.41)	SRE (.55)	Blood pressure (.65)	SIP (.97)
Social Support Questionnaire (SSQ; .97) (Sarason et al., 1983)	.278	.32*	.37*	.40*	.48*
Social Support Satisfaction Scale (short form; .69) (Blaik & Genser, 1980)	.24	.27	.31*	.33*	.41*
Arizona Social Support Interview Schedule (ASSIS; support need measure; .52) (Barrera, 1981a)	.20	.23	.27	.29*	.36*
Social Support Scales (Lowenthal & Haven items; .28) (Dean et al., 1981)	.15	.17	.20	.21	.26

* $p < .05$ (.2818)

between social support and health outcomes involve measures with less than optimal reliability.

Despite this, a researcher may select a measure with moderate or even low reliability because of the instrument's ease of administration, simplicity in scoring, or appropriateness to the variable being examined. Although determining the reliability of scales is a well-established standard in terms of assessment, it may be legitimate to use a less reliable tool if the dependent measure has high enough reliability on its own to make the inquiry worthwhile. In any case, the researcher should be cognizant of the approximate reliability of measures and the potential impact of low reliability on observed correlations. Other outcome measures used in social support research (e.g., mortality) are considerably more reliable; however, they may also occur with a relatively low probability. In a prospective study, only a small proportion of the participants will die in any defined time period. As a result, sample sizes for observational studies that use mortality as an outcome often need to be in the thousands or even tens of thousands.

Summary

Several chapters in this volume describe problems in the conceptualization and measurement of social support. Lack of a consensual definition of social support has made it difficult, if not impossible, to compare studies linking social support to stress, health outcomes, and general psychological and physical well-being. Many authors attribute the problems in this field to the inadequate definitions of social support (Heitzmann & Kaplan, 1988); however, there are equally serious problems in the conceptualization and measurement of health status. The definition of health has been ambiguous for several centuries. Within the last 25 years, several groups have attempted to define health status using quantitative measures. Although progress is being made, large conceptual problems still remain.

Perhaps the best consensus on a measure of health status is for mortality. Indeed, there is some convincing evidence that social support is associated with a lower rate of mortality from cardiovascular disease. Beyond mortality, studies linking social support to health become very problematic. The number of reported symptoms is not a strong outcome measure, because symptom reporting is highly subjective, unreliable, and subject to various biases.

Another problem in research on social support and health concerns the ratio of the number of subjects to variables. It is not uncommon for

investigators to capture health status by measuring or tabulating an enormous number of indicators; some studies, for example, use as many as 100 indicators of life quality. Such approaches greatly inflate the probability of spurious findings. Studies focusing on one well-defined outcome, such as mortality, may also encounter problems in sample size if only a small number of cases actually die.

In summary, the problems in the conceptualization and measurement of social support are well recognized. As noted above, though, there are equally, if not more serious, problems in the conceptualization and measurement of health status. The literature on health status is well documented (Patrick & Erickson, 1988; Walker & Rosser, 1988). Future studies should embrace state-of-the-art technologies for both health status and social support measurements.

References

- Anderson, J. P., Bush, J. W., & Berry, C. C. (1986). Classifying function for health outcome and quality of life evaluation: Self-versus-interviewer modes. *Medical Care*, 24, 454-469.
- Barrera, M. (1981a). Preliminary development of a scale of social support. *American Journal of Community Psychology*, 9, 435-447.
- Barrera, M. (1981b). Preliminary development of a scale of social support. *American Journal of Community Psychology*, 9, 435-447.
- Benowitz, N. L., Kuyt, T., & Jacobs, P. (1984). Influence of nicotine on cardiovascular and hormonal effects of smoking. *Clinical Pharmacological Therapeutics*, 36, 74.
- Bergner, M. (1985). Measurement of health status. *Medical Care*, 23, 696-704.
- Bergner, M., Bobbitt, R. A., Carter, W. B., & Gilson, B. S. (1981). The Sickness Impact Profile: Development and final revision of a health status measure. *Medical Care*, 19, 787-786.
- Berkman, L. F. (1982). Social networks and analysis and coronary heart disease. *Advances in Cardiology*, 29, 37-49.
- Billings, A. F., & Moos, R. H. (1982). Social support and functioning among community and clinical groups: A panel model. *Journal of Behavioral Medicine*, 5, 295-311.
- Blaik, R., & Genser, S. G. (1980). Perception of Social Support Satisfaction: Scale development. *Personality and Social Psychology Bulletin*, 6, 172.
- Brandt, P. A., & Weinert, C. (1981). The PRQ—a social support measure. *Nursing Research*, 30, 277-280.
- Broadhead, W. E., Kaplan, B. H., James, S. A., Wenger, E. H., Schoenbach, V. J., Grimson, R., Heyden, S., Tibblin, G., & Gehlbach, S. H. (1983). The epidemiologic evidence for a relationship between social support and health. *American Journal of Epidemiology*, 117, 521-537.
- Brook, R. H., Goldberg, G. A., Harris, L. J., Applegate, K., Rosenthal, M., & Lohr, K. N. (1979). *Conceptualization in measurement of physiologic health in the health insurance study*. Santa Monica, CA: RAND Corporation.
- Bush, J. W. (1984). Relative preferences versus relative frequencies in health-related quality of life evaluations. In N. K. Wenger, M. E. Mattson, C. D. Furberg, & J. Elinson (Eds.), *Assessment of quality of life in clinical trials of cardiovascular therapies* (pp. 118-139). New York: LaJacc.

- Cohen, S., Mermelstein, R., Kamarck, T., & Hoberman, H. N. (1985). Measuring the functional components of social support. In I. Sarason & B. Sarason (Eds.), *Social support: Theory, research, and applications* (pp. 73-94). Dordrecht, Netherlands: Martinus Nijhoff.
- Cohen, S., & Syme, S. L. (1985). *Social support and health*. New York: Academic Press.
- Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin*, 98(2), 310-357.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281.
- Croog, S. H., Levine, S., Testa, M. A., Brown, D., Bulpitt, C. J., Jenkins, C. D., Klerman, G. L., & Williams, G. H. (1986). The effects of anti-hypertensive therapy on quality of life. *New England Journal of Medicine*, 314, 1657-1664.
- Davidson, T. N., Bowden, L., & Tholen, D. (1979). Social support as a moderator of burn rehabilitation. *Archives of Physical Medicine and Rehabilitation*, 60, 556.
- Dean, A., Lin, N., & Ensel, W. M. (1981). The epidemiological significance of social support systems in depression. *Research in Community Mental Health*, 2, 77-109.
- Diabetes Control and Complications Group (1993). The effect of intensive treatment of diabetes on the development and progression of long term complications in Insulin Dependent Diabetes Mellitus. Results of the Diabetes Control and Complications Trial. *New England Journal of Medicine*, 329 (14), 977-986.
- DiMatteo, M. R., & Hays, R. (1981). Social support and serious illness. In B. H. Gottlieb (Ed.), *Social networks and social support* (pp. 117-148). Beverly Hills, CA: Sage.
- Dupee, H. J. (1969). *The psychological examination: Cycle IV*. Washington, DC: National Center for Health Statistics.
- Fanshel, S., & Bush, J. W. (1970). A health-status index and its applications to health-services outcomes. *Operations Research*, 18, 1021-1066.
- Fillenbaum, G. C. (1978). Validity and reliability of the multidimensional functional assessment questionnaire. In *Multidimensional functional assessment: The OARS methodology* (pp. 25-35). Durham, NC: Duke University, Center of the Study of Aging and Human Development.
- Fishburn, P. (1964). *Decision and value theory*. New York: Wiley.
- Follick, M. J., Gorkin, L., Smith, T., Capone, R. J., Visco, J., & Stabein, D. (1988). Quality of life post-myocardial infarction: The effects of a transtelephonic coronary intervention system. *Health Psychology*, 7, 169-182.
- Gordon, T., Kannel, W. B., & McGee, D. (1974). Death and coronary attacks in men after giving up cigarette smoking. *Lancet*, 2, 1345-1348.
- Gotto, A. M. (1986). Interaction of the multiple risk factors of coronary heart disease. *American Journal of Medicine*, 80, 48-55.
- Greene, S. B., Aavedel, M. J., Tyroler, H. A., Davis, C. E., & Hames, C. G. (1977). Smoking habits and blood pressure change: A seven year follow-up. *Journal of Chronic Diseases*, 30, 410.
- Heitzmann, C. A., & Kaplan, R. M. (1988). Assessment of the methods for measuring social support. *Health Psychology*, 7, 75-109.
- Henderson, S., Duncan-Jones, P., Byrne, D. G., & Scott, R. (1980). Measuring social relationships: The Interview Schedule for Social Interaction. *Psychological Medicine*, 10, 723-734.
- Holbrook, J. H. (1986). *Cigarette smoking*. Washington, DC: The Brookings Institution.
- Holmes, T. H., & Rahe, R. H. (1967). The Social Readjustment Rating Scale. *Journal of Psychosomatic Research*, 11, 213-218.
- Hunt, S. M., & McEwen, J. (1983). The development of a subjective health indicator. *Sociology of Health and Illness*, 2, 231-245.
- Hypertension Detection and Follow-Up Program Cooperative Group. (1979). Five years

- findings of the Hypertension Detection and Follow-up Program: I. Reduction in mortality of persons with high blood pressure, including mild hypertension. *Journal of the American Medical Association*, 242, 2562-2571.
- Kaplan, N. M., & Stamler, J. (1983). *Prevention of coronary heart disease*. Philadelphia: Saunders.
- Kaplan, R. M. (1985a). Quality of life measurement. In P. Karoly (Ed.), *Measurement strategies in health psychology* (pp. 115-146). New York: Wiley-Interscience.
- Kaplan, R. M. (1985b). Quantification of health outcomes for policy studies in behavioral epidemiology. In R. M. Kaplan & M. H. Criqui (Eds.), *Behavioral epidemiology and disease prevention*. New York: Plenum.
- Kaplan, R. M., & Anderson, J. P. (1988). A general health policy model: Update and applications. *Health Services Research*, 23(2), 203-235.
- Kaplan, R. M., & Bush, J. W. (1982). Health-related quality of life measurement for evaluation research and policy analysis. *Health Psychology*, 1, 61-80.
- Kaplan, R. M., Bush, J. W., & Berry, C. C. (1976). Health status: Types of validity for an index of well-being. *Health Services Research*, 11, 478-507.
- Kaplan, R. M., & Saccuzzo, D. S. (1993). *Psychological testing: Principles, applications, and issues* (3rd ed.). Pacific Grove, CA: Brooks-Cole.
- Katz, S. T., Downs, H., Cash, H., & Grotz, R. (1970). Progress and development of an index of ADL. *Gerontologist*, 10, 20-30.
- Katz, S. T., Ford, A. D., Moskowitz, R. W., Jackson, B. A., & Jaffe, M. W. (1963). Studies of illness in the aged: The index of ADL. *Journal of the American Medical Association*, 185, 914-919.
- Liang, N. H., Cullen, K., & Larson, M. (1982). In search of a more perfect mouse trap (health status or a quality of life instrument). *Journal of Rheumatology*, 9, 775-779.
- Lin, N., Simeone, R., Ensel, W., & Kuo, W. (1979). Social support, stressful life events and illness: A model and an empirical test. *Journal of Health and Social Behavior*, 20, 108-119.
- McDowell, I., & Newell, C. (1987). *Measuring health: A guide to rating scales and questionnaires*. New York: Oxford.
- McFarlane, A. H., Neale, K. A., Norman, G. R., Roy, R. G., & Streiner, D. L. (1981). Methodological issues in developing a scale to measure social support. *Schizophrenia Bulletin*, 7, 90-100.
- Norbeck, J. S., Lindsey, A. M., & Carrieri, V. L. (1981). The development of an instrument to measure social support. *Nursing Research*, 30, 264-269.
- Paffenbarger, R. S., Thorne, M. C., & Wing, A. L. (1968). Chronic disease in former college students: Characteristics in youth predisposing to hypertension in later years. *American Journal of Epidemiology*, 88, 25.
- Patrick, D. L., & Erickson, P. (1988). Assessing health-related quality of life in clinical decision making. In S. R. Walker & R. M. Rosser (Eds.), *Quality of life: Assessment and applications* (pp. 9-19). London: MTP Press.
- Porrit, D. (1979). Social support in crisis: Quantity or quality. *Social Science and Medicine*, 13, 715-721.
- Procidano, M. E., & Heller, K. (1983). Measures of perceived social support from friends and from family: Three validation studies. *American Journal of Community Psychology*, 11(1), 1-24.
- Rice, R. M. (1984). Organizational work and the overall quality of life. In S. Osgamp (Ed.), *Applied social psychology annual: Applications in organizational settings*, vol. 5 (pp. 155-178). Beverly Hills, CA: Sage.
- Russell, L. B. (1986). *Is prevention better than cure?* Washington, DC: Brookings Institution.
- Sarason, I. G., & Sarason, B. R. (1985). *Social support: Theory, research, and applications*. Boston: Martinus Nijhoff International.

- Sarason, I. G., Levine, H. M., Basham, R. B., & Sarason, B. R. (1983). Assessing social support: The Social Support Questionnaire. *Journal of Personality and Social Psychology*, 44(1), 127-139.
- Schaefer, C., Coyne, J. C., & Lazarus, R. S. (1981). Health related functions of social support. *Journal of Behavioral Medicine*, 4, 381-406.
- Schafer, L. C., McCaul, K. D., & Glasgow, R. E. (1984). *Supportive and nonsupportive family behaviors: Relationships to adherence and metabolic control in persons with Type 1 diabetes*. Unpublished manuscript, North Dakota State University, Fargo.
- Seltzer, C. C. (1974). Effect of smoking on blood pressure. *American Heart Journal*, 87, 558.
- Stewart, A. L., Ware, J. E., Brook, R. H., & Davies-Avery, A. (1978). *Conceptualization and measurement of health for adults: vol. II: Physical health in terms of functioning*. Santa Monica, CA: RAND Corporation.
- Sullivan, D. F. (1966). Conceptual problems in developing an index of health. Office of Health Statistics, National Center for Health Statistics, Monograph Series II, No. 17.
- Surgeon General of the United States. (1979). *Smoking and health*. (Publication No. 79-50066). Washington, DC: Department of Health, Education, and Welfare.
- Tchobroutsky, G. (1978). Relation of diabetic control to development of microvascular complications. *Diabetologia*, 15, 143-152.
- Toshima, M. T. (1987). *Cigarette smoking cessation*. Unpublished manuscript, San Diego State University, San Diego, CA.
- Walker, S., & R. M. Rosser (Eds.). (1988). *Quality of life: Assessment and applications*. London: MTP Press.
- Wallston, B. S., Alagna, S. W., DeVellis, B. M., & DeVellis, R. F. (1983). Social support and physical health. *Health Psychology*, 2, 367-391.
- Ware, J. E., Brook, R. H., Davies-Avery, A. R., et al. (1981). Choosing measures of health status for individuals in general populations. *American Journal of Public Health*, 71, 620-625.
- Ware, J. E., & Karmos, A. H. (1976). *Development and validation of scales to measure perceived health and patient role propensity*, Vol. 2. Carbondale: Southern Illinois School of Medicine.
- Weinstein, M. C., & Stason, W. B. (1976). *Hypertension: A policy perspective*. Cambridge, MA: Harvard University Press.
- Wenger, N. K., Mattson, M. E., Furberg, C. D., & Elinson, J. (1984). *Assessment of quality of life in clinical trials of cardiovascular therapies*. New York: LaJacq.
- World Health Organization. (1948). *Constitution of the World Health Organization*. Geneva: Author.