Given that the assessment of multiple dimensions of health/well-being can create respondent burden, assessment scales that are both effective and brief hold great attraction. In this study, we used Cronbach’s alpha and correlational methods, including factor analysis, to evaluate the performance of four short scales measuring psychosocial aspects of well-being (depression, quality of life, sense of coherence, social support) in two samples of community-dwelling persons aged 75 and over ($n = 414, n = 50$). All four scales exhibited good range, high internal consistency, strong temporal reliability, and reasonable levels of construct validity. We conclude that they are practical contributors to measuring health in community-based older adults.

Key Words: Psychometrics, Depression, Quality of life, Sense of coherence, Social support

Measuring Psychosocial Aspects of Well-Being in Older Community Residents: Performance of Four Short Scales

Andrea Steiner, PhD, Kristiana Raube, PhD, Andreas E. Stuck, MD, Harriet U. Aronow, PhD, David Draper, PhD, Laurence Z. Rubenstein, MD, MPH, and John C. Beck, MD

Whether termed health, well-being, or, more colloquially, “the good life” (Lawton, 1983), the way that people feel and function is a complex construct with multiple dimensions (e.g., see Rubenstein et al., 1989). Among those dimensions are physical health, social behavior, emotional or psychological well-being, cognitive status, economic prosperity, life satisfaction, role functioning, and overall quality of life or well-being. Within each dimension, an array of subdomains have been described (Wenger, Mattson, Furberg, & Elinson, 1984).

It is easy to accept, in principle, the advantages of using multiple measures of health and well-being in a single study (Dew & Simmons, 1990). However, real world constraints of time and money, as well as concerns about respondent burden, often make it infeasible to include a full set. This is especially so in observational studies, where quality of life, for instance, may be only one of a much larger set of variables under study.

For these reasons, short scales hold a great attraction. First, they enable researchers to examine a larger number of health domains than would otherwise be possible. This allows for comparisons across domains, examination of their interactions, and assessment of the differential effects of particular interventions. Finally, they can reduce respondent burden, which may be particularly important with older subjects (Arnold, 1991). All such claims, however, are contingent on the scales’ reliability and validity.

In this article, we focus on the performance of four short scales that, following Lawton’s metaconstruct of the good life (see Figure 1), purport to measure some component of either psychological well-being or perceived quality of life. They are:

(a) The 13-item (short) version of Antonovsky’s Sense of Coherence scale (Antonovsky, 1987; hereafter SSOC);

(b) A 6-item version — never used before — of the social support scale developed as part of RAND’s Medical Outcomes Study (Sherbourne & Stewart, 1991; hereafter SSUP);

(c) The 11-item Reintegration to Normal Living scale (Wood-Dauphinee, Opzoomer, Williams, Marchand, & Spitzer, 1988; hereafter RNL), which has been proposed as a proxy for quality of life in
older people (Wood-Dauphinee & Williams, 1987); and

(d) The 15-item version of the Geriatric Depression Scale (Sheikh & Yesavage, 1986; hereafter SGDS).

On face validity grounds, both the depression measure and the sense of coherence scale belong under the rubric of “psychological well-being.” The others seem to contribute to “perceived quality of life.” In contrast, something like the Basic and Instrumental Activities of Daily Living (BADL and IADL) scales (Lawton, Moss, Fulcomer, & Kleigan, 1982) would belong in the category of “behavioral competence” and, presumably, a social network scale such as Lubben’s (1988) would be a part of an “objective environment” category. The instruments under review here are not the sole measures of these health dimensions, but they do share the virtue of brevity.

In this study, we analyzed the scales’ performance in a cohort of relatively healthy community dwellers aged 75 and over, primarily by assessing their variability, reliability and validity. Where relevant, we also considered whether they perform as well (a) in their short forms as in the longer versions; (b) in a sample aged 75-plus as in younger cohorts; and (c) in mainly healthy people as in patient populations.

Methods

Instruments

Like the 29-item original version, the 13-item SSOC purports to tap a stable psychological orientation hypothesized to be associated with both vitality and the capacity to identify and draw upon internal or external resources in the face of stressful situations. The scale’s three sub-dimensions — comprehensibility, manageability, and meaningfulness — are retained in the shorter form, going from 11 to 5, 10 to 4, and 8 to 4 items, respectively. The SOC scales have been used around the world on a wide range of populations (Antonovsky, 1993), but studies of elderly people have been infrequent (Antonovsky, Adler, Sagy, & Visel, 1990; Coe, Romeis, Tang, & Wolinsky, 1990; Sagy & Antonovsky, 1990; Sagy, Antonovsky, & Adler, 1990). All of the studies used the longer (29-item) version of the scale rather than the 13-item short form, and defined “elderly” either as retirement age or 55 years old or older. Moreover, Coe et al.’s sample included only men.

Like the 19-item reference instrument, the SSUP assesses social support by measuring its perceived availability, with a focus on different functional, rather than structural, aspects of support. Disaggregated, it measures tangible support (e.g., a ride to the doctor if needed), affection, positive interaction (e.g., someone with whom to do something fun), informational support, and emotional support. All the subscales in the Medical Outcomes Study (MOS) have extremely high reliability and validity; however, we are the first to use a 6-item version of the social support scale. These are items 2, 5, 16–18, and 20 (note that item 1 refers to social network size, not to social support). One item was selected for each aspect of support. The MOS scales were primarily designed for, and tested on, a somewhat younger population (aged 18 and over; mean age, 55) than observed here.

With its focus on functional capacity and compensation for loss, the RNL would seem to have excellent face validity for an elderly sample. However, it was initially developed for younger patients who had suffered a disabling trauma from which they would not fully recover (Wood-Dauphinee et al., 1988) — a construct that may differ importantly from the more gradual declines associated with aging. The 11-item RNL is the original rather than a shortened version. Arnold (1991) notes that although the scale appears to be promising for use with elderly populations, more work is needed to establish its reliability and validity. For instance, the number of elderly used in initial assessments of performance was quite low. Test-retest reliability was not reported at all.

Finally, the 30-item Geriatric Depression Scale (Yesavage & Brink, 1983) is a well-known and much-studied instrument developed expressly to screen for clinical depression in the elderly. A partial list of tested samples includes: community-based elderly (Dunn & Sacco, 1989), institutionalized elderly (Parmalee, Lawton, & Weinstein, 1988), people diagnosed with mild to moderate levels of dementia (O’Riordan et al., 1990), hospitalized elderly with medical illness (Lyons, Strain, Hammer, Ackerman, & Fulop, 1989), elderly receiving outpatient care (Norris, Gallagher, Wilson, & Winograd, 1987), and the “old-old” (Weiss, Nagel, & Aronson, 1986). The SGDS was developed by selecting the 15 questions that had the highest correlation with depressive symptoms in the original depression scale validation studies. However, its own validation study used a sample of only 35 subjects, half of whom had been diagnosed with major depression. Thus its applicability to the sample of interest here, or to research for purposes other than patient screening, cannot be assumed.
Sample

To complete this study, we enrolled two samples. Both groups were composed of residents of Santa Monica, California, a community with a population of 80,000 adjacent to Los Angeles and itself fairly urban and also relatively rich in community resources. The research occurred in the context of a community-based randomized controlled trial that assessed the effects of a three-year in-home preventive health care intervention emphasizing comprehensive geriatric assessment and self-care (Rubenstein et al., 1994). Inclusionary criteria were determined according to subjects’ potential to benefit from a preventive intervention.

Participants had to be at least 75 years old, not living in a nursing home or board and care facility, not diagnosed with a terminal illness, and not severely impaired either functionally or cognitively. We identified people as “severely impaired” if they missed three or more of the ten questions on the Mental Status Questionnaire (Kahn, Goldfarb, Pollack, & Peck, 1960) or could not perform two or more of the five Basic Activities of Daily Living (Lawton et al., 1982).

The first sample, hereafter referred to as the “primary” sample, consisted of all participants in the above-mentioned study. Recruitment goals were to get a representative sample via a health-neutral source. Although a central population-based registry, such as exists in some other countries (Hendriksen, Lund, & Stromgard, 1984; Vetter, Jones, & Victor, 1984) would have been ideal, we had no such option and chose instead to recruit from two accessible population subgroups.

The first were those individuals listed as aged 75 or older on the current voter registration list. Voters were first approached by telephone (966 functioning numbers; response rate 37%); letters of invitation were sent to the rest (765 delivered successfully; response rate 11%). The second source was the remaining population of 2,790 Santa Monica residents aged 75 and over, who either were not registered voters or had declined to state their age, and who volunteered to participate in the project.

After exclusions, 414 people formed the primary sample. Mean age was 81.2 years. Nearly 70% were female, 64% lived alone, and 96% were Caucasian, which reflected the racial distribution in Santa Monica for this age group. Some 80% had completed high school; however, 37% reported incomes below the area’s poverty line. Subjects reported an average of 3.2 chronic conditions, with a range from 0 to 12. IADL scores tended to be high (M = 23.2 out of 27) but ranged as low as 12 (5% of primary sample).

We also recruited a second group, hereafter called the “reliability sample,” specifically to examine the instrument’s test-retest reliability. Reliability sample subjects (n = 50) were selected according to three approaches. First, we solicited participants from a local senior center’s health education program. Second, we directly approached visitors to, and volunteers at, that center. This group included patients seeking medical or social services at the center’s various clinics. Finally, we made recruitment speeches at church group meetings, nutrition sites, and retirement homes with congregate dining. In this way, we were able to achieve a mix in terms of the subjects’ health and functional status comparable to what exists in the community. Our choice of a sample size for this part of the research was dictated by precedents in the literature (e.g., Nelson, Landgraf, Hays, Wasson, & Kirk, 1990; Parkerson, Broadhead, & Tse, 1990).

Primary and reliability samples matched each other well except in the area of gender, where the reliability group was disproportionately female. Reweighted analyses produced substantially the same results as unweighted analyses; for simplicity we present unweighted findings below.

Data Collection

In both cases, data were collected by trained interviewers who followed a structured format in a face-to-face meeting. All primary sample subjects (n = 414) were interviewed a single time, in their homes. They were asked a range of questions about health-related topics, of which the scales under study here were a subset. The entire interview lasted approximately 60 minutes. Reliability sample subjects (n = 50) were interviewed twice, by the same interviewer at the same site, with 7 to 14 days between interviews. Usually that site was the senior center; occasionally it was the respondent’s home. These interviews consisted solely of the four scales discussed in this article. All four instruments were administered at each interview, in the same order. To screen for appropriate change in subjective ratings, subjects were asked whether the time between interviews had been unusual, with a positive response prompting a probe. No observations had to be excluded as a result. Reliability sample interviews were timed. Because the scales were interspersed throughout the primary sample’s longer interview, no precise timings were obtained there.

The level of missing data was very low, perhaps because all data were obtained in face-to-face interviews. For the reliability sample, every subject responded to each item of all four scales. For the primary sample, 2.8, 0.9, 1.4, and 1.9% of respondents, respectively, missed more than two items from the SOSC, SSUP, RNLI, and SGDS. Anecdotally, interviewers reported some difficulties administering the SOSC, primarily because the positive and negative anchors varied from item to item.

Analyses

To evaluate whether these scales were capable of eliciting a full range of values from a community-based sample aged 75 and older, we compared the actual ranges observed in both our samples to the possible ranges defined by the scales’ outer limits. (For ease of interpretation, all instruments have been transformed to a 100-point scale, with 100 representing the healthier end of the spectrum.) To assess whether the data were skewed, we examined the
scales' distributions, comparing mean scores to medians and modes.

Next, to examine the scales' reproducibility, we looked at internal consistency and test-retest reliability. To assess internal consistency in both the primary and reliability samples, we used Cronbach's alpha coefficient (Cronbach, 1951). For the SGDS, which has dichotomous outcomes, we used the Kuder-Richardson formula 20 (Carmines & Zeller, 1979). KR20 is a special case of Cronbach's alpha that evaluates measures with zero/one outcomes. By convention, alpha values between .7 and .9 are considered highly reliable for group comparisons (Stewart, 1990).

Cronbach's alpha increases as either or both of the average inter-item correlation and the number of items in the instrument increase. Thus, longer instruments can compensate for some weaknesses simply by adding more items that are related to the concept being measured. In contrast, and virtually by definition, short instruments cannot do the same. Instead, they must have especially strong inter-item correlations.

We used two methods to examine temporal reliability. In both cases, the analyses were confined to the reliability sample, so as not to disrupt the larger study's research protocol. We began by obtaining the correlations between test and retest (time 1 and time 2, hereafter T, and T1). We also divided the sample according to age group to see whether a measure's performance was diminished in respondents aged 85 and older. In these analyses, we identified two outliers — one for the SSUP (score changed 10 points out of a possible 30), another for the SGDS (score changed 7 points out of a possible 14). Given that another sample of 50 could, conceivably, include a few subjects with unusual responses, we present results in the form of a sensitivity analysis, both with and without the two anomalous values (Freedman, Pisani, & Purves, 1978).

Test-retest correlations are the conventional criteria of temporal reliability, but can overestimate the true strength of the relationship (Bland & Altman, 1986) by confounding within-subject change scores (T1 - T2) with differences between subjects in their "true" scores (estimated by (T1 + T2)/2). Therefore, we also examined the correlations between these quantities. Whereas strong correlations are to be desired between T1 and T2, weak correlations are preferred in the "true" vs change score analyses, because they would indicate that respondents' starting values — having a weak rather than a strong SSOC, for instance — did not influence the stability of their answers.

Similarly, we compared the standard deviations (SDs) of respondents' "true" and change scores. Here, a measure has greater reliability if the "true" score SD is larger than the change score SD; that is, the expectation is that subjects will differ from each other, but it is hoped that they will themselves be relatively consistent in their responses. It is also hoped that the magnitude of change will itself be small. We verified this by examining the means of both T1 - T2 and the absolute values of T1 - T2.

Finally, we used the primary sample to assess construct validity because we could take advantage of the longer battery of questions. To explore the relationships among the scales, and to see how they fit into the "good life" paradigm, we performed a factor analysis on the correlation matrix for seven variables: the four scales studied here, BADLs and IADLs, and Lubben's social network scale. The latter three variables were included to help clarify the distinctive contributions made by the scales of principal interest. To assess the sensitivity of the findings to the choice of factoring method, we used both principal components analysis and maximum likelihood.

To assess construct validity within each scale, we used what Read, Quinn, and Hoefler (1987) called differential convergent validity, and Stewart (1990) termed convergent/discriminant validity. In this form of validity-testing, correlations between the measure of interest and other constructs or variables are examined. If valid, the measure of interest should be more strongly correlated with related constructs than with unrelated ones. On the basis of literature review and clinical expertise, we formed the following hypotheses:

1. With the SSOC, where an individual's orientation to life and its stressors is hypothesized to play a significant role in one's "location and movement on the health ease/dis-ease continuum" (Antonovsky, 1993), we expected the strongest correlations to occur with measures of physical health — especially subjective or attitudinal ones. We expected positive correlations between SSOC and both perceived health and physical function and negative correlations between SSOC and number of both bed days and chronic medical conditions.

2. With SSUP, we hypothesized that the structural variables of age, gender, and living arrangement, and the attitudinal variable of perceived health status, would be more strongly correlated with this measure of social support than the physical health measures of bed days or functional status (Kasper, 1988). Based on Sherbourne and Stewart's finding (1991) that the original 19-item MOS social support scale was significantly and negatively correlated with a measure of "physical symptoms," we hypothesized a similar — negative, relatively strong — relationship between SSUP and the number of chronic medical conditions. Those in poor health might need more social support, but it does not necessarily follow that they will receive it (Stoller & Earl, 1983).

3. For both the RNL and SGDS, we hypothesized stronger correlations between these scales and health-related variables than with age or gender. We also hypothesized that people living alone would report worse quality of life and more depression than those not living alone (Aneshensel, Frerichs, & Huba, 1981; Gurland, Wilder, & Berkman, 1988; Kennedy et al., 1989; Mirowsky & Ross, 1992).

Table 1 lists the specific hypothesized relation-
ships between the scales under study here and these other variables. Although the SGDS is generally scored so that a higher score is associated with more depression, we have reversed this so that the four scales are consistent in direction. Thus, as presented in Table 1 and elsewhere in the paper, a higher SGDS score signifies less depression.

**Results**

**Range and Distribution of Responses**

Table 2 presents the univariate findings that pertain to variability and distribution. Reliability sample scores \( (n = 50) \) refer to the initial measurement point. On the whole, the two study samples exhibited fairly similar values and distributions, although, not surprisingly, the ranges and SDs were somewhat smaller in the reliability sample. Of these four scales, the SSOC was closest in distribution to symmetry. In general, the scales’ median and modal values were higher than their means, indicating that subjects’ responses tended to concentrate at the healthier end of the spectrum. Even so, variability was high in all scales.

**Reliability.** — As columns 3 and 4 of Table 3 indicate, the alpha values for all four scales were high, ranging from .75 to .86 in the two groups (columns 5 and 6 will be discussed later). Regarding temporal reliability, test-retest correlation coefficients were also high (see Table 4). Whether the two single outliers were included or set aside, the correlations between Time 1 and Time 2 were robust in the SSOC, SSUP, and RNL. Subgroup analyses by age (Table 5) showed that the scales performed as well for the oldest-old — those aged 85 and over — as for the two older age groups.

### Table 1. Convergent/Discriminant Validity: Hypothesized Relationships Among Variables

<table>
<thead>
<tr>
<th>Scale</th>
<th>Age</th>
<th>Male</th>
<th>Adult</th>
<th>Bed</th>
<th>Days</th>
<th>IADLs</th>
<th>Chronic Conditions</th>
<th>Perceived Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSOC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSUP</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>RNL</td>
<td>0</td>
<td>0</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SGDS</td>
<td>0</td>
<td>0</td>
<td>−</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
</tbody>
</table>

Notes: 0 stands for weak correlation hypothesized; +, for stronger positive correlation hypothesized; −, for stronger negative correlation hypothesized. SGDS scores are higher when subject is less depressed, i.e., higher scores signify better health.

### Table 2. Variability and Distribution of Study Scales (Primary and Reliability Samples)

<table>
<thead>
<tr>
<th>Scale ( (n) )</th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>SD</th>
<th>Observed Range</th>
<th># Possible Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSOC ( (n = 414) )</td>
<td>77.6</td>
<td>80</td>
<td>77</td>
<td>13.8</td>
<td>31–100</td>
<td>79</td>
</tr>
<tr>
<td>SSOC ( (n = 50) )</td>
<td>72.6</td>
<td>75</td>
<td>78</td>
<td>13.6</td>
<td>35–95</td>
<td></td>
</tr>
<tr>
<td>SSUP ( (n = 414) )</td>
<td>66.7</td>
<td>71</td>
<td>100</td>
<td>27.5</td>
<td>0–100</td>
<td>25</td>
</tr>
<tr>
<td>SSUP ( (n = 50) )</td>
<td>67.1</td>
<td>75</td>
<td>79</td>
<td>24.2</td>
<td>4–100</td>
<td></td>
</tr>
<tr>
<td>RNL ( (n = 414) )</td>
<td>84.0</td>
<td>90</td>
<td>100</td>
<td>17.6</td>
<td>13–100</td>
<td>100</td>
</tr>
<tr>
<td>RNL ( (n = 50) )</td>
<td>88.5</td>
<td>91</td>
<td>100</td>
<td>5.9</td>
<td>59–100</td>
<td></td>
</tr>
<tr>
<td>SGDS ( (n = 414) )</td>
<td>80.3</td>
<td>87</td>
<td>100</td>
<td>18.6</td>
<td>20–100</td>
<td>16</td>
</tr>
<tr>
<td>SGDS ( (n = 50) )</td>
<td>86.7</td>
<td>93</td>
<td>93</td>
<td>12.7</td>
<td>33–100</td>
<td></td>
</tr>
</tbody>
</table>

Note. To facilitate comparisons, all scales were transformed to a 0–100 point range using the transformation \( (actual \text{ score} - \text{min}) / (\text{max} - \text{min}) \), where min and max define the possible (not actual) range of the scale. Before transformation, actual SGDS scores were subtracted from the maximum possible so as to reverse the direction of the scoring. In all scales, then, a higher score indicates greater well-being.

### Table 3. Reliability of Scales: Internal Consistency, Cronbach’s Alphas

<table>
<thead>
<tr>
<th>Scale</th>
<th>Primary Sample ( n = 414 )</th>
<th>Reliability Sample ( n = 50 )</th>
<th>Short Version, Other Samples</th>
<th>Long Version, All Ages</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSOC</td>
<td>13</td>
<td>.77</td>
<td>.80</td>
<td>.89</td>
</tr>
<tr>
<td>SSUP</td>
<td>16</td>
<td>.65</td>
<td>.86</td>
<td>.97</td>
</tr>
<tr>
<td>RNL</td>
<td>11</td>
<td>.63</td>
<td>.76</td>
<td>.90</td>
</tr>
<tr>
<td>SGDS</td>
<td>15</td>
<td>.78</td>
<td>.75</td>
<td>.91</td>
</tr>
</tbody>
</table>

*Unweighted mean of findings from 16 short-form studies (range = .74–.91) and from 26 long-form studies (range = .82–.95).

*Sherbourne and Stewart (1991) assessed the 19-item MOS social support scale on an ambulatory patient population \( n = 2,987 \).

*Wood-Dauphinee et al. (1988) reported a Cronbach’s alpha for 80 patients.

*Unweighted mean of findings reported by Dunn and Sacco (1989); Lyons et al. (1989) (2 tests); Parmalee et al. (1988); and Yesavage and Brink (1983).

### Table 4. Temporal Reliability of Scales: Correlation between Test and Retest

<table>
<thead>
<tr>
<th>Scale</th>
<th>Reliability Sample ( n = 50 )</th>
<th>2 Outliers Dropped ( n = 48 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSOC</td>
<td>.86</td>
<td>.85</td>
</tr>
<tr>
<td>SSUP</td>
<td>.89</td>
<td>.91</td>
</tr>
<tr>
<td>RNL</td>
<td>.83</td>
<td>.84</td>
</tr>
<tr>
<td>SGDS</td>
<td>.75</td>
<td>.85</td>
</tr>
</tbody>
</table>

### Table 5. Test-Retest Correlations by Age Group (Reliability Sample)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Correlation for Total ( n = 22 )</th>
<th>Age 75–79 ( n = 14 )</th>
<th>Age 80–84 ( n = 14 )</th>
<th>Age 85+ ( n = 14 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSOC</td>
<td>.85</td>
<td>.68</td>
<td>.84</td>
<td>.91</td>
</tr>
<tr>
<td>SSUP</td>
<td>.91</td>
<td>.91</td>
<td>.97</td>
<td>.82</td>
</tr>
<tr>
<td>RNL</td>
<td>.84</td>
<td>.82</td>
<td>.93</td>
<td>.76</td>
</tr>
<tr>
<td>SGDS</td>
<td>.85</td>
<td>.71</td>
<td>.93</td>
<td>.82</td>
</tr>
</tbody>
</table>

\( *n = 13 \); with single outlier included, correlation = .89.

\( *n = 21 \); with single outlier included, correlation = .37.

58 The Gerontologist
younger age groups. In fact, one of the strongest correlations observed was between test and retest of the SSOC among the 85-plus. However, these results should be taken as suggestive rather than definitive because the numbers of observations in each age comparison group are small.

The more conservative tests of temporal reliability support the finding that all the scales reviewed perform well in this regard (Table 6). The magnitude of change was small, with means close to zero. (The average absolute deviations ranged narrowly from 5 to 7 points on a common 100-point scale.) Change score SDs were consistently and substantially smaller than “true” score SDs, and correlations between respondents’ estimated true scores and their change scores were low.

Validity. — In the factor analysis, the principal components (PC) and maximum likelihood factoring methods produced similar results, except for a minor variation in the order of the factors extracted: three factors retained, with the PC approach capturing 77% of the overall standardized variation in the seven relevant variables. Table 7 presents results from PC using varimax rotation.

The first factor is essentially a weighted sum of three of the four scales studied here — SSOC, SSUP, and RNL — perhaps placing it conceptually at the intersection of the upper two circles in Figure 1 (psychological well-being and perceived quality of life). The second factor consists mainly of the BADL and IADL scales, locating it firmly in the circle representing behavioral competence. The third factor combines SSUP and social network, presumably at the intersection of the two right-hand circles (perceived quality of life and objective environment).

Table 8 presents the results of the differential convergent validity assessment. This may be compared directly with Table 1, which listed the relationships hypothesized. In general, though not for SSUP, we expected the variables at the left hand side of the table to have weaker correlations than those on the right. For the most part, we were able to confirm our hypotheses. The exceptions are as follows (there was one per instrument):

- An unpredicted negative correlation between age and RNL. Though not very large, its magnitude was greater than that of other variables expected to be unrelated to the scales, and was statistically significant.
- No significant relationship between gender (being female) and SSUP.
- No significant relationship between living arrangement and SGDS; we had expected living alone to be positively associated with depression.
- A rather strong correlation between chronic medical conditions and SSOC, but positive rather than negative as had been expected.

In all other cases — 24 out of 28 — the correlations were of the relative magnitudes hypothesized, with the associations in the expected direction. Furthermore, since subject recruitment criteria excluded extremely frail individuals, the size of the correlations observed here may be taken as a likely lower bound for the scales' construct validity in the wider community-based older population; that is, the magnitude of these correlations could well increase considerably if more liberal selection criteria had been used. Thus, we conclude that all four scales have reasonably good construct validity in an elderly community-based population.

Comparison with Longer Forms/Other Populations. — Finally, wherever possible, we compared our results with those reported for longer versions of the scales, or for different populations. In both cases and for all four scales, we found that, on average, internal consistency coefficients were somewhat higher in those situations than those reported above.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Rotated Factor Loadings</th>
<th>Uniqueness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>SGDS</td>
<td>+.80 (.24)</td>
<td>+.27 (.27)</td>
</tr>
<tr>
<td>SSUP</td>
<td>+.24 (.01)</td>
<td>+.86 (.20)</td>
</tr>
<tr>
<td>SSOC</td>
<td>+.89 (.09)</td>
<td>+.91 (.19)</td>
</tr>
<tr>
<td>RNL</td>
<td>+.64 (.43)</td>
<td>+.37 (.27)</td>
</tr>
<tr>
<td>BADL</td>
<td>-.02 (.85)</td>
<td>+.10 (.27)</td>
</tr>
<tr>
<td>IADL</td>
<td>+.24 (.82)</td>
<td>-.03 (.27)</td>
</tr>
<tr>
<td>NETWORK</td>
<td>+.12 (.07)</td>
<td>+.88 (.20)</td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>3.04</td>
<td>1.36</td>
</tr>
</tbody>
</table>

| Scale | Age Male Living Alone Bed Days IADLS Chronic Conditions Perceived Health |
|-------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| SSOC  | .03 .05 -.08 -.12 .17 .31 .34 | SSUP -13*.07 -.36 -.03 .07 -.22 .23 | RNL -11*.04 -.14 -.16 .47 -.32 .45 | SGDS .09 .06 -.05 -.15 .32 -.32 .52 |

*Denotes statistical significance at p = .05 or less.
(see columns 5 and 6 of Table 3). Evidence regarding temporal reliability was scanty. Of the four scales, only the SGDS had published test-retest data. Using a ‘within one month’ time frame between T, and T2, Cwikel and Ritchie (1989) reported a correlation of .40, which was considerably weaker than our results.

We found greater opportunity for comparison in the area of validity. In particular, there is good published work on the convergent/discriminant validity of the SOC and MOS social support scales. Antonovsky (1993) reported findings from 11 studies that examined correlations between SOC and measures of health and well-being, including global health status, functional ability, and bed days. Correlations between SOC and health status were all positive, ranging from .19 to .46 (consistently with our result of .34). The reported correlation between SOC and functional ability (Carmel, Anson, Levenson, Bonneh, & Maoz, 1991) was .12, again quite similar to our .17. Finally, Coe et al. (1990) reported a correlation of -.31 between SOC and bed days, as compared to our weaker, but also negative, correlation of -.17. There were no reports on the correlation between SSOC scores and chronic medical conditions. These results obtain for long and short versions alike, with a wide spectrum of study populations.

For the SSUP, the primary reference article is Sherbourne and Stewart’s analysis (1991) of the original MOS scale’s psychometric properties. The authors reported associations between social support and physical functioning, physical symptoms, and current health, of .11, -.23, and .22, respectively. Our correlations between the 6-item SSUP and physical functioning, chronic conditions, and global health were remarkably similar (.07, -.22, and .23). Although their sample was generally younger than ours, it was composed of patients with chronic conditions — a group perhaps not very different from our own target population.

Wood-Dauphinee et al. tried to predict RNL score distribution from a set of variables including living arrangement, which we also examined. Using a sample of younger patients, they found no significant relationship between the variables, whereas we found a small but statistically significant correlation between living alone and RNL score.

For the SGDS, validity-testing generally pertained to criterion-related validity (Aldey, Austin, & Sturgeon, 1989; Cwikel & Ritchie, 1988, 1989; Sheikh & Yesavage, 1986), so cannot be directly compared with our analysis. However, our findings bear a striking resemblance to the results of Blazer, Hughes, and George’s (1987) epidemiological study of the prevalence of depression in community-based elderly. With a larger sample and using measures of depression other than the GDS, that study found that among those aged 75 and older, 8.5% suffer from either major depression, dysthymia, or symptomatic depression, and 19.8% have either dysphoria, mild depression, or anxiety. In our primary sample, 8.2% scored 8 or higher on the SGDS, indicating serious depression, and 18.8% scored higher than 5 but less than 8, suggesting mild or moderate levels of depression. Thus, it would appear that the SGDS does a good job of detecting depression in an elderly community-based population.

As for time savings, Antonovsky (1993) reported that the 29-item SOC scale takes between 15 and 20 minutes to complete, whereas the 13-item version takes 10 to 15 minutes. Hyer and Blount (1984, in Aldey et al., 1989) report that the GDS takes 8 to 10 minutes to complete, while Kavan, Pace, Potonrotto, and Barone (1990) suggested a range of 10 to 15 minutes. In contrast, Sheikh and Yesavage (1986) reported that, using an interview format, 5 to 7 minutes were needed to complete the SGDS. (No estimates were available for SSUP or RNL.) Our research confirms time savings reported elsewhere: on average, the four scales took 18 minutes to administer (the range was 12–23 minutes).

Discussion

We examined the psychometric characteristics of four short scales that measure different aspects of psychological well-being and perceived quality of life, which until now had not been assessed for their performance in a sample of community-based people aged 75 and over. Short instruments have great potential in gerontological health research, because they may reduce respondent burden and thus make measurement of complex concepts more feasible. In this section, we comment on the central findings.

In general, our subjects varied across a large proportion of the possible response range in all four scales, although values clustered to some extent at the high end of the continuum. This is not necessarily problematic, since the scales are simply reflecting the generally better health of community-based older persons relative to institutionalized or patient populations. It might be difficult, using only the scales studied here, to document modest health gains or decrements in small community-based samples, but that would be true of any measures designed to assess well-being in a broad spectrum of people (Ware, Nelson, Sherbourne, & Stewart, 1992). Indeed, Kutner, Ory, Baker, Schechtman, Hornbrook, and Mulrow (1992) advise the use of multiple measures in part to cope with this specific issue.

We found that the scales were quicker to complete than longer forms, and that reliability was maintained at sufficiently high levels to justify their use in group comparisons. We found no evidence that subjects aged 85 and older had any more difficulty providing consistent responses to the questionnaires than younger subjects.

Regarding validity, we found that differential convergent validity was generally good for all four measures, insofar as 24 out of 28 correlations were stronger with variables hypothesized to be directly related to well-being than with variables expected not to be. The moderately strong correlation between quality of life and functional status meets face validity criteria, but may be particularly noticeable with this instrument, because the RNL was devel-
oped in a rehabilitation context. Thus the scale may value functional ability more highly than other aspects of quality of life.

On the lack of association between depression and living alone, we would speculate that in this age group (75-plus) living alone provides a sense of greater independence, which may in turn outweigh potential feelings of isolation. Our finding that men and women did not differ in their perceived levels of social support may be due to the fact that a higher proportion of men in our study than women were married, with spouses presumably providing significant amounts of social support. Simple correlation analyses would not adjust for such a difference in sample composition. Nor would they prevent the age variable from proxying for some aspects of health and function; this might explain the observed negative association between age and RNL.

The "good life" paradigm seems a suitable conceptual framework for assessing a complex concept; however, it interested us that in this study, much of the action was in the overlap between circles. For instance, the social support scale did not differentiate itself from the social network scale, but met at the intersection of objective and subjective assessments of a common phenomenon. On the other hand, all four scales were clearly distinguished from measures of physical health and other objective indicators of well-being.

A major implication of this work is that researchers whose operational definition of "older" is shifting upward can be reassured that at least some of the measures developed for younger populations apply well to the 75-plus population. It is logical enough for the scientists who create measures of well-being to target fairly broad populations. At the same time, reliability and validity tested on samples with only small numbers of the older old provide little assistance to those who must decide which instruments to select in their own studies of health and aging in long-lived individuals.

These encouraging results should be interpreted with caution. The study samples matched the national household population in terms of age and gender distribution, and the local population aged 75 and over in terms of racial balance, but were predominantly Caucasian and better educated than national norms would predict. To extend our findings, it would be useful to test our results in other environments and on different subgroups of older people.

We did not examine the scales' responsiveness, i.e., their capacity to capture changes; this was beyond the scope of the study. There is a further issue of how to interpret observed changes in the absence of a "gold standard" criterion (Fitzpatrick et al., 1992). Advances in this area would be most welcome. Spector, Katz, Murphy, and Fulton (1987) have made progress in giving a clinically meaningful hierarchical structure to BADL and IADL information. It is more challenging, but equally important, to find tangible and socially relevant translations of change information in scales like these.

It must be reiterated that all results reported here pertain to the usefulness of these measures as psychometric, not clinical, tools. The high reliability coefficients (consistently over .7) are considered excellent for making group comparisons. For individual comparisons (i.e., for clinical application), reliability must consistently exceed .9. Cwikel and Ritchie (1988) suggested that for purposes of clinical screening, even slight indications of depression according to the SGDS should trigger more careful professional assessment.

In considering how close older people come to realizing "the good life," researchers, analysts, and decision-makers will probably continue to depend on both multidimensional instruments, which cover many facets of health in a single questionnaire, and on a set of unidimensional measures that cover particular domains relevant to the questions under study (Fletcher, Dickinson, & Philip, 1992). On one hand there might be, as an example, the quality adjusted life year (Williams, 1985). The reduction embodied by this measure, in which quality of life is quantified as a number between 0 and 1, with death as 0 and perfect health as 1, should be enormously appealing to policy makers who would be responsible for difficult rationing decisions at the local or national level. On the other hand, the use of multiple assessment tools allows researchers to go where no single "health number" can: to understand the dynamics of well-being under a variety of circumstances, to sort out which aspects of health different interventions actually target, and to develop successful preventive and therapeutic approaches to caring for the elderly. Our results indicate that short scales hold promise to support reliable and valid multidimensional assessment of complex constructs without undue respondent burden.

References


 properties of the geriatric depression scale among the institutionalized aged. Paper presented at the annual conference of the American Psychological Association.


Received November 27, 1994
Accepted August 29, 1995