Proxy Sources for Information on Nursing Home Residents’ Quality of Life

Robert L. Kane,1 Rosalie A. Kane,1 Boris Bershadsky,1 Howard Degenholtz,2
Kristin Kling,3 Annette Totten,1 and Kyoungae Jung1

1University of Minnesota, Minneapolis. 2University of Pittsburgh, Pennsylvania. 3St. Cloud State University, Minnesota.

Objectives. This study explores how well staff and family proxies’ reports on selected quality-of-life (QOL) domains (comfort, dignity, functional competence, privacy, meaningful activity, food enjoyment, relationships, security, and autonomy) correspond to residents’ own reports.

Methods. We compared QOL domain scores for nursing home residents and 1,326 staff proxies and 989 family proxies at the individual and facility level using means, Pearson correlation statistics, and intraclass correlations. Regression models adjusted for residents’ age, gender, length of stay, ability to perform activities of daily living, and cognition.

Results. For each domain in more than half the cases, proxy means were within 1 SD of the resident means. Resident and family proxy individual reports for selected domains were correlated at 0.14 to 0.46 (all p < .000). Resident and staff proxy individual reports were correlated at 0.13 to 0.37 (all p < .000). Correlation of mean levels by facility for staff proxies was 0.26 to 0.64 (generally p < .05) and for family proxies 0.13 to 0.61 (p < .01 except for one domain).

Discussion. Although staff and family proxy domain scores are significantly correlated with resident scores, the level of correlation suggests they cannot simply be substituted for resident reports of QOL. Determining how proxy reports can be used for residents who cannot be interviewed at all remains an unresolved challenge.
functioning (although proxies tended to identify more impairment than older people) and that agreement was also good with regard to assessments of overall health, chronic physical conditions, and physical symptoms. Agreement on what the older person preferred for health states, the older person's experience on depressive symptoms, and psychosocial well-being was low to moderate.

Numerous studies have compared measures of functioning, mood, preference, and QOL for older people with specific conditions and in specific circumstances to proxy assessors, including for cancer patients (Curtis & Fernsler, 1989; Moinpour, Lyons, Schmidt, Chansky, & Patchell, 2000; Sigurdardottir, Brandberg, & Sullivan, 1996; Snoeuv et al., 1997; Snoeuv, Aaronson, Sprangers, Detmar, Wever, & Schornagel, 1999), epilepsy (Hays et al., 1995), stroke (Dorman, Waddell, Slattery, Dennis, & Sandercock, 1997), coronary surgery (Page, Verhoef, & Emes, 1995), intensive care units (Capuzzo, Grasselli, Carrer, Gritti, & Alvisi, 2000), elderly outpatients in general (Epstein, Hall, Tognetti, Son, & Conant, 1989), persons with developmental disability (Stancliffe, 1999), and even persons with Alzheimer's disease (Hickey & Bourgeois, 2000; Novella et al., 2001). A meta-analysis comparing physician and patient judgments about QOL showed substantial discrepancies (Janse, Gemke, Uiterwaal, van der Tweel, Kimpen, & Sinnema, 2004). QOL outcome measures used included disease-specific measures and general health-related QOL measures. The studies varied greatly in their criteria for defining proxies. Some required a proxy to be “someone who knows the study subject well” in a nonprofessional capacity, which could be a family member, next of kin, friend, or informal caregiver; some asked the older person to nominate the proxy; a few accepted caregiving staff; and some used either a family member or a provider, depending on availability. Four studies included both a family member and a health care provider. From these studies, we find varying amounts of agreement between subject and proxy with agreement most pronounced when the items were concrete and verifiable. Even when general agreement was high for sample means, striking discordance between pairs was found in specific cases.

Although proxies are frequently used, their validity is rarely discussed in depth. For example, in the National Hospice Study, a substantial amount of the reports on pain control came from proxies because the patients were either too sick to respond or already dead; however, no evidence was presented to show that a third party could describe another person’s pain (Greer, Mor, Morris, Sherwood, Kidder, & Bimbaum, 1986). Typically, most literature portrays the study subject as a gold standard, considering proxy responses that differed from the subject's to be wrong. But some commentators also discuss the possibility that the person's own self-reports could also be wrong and lament the lack of a gold standard or proposed that some other source be used as a criterion, such as a clinical assessment or administrative records. In general, there has been a tendency for more investigators in long-term care to attempt to gather direct perceptions from residents, including those with substantial cognitive impairment (Albert et al., 1999; Brod, Stewart, Sands, & Walton, 1999; Logsdon, Gibbons, McCurry, & Teri, 1999), and some investigators have experimented with different ways of framing questions to maximize the likelihood of getting responses directly in the residents’ own voices (Simmons, Babineau, Garcia, & Schnelle, 2002; Simmons & Schnelle, 1999).

We embarked on this study at a time when interest in measuring QOL of nursing home residents and frail older people in other residential settings was high. States have been interested in report cards comparing the relative ability of various licensed care providers to deliver a good QOL to residents (Zimmerman & Bowers, 2000). The National Quality Forum identified QOL and resident satisfaction as important but presently omitted indicators of quality in nursing homes (National Quality Forum, 2004). Given how much easier data collection from family members would be, a debate is ongoing about whether family members could be solicited either instead of residents or when residents cannot easily be polled.

**Quality-of-Life Measures in Nursing Homes**

QOL is widely agreed to be a multidimensional concept, though theorists vary in the range of domains they attempt to tap (Abeles, Gift, & Ory, 1994; Birren, Lubben, Rowe, & Deutchman, 1991; Lawton, 1991; Stewart & King, 1994). Between 1998 and 2003, the Centers for Medicare and Medicaid Services (CMS) funded the iterative development of self-report measures of QOL tailored to relevant domains for NF residents. Because relocation to an NF is such a drastically life-altering change affecting all aspects of everyday life activities and relationships (Agich, 1993; Kane & Caplan, 1990; Tobin, 1999), QOL measures needed to be considered broadly and to embrace psychological and social outcomes. Based on previous work, literature review, and focus groups, QOL measures developed were designed to tap 11 QOL domains identified as important: comfort, autonomy, privacy, dignity, meaningful activity, relationships, food enjoyment, security, functional competence, spiritual well-being, and individuality (Kane, 2001, 2003). The developed scales, in their short form, contained three to six items per scale. As described elsewhere, these scales had good alpha reliabilities, except for the individuality domain, which was therefore dropped from the analyses (Kane et al., 2003). The remaining 10 domains were shown through confirmatory factor analysis to be independent constructs related to a common latent QOL construct. The three-item enjoyment scale in its final iteration related only to enjoyment of food, other items initially posited as “enjoyment” items having migrated to other domains or failed to load with any. The QOL measures also correlated well with a scale adapted from Brod and Stewart that was used to assess the resident’s affective or mood state (Brod et al., 1999).

**METHODS**

**Sample and Data Collection**

Residents.—The analyses reported here were drawn from the first wave of data collection for the CMS QOL study, collected from 40 nursing homes in five states in 2000. Eight facilities were selected from each of five states (California, Florida, Minnesota, New Jersey, and New York). A random stratified sampling procedure was used to generate 50 residents from each NF, 50% at the upper half and 50% at the lower half in cognitive functioning based on the cognition items from the MDS report.
closest to sampling. After excluding residents under age 65 years and residents who were comatose or in a vegetative state, we attempted to interview all residents who were sampled, regardless of their recorded cognitive levels. As shown elsewhere, about 60% of residents approached could complete the QOL interview.

**Staff.**—At the same time as the field work with residents was done, we interviewed a NF direct care staff member for each resident. To select the staff interviewee, who was usually a certified nursing assistant, we identified staff members on the day or evening shift likely to know each resident well. If a primary nursing system was in place, we interviewed someone primarily assigned to the resident. To qualify to be interviewed, the staff member must have cared for that resident for at least 2 weeks. The same staff member was permitted to report on up to five residents. Staff were interviewed in person and given a $10 gift certificate for each interview.

**Family.**—With the help of the NF social worker or at times the unit coordinators, we undertook a procedure to identify up to three family members currently involved with the resident in rank order based on the amount of contact with the resident. We mailed an explanatory letter and questionnaire to each identified family member. If no family member replied by mail, we followed up by phone. Only 13% of the family proxies were interviewed by phone. Responses from the family respondent with the highest contact rating were used in these analyses.

**Variables**

Resident data used for this study include the QOL subscales for which we had parallel information from staff and/or family and the emotion rating questions. As indicated above, domains were measured by three to six items. Almost all items used the response set "often, sometimes, rarely, never," which were in turn scored as 4, 3, 2, and 1, respectively; reverse coding was done so a positive response always meant a better QOL. Respondents who had difficulty with the complexity of that response set were offered the response options "mostly yes" or "mostly no." After empirically testing a variety of scoring methods, we interpolated the binary responses into the Likert responses, using a score of 3.8 for a binary positive response and 1.5 for a binary negative response. For comparability, all QOL scales were summed and divided by the number of items, so that all subscales ranged from 4 to 1, with 4 representing the better QOL.

Staff and family questionnaires were developed to parallel the questions posed to residents; in all cases, the staff or family respondents were instructed to answer the questions as they thought the residents would answer them, basing responses on their best guess as to the resident’s appraisal rather than their own opinion. Because staff and family questionnaires were fielded along with the resident developmental questionnaire, the QOL domains had not yet been established. In the family and staff questionnaires, we omitted asking the respondents to judge the spiritual well-being items because we thought they would be unlikely to be known by a third party. We also omitted the dignity and relationship items from the staff interview because the items could call for substantial self-criticism from staff (e.g., Do staff treat Mrs. Jones roughly while giving her care? Do staff treat residents with respect?). Thus, we obtained nine matched domain scores for resident–family comparisons (comfort, functional competence, privacy, dignity, meaningful activity, enjoyment, security, relationships, and autonomy) and seven matched domains for resident–staff comparisons (omitting dignity relationships). We collected data on residents’ current emotions from family and staff respondents.

Residents in each of the cognitive groups provided responses, although the rate of responses complete enough to allow calculation of domain scores was higher among those more cognitively intact. For example, domain scores could be calculated for 78% or more of those in the highest two cognitive groups compared with 35% in the lowest two groups. We were able to obtain a staff respondent for each of 1,326 resident respondents. We obtained family proxies for 989 residents. Table 1 shows the mean values for the three types of respondents and the test comparing the residents’ scores with those for each type of proxy. The extent of the difference is estimated by dividing the difference between the resident and proxy scores by the standard deviation of the resident scores. All but one of the means were significantly different; however, the magnitude of the differences varied. The possible range for the means is from 1 to 4, and in 6 of the 15 comparisons, the differences in the means were small (< .1), whereas in 9 comparisons, the difference was ≥ .1. When the matched paired results were compared, the staff reports significantly (p < .01) differed from those of the residents on all domains.
but the pattern of difference was not consistent. In six instances, the staff mean was higher. The staff mean was within 1 SD of the resident mean for 55% of all cases and within 60% of the cases for all but three items (meaningful activity, security, and autonomy). Matched family proxy means were significantly different from resident means for all domains except food enjoyment and relationships, but again the patterns differed. Family means were lower for four domains (comfort, functional competence, meaningful activity, and security) but higher for three domains (privacy, autonomy). Matched family proxy means were significantly different from resident means for all domains except food enjoyment and relationships, but again the patterns differed. Family means were lower for four domains (comfort, functional competence, meaningful activity, and security) but higher for three domains (privacy, autonomy).

The family mean was within 1 SD of the resident mean for 55% of all cases and within 60% of the cases for all but two items (comfort and relationships). Matched family proxy means were significantly different from resident means for all domains except food enjoyment and relationships, but again the patterns differed. Family means were lower for four domains (comfort, functional competence, meaningful activity, and security) but higher for three domains (privacy, autonomy). The family mean was within 1 SD of the resident mean for 55% of all cases and within 60% of the cases for all but two items (comfort and relationships).

Table 2 examines the correlations between resident and staff proxy ratings. Two different forms of the analyses are shown: the Pearson correlation coefficients and the ICCs. In all cases, the results are highly statistically significant, but the actual values do not show high levels of correlation. Only one of the Pearson coefficients for staff exceeds .3. The Spearman coefficients perform better (not shown). All but one of the ICC values is > .2, but only one is > .4.

Another test of agreement used the measures of emotions. As shown in Table 3, the correlations between staff and residents were generally low; none of the Pearson coefficients or the ICC values reached .2. The correlations with family proxies were somewhat better. Three Pearson coefficients and two ICC values were > .2.

To test the effect of residents’ cognitive status on agreement with proxies, we compared the mean percentage agreement across 56 items common to the resident, family, and staff surveys for each level of cognition in the CPS. Figures 1 and 2 show the patterns for family and staff proxies, respectively. There is a significant difference across the levels of cognition in both cases, with a slightly U-shaped distribution. The mean difference for the most cognitively impaired group was greater than the rest. For resident–staff, the F value was 3.21 (p = .007); for resident–family, it was 5.81 (p < .001). However, the actual differences in means are modest. For staff proxies, the highest mean correlation was .56 for the most impaired group and the lowest was .49 for those with a modified CPS score of 2 (midrange). For family proxies, the most impaired group again showed the highest correlation (.59) and the lowest was again in group 2 (.48).

To explore whether the characteristic of the residents affected the relationship of proxy and resident reports, we regressed the Spearman coefficients perform better (not shown). All but one of the ICC values is > .2, but only one is > .4.

Another test of agreement used the measures of emotions. As shown in Table 3, the correlations between staff and residents were generally low; none of the Pearson coefficients or the ICC values reached .2. The correlations with family proxies were somewhat better. Three Pearson coefficients and two ICC values were > .2.

To test the effect of residents’ cognitive status on agreement with proxies, we compared the mean percentage agreement across 56 items common to the resident, family, and staff surveys for each level of cognition in the CPS. Figures 1 and 2 show the patterns for family and staff proxies, respectively. There is a significant difference across the levels of cognition in both cases, with a slightly U-shaped distribution. The mean difference for the most cognitively impaired group was greater than the rest. For resident–staff, the F value was 3.21 (p = .007); for resident–family, it was 5.81 (p < .001). However, the actual differences in means are modest. For staff proxies, the highest mean correlation was .56 for the most impaired group and the lowest was .49 for those with a modified CPS score of 2 (midrange). For family proxies, the most impaired group again showed the highest correlation (.59) and the lowest was again in group 2 (.48).

To explore whether the characteristic of the residents affected the relationship of proxy and resident reports, we regressed the Spearman coefficients perform better (not shown). All but one of the ICC values is > .2, but only one is > .4.

Another test of agreement used the measures of emotions. As shown in Table 3, the correlations between staff and residents were generally low; none of the Pearson coefficients or the ICC values reached .2. The correlations with family proxies were somewhat better. Three Pearson coefficients and two ICC values were > .2.

To test the effect of residents’ cognitive status on agreement with proxies, we compared the mean percentage agreement across 56 items common to the resident, family, and staff surveys for each level of cognition in the CPS. Figures 1 and 2 show the patterns for family and staff proxies, respectively. There is a significant difference across the levels of cognition in both cases, with a slightly U-shaped distribution. The mean difference for the most cognitively impaired group was greater than the rest. For resident–staff, the F value was 3.21 (p = .007); for resident–family, it was 5.81 (p < .001). However, the actual differences in means are modest. For staff proxies, the highest mean correlation was .56 for the most impaired group and the lowest was .49 for those with a modified CPS score of 2 (midrange). For family proxies, the most impaired group again showed the highest correlation (.59) and the lowest was again in group 2 (.48).

To explore whether the characteristic of the residents affected the relationship of proxy and resident reports, we regressed the Spearman coefficients perform better (not shown). All but one of the ICC values is > .2, but only one is > .4.

Another test of agreement used the measures of emotions. As shown in Table 3, the correlations between staff and residents were generally low; none of the Pearson coefficients or the ICC values reached .2. The correlations with family proxies were somewhat better. Three Pearson coefficients and two ICC values were > .2.

To test the effect of residents’ cognitive status on agreement with proxies, we compared the mean percentage agreement across 56 items common to the resident, family, and staff surveys for each level of cognition in the CPS. Figures 1 and 2 show the patterns for family and staff proxies, respectively. There is a significant difference across the levels of cognition in both cases, with a slightly U-shaped distribution. The mean difference for the most cognitively impaired group was greater than the rest. For resident–staff, the F value was 3.21 (p = .007); for resident–family, it was 5.81 (p < .001). However, the actual differences in means are modest. For staff proxies, the highest mean correlation was .56 for the most impaired group and the lowest was .49 for those with a modified CPS score of 2 (midrange). For family proxies, the most impaired group again showed the highest correlation (.59) and the lowest was again in group 2 (.48).

To explore whether the characteristic of the residents affected the relationship of proxy and resident reports, we regressed the Spearman coefficients perform better (not shown). All but one of the ICC values is > .2, but only one is > .4.
resident domain score against the proxy domain score incorporating measures of residents’ age, gender, disability level in ADLs, cognitive status, and length of stay. As shown in Table 4, adjusting for resident characteristics did not affect the relationship between the resident and the proxy reports; nor did the full model account for much of the variance in the resident reports. For staff reports, the best models could account for only 7% to 8% of the variance. For families, the best model (functional competence) accounted for 23% of the variance because of the strong role played by the ADL adjuster. The models for two other domains accounted for 14% (food enjoyment) and 15% (meaningful activity) of the variance in resident reports. It is noteworthy that the effects of the cognitive variable were mixed: In some instances, it was positive and in others negative; the pattern was constant across the two sets of analyses. By contrast, the ADL dependency measure score was consistently negatively associated with the residents’ rating.

To simulate how proxy reports might be used in benchmarking exercises, the proxy ratings were used in a second mode, creating a mean score for each facility by averaging the proxy ratings for that facility. When these mean scores were compared with the mean scores for the corresponding residents, the levels of agreement were much higher than the correlations at the individual level. Table 5 shows the Pearson correlations and ICC values for the mean resident facility ratings for each applicable domain and the family member and staff means. The results are much stronger. Among the staff, four of the possible seven Pearson correlations and three ICC values are above .4, and all but one is statistically significant despite the much smaller sample size \( N = 40 \). For family members, the pattern is stronger. For all but one domain (meaningful activity), the Pearson correlations and the ICC values are ≥ .4.

**DISCUSSION**

The results extend earlier work in other areas to suggest that neither family members nor staff can confidently be expected to report the QOL of NF residents the same way as the residents do themselves. The relationship between agreement with proxies and resident cognitive status is basically U shaped. When residents are both more and less cognitively impaired, agreement is slightly better.

Confidence in proxies depends in large part on the nature of the data being sought and the way it will be used. Proxies can provide information about factual events, such as hospital admissions or even falls, where they have the opportunity to

### Table 3. Correlations Between Proxies and Residents for Emotion Items

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Residents vs Staff Pearson</th>
<th>Significance</th>
<th>ICC</th>
<th>Significance</th>
<th>Residents vs Family Pearson</th>
<th>Significance</th>
<th>ICC</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lonely</td>
<td>.085</td>
<td>.001</td>
<td>.0831</td>
<td>.0008</td>
<td>.159</td>
<td>.000</td>
<td>.1526</td>
<td>.0000</td>
</tr>
<tr>
<td>Happy</td>
<td>.149</td>
<td>.000</td>
<td>.1405</td>
<td>.0000</td>
<td>.219</td>
<td>.000</td>
<td>.2151</td>
<td>.0000</td>
</tr>
<tr>
<td>Bored</td>
<td>.071</td>
<td>.009</td>
<td>.0690</td>
<td>.0052</td>
<td>.206</td>
<td>.000</td>
<td>.1969</td>
<td>.0000</td>
</tr>
<tr>
<td>Angry</td>
<td>.152</td>
<td>.000</td>
<td>.1520</td>
<td>.0000</td>
<td>.188</td>
<td>.000</td>
<td>.1860</td>
<td>.0000</td>
</tr>
<tr>
<td>Contented</td>
<td>.122</td>
<td>.000</td>
<td>.1176</td>
<td>.0000</td>
<td>.191</td>
<td>.000</td>
<td>.1887</td>
<td>.0000</td>
</tr>
<tr>
<td>Worried</td>
<td>.113</td>
<td>.000</td>
<td>.1119</td>
<td>.0000</td>
<td>.160</td>
<td>.000</td>
<td>.1567</td>
<td>.0000</td>
</tr>
<tr>
<td>Interested</td>
<td>.176</td>
<td>.000</td>
<td>.1744</td>
<td>.0000</td>
<td>.228</td>
<td>.000</td>
<td>.2273</td>
<td>.0000</td>
</tr>
<tr>
<td>Sad</td>
<td>.087</td>
<td>.001</td>
<td>.0850</td>
<td>.0007</td>
<td>.169</td>
<td>.000</td>
<td>.1632</td>
<td>.0000</td>
</tr>
<tr>
<td>Afraid</td>
<td>.117</td>
<td>.000</td>
<td>.1164</td>
<td>.0000</td>
<td>.101</td>
<td>.002</td>
<td>.1006</td>
<td>.0008</td>
</tr>
</tbody>
</table>

Note: ICC = intraclass correlation.
witness the event or are likely to know of its occurrence. However, relying on proxies to provide information about another person’s emotions or subjective judgments on QOL raises much greater concerns.

Although the choice of correlation statistic should be based on the nature of the data being analyzed, in this study we found little difference in the results regardless of the statistic used. This low level of agreement urges caution in using either family or direct staff as proxies for resident reports when we have any possibility of getting feedback from residents themselves.

Our results were similar to those reported in other settings. Although the choice of correlation statistic should be based on whether individuals with low levels of cognition are capable of assessing their own QOL. In our data, however, the congruence between residents and proxies was not much better when we looked only at residents with good cognitive functioning.

Table 4. Regression Coefficients for Resident Ratings by Domain

<table>
<thead>
<tr>
<th>Variable</th>
<th>Comfort</th>
<th>Functional</th>
<th>Meaningful</th>
<th>Privacy</th>
<th>Activity</th>
<th>Enjoyment</th>
<th>Security</th>
<th>Autonomy</th>
<th>Dignity</th>
<th>Relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resident vs staff</td>
<td>(Constant)</td>
<td>1.840</td>
<td>2.592</td>
<td>3.042</td>
<td>2.145</td>
<td>1.754</td>
<td>2.652</td>
<td>2.793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proxy</td>
<td>0.262***</td>
<td>0.292***</td>
<td>0.133***</td>
<td>0.235***</td>
<td>0.268***</td>
<td>0.122***</td>
<td>0.128***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOS</td>
<td>0.073</td>
<td>0.037</td>
<td>0.000</td>
<td>0.075</td>
<td>0.009</td>
<td>0.140**</td>
<td>0.037</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>0.009</td>
<td>0.037</td>
<td>-0.010</td>
<td>0.053</td>
<td>-0.030</td>
<td>-0.008</td>
<td>-0.026</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.003</td>
<td>-0.004</td>
<td>0.000</td>
<td>-0.002</td>
<td>0.006*</td>
<td>0.056**</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADLs</td>
<td>-0.017</td>
<td>-0.086***</td>
<td>-0.047**</td>
<td>-0.041</td>
<td>-0.020</td>
<td>-0.024</td>
<td>-0.079***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognition</td>
<td>0.032**</td>
<td>0.037**</td>
<td>-0.059</td>
<td>0.023</td>
<td>0.051***</td>
<td>-0.005</td>
<td>-0.011</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.072</td>
<td>0.157</td>
<td>0.059</td>
<td>0.079</td>
<td>0.058</td>
<td>0.036</td>
<td>0.056</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resident vs families</td>
<td>(Constant)</td>
<td>1.740</td>
<td>2.477</td>
<td>2.719</td>
<td>1.548</td>
<td>1.663</td>
<td>2.281</td>
<td>3.137</td>
<td>2.590</td>
<td></td>
</tr>
<tr>
<td>Proxy</td>
<td>0.276***</td>
<td>0.344***</td>
<td>0.191***</td>
<td>0.371***</td>
<td>0.366***</td>
<td>0.240***</td>
<td>0.154***</td>
<td>0.239***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOS</td>
<td>0.090</td>
<td>0.038</td>
<td>0.056</td>
<td>0.205**</td>
<td>0.025</td>
<td>0.061</td>
<td>0.057</td>
<td>0.144*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>0.047</td>
<td>0.010</td>
<td>-0.009</td>
<td>0.038</td>
<td>-0.129*</td>
<td>-0.033</td>
<td>-0.038</td>
<td>0.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.004</td>
<td>-0.004</td>
<td>0.001</td>
<td>0.000</td>
<td>0.004</td>
<td>0.005</td>
<td>0.001</td>
<td>-0.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADLs</td>
<td>-0.021</td>
<td>-0.076***</td>
<td>-0.033</td>
<td>-0.027</td>
<td>-0.019</td>
<td>-0.020</td>
<td>-0.031**</td>
<td>-0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognition</td>
<td>0.025*</td>
<td>0.055***</td>
<td>-0.052***</td>
<td>0.045*</td>
<td>0.027</td>
<td>-0.005</td>
<td>-0.018</td>
<td>-0.017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.080</td>
<td>0.233</td>
<td>0.076</td>
<td>0.152</td>
<td>0.138</td>
<td>0.073</td>
<td>0.042</td>
<td>0.081</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: LOS = length of stay; ADLs = activities of daily living.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 5. Correlations of Facility Level Domain Scores Between Residents and Proxies

<table>
<thead>
<tr>
<th>Domain</th>
<th>Residents vs Staff</th>
<th>Pearson</th>
<th>Significance</th>
<th>ICC</th>
<th>Significance</th>
<th>Pearson</th>
<th>Significance</th>
<th>ICC</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comfort</td>
<td>.6355</td>
<td>.0000</td>
<td>.6352</td>
<td>.0000</td>
<td>.4136</td>
<td>.0080</td>
<td>.4134</td>
<td>.0036</td>
<td>.0002</td>
</tr>
<tr>
<td>Functional competence</td>
<td>.4296</td>
<td>.0057</td>
<td>.3692</td>
<td>.0088</td>
<td>.6385</td>
<td>.0000</td>
<td>.5203</td>
<td>.0002</td>
<td>.0001</td>
</tr>
<tr>
<td>Privacy</td>
<td>.5863</td>
<td>.0001</td>
<td>.5539</td>
<td>.0001</td>
<td>.6125</td>
<td>.0000</td>
<td>.5488</td>
<td>.0001</td>
<td>.0001</td>
</tr>
<tr>
<td>Dignity</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.5378</td>
<td>.0003</td>
<td>.5377</td>
<td>.0001</td>
<td>.0004</td>
</tr>
<tr>
<td>Meaningful activity</td>
<td>.4053</td>
<td>.0095</td>
<td>.4024</td>
<td>.0046</td>
<td>.1267</td>
<td>.4359</td>
<td>.1247</td>
<td>.2187</td>
<td>.0001</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>.2587</td>
<td>.1069</td>
<td>.2457</td>
<td>.0607</td>
<td>.5029</td>
<td>.0009</td>
<td>.5022</td>
<td>.0004</td>
<td>.0001</td>
</tr>
<tr>
<td>Relationships</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.5014</td>
<td>.0010</td>
<td>.4054</td>
<td>.0043</td>
<td>.0001</td>
</tr>
<tr>
<td>Security</td>
<td>.3317</td>
<td>.0365</td>
<td>.3307</td>
<td>.0174</td>
<td>.4337</td>
<td>.0052</td>
<td>.5504</td>
<td>.0004</td>
<td>.0001</td>
</tr>
<tr>
<td>Autonomy</td>
<td>.3162</td>
<td>.0468</td>
<td>.3069</td>
<td>.0255</td>
<td>.4501</td>
<td>.0036</td>
<td>.4317</td>
<td>.0024</td>
<td>.0001</td>
</tr>
</tbody>
</table>

Note: ICC = intraclass correlation.
so acculturated to their environment as to lose perspective. However, we hold that that this acculturation, which some might label as coping, is a critical element in determining QOL. Others support this need for a resident-centered approach and the incorporation of the resident voice or, more generally, the consumer voice in the appraisal of long-term care (Appelbaum, Straker, & Geron, 2000; Rubinstein, 2000; Schnelle, 2003).

Our findings have important implications for the immediate task of deciding how to assess QOL in nursing home residents and in the larger context of how much one can rely on proxies in general. With regard to the former, if we reject using proxies, we are left with the dilemma of how to address the QOL for those residents who cannot speak for themselves. Two approaches are possible: extrapolating from the views of residents who can respond, or developing a method to extrapolate QOL based on observations. Observations, even structured observations of residents’ affect (Lawton, Van Haistma, & Klapper, 1966), are simply a variation on the proxy theme. Extrapolating from the reports of those residents who can respond entails other risks that the experiences of those who can communicate verbally are different from those of the severely cognitively impaired. In a regulatory environment, there is the risk that providers might focus on those who can report their QOL and neglect those who cannot. An alternative is to use a proxy approach, acknowledging and trying to minimize its flaws. The fact remains that when trying to get a sense of the residents’ QOL for many possible uses, including reporting to potential consumers, the question of how to incorporate the experience of those unable to self-report, frequently owing to cognitive impairment, remains a perplexing problem.

For some purposes, the problem may be less critical. If one is simply trying to develop an overall score at the facility level for use in a quality-reporting system or benchmarking, for example, aggregated data might suffice. We have used the residents’ reports of QOL to show differences among NFs (Kane et al., 2004). Aggregated reports, based on aggregated mean values, have much better concordance. Here the individual pairs need not agree, as long as one high score in one group offsets a similar score in the other.

Some limitations of this study must be acknowledged. Because of the overall design of the parent study, the questions used for the residents and the proxies differed slightly. The response metric for the residents was designed to maximize participation. A statistically determined crosswalk was developed (Kane et al., 2003) and used here.

Further research is warranted. We intend to do additional work with both the family and the staff reports to see if other techniques might be helpful for approximating QOL for residents unable to provide self-reports and to see if any family or staff characteristics predict better congruence with residents. Also of interest is whether certain items or domains garner greater proxy congruence than others, perhaps leading us to a theory of how an individual’s reality can be understood by others. This general topic may also lend itself to qualitative work to help explain the discrepancies in quantitative results and to assess the characteristics of proxies or their experiences vis-à-vis the person in whose stead they reply that might make them more suited to the task.

Some clinical implications can also be cautiously drawn. The fact that staff, on whom residents are so dependent, respond to questions about QOL differently from the residents under their care suggests a difference in perception that underscores the need for resident-derived data on which to base efforts to improve resident QOL. Perhaps if QOL were included in the MDS as a basis for NF accountability, staff might be more motivated to discuss aspects of residents’ lives and emotions and to listen carefully to the answers. Staff might then better anticipate the needs and preferences of residents and provide them more sensitive care.

Acknowledgments
Annette Totten is now at Boise State University, Idaho.
Address correspondence to Robert L. Kane, MD, Division of Health Services Research and Policy, University of Minnesota School of Public Health, 420 Delaware St. SE, D351 Mayo (MMC 197), Minneapolis, MN 55455. E-mail: kanex001@umn.edu

References
Janss, A. J., Gemke, R. J. B. J., Uiterwaal, C. S. P. M., van der Tweel, L,


