Executive Summary
"*Measures, Indicators, and Improvement of Quality of Life in Nursing Homes*."
Rosalie A. Kane, PhD, Principal Investigator

In May 1998, the Centers for Medicare & Medicaid Services (CMS) awarded the University of Minnesota a contract: *Measures, Indicators, and Improvement of Quality of Life in Nursing Homes* (RFP: HCFA-98-002/PK0, which is being conducted under Master Contract Number 500-96-0009 between CMS and the University of Minnesota.. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. The general objectives of the study were: 1), to develop and test measures and indicators of quality of life (QOL) for nursing home residents, and 2), to study how the physical environment, including private rooms, affected resident QOL. The resulting measures, tools, approaches and findings were meant to assist regulators for quality assurance efforts, providers for quality improvement efforts, and consumers, consumer advocates, and third-party purchasers seeking to compare nursing homes. The study, which concluded in July 2003, entailed 2 major waves of national data collection and many distinct areas of study and analysis. Volume 1 of this report summarizes the procedures and findings undertaken in the 5-year study period. Data collection instruments are appended to the report in Volume 2, which consists of questionnaires, interview protocols, and related materials used for Wave 1 or Wave 2 of the Quality of Life study or for the Transferability Study, which could be useful templates, tools and examples for providers.

Key Findings

1. It is feasible to collect information directly from nursing home residents about their own QOL.
   a. Even residents with substantial cognitive confusion completed QOL interviews.
   b. On average, researchers were able to collect first-hand data from 60% of residents in a nursing home.
2. The 11 QOL domains that researchers identified (comfort, security, functional competence, relationships, enjoyment, meaningful activity, dignity, individuality, privacy, autonomy, and spiritual well-being) could be tapped in 54 questions.
   a. The scales had adequate scale properties in 2 large field tests.
3. Family members and staff members were poor proxies for residents themselves.
   a. When researchers had multiple sources of reporting, family and staff reports were statistically significantly associated with resident reports but correlations on scales were rarely better than 0.3.
   b. Family members were somewhat more congruent with residents than staff.
4. QOL average domain scores for facilities are capable of differentiating one facility from another.

a. At both Wave 1 and Wave 2[1], researchers were able to discern patterns whereby some nursing homes fell several standard deviations above or below the mean on multiple QOL domains.
5. Although QOL varies across facilities, researchers also see variation across residents.
   a. Resident ADL, cognitive status, and sensory impairment are predictors of better QOL.
6. During Wave 1, facility-wide observation of resident and staff interactions identified several observable phenomena that were associated with better or worse resident QOL.
   a. At both Waves, the results of the observations distinguished facilities from each other, suggesting some promise for observational protocols as an auxiliary way of assessing facility QOL, especially as it related to those who cannot be interviewed.
7. QOL measures could be satisfactorily conducted by staff members and surveyors under research conditions.
   a. No systematic differences in accuracy, concordance between research interviewers, or attitudes to the process could be attributed to whether the staff assessor was a nurse, on the one hand, or a member of the social work or activities staff, on the other.
8. No strong pattern emerged to connect off-site indicator data from OSCAR and MDS to QOL.
9. Researchers observed an enormous range in the way only 40 nursing homes arrange their nursing services, activities, and a myriad of other functions that could have an impact on the QOL domains.
   a This variation has implications for measurement of any new indicators. Records for activities, room changes, in-service education, care planning, and other kinds of archival data varied also in both level of detail and format.
10. Researchers developed new approaches to measuring physical environments that had high inter-rater reliability and did not require specialized knowledge or subjective judgments.
    a. Researchers found many problems with physical environments in the 40 Wave 1 nursing homes. Private rooms are associated with better QOL.
    b. Researchers have begun a cluster analysis approach that is promising in that the clusters seem associated with QOL.
       i. Our new approach to environmental assessment pinpoints the environment for each individual resident, enabling use of more powerful hierarchical analysis to examine the effects of the nested environments of rooms, units, and facilities.

Introduction

For the purposes of the study, a QOL *measure* was defined as any feature of the program, staffing, physical environment, or policy that might be associated with QOL outcomes for all or a subset of nursing-home residents. CMS also requested that we distinguish between *off-site indicators* which could be derived from currently available sources such as the On-Line Survey

---

[1] See Appendix I for a comparison of Wave 1 and Wave 2 testing.

Assessment and Review (OSCAR) data set or the nursing home Minimum Data Set (MDS), on the one hand, and *on-site indicators,* which refer to indicators that would require special data collection to generate.

By CMS intent, our work emphasized psychological and social aspects of QOL. Although health status, sensory status, cognitive status, pain, and emotional status are themselves often perceived as dimensions of quality of life, we were charged with identifying and examining other domains of QOL not emphasized in the MDS. We also included domains of QOL (such as dignity and individuality) that are emphasized in current nursing home standards. Further, our focus was narrowed to older nursing home residents, mostly well over age 65. Therefore, our work cannot be generalized to younger nursing home residents, who often have developmental or mental health disabilities. We excluded from our purview nursing homes strictly targeted for people with mental retardation, developmental disability, or psychiatric conditions. We also eliminated swing beds in hospitals, nursing home care units in Veteran's Administration medical centers, states' Veterans Homes, and nursing homes that accepted <u>neither</u> Medicare or Medicaid.

We assumed that residents are the best source of information about the quality of their own lives because of the subjective nature of QOL outcomes. Accordingly, one goal was to determine to what extent people with dementia can report reliably on their quality of life, and how well the ability to collect QOL measures corresponded to cognitive impairment as noted on the MDS. Given that some residents would be unable to be interviewed about their lives because of cognitive disability, we also assumed a need to supplement resident self-report with reports from other sources and direct observations. Finally, along with other writers and researchers, we perceive QOL as a multidimensional construct.

Field Work Methods

For the over-arching objective to develop and test measures and indicators, we identified 11 tasks: 1) review literature and existing QOL measures; 2) identify domains for psychological and social aspects of resident QOL, 3) identify an item pool to measure the domains; 4) develop reliable and valid measures of QOL outcomes; 5) determine the extent to which QOL can be measured directly in interviews with nursing home residents, including those deemed cognitively impaired; 6) determine how many residents need to be interviewed to develop stable estimates of average QOL in a nursing home; 7) determine how a sample of facilities differ in resident-reported QOL; 8) determine how resident and facility characteristics affect QOL at the individual and facility level; 9) test proxies for resident self-report, including reports from direct care giving staff, reports from family members, and observation of resident affect; 10) develop and test facility-level observation tools for resident and staff interactions that might be a proxy for QOL for residents who cannot be interviewed; and 11) identify possible indicators of QOL based on existing data sets from OSCAR or the MDS or from additional information that could be collected from or reported on by facilities. Data collection proceeded in two major waves with piloting before each wave; the first wave fielded more inclusive instruments with the intent of testing more parsimonious approaches suggested by Wave 1 analysis in Wave 2. Special additional studies on development of a personality measure for nursing home residents, inter-rater reliability testing, and a transferability study were also undertaken.

<u>Wave 1</u>

After pre-testing questionnaires, we drew the Wave 1 sample from 40 nursing homes in 5 states. The states (California, Florida, Minnesota, New Jersey, and New York) were selected with CMS involvement to reflect a range of policy environments based on variations in nursing home reimbursement levels, ownership patterns, regulatory history, and workforce characteristics. California, Florida, and New York were further narrowed into smaller areas, including, respectively, San Francisco and northern California, central Florida (an area comprising Tampa and St. Petersburg and Orlando and surrounding rural areas), and western New York (and area comprising Buffalo, Rochester, Syracuse, and Utica and surrounding rural areas). The nursing homes were randomly selected from the roster of eligible facilities so as to equally include large and small nursing homes and urban and rural nursing homes. Because of CMS interest in single rooms, we also over-sampled in each state to include, when possible, at least 2 nursing homes with 75 per cent or greater residents in private rooms. We excluded facilities with less than 50 beds. Given the extensive commitment from participating facilities for Wave 1 and our desire to include nursing home personnel as partners in the efforts, the sampled facilities were allowed to decline participation. Only a few facilities did so, typically for reasons with which we concurred (e.g., a major construction project underway, or management changes and turmoil).

In each nursing home, we chose up to 5 nursing units, selecting any Alzheimer's Special Care Unit or rehabilitation unit, and randomly selecting from the remainder. (The average nursing home in the sample had 3 units, and only five (5) exceeded 5 units.) We then selected residents from those units using a computer program so as to result in a random stratified sample of 50 residents per facility evenly divided by residents with better or poorer cognitive functioning as defined by an MDS-derived cognitive measure. To divide the sample by cognitive status, we abstracted records 2-3 weeks prior to data collection and used a simple MDS-derived 6-point scale using short-term memory, long-term memory, and decision-making capability. We also over-sampled for residents in private rooms, allowing the sample to include as close as possible to 20% people in private rooms. Finally, as much as feasible, we divided residents evenly across units, selecting at least 10 residents from each unit, excluding residents in comas or vegetative states and younger residents from the sample. In a few small nursing homes, the exclusions combined with low occupancy resulted in a sample of fewer than 50 residents.

The purposes of Wave 1 data collection were to conduct the psychometric work needed to develop QOL scales in terms of scale consistency, reliability, concurrent validity, and factor analysis, to determine the adequacy of proxy approaches for resident self-report, and to determine the sample size of residents needed to make facility-level estimates. During Wave 1, we also: collected extensive data on physical environments; collected quantitative and qualitative data on policies, procedures, and staffing in each facility (to explore possible QOL indicators and how to collect the data for them); and we piloted a facility-level observational approach to characterize interactions among residents and staff as a possible supplement to QOL measures.   Wave 1 data ultimately included 1988 residents in 131 units in 40 facilities.

Prior to drawing the Wave 1 facility sample, we mailed a survey to all nursing homes in our study areas to determine the proportion of single rooms and the kind of quality of life initiatives then taking place. CMS was interested in whether facilities were using model tools developed by

the two major nursing home trade associations. The proportion of the 987 responding facilities using these trade association tools varied from 3% in California to 19% in New Jersey. Many more facilities routinely collected feedback from residents using their own corporation's tools (varying from 49% of New Jersey respondents to 73% of California respondents). Proportions of residents in private rooms were low. The percent of responding nursing homes with nobody in a private room ranged from 33% in California to 5% in New York and New Jersey, and those with fewer than 25% in private rooms ranged from 86% in New Jersey to 58% in California. Only a tiny proportion of the responding facilities had 75% or more of their residents in private rooms: 6% in California, 3% in Florida and Minnesota, 4% in New Jersey, and 5% in New York.

Wave 2
    The purposes of Wave 2 data collection were to confirm the factors and properties of the QOL measures in a new sample, to examine individual factors associated with QOL in a sample that would include more racial variation and more short-stay residents than the Wave 1 procedures permitted, to examine facility factors associated with QOL in a more representative sample of facilities using extant data and data supplied by each administrator, and to test a shortened version of facility level observational tools.

    We again sampled 5 states (substituting Maryland for New Jersey). We increased the number of facilities per state to 12 and sought efficiency by building the samples within a 30 mile radius of a major urban area. We also chose urban areas to increase the number of African American and Hispanic residents in the sample: the Wave 2 sample was centered in zip codes in central Los Angeles, central Baltimore, North Miami, Minneapolis, and Albany, NY. We arrayed facilities in these areas using an algorithm based on positive and negative extremes on regulatory citation history and staffing levels, and recruited facilities from the top and bottom of that distribution. Rather than stratifying residents based on cognition (which required an earlier step of sorting the residents by cognitive strata several weeks before the anticipated field visit, thus undercounting short-stay residents), we interviewed a random sample of residents. Unlike with Wave 1, we positioned participation in the 2nd wave as mandatory, with us conducting a "developmental survey" as an agent of CMS.

    From Wave 1, we learned that completed interviews from 28 residents were ample to describe the facility on all the QOL domains we measured. Accordingly, we sought a sample of 28 residents per facility with completed interviews. Residents were approached in random order; those who could not complete an interview were excluded and the sampling continued until 28 interviews were produced. The Wave 2 sample included 1680 residents from 60 nursing homes.

Transferability Study

    After Wave 2 data were analyzed we conducted a separate study in Minnesota to determine the extent to which nursing home staff would obtain the same results as University of Minnesota researchers in resident interviews and facility wide observations of resident and staff interactions. This component was conducted as a quasi-experiment wherein we examined whether the extent of training of facility personnel and the staff member's disciplinary background (nursing versus social work or activities) affected how congruent staff results were with those of researchers. This study took place in 8 nursing homes and entailed 4 staff assessors

and 24 residents per facility for a total of 16 staff assessors and 192 residents.   We conducted a similar transferability study with a sample of Maryland surveyors

Date Collection and Procedures

At each Wave, we hired 40 local interviewers (8 per state) and conducted 40 hours of training. Interviewers needed to achieve high inter-rater reliability with their trainers on environmental observations and on observations of resident and staff interactions before they went into the field. At Wave 2, we used a data collection form that could be entered in a scanning procedure. Wave 1 data collection was labor-intensive, entailing an average of three interviewers spending two to three weeks in the facility. Wave 2 procedures were streamlined, entailing only resident interviews and observation protocols; interviewer teams tended to complete them within a week in a facility.

Resident Self-Report Measures

As a result of an extensive literature review and interview process as well as focus groups, we ultimately chose to measure 11 domains of QOL. These were comfort, security, meaningful activity, relationships, enjoyment, functional competence (defined to mean that within their physical and cognitive limitations residents were as independent as they wanted to be), autonomy, dignity, privacy, individuality, and spiritual well-being. Each of these domains was construed as a resident outcome. For each, we identified a pool of candidate items of resident self-report that might reflect each domain. We recognized that these outcomes would be sensitive to physical environments, facility policies and practices, and staffing patterns, and the team generated hypotheses about potential indicators related to the outcomes.

In Wave 1, we fielded a questionnaire with 4 to 13 items per domain that took 45 to 90 minutes to complete. The intent was to develop scales of more practical length from subsets of these items. The QOL items were embedded in a questionnaire that also included: demographic items, residents self-report on their experience of 10 emotional states (adapted from Meryl Brod's Dementia QOL measure), residents self report on 4 general satisfaction items, and a series of 12 summary ratings by residents on their QOL (one for each domain and "quality of life as a whole"). For most of the QOL items, we posed questions using 4-point Likert scales ("often, sometimes, rarely, never" or, for summary ratings "excellent, good, fair, poor"). Recognizing that some residents might be unable to use Likert responses, we allowed interviewers to revert to a dichotomous response pattern ("mostly yes or mostly no") when a resident could not reply with the Likert categories after 3 attempts. If a resident had repeated problems with the Likert response categories, the interviewer was allowed to conduct the entire interview using the binary categories. For only 5% of residents did we use 2 or more sittings to complete the questionnaire; most of these interviews were divided to accommodate the interviewer's or the resident's scheduling convenience, but in some instances the breaks accommodated resident fatigue.

All English-speaking residents over 65 were eligible for the study unless comatose or in a vegetative state. We ruled out 325 residents for an interview based on a screening determination that they were completely confused or incoherent or unable to rouse. We eliminated other

residents who gave incoherent or non-responsive replies to 8 of the 13 comfort items at the beginning of the questionnaire. Sixty per cent of all 1988 residents in the sample (1188 residents) were able to respond sufficiently for us to complete the bulk of the interview and calculate most scales. (The exact N varies by scale). The ability to calculate QOL scores decreased as the respondent's cognitive score result worsened, but we could calculate most QOL scores for many respondents with poor MDS-derived cognitive scores and were unable to calculate scores for a few residents with perfect cognitive scores.

We tested various ways of interpolating the binary responses into the Likert responses. We applied z-transformations for each item, separately for the Likert and for the binary response options and then combined the responses into one variable regardless of the response option used. For the majority of items, a score of 3.8 corresponded to "mostly yes" and 3.5 to "mostly no" responses, and we used this scheme throughout. A few items called for a dichotomous response from the beginning (e.g., having any resident as a close friend, having a confidante, developing a new interest since coming to the nursing home). So as not to under weight these questions, we coded "mostly yes" as a "4" and "mostly no" as a "1." We did our original scale development on subjects who had complete responses using Likert response options, and then tested whether and how interpolating the binary responses changed the results. In Wave 1, we required that 75% of items be completed for scales with 4 items and that 66% of items be completed for scales with 5 or more items; in such cases, missing items were recorded at the mean of the individual resident's responses for the remainder of the particular scale.

The first scales generated, which we refer to as the long QOL scales ranged from 3 to 13 items. The shortest scale, Enjoyment, was comprised of 3 food-related items, and is more properly dubbed Food Enjoyment. Scale consistency using Chronbach's alpha was acceptable for all but the Individuality scale. We looked at the inter-correlations and independence of the scales; these procedures tended to eliminate items that may have been associated with a high rating for overall quality of life but that loaded with multiple scales. We tested concurrent validity by testing the hypothesis that each QOL domain would be positively associated with overall satisfaction and better emotional well-being. We also tested the hypothesis that each domain would be associated with the single summary rating for that domain by regressing each domain score against all summary items. These concurrent validity tests worked as expected. In some instances more than one summary item was related to the scale score but the intended summary item was most associated with the overall scale.

We conducted confirmatory factor analysis (CFA) on the remaining 76 items, finding that the individuality and relationship domains were practically indistinguishable. The individuality scale also had poor internal consistency. When we dropped individuality, a resultant 10-factor solution fit the data well.

We then relied on cluster analysis to somewhat reduce the scales to shorter versions, selecting the items within each domain scale that best preserved the integrity of the scale and discriminated that domain score from the others. We repeated all analyses on the short scales, which ranged from 3 to 6 items per scale. Results were similar to those achieved for the long scales. Once again, CFA resulted in a 10-factor solution, eliminating individuality. Alpha

reliabilities fell slightly with the shorter scales, but only the individuality scale fell below 0.63 or greater. The inter-correlations among domains dropped as a result of the shortening procedure.

For the most part QOL was the same for the individuals with the better or worse cognitive functioning. Where we found significant differences, QOL was better in some domains for the persons with higher cognitive functioning and was better in other domains for the persons with worse cognitive functioning.

We fielded a revised instrument at Wave 2. We used only the short-scale items, and substituted a new 6-item scale for "individuality," which was tested prior to fielding Wave 2. We approached residents in random order, did not interview residents who failed a simple conversational screen (i.e., tracking with a greeting and introduction of the study), and also terminated interviews with residents who failed to give coherent or usable responses to 4 of the 6 comfort questions at the beginning of the interview. In one facility we obtained fewer than 20 completed questionnaires and replaced the facility. In all others we were able to obtain our desired 28 completed questionnaires, although often facility staff had told us that only a handful of residents could be interviewed.

As with Wave 1, we were able to interview residents with poor cognitive performance. In Wave 2 a randomly selected sample of 1680 individuals completed the questionnaires, with 78.5% in the 3-5 cognitive range, connoting poorer cognitive functioning. The mean resident length of stay was almost 2 years, with 8.9% having stayed one month or less, and 6.2% having stayed between 1-2 months; 21.5% having stayed 3 months or less. The Wave 2 sampling approach made it possible to compare QOL of longer and shorter stay residents.

When we applied Confirmatory Factor Analysis, this time we were able to confirm an 11-factor solution that fit the data well. The scales, including the new Individuality scale, had adequate scale properties and concurrent validity, using the same procedures as in Wave 1. At Wave 2 we found it was possible to develop a sample that was balanced in cognition without an elaborate prior sampling scheme, and, thus, include short-stay residents in the sample.

Family and Staff Proxies for Resident QOL

We conducted a test in Wave 1 data to examine how family and direct caregiver appraisals of residents' QOL concur with the appraisals of the residents themselves and how consistent family and direct care staff are with each other when no direct resident appraisal is possible. We conducted an in-person interview pertaining to each of the 1988 residents with a line staff member best likely to know the resident.

We sent mailed questionnaires to up to 3 family members per resident, rank-ordered by their estimated level of involvement with the resident. If more than one family member returned the questionnaire, we used the one with the greatest current involvement in the primary analyses. When no questionnaires were returned, we used telephone follow-up to remind respondents to return the survey, or, if they preferred, interviewed them by phone. This procedure resulted in our securing at least one family questionnaire for 1471 residents.

Our questionnaires used items that paralleled the QOL items administered to the residents. We excluded some domains in the proxy interviews. Neither family nor staff members were asked the items on spiritual well-being or relationships because they are so subjective to the residents. On the staff interviews, we also omitted the dignity items because that scale was comprised of items about how staff treated the resident and seemed to require a great deal of self criticism. We asked for the emotions ratings and the summary ratings from both proxies. As control variables we asked both staff and family questions about frequency of contact and familiarity with the resident, and staff members were also asked about educational levels and work shift.

We compared domain scores of various respondents by simple correlations of scale values, by comparing mean values across groups using ANOVA for independent samples, and by comparing matched samples of each proxy and the corresponding residents with t tests. To test the correlations, we used 3 different correlation statistics: Pearson, Kendall, and Spearman. We also calculated Kappa statistics and inter-class correlations (ICC). To obtain Kappas, we dichotomized the continuous scores at both the 20th and the 25th percentiles.

In all correlations between staff and resident, the results are highly statistically significant, but the actual correlations show low values. Only 4 of Pearson correlations exceed 0.2; the Spearman coefficients are slightly higher and the Kendall slightly lower. The Kappa values, which improve if the 25th percentile is used for dichotomizing, are all below 0.2. Only 4 ICC values were above 0.2.

Family comparisons with residents are similar though the agreement is a little better than between staff and residents. Three Pearson coefficients exceed 0.2 though none reach 0.5. Neither the Kendall nor the Spearman perform better. All but one ICC value is above 0.2 but only one is above 0.4.

To explore whether resident characteristics affected the concordance of resident and proxy reports, we regressed the resident ratings against each proxy type rating controlling for resident's age, gender, ADL level, cognitive status, and length of stay. These adjustments made no difference. We also were unable to improve concordance very much by incorporating proxy characteristics and seeking a "best" model.

We also created QOL scores at the facility level for each proxy type by aggregating the scores and dividing by the number of respondents. When we compared these aggregate staff and family scores to aggregate resident ratings, we achieved much higher concordance then we did at the individual level. For staff, 5 of the possible 8 Pearson correlations with residents and 4 of the ICC values are above 0.4 and all are highly statistically significant despite the much smaller sample size when the N is reduced to 40 facilities. The pattern of concordance between resident average scores and family average score in a nursing home is even stronger. For all but one domain (meaningful activity), the Pearson and ICC values are 0.4 or greater.

We concluded that staff and family are poor proxies for residents who cannot respond. Further, in a small reliability test of staff wherein we selected a second staff member equally

familiar with the resident for 62 residents in 30 facilities, we found that staff members were poorly correlated with each other regarding the same resident. Thus, the idiosyncratic selection of a staff respondent may heavily influence results when staff proxies are used. Although aggregation of data to the facility level improves the correlations between proxy respondents and residents, the aggregation procedure conceals the discordance within individual dyads.

Resident Personality

We developed personality measures suited for oral administration to nursing home residents and tested their effects on QOL self-ratings. Prior work on personality assessment tended to rely on rather long self-completed questionnaires, and the inclusion of many items that refer to work place performance. Self-completion is impractical for nursing home residents, some of the items are inappropriate, and the length of the batteries is a problem, especially if a personality inventory is to be combined with other assessment items.

We used the Big Five Personality Inventory (BFI), developed at University of California at Berkeley, as the starting point for developing an approach for nursing home residents. (We chose this instrument because its items were shorter and clearer than those in other available tools.) The BFI yields a score with reference to 5 personality traits: Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness to Experience.

We first conducted a small pilot test with 200 residents after Wave 1 data collection wherein we administered all 44 of the BFI items in interviews with residents with only minor modifications to fit a nursing home population. For example, we introduced an introductory instruction "to consider the kind of person you have been for most of your life" and modified items referring to employment. Analyses indicated that we could shorten the scales to 27 items.

In a second phase, we administered the 27 items to all residents in Wave 2. Analysis of Wave 2 data indicated that 2 of the 27 items did not perform as well as they did in the pilot and they were dropped. The resulting personality inventory consisted of 25 items, 5 each for Neuroticism, Extroversion, and Agreeableness, 6 for Openness, and 4 for Conscientiousness. Alpha reliabilities were all above 0.6.

We then conducted separate regressions for each QOL outcome measure to determine whether resident personality affected resident reported QOL after we controlled for resident functional status and cognition. QOL subscales differed in the extent to which they were linked with personality. Much more of the variance on Spiritual Well-being and Security were attributable to personality than for other domains, whereas Autonomy and Functional Competence were least affected by personality. High levels of Agreeableness were associated with higher levels of QOL on Comfort, Relationships, Dignity, Meaningful Activities, and Spiritual Well-being. High levels of Neuroticism were associated with lower scores on many QOL domains including Comfort, Meaningful Activities, Security, and Autonomy.

Facility-Level QOL Data

We aggregated QOL at the facility level by domain and applied case-mix adjustments, using the QOL data we had collected, basic demographic data on the resident questionnaire (gender, race, and marital status), and MDS data. From the MDS data we created variables for ADL functioning, cognition, and length of stay since admission. Facility descriptors such as size and urban/rural location were drawn from our data base.

Analyses were conducted separately for Wave 1 and Wave 2. Two-way analysis of variance was used to compare difference of distributions for a given QOL domain within each facility and across facilities. Chi square tests were used to compare the proportion of outliers by facility characteristics. We used mixed-effect hierarchical linear models with main effects to fit the data. Ten (10) QOL domains were used at Wave 1 and Eleven (11) at Wave 2. Independent variables included combinations of random factors (Facility, Interviewer) and covariates (length of stay, age, education, race, gender, having adult children, and MDS-derived cognition and ADL measures).

In 23 of the 40 Wave 1 facilities, 2 or more domains were positive or negative by at least one standard deviation from the mean. Ten (10) facilities showed a consistently positive pattern and 13 a consistently negative pattern. Risk adjustment for patient characteristics made no dramatic changes. Adjusting for facility characteristics increased the variance in some domains and decreased it in others. When Interviewer was added as a random factor, 8 domains demonstrated a loss of relative variance and only 2 showed no change. At Wave 1, we found no difference in QOL by resident length of stay.

The pattern of facilities showing either negative or positive patterns of variation held up in Wave 2. Although Wave 2 resident sampling provided a much larger number of short stay residents, we found, as in Wave 1, that resident length of stay had no effect on QOL domains.

We estimated the number of residents needed to show a significant facility effect. At Wave 1, we had conservatively included 50 residents per facility. If we set the alpha level at .05 and power at 0.80, we needed 7 to 12 residents in the sample, depending on the QOL domain being measured. When we set more stringent standards of an alpha level of .01 and power of 0.90, we could detect a difference in all but 2 domains with a sample of 20 residents, but needed 23 residents for Functional Competence and Dignity. Therefore, we conservatively decided to interview 28 residents per facility in Wave 2. When we repeated the analyses with the Wave 2 sample, we found a sample of 13 residents would be adequate for an alpha of .05 and power of 0.80 on all domains. For the more conservative alpha of .01 and power of 0.90, a sample of 20 was sufficient for all domains but dignity, and a sample of 25 was sufficient for all domains.

Shortened QOL Screeners

CMS requested us to suggest a small number of items to screen for QOL that might be incorporated into MDS 3.0, which was about to be tested. Because for some purposes, CMS might be interested in an approach that was binary for all residents (e.g., to increase response rates and simplify administration), we worked within the existing data set to explore the best way to extrapolate the 4-point scales into the 'mostly yes" and "mostly no" binary scales. We tested the three possible ways of reducing the Likert scales to binary scales, and found the best solution

was to treat "often and sometimes" as "mostly yes," and "rarely" and "never" as "mostly no." Fielding the binary versions, however, was beyond our scope, and, thus, we are unsure about the distribution that would be yielded if binary versions were used at the outset. Nor are we certain of the effect on responses if residents are asked only a few questions about QOL rather than a full-length QOL interview.

We developed a 34-item multidimensional approach (QOL-MD-34) that yields 9 domain scores. In these procedures, we included the summary items for each domain (11 items) in the item pool. The 34-item solution drops the Relationship domain because it was highly related to the Individuality domain, and combined items from the Meaningful Activity and Spiritual Well-being scales into an Activity Domain that includes participation in and enjoyment of religiously oriented activity but omits the items from Spiritual Well-Being that relate more to spiritual fulfillment. The Security scale uses the summary item (in its binary form: Do you feel safe and secure?) as one of its 3 items.

We then proceeded to see how much multidimensionality could be achieved while producing a version with less than 20 items. The QOL-MD-14 is a 14-item scale that contains 14 items and taps Food Enjoyment (3 items), Security (3 items), Privacy (4 items) and the new Activity domain (4 items), the latter identical to the QOL-MD-34.

We also developed a short 14-item unidimensional QOL (QOL-14) scale that emphasized using items with a good ability to differentiate among facilities. Using one-way ANOVA followed by Tukeys B post hoc test for homogenous subgroups, we selected those 7 domains that distinguished among facilities at both waves of data collection. We selected the best pairs of items for those 7 domains by estimating inter-class correlations for all the pairs. The resultant 14-item scale has a good alpha reliability as a single scale, and contains 2-items each from the Privacy, Meaningful Activity, Individuality, Enjoyment, Security, Relationships, and Spiritual Well-Being scales. (We also proposed 2 items from the Dignity scale to produce a QOL-16 scale. Dignity differentiated facilities in Wave 1, shows within-facility variation, and is intrinsic to the QOL standards in the federal nursing home regulations.)

Facility-Level Protocols to Observe Staff and Resident Interactions

We developed observational protocols for resident and staff interactions, and tested how such observations were related to resident-reported QOL. We identified observable positive and negative phenomena that might be seen during a meal, an activity, or a slow walk through the entire facility. Each interviewer performed 2 walkthroughs (one on a weekend), two breakfast observations, and two dinner observations in each facility at Wave 1.

Some frequently-observed behaviors did not differentiate facilities (e.g., the negative observations of residents sitting at the nursing station doing nothing or residents sitting at a dining table saying nothing, or the positive observations of staff pausing to answer resident questions, or residents visibly enjoying activities). Others occurred too infrequently to be useful for scales, though they are telling when they are observed. For example staff members were seen talking baby talk to residents at meals in 5% of 135 pairs of activity observations, staff members were seen talking roughly to a resident or threatening him/her in 5.1% of activity observations,

and staff asked residents about weight, bowel movements or continence during an activity in 1.5% of observations. During meals, intimidating or threatening staff behavior was observed more frequently (seen in 11.8% of the 135 pairs of meal observations), quarrels among residents were observed in 9.6% of the dining observations; and staff queries about weight, bowel movements or continence at meals in 4.4%. A more positive observation, seeing a private dining room or area used by a resident and his/her guest, was noted in 6.7% of the observations. Based on these distributions, we dropped items from future iterations.

We correlated counts of positive and negative observations with domain scores. Significant relationships between total positive and negative counts in the expected directions were found for 20 of the 22 correlations. Counterintuitive finds were noted in two cases where a positive behavior count was negatively associated with the comfort domain and the autonomy domain.

We also created an overall QOL score combining domains. At Wave 1, several items predicted overall QOL. From the meal observations, 6 of the 24 items observed were statistically significant (including residents sitting at a table in silence, staff moving resident wheelchairs without discussion or permission, staff talking to each other over resident's head without involving the residents, residents being fed in a messy way, residents observed expressing displeasure, and residents observed calling out in distress). From the activity protocol, 3 of the 24 items were significantly related to overall QOL, (namely all residents silent, staff talking with each other over resident's head, resident calling out in distress, and staff pursuing their own discussions during resident activities). From the 36 walkthrough items observed, 5 were significantly related to QOL (namely: staff moving resident's wheelchairs without discussion, staff talking to each other over resident's head, resident expressing displeasure, and staff seen assisting a family member). We aggregated common features across all 3 forms of observation, finding 5 consistent predictors of QOL: staff talking over resident's head, staff moving resident's wheelchair without asking or explaining; resident expressed displeasure, resident calls out in distress and is ignored, and resident is disengaged.

One behavior that we hypothesized as positive (staff explaining a rule or policy to a resident) was negatively correlated with QOL. On debriefing observers, we found that such explanations were usually associated with telling residents that they were not allowed to do something and enforcing a restriction, which may account for the counter-intuitive finding.

For Wave 2, we streamlined our observational protocol to 4 walkthroughs at planned times in each facility; we incorporated timed observations of meals during the breakfast and dinner-time observations and organized activities during the morning and afternoon observations. We reduced the number of items to 16, relying heavily on those that were significantly related to QOL at Wave 1 and a few other items that seemed conceptually important, such as observing staff threatening residents. We included "staff imposing a restriction on resident," a negative observation, rather than the earlier positive "staff member is explaining a rule or policy." The frequency of observing any of the 16 items in a facility ranged from seeing a resident engaged in a solo activity in 77.6% of the facilities, staff answering resident questions in 61.5% of the facilities, and residents disengaged at the nursing station in 52.5% of facilities to less frequently observed behaviors such as staff moving wheelchairs without discussion (seen in 14.3% of facilities), staff imposing a restriction on a resident (13.1%), staff discuss resident's private

business in public (9.8%), staff speaking roughly or threatening residents (4.1%), and resident being fed messily (4.1%). At Wave 2, we allowed for an actual count of occurrences up to 9 times during an observation, but found that the additional information conveyed by the laborious and perhaps distracting effort to count the behaviors yielded little useful additional information.

The correlations between QOL domains and observations were much less impressive at Wave 2. The only statistically significant relationship was between the count of positive behaviors and the resident-reported privacy domain, and some non-significant correlations were in the opposite of the direction expected We concluded that we had overly constrained the procedure in Wave 2 by limiting the number of observation occasions to 4 per facility and dramatically reducing the items to be observed.

Facility-level observations encompassed the entire facility, including dementia SCUs. Thus, we could potentially observe interactions involving residents who were not in the sample, and some of the sampled residents may not have been observed. Accordingly, the observation strategies could be pursued as a supplementary way of assessing QOL. To this end, we used one-way ANOVAs to explore which items usefully discriminated among facilities. With the 40 Wave 1 facilities, 7 items were highly discriminating in the Wave 1 meal observations, 8 in the Wave 1 activity observations, and 6 in the Wave 1 walk-through procedures. At Wave 2, we performed the analysis of the discriminating power of the 16 items using 61 facilities (we were able to count the observations in the facility with too small a sample of residents to use in the QOL analyses). Eleven (11) of the 16 items were able to discriminate among these 61 facilities, 8 with ANOVAs significant at the 0.000 level.

On-Site Potential Indicators

During Wave 1 of the study, we developed an extensive data base on the policies, practices, and staffing patterns of each of the 40 nursing-homes. The project director interviewed administrators, directors of nursing, directors of social work, and directors of activities (or designated members of their departments). We also examined archival information, including personnel records, activity department records, care planning records, resident council records, in-service education programs, and records of room changes. The ultimate goal was to identify structural and process aspects of the nursing homes that might prove to be indications of better or worse resident-reported QOL. We utilized the labor intensive approach of individual interviews with multiple informants to determine the feasibility of gathering various sorts of information in a consistent way. Based on the experience at Wave 1, we developed a single 4-page questionnaire on facility structure and process that was completed by the administrator or his/her designee when data collectors were in the facility for Wave 2.

We performed separate regression analyses with Wave 1 data (N = 40 NFs) and Wave 2 data (N=57) to seek relationships between the potential indicator and the facility-level QOL score by domain. When we had consistent data for both Waves, we performed a combined analysis, raising the N to 97 NFs. From this large number of regressions, we found a few features of NFs that were related to significantly better QOL on four or more QOL domains. In future work, we plan to develop composite indices reflecting topics (e.g., autonomy policies; rehabilitation and

health care; activity and stimulation; continuity of staff; experience of staff) and to use cluster analyses to determine relationships to QOL.

In the 40 nursing facilities where we did the intensive data collection, we noted great variation in almost all the parameters we included. Attributes thought to affect QOL domains (e.g., permanent assignment of CNAs to residents; vigorous, individualized activity programs; individualized care-planning and routines) will need to be carefully defined before large numbers of facilities can be compared on this parameter. Even such well-established indicators as nursing staff ratios are subject to measurement error and interpretation.

Our analysis of the relationship between these process and structural indicators to date entailed univariate regressions between the putative indicator and the QOL domains; such regressions were performed with and without adjustment for resident characteristics (age, cognition, ADL, gender, and length of stay). Substantial numbers of the possible indicators were associated with better or worse QOL domain scores in four or more domains even after adjusting for resident characteristics. In future analyses (beyond the scope of this contract), we plan to aggregate the many items into clusters that signify patterns in a facility related to such factors as rehabilitation focus, active activity program, individualization of routines, consistency of personnel, and experience of personnel.

Off-Site External Indicators

To examine potential correlations between external indicators and QOL we relied on data from MDS and OSCAR. The former included demographic and clinical data as well as the quality indicators developed by CHSRA. The OSCAR addressed various staffing measures and relevant deficiency rates on the annual surveys. We used Hierarchical Linear Modeling (HLM) to account for the nested nature of the resident data within facilities. Some resident characteristics were associated with variations in some QOL domains but no useful patterns emerged. Likewise the relationship between QIs and QOL is varied; the associations go in both directions such that better QOL is not consistently linked to better quality of care. Higher levels of staffing are not consistently linked to better QOL. Different types of staff are associated with better outcomes in different QOL domains. For example, activities staffing is tied to Privacy and Meaningful Activity; administrative staff is tied to Functional Competence; LPN ratios are associated with lower Dignity scores. In general facilities with greater numbers of citations have lower QOL, but this trend is statistically significant for only Privacy and Autonomy.

Transferability

Part of the project entailed determining the extent to which facility staff and surveyors could apply the interview tools and the observational tools with results similar to researchers. Two general factors might influence this concordance: first, the deliberate approaches used for data collection in research are inconsistent with the way that practitioners and surveyors conduct observations; and second, residents might behave and respond differently when questioned by staff or surveyors. CMS built into the project the question of whether nursing staff or psychosocial staff would perform better in conducting QOL assessments. Later we also decided

to build in an exploration of whether staff could conduct the QOL evaluations with training materials as opposed to 2 days of in-person training.

We conducted the staff transferability study in 8 nursing homes in Minnesota. In each home, 2 nurses and 2 psychosocial staff members (social workers or activities personnel) participated. Each participant conducted 6 interviews that were paired with researcher interviews; interviews were performed 2-5 days apart and the order of interviews (staff first versus researcher first) was randomly altered. We built in a question to tap whether the respondent believed his or her quality of life had changed in the last few days. Staff and researchers also conducted walk-through observations together, each completing the protocols independently. Half the nursing homes were randomly assigned to having their staff receive in-person training; staff from all nursing homes received a manual and an audio-tape on how to conduct the interviews and a number to phone with questions. This procedure resulted in 196 pairs of interviews, half done with a nurse and half with a psychosocially oriented staff member, and half with and half without in-person training. Analyses for congruence were done at the item and scale level. Qualitative de-briefing sessions were also conducted at the participating nursing homes.

Briefly, we found that researcher-staff member congruence at the scale level was as good as the test-retest congruence when both interviews were done by researchers. We found that discipline of staff member had no effect on congruence. In-person training decreased the number of missing items and the amount of use of binary rather than Likert scales but had no effects on concordance. Debriefing revealed that many staff members ignored the training materials or skimmed them, that some staff members were negative about the efforts (perceiving them as a waste of time and respondents as inaccurate), and that some staff members were positive about the experience (perceiving it as an opportunity to truly speak to residents in depth and glean their feelings, and to challenge themselves to improve a resident's QOL). Some staff believed that they had learned a lot from the experience. Positive or negative staff reactions were unrelated to discipline but seemed related to the roles and alternative time pressures they experienced; several MDS coordinators and charge nurses who participated were negative, whereas staff development nurses and part-time nurses were more positive, for example.

The small surveyor trial, which was conducted with 6 Maryland surveyors and entailed 24 paired interviews and observational protocols, had similar results. Surveyors achieved satisfactory concordance with researchers on both the interviews and observations.

This quasi-experiment suggested that under the right conditions staff can conduct QOL assessments. It is possible, however, that if such assessments were used for public reporting or regulatory purposes, the dynamic would be changed and the results affected.

Physical Environments

Physical environments, for this study, were defined as the built environment and grounds, including furnishings, fixtures, decorations, and equipment, and the use of space. We developed a detailed approach to assessing elements of the physical environments at the level of the resident's room, the nursing unit where the resident resided, and the facility as a whole. This

approach allowed us to link analytically the individual resident's characteristics, including his or her QOL, and his or her specific physical environment. The assessment tools were designed as checklists of easily observed phenomena that could be noted by an interviewer without prior expertise in design or environments. Global judgments (e.g. ratings of homeliness or functionality) were not required.

All environmental data were collected at Wave 1. The 40 research observers achieved excellent inter-rater reliability and concordance with their trainers before going into the field to conduct the room and bath evaluations, and spot checking throughout the data collection showed they maintained that standard. All facility and unit level environmental assessments were done by one team member (Lois J. Cutler), who also applied light meter readings and identified innovative and problematic environmental features for later detailed qualitative study.

Descriptive analysis showed many environmental deficits, including inadequate showers, few knobs and switches that were operable by residents, cluttered corridors, closet rods out of resident's reach, lighting levels that approximated conditions of blindness, lack of horizontal work space for residents, lack of a resident's bedside chair in many instance, access to the outdoors, distances that residents needed to traverse, and many other substandard features. The more people sharing a resident room, the less likely the resident was to have ample spaces for privacy and activity at the unit and facility level.

We created composite indices of environmental attributes such as Visual Privacy, Function Enhancing Features, Life Enriching Features, and Environmental Controls for Residents. These Indices were constructed at Room, Unit, and Facility level, and an index of the Dining Environment was created on the basis of both unit and facility level observations. We present preliminary findings from cluster analysis that generates Environment Types, using data from all 3 levels. Since this study found that being in a private room had a positive effect on QOL domains when controlling for resident characteristics, a NIH Career Award to Howard Degenholtz at the University of Pittsburgh permits further use of hierarchical modeling (resident in room in unit in facility) on how environmental cluster membership affects quality of life.

Conclusions and Recommendation

This project resulted in a variety of tools that can be used to monitor and study QOL in nursing homes and other residential settings where older people receive care. We are encouraged by the extent to which residents can report on their own QOL. We are also encouraged by the new environmental tools that were developed for this project, and the prospect of further research in this area. Facility staff can collect QOL information under the right conditions, and the very act of collecting such data is likely to improve their awareness of resident need and preferences. We suggest a multi-faceted research agenda aimed both at improving QOL measures, including establishing cross-cultural relevance and exploring the sensitivity of such tools through longitudinal study, and also aimed at learning more about the determinants of QOL.

Implications

This report demonstrates the feasibility of collecting information from nursing home residents about their quality of life (QOL). Given that QOL data is collected directly from residents, such contact provides an opportunity to collect other salient information at the same time on relevant topics like affect and satisfaction. This information can be used for a variety of purposes including quality assessment and assurance, consumer information, and quality improvement. It can also be employed as a building block in a payment system that re-enforces better quality by paying more for it.

Underlying this work is the growing appreciation that QOL and satisfaction are major issues of concern to current and potential users of long-term care. The historical preoccupation with technical elements of quality of care, focused largely around nursing care issues, needs to be tempered with more attention to this important area. Plans to revise the MDS 3.0 reflect recognition of this shift in priorities.

The measure developed (in both its full form and subsequent shorter versions) was able to distinguish among NHs. Although resident characteristics explained the majority of the explained variance, the proportion accounted for by facilities was sufficient to permit discrimination and to serve as the basis for rewarding desired behavior in this area. It would be useful to field these measures in new and larger samples of nursing homes. In their further development, they also need to be tested for cross-cultural relevance.

Making QOL measurement in the nursing home context fully operational probably requires a more comprehensive list of domains that those assessed in this project. Salient elements, such as functioning, health, cognition, and affect, were specifically eliminated from this study under the terms of the contract, but measures of these domains should be included in final efforts to field a QOL battery. M. Powell Lawton, an early pioneer in defining QOL (and an original collaborator on this project until his death), exhorts researchers to recognize that QOL must represent both subjective and objective aspect of quality (Lawton, 2001). On the other hand, Lawton summarized a desirable subjective multidimensional QOL construct with great attention to the domains studied here. Aggregating "investigators' top-10 lists" from previous research, Lawton proposed 11 human needs that must be satisfied in nursing homes to create a better QOL: autonomy, privacy, dignity, social interaction, meaningful activity, individuality, enjoyment versus aversive stimulation, safety and security, spiritual well-being, clarity of structure and functional competence. He argues that each of these 11 needs is universal and each "can be served or frustrated by caregivers; and may be represented by its fulfillment for each individual and by its capacity to be fulfilled by the facility (Lawton, 2001, p.147)."

Our work shows that the data derived from proxy reports is not sufficiently comparable to the reports of residents themselves to allow ready substitution at the individual level, although the mean values of the proxy respondent groups were closer to the means of the residents and could therefore be used to develop NH level information. The poor performance of proxies suggests that whenever a resident can be interviewed directly, proxies should not be used in lieu of resident self-report.

Determining QOL for those who cannot express themselves verbally, even in short and simple interviews remains a problem. Many residents with substantial dementia according to the

MDS could complete usable QOL responses; we, therefore, recommend against any MDS-based screening prior to conducting QOL interviews. That being said, many people who are unable to respond lack a voice, and in some instances we cannot really infer their QOL from any non-verbal clues. One could rely on those who can respond to serve as the sentinels for those who cannot, but such a policy is uncomfortable. More work is needed to identify ways to give the resident who cannot communicate a better voice. The observational efforts at the facility level were designed to observe phenomena that have face validity for being associated with a better or worse QOL for the residents observed, particularly residents with dementia. At the same time, this study shows the dangers of arbitrarily excluding resident respondents on the basis of their cognitive performance. Although the level of participation declined with more severe cognitive impairment, even some severely impaired residents were able to participate meaningfully.

Substantial efforts went into developing a QOL tool that was psychometrically sound. As a result it had reasonable internal reliability, especially for a measure of amorphous constructs. The level of reliability was maintained in the shorter versions, which did not capture as many domains. Various tests showed that there was good inter-rater reliability and that the questionnaire could be used to good effect by both nursing home staff and surveyors. Although the performance of both types of staff under artificial test conditions may be better than in actual practice, when other factors may influence their performance, these findings are encouraging.

Whereas the QOL appears to have sufficient reliability to make it a viable tool, the issue of validity is much harder to address. There is some evidence that conceptually related domains appear to be better correlated with each other than those expected to be less related, but in the absence of an independent gold standard for QOL there is no specific way to demonstrate validity. The sample of nursing homes was designed to provide some variation, but again there was no way to create an independent measure along which to array the homes. Subsequent work might well examine how well the measure discriminates among homes believed by some independent criterion to represent better and worse QOL. For example, facilities that are making an active effort to create a social and physical environment designed to enhance QOL might be compared to more typical facilities. During this project, we struggled with the problems, however, of using facility reputation for good QOL as a source of validation of QOL measures, and for using expert raters. There is no reason to think such criterion measures should have greater weight than the measures derived from residents themselves.

Another line of investigation might explore more carefully the correlates of better QOL scores. While the easily accessible structural measures drawn from interviews and that OSCAR did not demonstrate a strong correlation, more subtle or simply other measures than more directly reflect ambiance and personal attention might be more useful. It is intriguing that the number of activities staff was a stronger correlate with QOL than was the amount of nursing staff (professional, nonprofessional or both). More attention might be usefully devoted to understanding the implications of this observation. The strongest relationship between environmental factors and QOL was the presence of single rooms, but much more work remains to be done in exploring the effects of environment. Future work might also look at interactions and the possibility that QOL is influenced by a specific confluence of resident characteristics within specific environments.

We are also eager to see longitudinal work undertaken with QOL measures to determine their sensitivity to change. If they are to be used in before-after tests of quality improvement efforts, they will need to demonstrate that sensitivity.

The QOL score has been largely handled as a series of domain measures, but work with a variety of constituencies suggests that the values applied to the individual domains are quite similar. Hence a simple summative summary score might be more easily understood by many users. More work could be done on the advantages of creating and using such a score. There are always tradeoffs, of course, between the detail involved in sub-scores and the efficiency of a single scale. For some purposes, the separate domains may be more conceptually useful, but in many instances a single score would likely encourage wider use of the construct. We caution again that the longer measures contain much information that might be useful for QI purposes. Similarly, short screeners, such as we developed as potential instruments for the MDS (the QOL-14 and the QOL-MD-14) and our 11-item list of summary statements in our QOL battery (which itself forms a scale and which we used to validate our domains) are examples of brief approaches.

Whether using a single score or domain score, issues arise about how various aspects of QOL should be weighted. In our developmental work, for example, we weighted all domains equally, but we recognize that such a stance is arbitrary. Once we recognize that different weighting schemes are feasible, the question arises as to whose values should inform the weighting and how, if at all, weightings can be keyed to individual preferences.

QOL measurement in nursing homes is in its infancy compared to the many decades of work on functional status and quality of care. Some individuals raise concerns about pursuing QOL too vigorously because they believe that many of the factors influencing QOL are out of facility control. They argue that if most QOL is influenced by health conditions and social circumstances, and even somewhat by personality, there is little room left for care to affect outcomes. We view such negativism as short-sighted. The very observation of variation across facilities suggests that positive efforts can affect QOL. It is well known that nursing home life can affect QOL negatively, and remedying these negative effects should itself produce positive effects. With greater knowledge and precision, professionals may be able to identify residents at higher risk of poor QOL because of certain physical or sensory impairments or because of social isolation and lack of family. It would then be feasible to target efforts to those individuals.

Systematically collecting and reporting information on QOL is itself likely to have a profound positive effect in NH care. Using the score in more deliberate ways to report on facility performance and even to encourage better QOL outcomes by linking payment to performance would likely have an even more profound effect. The question always arises about when is an instrument ready for broad application. Almost any tool can be improved. Undoubtedly this measure will evolve over time but the fundamental building blocks are in place. It is certainly as good as many such instruments in common use. The needs for emphasizing these vital aspects of nursing home care outcomes argue strongly for active implementation, even while additional refinements are pursued.

Appendix I. Comparison of Wave 1 and Wave 2

Parameter	Wave 1	Wave 2
Purposes	Field test measures for psychometric work. Determine who can respond by cognition. Test proxy informants on resident QOL. Develop extensive on-site programmatic data for possible indicator development. Collect data for environmental component of study. Proxy data collected on all residents, including those who could not be interviewed.	Confirm QOL measures. Examine resident correlates of QOL including personality and length of stay. Test shortened observational protocol.
States	CA, FL, MN, NJ, NY	CA, FL, MD, MN, NY
NH sample criteria	Stratified by size and rural/urban, and oversampled for 75%+ in private rooms. Entire state (MN, NJ) or substate (CA, NY, FL) catchment areas, the latter with at least 300 nursing homes. Volunteer NHs. Sample size 40 facilities.	Facilities enumerated in driving range of urban areas (Albany, NY; Baltimore, MD; Los Angeles, CA; Miami/Fort Lauderdale, FL; Minneapolis/St. Paul, MN); selected from top and bottom of a list arrayed by citation and staffing features thought to influence quality. Sample size 60 facilities.
Resident sample criteria	Facility census enumerated by unit, cognitive status from MDS, and single room status; sampled so as to select up to 20% in single rooms if possible, an even division between high and low cognitive functioning, and even representation of up to 10 residents from each nursing unit.	Resident census approached in random order.
Sample size	50 residents per facility (actually total sample was 1988 because several facilities had fewer than 50 eligible, consenting residents after exclusions. Approximately 1200 interviewees with QOL data; N varied by domain.	28 residents actually completing QOL questionnaires; total sample 1680.
Other data collected or used.	Staff interview for each resident. Family questionnaire when possible. Apparent-affect rating scale for 20% random sub-sample and all who could not be interviewed. Detailed programmatic data, staffing data, and environmental data at room, unit, and NF level. Facility level observational protocols. MDS data for sample and facility and OSCAR data for facility.	Streamlined facility level observational protocol. Administrator self-report questionnaire on facility programmatic data. MDS data for sample and OSCAR data for facility.