NIH Public Access

Author Manuscript

Med Care. Author manuscript; available in PMC 2010 June 1.

Published in final edited form as:

Med Care. 2009 June ; 47(6): 651–661. doi:10.1097/MLR.0b013e31819241b9.

Responsiveness and construct validity of the Health Utilities Index (HUI) in patients with dementia

Abstract

Background—Assessment of health-related quality of life (HRQOL) is important for cost-effectiveness analyses, but the validity of generic HRQOL instruments has not been adequately evaluated in persons with dementia.

Objective—To evaluate the validity (including responsiveness to change) of the Health Utilities Index Mark 2 (HUI2) and Mark 3 (HUI3), two commonly used generic HRQOL measures, in patients with dementia.

Subjects—408 patient-caregiver dyads in an 18-month dementia care management trial.

Methods—We assessed construct validity by evaluating correlations of proxy (caregiver)-reported HUI2 and HUI3 with the Blessed Dementia Rating Scale (BDRS), the Charlson Comorbidity Index, and a behavior rating scale. Responsiveness was estimated using effect size (ES) statistics for behavior scale change (unchanged, small, medium, large change) and for residential status change (home to skilled nursing facility), as a global external change criterion.

Results—The HUI2 and HUI3 were responsive to behavioral worsening (multi-attribute ES range: −0.48 to −0.78) and global decline (multi-attribute ES range: −0.50 to −0.76), but not improvement. The HUI2 was more responsive than the HUI3. Correlations with the BDRS (r=−0.69 with both HUI2 and HUI3 multi-attribute scores) and behavior scale (r=0.44 and 0.41, respectively, for HUI2 and HUI3 multi-attribute scores) supported the validity of the HUI in patients with dementia.

Conclusions—Our results support the construct validity of the proxy-rated HUI2/3 in patients with moderate to severe dementia, but responsiveness results were mixed. Further studies are needed of the HUI2/3's validity, including responsiveness, in patients across the full range of dementia severity, using both self and proxy report, with particular attention to the impact of general population preference weights. When possible, multiple HRQOL measures need to be used to confirm the robustness of the findings. The proxy-rated HUI should be used in patients with moderate to severe dementia, but the self-rated HUI may be appropriate for subjects with milder cognitive impairment.

Keywords

Alzheimer's; dementia; quality of life; health status; Health Utilities Index; HUI2; HUI3; clinical outcomes; health economics; responsiveness

INTRODUCTION

Alzheimer's Disease (AD) affects 3–4 million persons in the United States, at an estimated cost of $100 billion annually, and the prevalence and cost of dementia are projected to rise significantly in the coming decades (1). To respond optimally to this growing public health problem, health policy makers must have valid assessments of the cost-effectiveness of available interventions for management of AD. Generic preference-based health-related quality of life (HRQOL) instruments, which assign numerical values (utilities) to health states based on preferences expressed in population surveys, are used to measure health benefits of clinical interventions across disease categories (2). However, since the validity of generic

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

HRQOL instruments in patients with dementia is still uncertain, recent cost-utility studies employing such measures have sparked controversy (3).

The Health Utilities Index (HUI) is a generic, utility-based HRQOL instrument applied in patients with a wide range of medical conditions in both clinical and general populations internationally. The HUI Mark 2 (HUI2) has seven attributes, each with 3–5 levels: sensation, mobility, emotion, cognition, self-care, pain and fertility. The HUI Mark 3 (HUI3) has eight attributes, each with 5–6 levels: vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain. Although the HUI3 was designed to address shortcomings of the HUI2, the scales are viewed to be complementary, and scores for both can be generated from the same instrument (4).

There are few published reports on the validity of the HUI in patients with AD and other dementias, and findings are mixed. In one cross-sectional comparison of the HUI2 and the HUI3 in patients with AD, scores on proxy-rated versions of both instruments discriminated well across dementia stages (questionable, mild, moderate, severe, profound and terminal) defined by the Clinical Dementia Rating scale in the expected direction (i.e., lower HRQOL in more advanced dementia). The greatest differences in HRQOL by dementia stage were observed in the HUI cognition and HUI2 self-care attributes (5). In contrast, in a sample of patients with mild to moderate AD, Naglie found no significant associations (Spearman correlations) between either patient or proxy-rated HUI3 scores and measures of physical function (Katz ADL, Lawton IADL), depressive symptoms (Geriatric Depression Scale) or cognition (Mini-mental State Exam) (6).

Evidence of the reliability of the HUI in patients with dementia is limited. Naglie et al found that test-retest reliability (at 2 weeks) of the HUI3 with proxy informants in mild and moderate dementia exceeded the standard for adequate reliability of 0.70 (intraclass correlation coefficient [ICC] = 0.81). Test-retest reliability was also acceptable for self-rated HUI3 in those with mild dementia (ICC = 0.75), but was poor for those with moderate dementia (ICC=0.25) (6).

The HUI's responsiveness to change in patients with dementia has not been previously reported, yet evaluation of the cost-utility of new therapies and of quality improvement interventions requires utility measures that are sensitive to meaningful change. In view of the lack of data and mixed findings on the validity of the HUI in patients with dementia, we conducted an assessment of construct validity (including responsiveness to change) of the HUI2 and the HUI3 in a longitudinal trial of dementia care management (7).

METHODS

Sample

Our analysis is drawn from data on 408 dementia patient-caregiver dyads enrolled in a care management trial, receiving care from one of three San Diego health care organizations. All patients identified in the organizations' administrative databases as having an ICD code for dementia (8) for a visit or hospitalization in the prior year, with diagnosis verified by each patient's primary care physician and confirmed through chart review, were eligible for study inclusion, and approximately 43% of eligibles enrolled (7). At baseline, 12 and 18 months, caregivers were mailed surveys that included study measures. All 408 enrolled dyads had baseline data; follow-up data at 12 and 18 months were available on 82% and 88%, respectively.

Measures

HRQOL—Caregivers completed a proxy version of the 15-item HUI, assessing patients' HRQOL over the preceding 4 weeks. (There are 16 English language versions of the HUI,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

differing in mode of administration (self versus interviewer), assessment viewpoint (self-assessment versus proxy), duration of health status assessment period (past 1-, 2- or 3-weeks) and length of questionnaire (15 versus 40-item)). Individual health domain scores (single attribute scores) range from 0.00 (maximum impairment) to 1.00 (no impairment). Multiattribute (HUI index) scores, a multiplicative function of individual attribute levels, range from −0.03 to 1.00 for the HUI2 and −0.36 to 1.00 for the HUI3, with anchors 0.00 = dead and 1.00=perfect health for both the HUI2 and HUI3 (4).

Variables used in construct validity assessment

Behavior—The California Dementia Behavior Questionnaire (CDBQ) is an 87-item caregiver survey for assessing behavioral disturbances in patients with dementia, with six caregiver mood items and 81 patient behavior items (9). Twenty-two CDBQ items from the three CDBQ subscales judged *a priori* by the trial's steering committee to be the most important targets for the intervention comprised the behavior rating scale in the caregiver survey. The three CDBQ subscales, defined previously in an unpublished principal components analysis (Dan Mungas, Ph.D., personal communication), were the following: depression (9 items, coefficient α = 0.82), anger/agitation (11 items, α=0.77) and physical aggression (2 items, α=0.92). Exploratory factor analysis of this abbreviated twenty-two item CDBQ with our patient sample confirmed the three factor structure: anger/agitation (eigenvalue 7.17; proportion variance explained=0.68; coefficient α=0.90); depressed mood (eigenvalue 1.90; proportion variance explained=0.18; coefficient α=0.82); physical aggression (eigenvalue 1.19; proportion variance explained=0.11; coefficient α=0.80). Fourteen items assessed symptom frequency, and eight items assessed severity over the preceding four weeks. Raw subscale scores were converted to a 0–100 possible range, with 0=most impaired and 100=least impaired (best) state.

Dementia severity—The Blessed Dementia Rating Scale (BDRS) (10) is a widely-used measure of dementia severity. The possible score range is 0–17, with higher scores indicating greater functional impairment. The coefficient α for the BDRS in our sample was 0.90.

Medical comorbidity—The Charlson Comorbidity Index (CCI) (11) is a measure of medical comorbidity. Prevalent medical conditions are assigned a weight from 1 to 6. A CCI score was calculated at baseline based on medical record abstraction assessing for medical conditions listed in the index. Higher scores indicate greater illness burden.

Global status—Initial and follow-up surveys inquired about the patient's residential status: home, assisted living (AL), board and care (B&C), and skilled nursing facility (SNF). We posited that residential status was a reasonable proxy indicator of global dementia severity (12), such that impairment was judged to be progressively worse across the residential levels in the following order: home, AL, B&C, and SNF.

Variables used in assessment of responsiveness—External criteria for defining change from baseline to 12 and 18 months were (1) change in the total behavior scale and subscales and (2) change in residential status.

Because the behavior scale yields continuous scores, we had to create categories defining "changed" and "unchanged" groups. Change categories were defined by first computing effect size (ES) statistics from behavioral scale change scores over the follow-up interval. ES's were then assigned ordinal change categories using Cohen's criteria: no change (|ES| <0.2), small change (0.2 ≤|ES| <0.5), medium change (0.5≤|ES|<0.8), or large change (|ES| ≥0.8) (13). Negative and positive ES's were grouped separately.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Residential change categories were “unchanged” (remained at home), and “changed” if the subject went from “home-to-SNF” between baseline and 12 months, or between baseline and 18 months. (Transitions to and from AL/B&C were excluded from this analysis because of the greater overlap in clinical status between patients in this group and those in the home and the SNF categories).

Expert panel judgements of responsiveness and construct validity measures— We made *a priori* predictions of the magnitude of association between criterion variables and the HUI. We used a modified delta method (14) with a 3-person expert panel consisting of two board-certified geriatric psychiatrists and one board-certified neurologist. Panelists were provided with copies of the HUI, BDRS, CCI, and the behavior scale but did not have access to the results of data analyses.

For the construct validity analysis, panelists predicted magnitudes of correlations between HUI2/3 scores and the BDRS, CCI, and CDBQ. Ordinal categories for the magnitude of predicted correlation magnitudes were defined using Cohen’s criteria (13):

0= no correlation: $|r|<0.10$

1=small correlation; $0.1\leq|r| <0.3$

2= medium correlation; $0.3\leq|r|<0.5$

3=large correlation; $|r|\geq0.5$

The panel was also asked to predict the presence/absence (1=presence/0=absence) of clinically meaningful differences in HRQOL (as indicated by mean HUI score) across residential levels (home, AL, B&C, SNF). Since the panel concluded that AL and B&C groups could not meaningfully be distinguished from each other, these groups were combined into the AL/B&C category.

For the responsiveness analyses, panelists predicted the magnitude of HUI2 and HUI3 ES statistics for all single attribute (except HUI2 fertility) and multi-attribute scores in patients defined as having moderate change (improvement or worsening) on the behavior scale and subscales at 12-month follow-up. We assumed that the magnitude of the ES would be independent of the direction of change. Predictions of the ES associated with residential change categories (“unchanged,” “home-to-SNF”) were also elicited. Ordinal categories of ES magnitudes were defined using Cohen’s criteria (13):

0=no significant effect: $|ES| < 0.2$.

1=small effect: $0.2\leq |ES| < 0.5$

2=medium effect: $0.5\leq |ES| < 0.8$

3= large effect: $|ES|\geq 0.8$

Panelists made independent predictions of the correlations, the presence/absence of group differences in HUI2/3 across residential levels, and ES statistics described above, and results were tabulated. Disagreements between the experts were resolved via a single phone consensus meeting.

Statistical Analysis—Construct validity was assessed with Pearson correlations between behavior, dementia severity, and comorbidity scale scores and the HUI. Mean HUI scores were compared across residential groups with ANOVAs, using Duncan’s multiple range test to identify significant pairwise differences.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Responsiveness was assessed with ES, standardized response mean (SRM), and Guyatt's statistic (GS)(15) across the behavior scale and residential status criterion variables at 12 and 18-month follow-up intervals.

We assessed the accuracy of expert panel predictions of construct validity and responsiveness measures using Cichetti-Allison weighted kappa statistics (using SAS software version 9.1). To compare the panel's ordinal scale predicted magnitudes of associations (scored from 0–3) with continuous values of observed correlations and ES's, the latter were converted to ordinal categories using Cohen's criteria. Since residential change was limited to decline, kappas for responsiveness results were computed only for subjects exhibiting decline. Kappas for correlation and ES results were computed by comparing 16 pairs of predicted and observed results (8 HUI3 single attribute plus 1 multi-attribute score; 6 HUI2 single attribute and 1 multi-attribute score) for each external criterion measure (i.e., BDRS, CCI, etc). Kappas were also estimated to assess correspondence of observed data with predictions of presence/absence of significant group differences in HUI across residential levels (n=48 comparisons, reflecting 3 pairs of residential levels [i.e., home vs. SNF] × 16 HUI predictions).

Following Landis and Koch, quality of agreement indicated by kappas was interpreted as follows: $\kappa<0.00$, poor; $\kappa=0.00–0.20$, slight; $\kappa=0.21–0.40$, fair; $\kappa=0.41–0.60$, moderate; $\kappa=0.61–0.80$, substantial; $\kappa=0.81–1.00$, almost perfect (16).

RESULTS

Baseline descriptive statistics (Table 1)

The range of BDRS scores was 0–17, indicating that our sample comprised patients with all stages (mild, moderate and severe) of dementia. The mean BDRS score of 5.9 (SD 3.7) is consistent with moderate dementia severity (17). Mean multi-attribute HUI2 and HUI3 scores were 0.54 (SD 0.23) and 0.17 (SD 0.31), respectively. By comparison, in a recent cross-sectional US population survey of non-institutionalized persons, mean HUI2 scores were 0.85 in those aged 65–74 and 0.83 in those 75–89; corresponding mean HUI3 scores were 0.80 and 0.75, respectively (18). The mean composite behavioral rating score in our sample was 85.0 (SD 14.1), reflecting mild-moderate levels of behavioral disturbance.

Convergent Validity

Cross-sectional correlations (r) with dementia severity, behavior, and comorbidity at baseline (Table 2)—Correlations with the BDRS were large for the multi-attribute HUI scores ($r=-0.69$, $p\leq0.001$ for both HUI2 and HUI3), and for HUI cognition, HUI2 self-care and mobility, and HUI3 ambulation attributes.

Correlations with the behavioral scale and subscales were generally moderate to large for HUI multi-attribute (HUI2, 0.44; HUI3, 0.41; $p\leq0.001$ for both) and emotion attribute scores and were larger for the HUI2 than the HUI3 emotion attribute (HUI2 range 0.20–0.66; HUI3 range 0.09–0.52).

Correlations between HUI multi-attribute scores and the CCI were small in magnitude but in the expected direction: HUI2, −0.12 ($p\leq0.01$); HUI3, −0.10 ($p\leq0.01$).

Cross-sectional mean differences in HUI scores across residential levels—Both HUI2 and HUI3 index scores differed between the three residential levels, such that subjects living at home had the highest scores, B&C/AL subjects had intermediate scores, and patients in SNFs had the lowest scores ($p<0.05$ for all pairwise comparisons). Scores on individual HUI attributes followed a similar pattern, but not all pairwise differences were significant. (These data are available from the corresponding author on request.)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

HUI2 and HUI3 single attribute scores were similar but multi-attribute and cognition scores were larger in the HUI2 than the HUI3 (HUI2 multi-attribute range across residential levels: 0.26 to 0.51; HUI3 multi-attribute range, −0.12 to 0.12; HUI2 cognition range, 0.26 to 0.51; HUI3 cognition range, 0.21 to 0.33).

Responsiveness of HUI to changes on the behavioral rating scale (Table 3)— Since results were similar with all three responsiveness statistics (ES, SRM, GS) and between the two follow-up intervals (baseline to 12 months and 18 months, respectively), we report only ES results from baseline to 12 months.

Responsiveness of the HUI depended on the direction of change in the external scale. In patients with behavioral worsening, ES's were negative with magnitudes generally proportionate to magnitude of behavioral change. While both HUI2 and HUI3 emotion attributes were especially responsive to behavioral worsening, ES's tended to be larger on the HUI2 than the HUI3. On the other hand, the sign and magnitude of HUI single and multi-attribute score ES's did not reflect the improvements in behavior. In general, ES's in behaviorally improved subjects were either trivial or small and negative.

To better understand the unexpected HUI responsiveness results among subjects with improvement on the behavioral criterion measure, we conducted a *post hoc* analysis to explore the possibility that these results were being driven by the HUI's weighting scheme. We focused on the HUI emotion attribute as it was the HUI dimension most directly linked with the behavior scale items. We computed ES's for the HUI emotion attribute using unweighted HUI scores (i.e., assuming equal distance between emotion attribute levels). The resulting ES's on the emotion attribute associated with small, medium, and large behavior improvements were 0.03, 0.46, and 0.28, respectively, for the HUI2, and 0.20, 0.05, and 0.19 for the HUI3; thus, all were small but were in the expected positive direction of change.

Responsiveness using change in living status as global change criterion (Table 4)—Responsiveness statistics (ES) differed between subjects with no change in residence and those with change from "home-to-SNF."

HUI2 ES's for the "home-to-SNF" group were generally medium to large (HUI2 multi-attribute ES, −0.76), while HUI3 ES's were small to moderate (HUI3 multi-attribute ES, −0.50). In contrast, single and composite ES's for subjects remaining at home indicated insignificant or small change.

Comparison of predicted and observed results—Weighted kappas comparing the expert panel's predicted versus observed results ranged from κ=0.55–0.61 for correlations between the HUI and both the behavioral scale and BDRS. Kappas were lowest for predictions about HUI-CCI correlations (κ=0.13) and for responsiveness results using the behavioral change criterion (κ=0.16). For all other outcomes, kappas ranged from 0.30 to 0.55.

DISCUSSION

This study provides new evidence regarding the construct validity (including responsiveness) of the HUI in community-based subjects with dementia. We report three primary findings: 1). The HUI2 and HUI3 were responsive to graded clinical decline, and HUI2 was more responsive than the HUI3 to global clinical decline and behavioral deterioration. 2). Responsiveness of the HUI was asymmetric, with good responsiveness to clinical worsening but poor responsiveness to improvement as defined by an external criterion of behavior change. 3). Support for the construct validity of the HUI in patients with dementia was found in associations with external indicators of HRQOL.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Responsiveness to clinical decline

Our study provides the first evidence of the HUI's responsiveness in subjects with dementia. Both the HUI2 and HUI3 were responsive to clinical worsening on a behavior rating scale. Effect sizes on the HUI2 and HUI3 emotion attributes were progressively larger negative values in subjects with small, medium and large worsening on the behavior scale. Multi-attribute ES values generally had a similar stepwise quality, but as composites of many other attributes, did not correspond as closely to behavioral change categories. Use of change in residential status as the external criterion provided evidence of the responsiveness of the HUI to global clinical *worsening* in patients with dementia. Notably, ES statistics tended to be greater in magnitude for the HUI2 than the HUI3.

Asymmetry in responsiveness results

On the other hand, HUI responsiveness statistics in behaviorally improved patients did not reflect their improvement, even in the emotion attribute. In the large behavioral improvement group, ES's on the HUI2/3 emotion attributes were either insignificant or small and negative. Our findings thus suggest that in patients with dementia, the HUI may be differentially responsive to improvement and decline, with greater responsiveness to the latter. Asymmetry in an instrument's responsiveness has been previously reported and underscores the need for caution in pooling subjects with similar magnitudes but opposite direction of change in assessing responsiveness (19).

Explaining the asymmetric responsiveness and differences in HUI2/3 responsiveness

The unexpected negative ES's in subjects with behavioral improvement resulted primarily from two factors: 1). differences in the "emotion" constructs captured by the HUI2, HUI3 and the behavior scale and 2). the HUI's preference weighting scheme.

The HUI2, HUI3 and the behavior scale capture different "emotion" constructs. The behavior rating scale consists of 22 questions, while the HUI2 and HUI3 both include only one "emotion" item. The behavior scale probes for a wider range of behavioral disturbances than the HUI2 and HUI3: namely, depressive verbalizations (i.e., expressed suicidal ideation or concentration difficulties) and caregiver observations of anxiety, anger, agitation, mood lability, motor restlessness, paranoia, and aggression. While the HUI2 emotion item assesses anger, agitation, anxiety and irritability, the HUI3 emotion item only assesses levels of happiness/unhappiness, which are terms not included in the behavior scale. The behavior scale's greater overlap with HUI2 than with HUI3 content may explain the former's superior responsiveness to behavioral change. At the same time, the discrepancies between HUI emotion items and the behavioral scale questions may explain how subjects with net improvement on the latter were rated as worsened or unchanged on the HUI2/3.

Our *post hoc* analysis of responsiveness results demonstrated the role of the HUI's preference weighting system in producing the unexpected responsiveness findings among subjects improved on the behavior scale. Each level of an HUI attribute has a preference weight, reflecting its perceived value in HRQOL. Weights for the five levels of the HUI3 emotion attribute are the following: level 1, 1.00 (i.e., perfect emotional health); 2, 0.91; 3, 0.73; 4, 0.33; and 5, 0.00 (20). When we computed ES's using equal spacing between levels (i.e., 0, 0.25, 0.50, 0.75, 1.00), ES's were consistently positive in subjects with improvement, though magnitudes were smaller than expected. The role of weighting was suggested by the finding that among those with large improvement on the external scale, a much greater proportion improved on the HUI emotion items than worsened (35% vs. 14% on HUI2; 35% vs. 18% on HUI3). Extreme values did not account for the results since the proportion with large ($\geq$2 levels) improvement on the HUI emotion items was at least equal to the proportion with large worsening (10% vs. 2% for HUI2; 10% vs. 10% for HUI3). Rather, among those with

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

behavioral improvement on the external scale, the subgroup with decline on the HUI emotion items experienced change between levels with a greater difference in weights than did those with improvement (difference between levels 3 and 4 is 0.40; between 4 and 5, 0.33; between 2 and 3, 0.18; and between 1 and 2, 0.09). To our knowledge, the impact of weighting on an instrument's responsiveness has not been reported previously and merits further study. With respect to our findings, it is possible that the HUI's general population weights do not reflect fully the health state valuations of persons with dementia or their caregivers.

Additional evidence of HUI construct validity in patients with dementia

Our study provided additional evidence of the construct validity of the HUI in patients with dementia. Scores on the composite HUI2/3 were strongly correlated with clinical measures of dementia severity (BDRS) and moderately correlated with behavioral status (CDBQ). As with our responsiveness results, the HUI2 emotion attribute had a stronger association with the behavior scale than did the corresponding HUI3 attribute. Our finding that HRQOL differed across residential levels in the expected direction also supported the HUI's construct validity.

On the other hand, the magnitude of the correlation between scores on the Charlson Comorbidity Index (CCI) and the HUI was small, even though our panel had predicted stronger associations. The small variance of CCI scores was likely responsible for this negative result, though it is possible that the medical conditions (i.e., hypertension) on the CCI are not significantly associated with HRQOL (21).

It is also possible that the weak correlations of the multi-attribute HUI2/3 with the CCI and only moderate correlations of the HUI2/3 with the composite behavioral scale reflect limitations of the HUI2/3 in capturing important aspects of HRQOL in persons with dementia. In a qualitative study of patients with mild to moderate dementia and their caregivers, Silberfeld et al found that the HUI2/3 included only 8 of 56 items identified by patients and caregivers as important to quality of life in persons with dementia (22), suggesting limitations of the content validity of HUI in patients with dementia.

Comparison of study results with other published reports

Differences between our validity findings and those of Naglie (6) may have resulted from the narrower range of clinical variability in the Naglie study, where patients had mild to moderate dementia and generally mild depression, whereas our sample included a broader spectrum of severity of dementia and of depression. It does not seem likely that choice of informant on the HUI contributed to differences between our study and Naglie's in associations between the HUI and measures of mood and behavior, since Naglie included proxy-rated as well as self-rated HUI scores. However, it is possible that differences in results were due to properties of the behavioral measures used in each study. Our behavioral rating scale included anger/ agitation and physical aggression subscales and was completed by the caregiver, whereas the Geriatric Depression Scale used by Naglie focused on depression and was completed by the patient (6).

Our results are consistent with those of Neumann et al, who found similar graduated differences in HUI scores dementia severity levels defined using an established clinical dementia rating scale (5). HUI2 and HUI3 multi-attribute score ranges in our subjects correspond to those reported by Neumann et al among patients with moderate to profound dementia (HUI2 multi-attribute range 0.53-0.27; HUI3 multi-attribute score range 0.19 to −0.08). The discrepancy between HUI2 and HUI3 scores in both Neumann's and our study is noteworthy. Neumann found that 60% of the difference between HUI3 and HUI2 scores could be attributed to differences in weighting of states worse than death. The lowest possible multi-attribute HUI3 score is −0.36 as compared with −0.03 on the HUI2, and Neumann found that differences

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

between HUI2 and HUI3 scores were greatest in patients with more advanced dementia. The differences in QOL scores between the HUI2 and HUI3 highlight the need for caution in comparing QOL assessments using different generic HRQOL instruments, and support the use of multiple HRQOL measures in cost-effectiveness studies. Comparative studies of generic HRQOL measures across the general population (18) and among specific patient groups (6, 23) have likewise found significant differences in HRQOL ratings of the same health state with different instruments. Most important in cost-effectiveness studies, though, is whether HRQOL change estimates differ significantly between measures. Differences in change scores between different HRQOL measures requires further study (23).

The magnitude of multi-attribute HUI2 and HUI3 ES's (−0.48 for both) observed in patients with small behavioral worsening are consistent with the ES's observed in responsiveness studies of other QOL instruments corresponding to small but clinically important differences. In a review of the literature on minimally important differences (MID) in HRQOL measures, Norman et al found that mean ES on QOL measures corresponding to a small clinically meaningful change was 0.495 (SD=0.155) (24). Reflecting the larger clinical change associated with change in residence from home to SNF, the corresponding ES on the HUI2 was −0.76 but was only −0.50 on the HUI3. Even smaller effect sizes are likely to be clinically meaningful (25).

Implications of responsiveness and validity results

Our findings have significant implications for analyses using the HUI in patients with dementia. First, both the HUI2 and HUI3 are responsive to grades of clinical deterioration and may offer important information regarding the utility of treatments that slow decline. Second, the HUI's poor responsiveness to behavioral improvement may result in underestimation of the utility of effective psychiatric interventions in subjects with dementia. Third, because of its greater sensitivity to a range of behavioral disturbances and inclusion of a self-care attribute with good validity and responsiveness, the HUI2 may be preferable to the HUI3 in cost-utility analyses studies in patients with dementia.

Strengths of study design

A particular strength of this study was the use of a formal method to generate *a priori* predictions about associations for our analysis of construct validity. In general, weighted kappa statistics indicated fair to moderate agreement (κ=0.30–0.55) between predicted and observed results. Predictions were least accurate (κ<0.20) for HUI-CCI scale correlations and for ES using the behavioral scale anchor. In the former case, as noted above, limited variance in CCI scores together with a floor effect produced unexpectedly modest correlations with the HUI, and the panel consistently overestimated the correlation. With the ES results using the behavioral anchor, predictions were based on the expected relationship between HUI and the behavioral scale items, but did not take into account the natural global decline in this group of patients with dementia. Thus, most mismatches between predicted and observed results in these responsiveness results resulted from subjects experiencing decline in attributes without an evident conceptual link with the behavioral scale items.

Our study has several additional strengths. We used well-accepted assessment instruments of dementia severity and medical comorbidity in the construct validity assessment. Multiple external indicators, namely the BDRS, CCI, behavioral scale and residential status, reflecting different aspects of dementia health status, were used to assess construct validity. Responsiveness was assessed using behavioral and global status measures, which, together with caregiver factors, are the most important indicators of HRQOL in patients with dementia (5,26,27). Also, our study has good external validity as the sample consisted of community-

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

based elders enrolled from all diagnosed dementia patients from three large medical practice groups.

Limitations of our study

We used an abbreviated behavior scale and a proxy measure of global status, but ideally assessment of instrument validity would have included additional external measures of important aspects of dementia, including a global rating and a more extensive behavioral assessment scale (19). While available evidence supports the use of residential status as a proxy for global dementia severity (12), we note that considerable overlap in dementia levels exists in patients across residential levels. In addition, although support for the test-retest reliability of the proxy-administered HUI3 in patients with mild to moderate dementia has been reported (6), reliability of caregiver-rated HUI2/3 scores in moderate to severe dementia has not been well-established.

The use of proxy informants must be considered in interpreting our findings. Reliability of proxies in assessment of QOL in advanced stages of dementia has not been established. Proxies tend to indicate greater impairment in QOL than do dementia patients in their self-report (27). Discrepancies between patient and proxy ratings are greater in more advanced dementia and in the more subjective HRQOL domains such as emotional status and pain. Also, evidence suggests that proxy ratings may be influenced by degree of relatedness to the patient (28). Nevertheless, as a practical matter, the unreliability and inaccuracy of self-report in patients with moderate to severe dementia necessitate use of proxy informants in HRQOL assessment with this patient population (29). While our study supports the construct validity of the proxy-rated HUI2/3 in patients with mild through severe dementia, further studies are needed to assess the validity (including responsiveness) of the self-rated HUI in subjects with mild to moderate dementia, for whom self-assessment may be feasible. Direct comparison of the reliability, validity, and agreement of proxy and self-reported HUI in patients with a range of dementia grades may help investigators to determine which HUI version (self vs. proxy) to use for which dementia level.

As with the HUI, our dementia rating instrument (BDRS) relied upon caregiver report of the patient’s functional status rather than direct memory assessment. Available evidence, however, including the original paper by Blessed et al (10) as well as more recent work by Jorm (30), indicates that informant-based assessments may better reflect dementia status than direct memory assessment in non-clinical settings.

Finally, our assessments of construct validity and responsiveness assumed linear relationships between HRQOL and behavioral symptoms, dementia severity and comorbidity, but in some chronic medical conditions, such as stroke and congestive heart disease, nonlinear models may better reflect the relationship between HRQOL and disease severity (31).

Conclusion

In summary, we have demonstrated good responsiveness of the caregiver-rated HUI to graded behavioral and global decline but poor responsiveness to improvement among patients with a wide range of dementia severity, including patients with moderate to severe dementia. Notably, the HUI2 may be more responsive to behavioral worsening and global decline than the HUI3 in patients with dementia. In addition, with two prior reports offering conflicting evidence regarding the validity of the HUI in subjects with dementia (5,6), this study provides additional support for the construct validity of the HUI as an HRQOL measure in patients with dementia. Additional studies of validity using multiple external criteria are needed to further gauge the validity, including responsiveness, of the HUI in patients with different grades of dementia, with particular emphasis on the role of weighting and choice of informant (self vs. proxy). In

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the meantime, since responsiveness of HRQOL measures in patients with dementia is not well-established, we recommend that researchers use multiple generic and disease-specific HRQOL measures in dementia trials to confirm the robustness of results using the HUI2/3. Investigators should use the proxy-rated HUI in persons with moderate to severe dementia, but it is possible that self-reported HUI may be appropriate in subjects with milder cognitive impairment.

Acknowledgments

The authors thank Dr. David Sultzer (geriatric psychiatrist) and Dr. Kristin Lueck (neurologist) for work as expert panel members in generating hypotheses regarding the HUI's validity in patients with dementia.

Support

Dr. Kavirajan received salary support through the Department of Veterans Affairs Ambulatory Care/Health Services Research Fellowship (Greater Los Angeles VA Medical Center). These analyses were also supported by the California Department of Public Health (contract 06-55314) for the Alzheimer's Disease Research Center of California. Dr. Hays was supported in part by a P01 grant (AG020679-01) from the National Institute on Aging and the UCLA Resource Center for Minority Aging Research/Center for Health Improvement in Minority Elderly (RCMAR/CHIME), NIH/NIA Grant Award Number P30AG021684.

References

1. Dekosky ST, Orgogozo JM. Alzheimer Disease: Diagnosis, Cost, and Dimensions of Treatment. Alzheimer Disease and Associated Disorders 2001;15(Suppl 1):S3–S7. [PubMed: 11669507]
2. Gold, MR. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996.
3. Knapp M. Economic outcomes and levers: impacts for individuals and society. International Psychogeriatrics 2007;19:483–495. [PubMed: 17391570]
4. Horsman J, Furlong W, Feeny D, et al. The Health Utilities Index (HUI): concepts, measurement properties and applications. Health and Quality of Life Outcomes 2003:54. [PubMed: 14613568]
5. Neumann PJ, Sandberg EA, Araki SM, et al. A comparison of HUI2 and HUI3 Utility Scores in Alzheimer's Disease. Medical Decision Making 2000;20:413–422. [PubMed: 11059474]
6. Naglie G, Tomlinson G, Tansey C, et al. Utility-based quality of life measures in Alzheimer's disease. Quality of Life Research 2006;15:631–645. [PubMed: 16688496]
7. Vickrey B, Mittman B, Connor K, et al. The effect of a disease management intervention on quality and outcomes of dementia care. Annals of Internal Medicine 2006;145:713–726. [PubMed: 17116916]
8. Pippenger M, Holloway R, Vickrey B. Neurologists' use of ICD-9CM codes for dementia. Neurology 2001;56:1206–1209. [PubMed: 11342688]
9. Victoroff J, Nielson K, Mungas D. Caregiver and clinician assessment of behavioral disturbances: the California dementia behavior questionnaire. International Psychogeriatrics 1997;9:155–174. [PubMed: 9309488]
10. Blessed G, Tomlinson BE, Roth M. The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. British Journal of Psychiatry 1968;114:797–811. [PubMed: 5662937]
11. Charlson ME, Pompei P, Ales K, et al. A new method of classifying prognostic validity in longitudinal studies: development and validation. Journal of Chronic Disease 1987;40:373–383.
12. Knopman DS, Berg JD, Thomas R, et al. Nursing home placement is related to dementia progression: experience from a clinical trial. Neurology 1999;52:714–718. [PubMed: 10078715]
13. Cohen J. A power primer. Psychological Bulletin 1992;112:155–159. [PubMed: 19565683]
14. Shekelle P. The Appropriateness Method. Medical Decision Making 2004;24:228–231. [PubMed: 15090107]
15. Hays, R.; Revicki, D. Assessing Quality of Life in Clinical Trials. Oxford: Oxford University Press; 2007. Reliability and validity (including responsiveness) In: Fayers P, Hays R, eds; p. 25-40.
16. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977;33:159–174. [PubMed: 843571]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

17. Davis P, Morris J, Grant E. Brief Screening Tests versus Clinical Staging in Senile Dementia of the Alzheimer Type. Journal of the American Geriatrics Society 1990;38:129–135. [PubMed: 2299116]
18. Fryback D, Dunham N, Palta M, et al. US norms for six generic health-related quality-of-life indexes from the national health measurement study. Medical Care 2007;45:1162–1170. [PubMed: 18007166]
19. Revicki D, Cella D, Hays R, et al. Responsiveness and minimal important differences for patient-reported outcomes. Health and Quality of Life Outcomes 2006:4. [PubMed: 16423298]
20. Feeny D, Furlong W, Torrance G, et al. Multi-attribute and single-attribute utility functions for the health utilities index mark 3 system. Medical Care 2002;40:113–128. [PubMed: 11802084]
21. Fortin M, Hudon C, Dubois M, et al. Comparative assessment of three different indices of multimorbidity for studies on health-related quality of life. Health and Quality of Life Outcomes 2005:1–7. [PubMed: 15634354]
22. Silberfeld M, Rueda S, Krahn M, et al. Content validity for dementia of three generic preference based health related quality of life instruments. Quality of Life Research 2002;11:71–79. [PubMed: 12003057]
23. Kaplan R. The Future of Outcomes Measurement in Rheumatology. American Journal of Managed Care 2007;13:S252–255. [PubMed: 18095788]
24. Norman G, Sloan J, Wyrwich K. Interpretation of Changes in Health-related Quality of Life: The remarkable universality of half a standard deviation. Medical Care 2003;41:582–592. [PubMed: 12719681]
25. Farivar S, Liu H, Hays R. Half standard deviation estimate of the minimally important difference in change scores? Expert Review in Pharmacoeconomics Outcomes Research 2004;4:515–523.
26. Shin I, Carter M, Masterman D, et al. Neuropsychiatric Symptoms and Quality of Life in Alzheimer's Disease. American Journal of Geriatric Psychiatry 2005;13:469–474. [PubMed: 15956266]
27. Karlawish JHT, Casarett D, Klocinski J, et al. The relationship between caregivers' global ratings of Alzheimer's disease patients' quality of life, disease severity, and the caregiving experience. Jourrnal of the American Geriatrics Society 2001;49:1066–1070.
28. Novella JL, Jochum C, Jolly D, et al. Agreement between patients' and proxies' reports of quality of life in Alzheimer's disease. Quality of Life Research 2001;10:443–452. [PubMed: 11763206]
29. Rabins PV, Kasper JD. Measuring quality of life in persons with dementia:conceptual and practical issues. Alzheimer Disease and Associated Disorders 1997;11:100–104. [PubMed: 9437454]
30. Jorm A. The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE). International Psychogeriatrics 2004;16:275–293. [PubMed: 15559753]
31. Ferrucci L, Baldesseroni S, Bandinelli S, et al. Disease severity and health-related quality of life across different chronic conditions. Journal of the American Geriatrics Society 2000;48:1490–1495. [PubMed: 11083330]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

TABLE 1

Baseline Sample Characteristics (n= 408)

	N (%) or Mean (SD) or Median (IQR)
Age, mean (SD), yrs	80 (7)
Female, N (%)	224 (55)
White, N (%)	353 (87)
Type of Dementia, N (%) [*]	
Alzheimer disease	304 (77)
Vascular dementia, multi-infarct dementia	31 (8)
All other	63 (16)
Caregiver Relationship, N (%)	
Spouse	224 (55)
Son/son-in-law or daughter/daughter-in-law	159 (40)
Other	25 (6)
Living in Home or apartment, N (%)	387 (95)
Caregiver lives with person with dementia, N (%)	287 (70)
Number of hours per day caregiver spends caring for person with dementia, median (interquartile range)	5.1 (2.0–20.0)
Charlson Comorbidity Index (CCI), mean (SD) [*]	2.7 (1.8)
Blessed Dementia Rating Scale (BDRS), mean (SD) [†]	5.9 (3.7)
Behavior Scales, mean (SD) [‡]	
Depressed mood	67 (11)
Physical aggression	74 (6)
Anger/Agitation	73 (18)
Composite behavior score	85 (14)
Health Utility Index 2 (HUI2), mean (SD) [§]	
Sensation utility score	0.64 (0.28)
Mobility utility score	0.79 (0.25)
Emotion utility score	0.85 (0.18)
Cognition utility score	0.58 (0.34)
Self-care utility score	0.73 (0.41)
Pain utility score	0.87 (0.20)
Multi-attribute utility score	0.54 (0.23)
Health Utility Index 3 (HUI3), mean (SD) [§]	
Vision utility score	0.84 (0.23)
Hearing utility score	0.76 (0.35)
Speech utility score	0.84 (0.23)
Ambulation utility score	0.73 (0.32)
Dexterity utility score	0.85 (0.27)
Emotion utility score	0.82 (0.21)
Cognition utility score	0.38 (0.29)
Pain utility score	0.81 (0.26)
Multi-attribute utility score	0.17 (0.31)

[*] Dementia type and Charlson Comorbidity Index score were based on medical record abstraction (n = 398). Higher scores indicate greater medical comorbidity.

[†] Observed & possible range: 0 – 17. Higher scores indicate more severe dementia.

[‡] Behavior Scales are scored 0 to 100, where 100 is the best possible state.

[§] Health Utilities Index (HUI) single attribute scores range from 0.00 to 1.00 for both HUI2 and HUI3, where 1.00 is the best possible state. Multi-attribute (index) scores range from −0.03 to 1.00 on the HUI2 and from −0.36 to 1.00 on the HUI3, where 1.00 is the best possible state.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

TABLE 2

Baseline Pearson Correlations of Health Utility Index (HUI) and the Charlson Comorbidity, Blessed Dementia, and Behavior Rating Scales[*]

	Charlson Comorbidity Index	Blessed Dementia Rating Scale	Composite Behavioral Score	Depressed Mood	Physical Aggression	Anger/Agitation
Health Utility Index 2 (HUI2) Utility Scores						
Sensation utility score	−0.07	−0.40[§]	0.26[§]	0.21[§]	0.13[‡]	0.31[§]
Mobility utility score	−0.20[§]	−0.56[§]	0.13[‡]	0.10	0.12[†]	0.14[‡]
Emotion utility score	−0.10	−0.19[§]	0.71[§]	0.66[§]	0.20[§]	0.64[§]
Cognition utility score	−0.0001	−0.66[§]	0.18[§]	0.21[§]	0.17[§]	0.21[§]
Self-care utility score	−0.08	−0.71[§]	0.11[†]	0.06	0.21[§]	0.17[§]
Pain utility score	−0.09[†]	−0.19[§]	0.31[§]	0.33[§]	0.03	0.26[§]
Multi-attribute utility Score	−0.12[‡]	−0.69[§]	0.44[§]	0.37[§]	0.21[§]	0.41[§]
Health Utility Index 3 (HUI3) Utility Scores						
Vision utility score	−0.06	−0.23[§]	0.25[§]	0.15[‡]	0.02	0.20[§]
Hearing utility score	−0.10[†]	−0.05	0.16[‡]	0.18[§]	0.01	0.13[‡]
Speech utility score	0.05	−0.42[§]	0.20[§]	0.10[†]	0.15[‡]	0.27[§]
Ambulation utility score	−0.20[§]	−0.57[§]	0.13[‡]	0.10	0.13[‡]	0.14[‡]
Dexterity utility score	−0.07	−0.46[§]	0.18[§]	0.15[‡]	0.13[‡]	0.17[§]
Emotion utility score	−0.04	−0.20[§]	0.52[§]	0.52[§]	0.09	0.41[§]
Cognition utility score	−0.002	−0.67[§]	0.24[§]	0.21[§]	0.17[§]	0.21[§]
Pain utility score	−0.13[‡]	−0.22[§]	0.26[§]	0.36[§]	0.04	0.22[§]
Multi-attribute utility score	−0.10[†]	−0.69[§]	0.41[§]	0.37[§]	0.19[§]	0.34[§]

[*] n= 392–408. Charlson Comorbidity Index score was based on medical record abstraction; higher scores indicate greater comorbidity burden. Blessed Dementia Rating Scale is scored 0 to 17, where higher scores mean more severe dementia. Composite behavior score, depressed mood, physical aggression and anger/agitation are scored 0 to 100, where 100 is best possible state. Health Utilities Index (HUI) single attribute scores range from 0.00 to 1.00 for both HUI2 and HUI3, where 1.00 is the best possible state. Multi-attribute (index) scores range from −0.03 to 1.00 on the HUI2 and from −0.36 to 1.00 on the HUI3, where 1.00 is the best possible state.

[†] $P \leq 0.05$

[‡] $P \leq 0.01$

[§] $P \leq 0.001$

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

TABLE 3

Score changes and responsiveness of the Health Utilities Index Mark2 (HUI2) and Health Utilities Index Mark 3 (HUI3) by change categories in 22-item behavior score (12 months minus baseline)*

	Large Improvement (n = 53)	Medium Improvement (n = 42)	Small Improvement (n = 43)	No Change (n = 87)	Small Decline (n = 35)	Medium Decline (n = 31)	Large Decline (n = 42)
HUI2 Utility Scores							
Multi-attribute							
Baseline score	0.44 (0.21)	0.52 (0.23)	0.58 (0.24)	0.59 (0.24)	0.64 (0.19)	0.62 (0.21)	0.54 (0.19)
Change in score	0.01 (0.22)	−0.07 (0.24)	−0.003 (0.19)	−0.07 (0.17)	−0.11 (0.17)	−0.15 (0.22)	−0.18 (0.19)
Effect Size	0.04	−0.30	−0.01	−0.30	−0.48	−0.65	−0.78
Sensation							
Baseline score	0.54 (0.32)	0.63 (0.31)	0.66 (0.30)	0.65 (0.30)	0.74 (0.18)	0.78 (0.13)	0.65 (0.27)
Change in score	−0.05 (0.33)	−0.02 (0.38)	−0.02 (0.35)	−0.05 (0.31)	−0.05 (0.24)	−0.14 (0.30)	−0.04 (0.31)
Effect Size	−0.18	−0.07	−0.07	−0.18	−0.18	−0.50	−0.14
Mobility							
Baseline score	0.75 (0.28)	0.80 (0.25)	0.81 (0.25)	0.79 (0.25)	0.87 (0.18)	0.85 (0.21)	0.81 (0.25)
Change in score	0.07 (0.32)	−0.12 (0.36)	0.03 (0.16)	−0.05 (0.20)	−0.08 (0.18)	−0.11 (0.24)	−0.16 (0.25)
Effect Size	0.28	−0.48	0.12	−0.20	−0.32	−0.44	−0.64
Emotion							
Baseline score	0.76 (0.19)	0.79 (0.19)	0.88 (0.18)	0.92 (0.13)	0.94 (0.07)	0.87 (0.14)	0.81 (0.17)
Change in score	−0.03 (0.21)	0.02 (0.18)	0.007 (0.15)	−0.02 (0.13)	−0.06 (0.11)	−0.12 (0.16)	−0.20 (0.22)
Effect Size	−0.18	0.12	0.04	−0.12	−0.36	−0.71	−1.18
Cognition							
Baseline score	0.50 (0.34)	0.60 (0.31)	0.60 (0.32)	0.63 (0.32)	0.65 (0.26)	0.64 (0.31)	0.59 (0.30)
Change in score	0.04 (0.42)	−0.10 (0.32)	−0.03 (0.26)	−0.08 (0.31)	−0.09 (0.38)	−0.18 (0.39)	−0.21 (0.35)
Effect Size	0.13	−0.32	−0.10	−0.26	−0.29	−0.58	−0.67
Self-care							
Baseline score	0.65 (0.44)	0.66 (0.44)	0.83 (0.34)	0.75 (0.41)	0.84 (0.32)	0.81 (0.37)	0.81 (0.37)
Change in score	0.14 (0.53)	−0.11 (0.53)	−0.05 (0.39)	−0.13 (0.38)	−0.19 (0.38)	−0.29 (0.45)	−0.30 (0.50)
Effect Size	0.35	−0.28	−0.13	−0.33	−0.48	−0.73	−0.76
Pain							
Baseline score	0.82 (0.22)	0.89 (0.14)	0.88 (0.20)	0.89 (0.21)	0.87 (0.22)	0.88 (0.22)	0.88 (0.18)
Change in score	0.04 (0.26)	−0.02 (0.16)	0.008 (0.18)	−0.01 (0.18)	−0.01 (0.18)	−0.01 (0.21)	−0.10 (0.31)
Effect Size	0.20	−0.10	0.04	−0.05	−0.05	−0.05	−0.50
HUI3 Utility Scores							
Multi-attribute							
Baseline score	0.03 (0.27)	0.15 (0.32)	0.18 (0.33)	0.24 (0.35)	0.27 (0.24)	0.32 (0.31)	0.13 (0.23)
Change in score	−0.02 (0.27)	−0.05 (0.28)	0.02 (0.22)	−0.09 (0.25)	−0.15 (0.24)	−0.21 (0.30)	−0.20 (0.18)
Effect Size	−0.06	−0.16	0.06	−0.29	−0.48	−0.67	−0.64
Vision							
Baseline score	0.76 (0.32)	0.76 (0.27)	0.83 (0.21)	0.90 (0.17)	0.88 (0.16)	0.92 (0.11)	0.83 (0.25)
Change in score	−0.06 (0.26)	0.05 (0.32)	−0.001 (0.22)	−0.03 (0.19)	−0.06 (0.19)	−0.06 (0.21)	−0.03 (0.25)
Effect Size	−0.26	0.21	−0.004	−0.13	−0.24	−0.26	−0.13
Hearing							
Baseline score	0.68 (0.39)	0.79 (0.34)	0.75 (0.36)	0.73 (0.36)	0.82 (0.27)	0.93 (0.17)	0.74 (0.36)
Change in score	−0.06 (0.46)	0.08 (0.35)	−0.01 (0.41)	−0.004 (0.36)	−0.04 (0.25)	−0.12 (0.37)	0.03 (0.31)
Effect Size	−0.17	0.23	−0.03	−0.02	−0.12	−0.35	0.09
Speech							
Baseline score	0.77 (0.26)	0.83 (0.22)	0.88 (0.20)	0.87 (0.18)	0.89 (0.15)	0.91 (0.16)	0.86 (0.16)
Change in score	0.01 (0.26)	−0.06 (0.25)	0.02 (0.21)	−0.06 (0.25)	−0.05 (0.16)	−0.09 (0.20)	−0.12 (0.24)
Effect Size	0.05	−0.30	0.10	−0.30	−0.25	−0.45	−0.60
Ambulation							

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

	Large Improvement (n = 53)	Medium Improvement (n = 42)	Small Improvement (n = 43)	No Change (n = 87)	Small Decline (n = 35)	Medium Decline (n = 31)	Large Decline (n = 42)
Baseline score	0.68 (0.34)	0.74 (0.30)	0.76 (0.30)	0.73 (0.31)	0.81 (0.23)	0.79 (0.29)	0.74 (0.33)
Change in score	0.08 (0.39)	−0.13 (0.42)	−0.01 (0.20)	−0.09 (0.25)	−0.12 (0.23)	−0.13 (0.31)	−0.19 (0.27)
Effect Size	0.02	−0.40	−0.02	−0.27	−0.37	−0.41	−0.60
Dexterity							
Baseline score	0.79 (0.28)	0.78 (0.29)	0.85 (0.25)	0.87 (0.27)	0.94 (0.14)	0.93 (0.17)	0.89 (0.23)
Change in score	0.13 (0.36)	−0.13 (0.38)	0.02 (0.27)	−0.07 (0.32)	−0.13 (0.31)	−0.15 (0.29)	−0.18 (0.37)
Effect Size	0.48	−0.47	0.06	−0.24	−0.47	−0.54	−0.66
Emotion							
Baseline score	0.73 (0.24)	0.79 (0.21)	0.80 (0.21)	0.88 (0.18)	0.91 (0.09)	0.88 (0.14)	0.80 (0.19)
Change in score	−0.04 (0.28)	−0.002 (0.16)	0.002	−0.01 (0.18)	−0.05 (0.18)	−0.11 (0.18)	−0.17 (0.24)
Effect Size	−0.18	−0.01	0.01	−0.06	−0.24	−0.51	−0.83
Cognition							
Baseline score	0.30 (0.26)	0.40 (0.29)	0.39 (0.29)	0.46 (0.31)	0.41 (0.23)	0.44 (0.30)	0.37 (0.26)
Change in score	0.01 (0.28)	−0.07 (0.23)	−0.01 (0.19)	−0.08 (0.26)	−0.07 (0.27)	−0.17 (0.27)	−0.19 (0.24)
Effect Size	0.02	−0.25	−0.02	−0.27	−0.23	−0.61	−0.65
Pain							
Baseline score	0.71 (0.33)	0.82 (0.21)	0.80 (0.27)	0.83 (0.26)	0.87 (0.20)	0.84 (0.27)	0.83 (0.23)
Change in score	0.04 (0.31)	0.02 (0.30)	0.04 (0.29)	−0.01 (0.25)	−0.08 (0.25)	−0.05 (0.20)	−0.19 (0.30)
Effect Size	0.15	0.07	0.13	−0.05	−0.30	−0.19	−0.70

[*] Change in score is defined as score at follow-up minus baseline score. Effect size (ES) is computed as follows: Change score/standard deviation at baseline for entire sample.

Behavioral change categories were defined according to the effect size (ES) of behavior scale change during the follow-up interval. No change (|ES| <0.2), small change (0.2 ≤|ES| <0.5), medium change (0.5≤|ES|<0.8), or large change (|ES| ≥0.8). Health Utilities Index (HUI) single attribute scores range from 0.00 to 1.00 for both HUI2 and HUI3, where 1.00 is the best possible state. Multi-attribute (index) scores range from −0.03 to 1.00 on the HUI2 and from −0.36 to 1.00 on the HUI3, where 1.00 is the best possible state.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE 4

Effect sizes on the Health Utilities Index Mark2 (HUI2) and Health Utilities Index Mark 3 (HUI3) by change in residential status (12 months minus baseline)*

	Effect Size No change - remained in home (n = 240)	Effect Size Change from home to skilled nursing facility (n = 29)
HUI2 Utility Scores		
Multi-attribute	−0.30	−0.76
Sensation	−0.13	−0.41
Mobility	−0.17	−0.70
Emotion	−0.39	−0.88
Cognition	−0.27	−0.36
Self-care	−0.27	−0.65
Pain	−0.05	−0.88
HUI3 Utility Scores		
Multi-attribute	−0.27	−0.50
Vision	−0.08	−0.23
Hearing	−0.03	−0.23
Speech	−0.25	−0.51
Ambulation	−0.19	−0.74
Dexterity	−0.12	−0.52
Emotion	−0.28	−0.53
Cognition	−0.26	−0.38
Pain	−0.11	−0.70

* Health Utilities Index (HUI) single attribute scores range from 0.00 to 1.00 for both HUI2 and HUI3, where 1.00 is the best possible state. Multi-attribute (index) scores range from −0.03 to 1.00 on the HUI2 and from −0.36 to 1.00 on the HUI3, where 1.00 is the best possible state.