ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

Patient-Reported Outcomes

How Should We Capture Health State Utility in Dementia? Comparisons of DEMQOL-Proxy-U and of Self- and Proxy-Completed EQ-5D-5L

Adam Martin, PhD,[1,*] David Meads, PhD,[1] Alys W. Griffiths, PhD,[2] Claire A. Surr, PhD[2]

[1]Academic Unit of Health Economics, Leeds Institute of Health Sciences, University of Leeds, Leeds, England, UK; [2]Centre for Dementia Research, School of Health and Community Studies, Leeds Beckett University, Leeds, England, UK

ABSTRACT

Background: Dementia-specific and proxy-completed preference-based measures have been proposed for use in intervention studies involving people living in residential care, in instances where generic, self-reported preference-based measures have been deemed inappropriate.

Objective: This study was conducted to investigate the construct validity, criterion validity, and responsiveness of DEMQOL-Proxy-U and of self- and proxy-completed EQ-5D-5L.

Methods: The analysis used a 3-wave, individual-level data set of 1004 people living with dementia in residential care that included self-completed EQ-5D-5L and formal-carer and informal-carer proxy-completed EQ-5D-5L and DEMQOL-Proxy-U utility values, in addition to other nonutility cognitive measures (Functional Assessment Staging [FAST], Clinical Dementia Rating [CDR], Cohen-Mansfield Agitation Inventory [CMAI]) and health-related quality of life (HRQOL) measures (nursing home version of the Quality of Life with Alzheimer's disease scale [QOL-AD-NH], Quality of Life in Late-Stage Dementia [QUALID] scale). Construct validity, criterion validity, and responsiveness were assessed using correlation, Bland-Altman plots, and panel data regression models.

Results: Self-completed EQ-5D-5L failed to reflect clinically important differences and changes in FAST, CDR, and CMAI but did capture the resident's own view of HRQOL (QOL-AD-NH). As dementia severity increased, collection of EQ-5D-5L-proxy and DEMQOL-Proxy-U data was more feasible than collection of self-completed EQ-5D-5L. These formal-carer and informal-carer proxy measures also better reflected changes in FAST, CDR, and CMAI but did not capture the resident's own view of HRQOL (QOL-AD-NH), despite adequately capturing the proxy's own view of the resident's HRQOL (QUALID). This indicates discrepancies between a proxy's view and resident's view of the impact that tangible declines in health, cognition, or functional abilities have on HRQOL. The EQ-5D-5L-proxy and DEMQOL-Proxy-U were generally poor substitutes. Regardless of which proxy completed it, the EQ-5D-5L-proxy was typically more responsive than the DEMQOL-Proxy-U to changes in CDR, FAST, and CMAI, indicating that use of the DEMQOL-Proxy-U is not always justified.

Conclusion: Disparities in the measurement properties of different utility measures mean that choices about how to measure utility in trials could affect economic evaluation outcomes and hence how resources are allocated for dementia care.

Keywords: dementia, DEMQOL, EQ-5D, preference-based measures.

VALUE HEALTH. 2019; 22(12):1417–1426

Introduction

The National Institute for Health and Care Excellence (NICE) Technology Assessment reference case states that health-related quality of life (HRQOL) should be measured using the self-completed EQ-5D, which has been shown to be valid, reliable, and responsive across numerous disease areas.[1] However, NICE may consider other preference-based measures (PBMs) to be informative, including when the self-completed EQ-5D has weak construct validity within a particular patient population. In these cases, proxy-completed or condition-specific utility measures may be used as supplementary evidence.

For people with dementia, proxy-completed PBMs may be justified owing to problems with communication recall, time perception, and insight.[2] One study showed almost half (48%) of respondents with dementia self-reported having no problems on any EQ-5D dimension (ie, "perfect health"),[3] which seemed unlikely to accurately reflect the respondents' true health state, given

---

[*] Address correspondence to: Adam Martin, PhD, Academic Unit of Health Economics, Leeds Institute of Health Sciences, University of Leeds, Leeds, England LS2 9TJ, UK. Email: a.martin1@leeds.ac.uk

https://doi.org/10.1016/j.jval.2019.07.002

the instrument's assessment of such areas as mobility and pain. However, proxy assessors can have different perceptions of another's capabilities and health state,[2,4] perhaps particularly for more unobservable or subjective dimensions (eg, pain or anxiety/depression vs functional ability), and exhibit additional sources of bias that could differ between individuals. For example, a relative's response might reflect additional emotional burden when compared with formal or employed carers. Ethical issues arise if the views of individuals about their own HRQOL differ from the views of their proxies, especially if these affect economic evaluation outcomes and hence resource allocation decisions affecting them.

Dementia-specific PBMs are designed to capture physical function and cognitive changes that are common in dementia but poorly reflected in generic measures and have been valued by general population samples, including in Australia and the UK.[5-7] Dementia researchers thus face a relatively unique challenge in choosing an appropriate measure (generic vs condition-specific) as well as respondent (self vs informal carer vs formal carer).

This study's main purpose was to examine and compare the measurement properties of proxy-completed versus self-completed PBMs and generic versus dementia-specific PBMs. Analyses of construct validity, criterion validity, and responsiveness will aid decisions about how best to capture utility (and change in utility) and thus support the interpretation and design of health economic analyses involving people with dementia.

This study used a large, 3-wave, individual-level data set of people with dementia living in care homes collected during a multisite trial (Enhancing Person Centred Care In Care Homes Trial).[8] The data include a rich combination of resident-completed, proxy-reported, generic, and dementia-specific PBMs in addition to other non-PBM cognitive and HRQOL measures. Although two Australian[4,9] studies have compared the included PBMs, this UK-based study used larger sample sizes, compared more cognitive health and HRQOL measures, and involved modeling temporal within-individual changes, using panel data regression techniques.

Methods

Data Set

Residents (n = 1004) were recruited from 50 UK care homes in an open cohort cluster randomized controlled trial that evaluated Dementia Care Mapping^{TM}, an intervention that aims to reduce agitation and improve HRQOL for people with dementia.[8,10] Residents (supported by researchers), a formal-carer proxy assessor (member of care home staff), and, where possible, a further informal-carer proxy assessor (a friend or relative who visited at least fortnightly) were invited to complete PBMs and other questionnaires at 1 (n = 425), 2 (n = 173), or 3 (n = 406) time points (when they entered the trial and at up to 2 follow-ups) between May 2014 and May 2017. The mean time between time points was 252 days (SD 55.9). Not all residents were assessed at all 3 time points mainly owing to death (n = 281), recruitment after trial baseline (n = 261), or leaving the care home (n = 47) (see Appendix Fig. 1 in Supplementary Materials found at https://doi.org/10.1016/j.jval.2019.07.002). Formal-carer proxies were replaced whenever they became unavailable (eg, because of staff turnover, sickness, or other absence). Not all residents had an informal-carer proxy either because their recruitment ceased in this study prior to the third time point or owing to a lack of regular visitors who wished to participate in the study.

Description of Variables

Table 1 summaries the resident- and proxy-completed questionnaires used in this study. Residents' health status was self-reported using the 5-dimension EQ-5D-5L and proxy-reported using the EQ-5D-5L-proxy (version 1)[11] and the 31-item DEMQOL-Proxy.[5] UK scoring algorithms were used to convert these to the EQ-5D-5L and DEMQOL-Proxy-U (using 4 items from the DEMQOL-Proxy—namely, appearance, memory, negative emotion, and positive emotion) utility scores.[5,12]

Residents' degree of cognitive and functional impairment was assessed by formal-carer proxies using the following:

- Functional Assessment Staging (FAST), which scores functional symptoms of dementia from 1 ("normal adult" without difficulties) to 7 ("severe dementia," eg, daily speech limited to ≤5 words).[13] In this study, scores ≤4 (ie, "mild dementia," eg, difficulties planning dinner for guests) were combined into a single category because recruitment to the trial required a score ≥4.
- Clinical Dementia Rating (CDR) scale, which assesses cognitive impairment in 6 categories (eg, related to memory or orientation) and generates a single severity score ranging from 0 ("normal adult") to 3 ("severe dementia").[14]
- Cohen-Mansfield Agitation Inventory (CMAI), which measures the frequency of 29 behaviors typically associated with agitation on a 7-point scale (from ”never” to "several times an hour") and generates a single score ranging from 29 to 203 (where changes ≥8 can be considered clinically meaningful).[15]

Residents' HRQOL was self-reported using the 15-item nursing home version of the Quality of Life in Alzheimer's Disease scale (QOL-AD-NH), which scores items including energy, mood, living situation, and memory on 4-point scales, and by both proxies using the 11-item Quality of Life in Late-Stage Dementia (QUALID) measure, which scores items including smiling, sadness, and enjoyment of interaction with others on 6-point scales.[16] In both cases, the sum of item-level scores generates a single measure on a 45-point scale.

Statistical Analysis

Descriptive analysis

Summary statistics (including the extent of missing data) were reported for variables of interest. Selected frequency distributions were examined using histograms.

Construct validity

Construct validity was examined by comparing utility scores. First, pairs of utility scores generated by different PBMs completed by the same proxy respondent at the same time point were compared (eg, informal carers' EQ-5D-5L-proxy and DEMQOL-Proxy-U responses) using Spearman's rank-order correlation (where correlation <0.3 was considered weak, 0.3 to <0.5 moderate, 0.5 to <0.6 strong, and ≥0.6 very strong[4]) and Bland-Altman plots (differences between 2 utility values generated by 2 different PBMs plotted against the mean of the utility values). Second, utility scores generated from the same PBM completed by different respondents at the same time point were compared (eg, informal-carer EQ-5D-5L-proxy vs formal-carer EQ-5D-5L-proxy or resident-completed EQ-5D-5L) using Spearman's rank-order correlation and Bland-Altman plots (first, differences between 2

Table 1. Description of questionnaires used in the study.

Description of questionnaire							Description of variable used in this analysis	
	What it measures	Resident	Formal-carer proxy	Informal-carer proxy	Perspective adopted by proxies	Timescale	Conversion of questionnaire data to numerical score	Range of possible values
Preference based measures (increasing in quality of life)								
EQ-5D-5L	Five dimensions: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression	✓	✓*	✓*	Their own opinion of the resident's health status	Resident's immediate situation—ie, their "health today"	Index utility score generated using a general population valuation of the health states[12]	−0.285 to 1
DEMQOL-Proxy	31 items including factors related to appearance, memory, positive emotions, and negative emotions	✗	✓	✓	Their view of what they think residents would provide themselves if they were willing and able	How the resident felt during the past week	Relevant components of the DEMQOL-Proxy converted to an index utility score (DEMQOL-Proxy-U) using a general population valuation of the health states[5]	0.363 to 0.937
Cognitive measures (increasing in severity of dementia or agitation)								
FAST	Functional severity of dementia on a scale from 1 (a "normal" adult with no difficulties) to 7 ("severe dementia")	✗	✓	✗	Not specified	Not specified	FAST scores ≤4 were combined into a single category (due to small sample sizes where FAST score <4)	Four categories: ≤4, 5, 6, 7
CDR	Cognitive impairment on 6 cognitive categories	✗	✓	✗	Not specified	Not specified	Converted to an overall severity rating ranging from 0 to 3. Ratings of 0, 0.5, and 1 were combined in this analysis (owing to small sample sizes where CDR<1)	Three categories: ≤1, 2, 3
CMAI	Frequency of 29 agitated or aggressive behaviors on a 7-point scale ranging from 1 ("never") to 7 ("several times an hour")	✗	✓	✗	Not specified	Based on observations during the previous fortnight	Frequency scores are summed	29 to 203
Other HRQOL measures (increasing in quality of life)								
QOL-AD-NH	15 aspects of quality of life including perception of energy, mood, living situation, and memory on a 4-point scale (poor, fair, good, excellent)	✓	✗	✗	N/A	N/A	Individual scores are summed	45-point scale

continued on next page

Table 1. Continued

Description of questionnaire							Description of variable used in this analysis	
	What it measures	Resident	Formal-carer proxy	Informal-carer proxy	Perspective adopted by proxies	Timescale	Conversion of questionnaire data to numerical score	Range of possible values
QUALID	Presence and frequency of 11 quality of life indicators including smiling, appearing sad, crying, and enjoying eating	✗	✓	✓	Not specified	Based on observations during the previous week	Individual scores are summed	45-point scale[†]

Note. Except for FAST and CDR, all variables were treated as continuous variables.

CDR indicates Clinical Dementia Rating; CMAI, Cohen-Mansfield Agitation Inventory; FAST, Functional Assessment Staging; HRQOL, health-related quality of life; N/A, not applicable; QOL-AD-NH, nursing home version of the Quality of Life in Alzheimer's Disease scale; QUALID, Quality of Life in Late-Stage Dementia scale.

[*] EQ-5D-5L proxy version 1.[11]

[†] To aid interpretation and for consistency with the QOL-AD-NH, but contrary to general usage, the QUALID scores were reversed in this study so that higher scores represented improved quality of life.

utility values recorded at the same time point were plotted against the mean, and second, differences between the change in utility values between 2 time points were plotted against the mean at the first time point).

Criterion validity

Criterion validity was examined by comparing utility scores and nonutility measures. Associations between each PBM and the non-PBM cognitive (FAST, CDR, and CMAI) and HRQOL measures (QUALID and QOL-AD-NH) (described earlier) were examined at the same time point using Spearman's rank-order correlation and panel data regression models with random effects (model 1). In the absence of evidence that other regression models would perform better in a dementia population, all regression analyses in this study used linear model specifications. This enabled model output comparisons across multiple outcome measures.

$$\text{Model 1: } Utility_{i,t} = \beta_1 + \beta_2 health_{i,t} + \varepsilon_{i,t},$$

where $utility_{i,t}$ is the utility score for each individual (i) at each time point (t) for a particular PBM-respondent combination (n = 5), and $health_{i,t}$ is a particular cognitive or HRQOL measure (n = 6) (this was a single continuous variable, except for the FAST and CDR scores, where a vector of categorical variables was created, with least impaired states in the reference category). This enabled model output comparisons across multiple outcome measures.

The model was run separately for each PBM-respondent combination (n = 5) and cognitive or HRQOL variable (n = 6). Missing data for particular utility scores meant sample sizes varied in each analysis. A sensitivity analysis used a restricted sample of observations where all 5 utility scores were completed.

The coefficient of interest (β_2) shows the association between an additional unit of cognitive health or HRQOL (eg, a 1-unit change in CMAI), or being in a particular health category compared with the reference category (eg, FAST score 6 compared with FAST score ≤ 4), and a particular utility score. We calculated 95% confidence intervals and adjusted for multiple comparisons using Bonferroni correction (n = 6 per outcome measure). Statistically significant β_2 values indicated criterion validity.

Responsiveness

The impact of changes in the cognitive or HRQOL measures (n = 6) on each utility score (n = 5) was evaluated using linear fixed effects panel data regression models (as model 1). The fixed effects approach means that only within-individual changes are included in the analysis. Hence, statistically significant β_2 values were deemed indicative of responsiveness.

Results

Descriptive Statistics

The sample included 1989 observations from 1004 residents and an average of 1.98 time points per resident (SD 0.91). Of these, 406 residents (40.4%) participated in the study at all 3 time points; the remainder of the residents were in the study for only 1 (42.3%) or 2 (17.2%) time points.

Residents ranged from 58.0 to 102.6 years of age on entering the data set (mean = 85.5 yr). They were more likely female (73.2%) than male and 3.3% had ethnic backgrounds other than white British (see Table 2 for descriptive statistics, with selected frequency distributions shown in Appendix Fig. 2 in the Supplementary Materials found at https://doi.org/10.1016/j.jval.2019.07.002).

PBMs

Sample sizes were largest for the formal-carer EQ-5D-5L-proxy and DEMQOL-Proxy-U, with at least 1 observation for 1003 individuals (100%) (N = 1980 for EQ-5D-5L-proxy and N = 1983 for DEMQOL-Proxy-U). Missing formal-carer–completed PBM data were rare (15 instances) and due to partial questionnaire completion (1-2 missing items, 13 instances) or noncompletion (2 instances).

At least 1 resident-reported EQ-5D-5L utility score was available for 558 individuals (56%) in the data set (N = 897, 45%), and of the resident-reported EQ-5D-5L utility scores, there was an average of 1.60 observations per individual. There were 67 instances (3.4%) of partial completion (1-4 missing items) and 1025

Table 2. Descriptive statistics for performance-based measures and other cognitive and health-related quality of life measures.

	Number of observations (N) and % of all observations	Number of individuals (n) and % of all observations	Number of observations per individual	Mean value	Range	Variation*		
						Overall	Between individuals	Within individuals
Performance-based measures (increasing in quality of life)								
EQ-5D								
Resident	897 (45.1%)	558 (55.6%)	1.607	0.861	−0.092 to 1	0.175	0.163	0.087
Formal-carer	1980 (99.5%)	1003 (99.9%)	1.974	0.658	−0.218 to 1	.241	0.221	0.118
Informal-carer	349 (17.5%)	184 (18.3%)	1.897	0.486	−0.281 to 1	0.249	0.224	0.105
DEMQOL-Proxy-U								
Formal-carer	1983 (99.7%)	1003 (99.9%)	1.977	0.751	0.363 to 0.937	0.117	0.100	0.071
Informal-carer	342 (17.5%)	184 (18.3%)	1.858	0.701	0.404 to 0.937	0.127	0.112	0.062
Cognitive measures (increasing in severity of dementia or agitation)								
CMAI								
Formal-carer	1964 (98.7%)	985 (98.1%)	1.994	45.114	29 to 129	16.570	15.518	7.693
FAST (categorical variable)								
Formal-carer	1954 (98.2%)	994 (99.0%)	1.966	n/a	1 to 4	192 (9.83%)	162 (16.3%)	64.3%
					5	156 (7.98%)	146 (14.7%)	56.05%
					6	1165 (59.62%)	728 (73.2%)	82.07%
					7	441 (22.57%)	278 (28.0%)	75.72%
CDR (categorical variable)								
Formal-carer	1957 (98.4%)	987 (98.3%)	1.983	n/a	0 or 0.5	75 (3.8%)	70 (7.1%)	58.3%
					1	450 (23.0%)	341 (34.6%)	69.6%
					2	767 (39.2%)	539 (54.6%)	70.4%
					3	665 (34.0%)	433 (43.9%)	76.1%
Other health-related quality of life measures (inceasing in quality of life)								
QOL-AD-NH								
Resident	665 (33.4%)	448 (44.6%)	1.484	42.466	17 to 60	5.793	5.667	2.666
QUALID								
Formal-carer	1958 (98.4%)	987 (98.3%)	1.984	19.970	11 to 49	6.598	5.894	3.632
Informal-carer	360 (18.1%)	186 (18.5%)	1.935	22.162	11 to 44	7.499	6.955	3.211

CDR indicates Clinical Dementia Rating; CMAI, Cohen-Mansfield Agitation Inventory; FAST, Functional Assessment Staging; QOL-AD-NH, nursing home version of the Quality of Life in Alzheimer's Disease scale; QUALID, Quality of Life in Late-Stage Dementia scale.

*For continuous variables, these 3 columns show standard deviation. For categorical variables (FAST and CDR), these 3 columns show the following:

- For the "Overall" column: The number of observations in a given category and as a percentage of all observations (total = 100%).
- For the "Between individuals" column: The number of individuals that were ever in a given category (eg, 70 individuals ever had CDR score = 0 or 0.5) and as a percentage of all individuals (eg, 7.1% of all individuals ever had CDR score = 0 or 0.5).
- For the "Within individuals" column: The fraction of time that individuals were in a given category, conditional on ever having been in that category (eg, conditional on having ever had a CDR score=0 or 0.5, 58.3% of an individual's observations were in that category).

instances (52%) of noncompletion (with reasons recorded in 1006 instances, typically related to the resident being unable or reluctant to communicate with the researcher). The likelihood of noncompletion increased with dementia severity, rising, for example, from 17% of observations in the "normal or very mild dementia" CDR category to 83% in the "severe dementia" category.

Sample sizes were smallest for the informal-carer PBMs (n = 184 individuals [18.3% of all individuals], with N = 349 EQ-5D-5L-proxy observations and N = 342 DEMQOL-Proxy-U observations) mainly owing to fewer eligible informal-carer proxies being recruited, but also due to 1-3 missing items on the PBM questionnaire (33 instances).

The mean value of DEMQOL-Proxy-U scores was higher than the mean value of EQ-5D-5L-proxy scores (ranging from 0.486 for informal-carer EQ-5D-5L-proxy to 0.751 for formal-carer DEMQOL-Proxy-U), with a lower standard deviation. The resident-completed EQ-5D-5L had the highest mean utility score (0.861). Across all PBM respondent combinations, within-individual deviation was smaller than deviation between-individuals.

Construct Validity

Appendix Figures 3-5 show the Spearman's rank order correlation coefficients and Bland-Altman plots. Appendix Figures 6-9 show comparisons between over-time changes in the resident-completed and 2 proxy-completed EQ-5D-5L utility scores. (These figures appear in the Supplementary Materials found at https://doi.org/10.1016/j.jval.2019.07.002.)

Different PBMs completed by the same proxy respondent

Correlation between informal-carer–completed utility scores was moderate (rho = 0.386) and between formal-carer–completed utility scores was weak (0.191) (Appendix Fig. 3[i] and [ii] in Supplementary Materials found at https://doi.org/10.1016/j.jval.2019.07.002). For both proxies, the DEMQOL-Proxy-U scores were higher on average than the EQ-5D-5L-proxy scores, although discrepancies between scores were lower at higher utility levels. Mean differences between EQ-5D-5L-proxy and DEMQOL-Proxy-U utility scores were greater among informal-carer (0.214) than formal-carer (0.092) proxies, with smaller discrepancies at higher utility levels.

Figure 1. Associations between utility scores and nonutility cognitive and health-related quality of life measures (criterion validity).

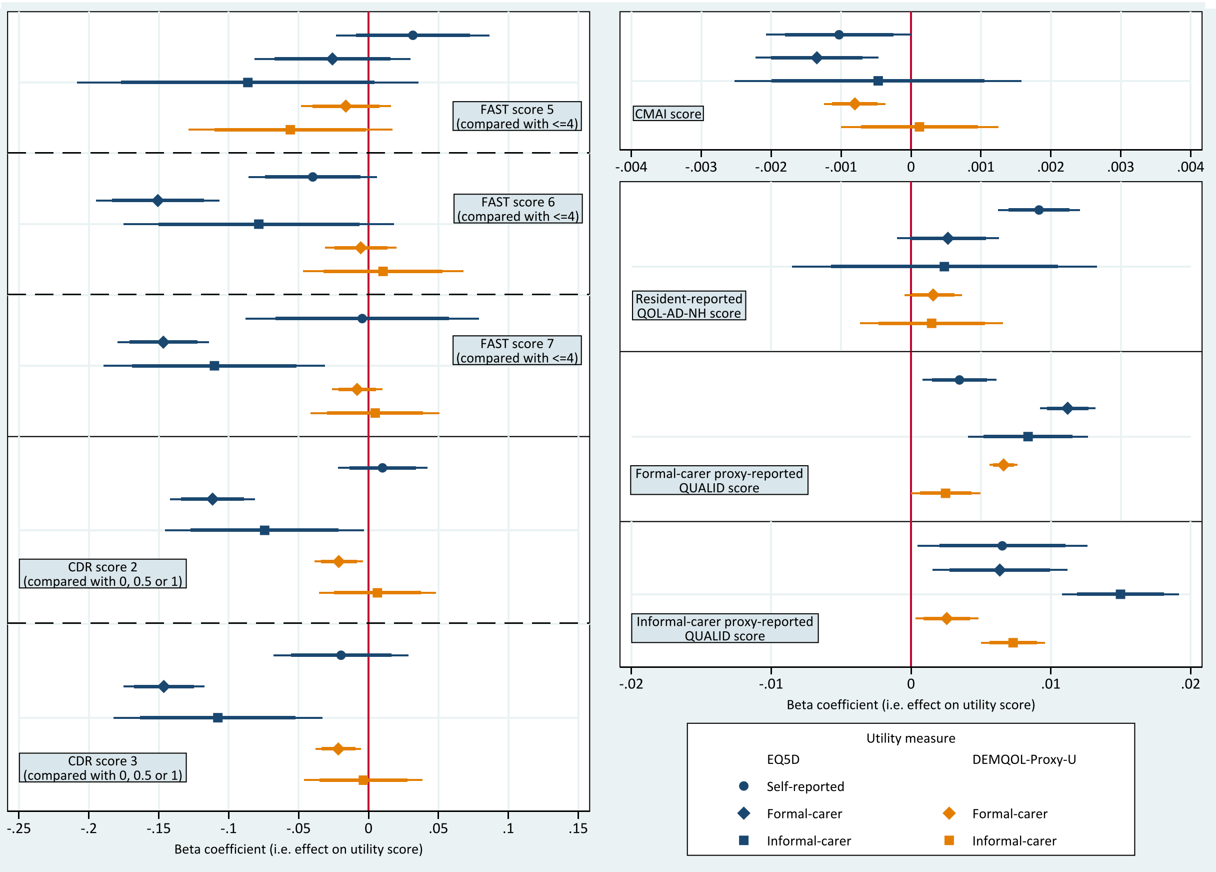

Figure shows the β_2 coefficients for model 1 run for each combination of utility (n = 5) and non-utility cognitive measures (n = 3). The FAST and CDR scores are categorical variables. CDR 0, 0.5 and 1 are in the reference category. FAST ≤4 is in the reference category. The CMAI (range 29-203), QOL-AD and QOL-AD-NH (45 point scales) scores are continuous variables. The thicker confidence intervals indicate $P < .05$. The thinner confidence intervals are after adjustment for multiple hypothesis testing using Bonferroni correction. CDR indicates Clinical Dementia Rating; CMAI, Cohen-Mansfield Agitation Inventory; FAST, Functional Assessment Staging; QOL-AD-NH, nursing home version of the Quality of Life in Alzheimer's disease scale; QUALID, Quality of Life in Late-Stage Dementia scale.

Different respondents completing the same PBM

Correlation between informal-carer and formal-carer proxy-completed utility scores was strong for EQ-5D-5L (rho = 0.528) and weak for DEMQOL-Proxy-U (0.088) (Appendix Fig. 3[iii] and [iv] in Supplementary Materials found at https://doi.org/10.1016/j.jval.2019.07.002). For both PBMs, formal-carer proxy-completed scores were higher on average than the informal-carer proxy-completed scores, although for EQ-5D-5L they were lower at higher utility levels and for DEMQOL-Proxy-U the discrepancies between scores were very small, especially at the highest and lowest utility levels.

Correlation between formal-carer EQ-5D-5L-proxy and resident-reported EQ-5D-5L utility scores was weaker (rho = 0.264) than that between informal-carer EQ-5D-5L-proxy and resident-completed EQ-5D-5L utility scores (0.354) (Appendix Figs. 4 and 5 in Supplementary Materials found at https://doi.org/10.1016/j.jval.2019.07.002). In both cases, correlation (rho) between the carer-reported EQ-5D-5L-proxy and resident-reported EQ-5D-5L utility scores decreased, and mean differences increased, with dementia severity (CDR score). On average, EQ-5D-5L-proxy scores were lower than the resident-completed EQ-5D-5L scores (eg, mean difference of −0.118 for formal carers and −0.275 for informal carers), with the greatest discrepancies between utility scores at lower utility levels. On average, over-time changes in utility were smaller for EQ-5D-5L-proxy respondent scores when compared with resident-reported EQ-5D-5L scores, although there was considerable heterogeneity between individuals (Appendix Figs. 6-9 in Supplementary Materials found at https://doi.org/10.1016/j.jval.2019.07.002).

Criterion Validity

Appendix Table 1 (in Supplementary Materials found at https://doi.org/10.1016/j.jval.2019.07.002) shows the Spearman's rank-order correlation results. Figure 1 shows regression results and Appendix Figure 10 (in Supplementary Materials found at https://doi.org/10.1016/j.jval.2019.07.002) shows the sensitivity analysis where data were complete for all 5 utility scores.

Cognitive measures

Relationship with resident-completed EQ-5D-5L scores. No relationship was identified in either the Spearman's rank or (Bonferroni adjusted) regression analysis between the resident-completed EQ-5D-5L and the CDR or FAST measures. Although before Bonferroni adjustment a statistically significant relationship was observed between the resident-completed EQ-5D-5L and the CMAI measure, the effect size was very small (an 8-unit increase in CMAI, ie, increased agitation, was associated with lower utility of 0.01).

Figure 2. Responsiveness of utility scores to changes in nonutility cognitive measures.

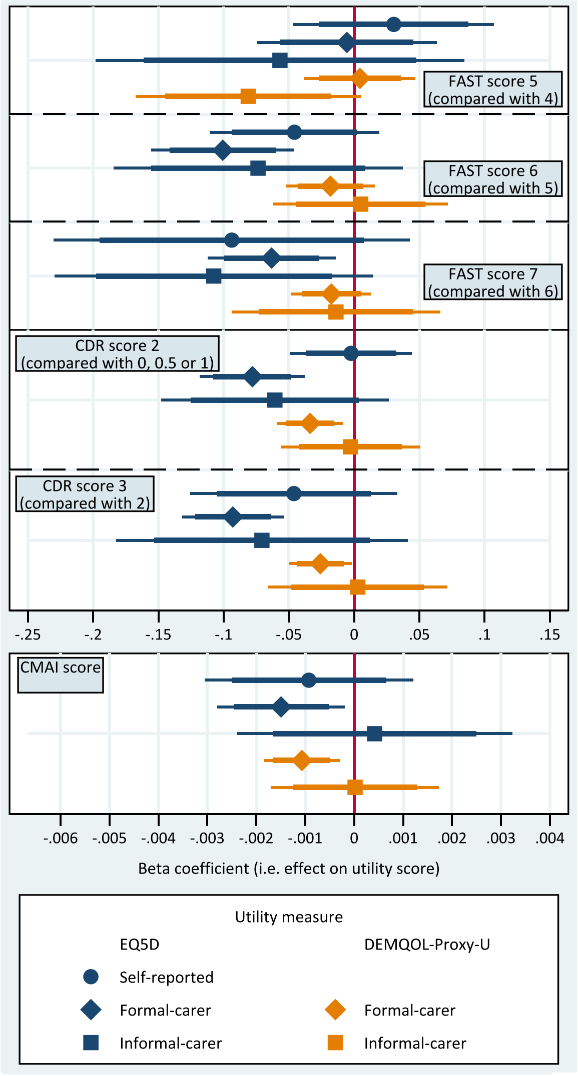

Figure shows the β_2 coefficients for model 1 run for each combination of utility (n = 5) and non-utility cognitive measures (n = 3). The FAST and CDR scores are categorical variables. The β_2 coefficient for FAST score 5 is compared here to the β_2 coefficient for FAST score 4, coefficients for FAST score 6 are compared to FAST score 5, and FAST score 7 is compared to FAST score 6. Similarly, CDR score 2 is compared to the reference category (i.e. CDR 0, 0.5 and 1), and CDR score 3 is compared to CDR score 2. The CMAI (range 29-203) is a continuous variable. The thicker confidence intervals indicate $P < .05$. The thinner confidence intervals are after adjustment for multiple hypothesis testing using Bonferroni correction. The responsiveness of utility scores to changes in non-utility HRQOL measures are shown in Appendix Figure 11. CDR indicates Clinical Dementia Rating; CMAI, Cohen-Mansfield Agitation Inventory; FAST, functional assessment staging; HRQOL, health-related quality of life.

Relationship with proxy-completed EQ-5D-5L scores. The relationship between both EQ-5D-5L-proxy utility scores and the FAST and CDR measures was shown by the Spearman's rank-order coefficients to be moderate (ranging from rho = −0.346 to −0.492) and, in the regression results, to be statistically significant and negative (with 1 exception). All regression coefficients increased in magnitude with dementia severity (eg, for formal carers, the coefficient for FAST score 7 was −0.323 compared with −0.177 for FAST score 6, both compared to FAST score ≤4). The size of the coefficients was generally comparable for both proxies (eg, coefficient for FAST score 6 compared with FAST score 4 was −0.18 in the formal-carer analysis and −0.16 in the informal-carer analysis).

The relationship between the formal-carer EQ-5D-5L-proxy utility scores and the CMAI measure was statistically significant in the regression results with larger coefficients than in the analyses of other utility measures, including the analyses of informal-carer EQ-5D-5L-proxy utility scores where coefficients were not statistically significant.

Relationship with DEMQOL-Proxy-U scores. The relationship between both proxy-completed DEMQOL-Proxy-U utility scores and the FAST and CDR measures was very weak in the Spearman's rank-order analysis. The only statistically significant effects observed in the regression analysis were the relationship between formal-carer DEMQOL-Proxy-U score and the CDR score, however the effect sizes were notably smaller than those observed in the analyses of EQ-5D-5L-proxy scores completed by the same proxy.

As with the EQ-5D-5L-proxy analyses, the relationship between the CMAI score and the formal-carer DEMQOL-Proxy-U score was of a higher magnitude than the informal-carer DEMQOL-Proxy-U score, which was not statistically significant.

Other HRQOL measures

The relationship between resident-reported QOL-AD-NH and EQ-5D-5L was weak in the Spearman's rank-order analysis (rho = 0.28), and in the regression analysis it was statistically significant. This contrasted with the finding of no relationship observed in the Spearman's rank-order analysis, and the finding of no statistically significant relationship observed in the regression analysis, between (resident-reported) QOL-AD-NH and the other 4 utility measures.

Conversely, in the Spearman's rank-order analysis, the formal-carer–completed QUALID had a weak to moderate relationship with all utility scores, except for the resident-reported EQ-5D-5L. The informal-carer–completed QUALID also had a weak to moderate relationship with both the informal-carer EQ-5D-5L-proxy and DEMQOL-Proxy-U.

In the regression analysis, the QUALID scores had a statistically significant relationship with all utility scores. The magnitude of the relationship between QUALID and a utility score was always greater when the QUALID and utility score were completed by the same proxy. For example, a 10-unit change in formal-carer QUALID was associated with a 0.112 change in the formal-carer EQ-5D-5L-proxy and a 0.083 change in the informal-carer EQ-5D-5L-proxy.

Responsiveness

Figures 2 and Appendix Figure 11 (in the Supplementary Materials found at https://doi.org/10.1016/j.jval.2019.07.002) show the coefficients from the regression models, outlined in the sections that follow.

Cognitive measures

Relationship with resident-completed EQ-5D-5L. No statistically significant relationships were observed among changes in the FAST, CDR, or CMAI measures and changes in the resident-completed EQ-5D-5L.

Relationship with formal proxy-completed PBMs. A statistically significant (negative) relationship was observed between changes in both the CMAI and CDR measures and changes in the formal-carer PBMs (EQ-5D-5L-proxy and DEMQOL-Proxy-U). The formal-carer EQ-5D-5L-proxy also captured a change to

FAST score 6 and FAST score 7 (indicating more cognitive and functional impairment); however, the formal-carer DEMQOL-Proxy-U was not responsive to any changes in FAST scores.

Relationship with informal proxy-completed PBMs. Statistically significant (negative) relationships were observed between all possible changes in the FAST score and the informal-carer EQ-5D-5L-proxy and DEMQOL-Proxy-U (with 1 exception). In the analyses of FAST scores, the magnitude of the relationships with informal-carer PBMs was always greater than the magnitude of the relationships with the formal-carer PBMs, whereas for the CDR and CMAI analyses, the coefficients were always larger for the formal-carer PBMs.

Other HRQOL measures

A statistically significant relationship was always observed between the HRQOL measures and PBMs where completed by the same respondent (eg, the relationship between changes in resident-reported QOL-AD-NH and resident-reported EQ-5D-5L was statistically significant).

The relationship between changes in the (proxy-completed) QUALID score and changes in the EQ-5D-5L-proxy score were also always statistically significant, regardless of which proxy had completed the assessment, but this was not the case in the analyses of the relationship between QUALID and DEMQOL-Proxy-U.

Discussion

This study identified disparities in the measurement properties of 2 different PBMs completed by up to 3 different respondents (residents, formal-carer proxy assessors, and informal-carer proxy assessors) about 1 individual.

How Do Proxy-Completed PBMs Compare to Resident-Completed EQ-5D for People With Dementia?

Consistent with existing studies, we found the likelihood of eliciting a self-completed EQ-5D response decreased substantially with dementia severity, with nonresponse rising to four-fifths of residents with “severe dementia” (FAST score 7). In contrast, EQ-5D-5L-proxy and DEMQOL-Proxy-U data collection via a formal-carer proxy was feasible for almost all residents, demonstrating the role of such proxies in eliminating a substantial risk of sample bias if only resident-reported measures are used. However, informal-carer proxies could be recruited for only a minority of residents (<20%), indicating that for a care home-based sample of people with dementia, staff member proxies would likely provide the largest number of responses.

Our finding that resident-completed EQ-5D-5L scores were of a greater magnitude than proxy-completed EQ-5D-5L scores was also consistent with existing studies of dementia[17-23] and some other conditions.[17] Researchers should consider any influence this might have on the results of a trial using proxy-completed data.

In tests of criterion validity and responsiveness, which NICE explicitly recommends for determining the appropriateness of EQ-5D for a particular population,[1] the self-completed EQ-5D-5L was poor at reflecting clinically important differences and changes in the FAST, CDR, and CMAI scores but rather better at capturing (the resident’s own view of) quality of life (QOL-AD-NH). This is consistent with some comparable studies involving people with dementia that identified correlation between resident-reported EQ-5D and QOL-AD-NH[24-26] but not between resident-reported EQ-5D and some cognitive measures (eg, the Psychogeriatric Assessment Scale–Cognitive Impairment scale [PAS-Cog] and the Neuropsychiatric Inventory-Questionnaire [NPI-Q]).[4,18,24,26]

Our finding that the EQ-5D-5L-proxy and DEMQOL-Proxy-U did capture changes in the FAST, CDR, and CMAI may justify their use in research, especially if the intervention under investigation is expected to affect these more objective, cognitive aspects of health. Nevertheless, we also found the formal-carer–completed and informal-carer–completed EQ-5D-5L-proxy and DEMQOL-Proxy-U were unable to capture the residents’ perceptions of their own quality of life (as measured by the QOL-AD-NH), despite adequately capturing the proxies’ views of the residents’ quality of life (QUALID). This might reflect discrepancies between proxy views and resident views of the impact that a tangible decline in cognition, functional abilities, or clinical aspects of health has on quality of life. Perhaps the resident did not notice that decline or felt it did not unduly interfere with his or her quality of life. For instance, despite declining physical mobility, a resident might still enjoy full access to the things he or she values because of support from staff and/or aids (eg, wheelchairs or hoists). The resident’s reference point, when considering what is meant by “usual activities,” for example, may also differ from that of the resident’s proxy, who might more likely think about younger, healthier people in the general population than fellow residents in the care home. Hence, without a fuller understanding of the degree to which residents’ cognitive impairments had prevented them from being able to appreciate their current situation reliably, researchers (and NICE) face a considerable challenge in terms of deciding when it is desirable to give less weight to a resident’s own stated views. After all, this could lead to interventions deemed of value by proxies being prioritized over other interventions that would have greater impact on those aspects of HRQOL residents themselves say they value the most.

If Researchers Opt for Proxy-Completed PBMs, Does It Matter Which They Use?

We found that EQ-5D-5L-proxy and DEMQOL-Proxy-U were generally weak or poor substitutes for measuring utility. Regardless of which proxy completed it, the EQ-5D-5L-proxy was typically more responsive than the DEMQOL-Proxy-U to changes in CDR, FAST, and CMAI. This probably reflects the relative content of the measures: EQ-5D examines the frequency and severity of symptoms and functional impairment directly, whereas DEQMOL-Proxy-U has a broader focus on the emotional impact of dementia across wider QOL domains. Nevertheless, because DEMQOL-Proxy-U was no better than the EQ-5D-5L-proxy at reflecting aspects of resident-reported HRQOL included in the QOL-AD-NH, this study provides no substantive evidence to justify using DEMQOL-Proxy-U over the default (and perhaps more straightforward[27]) EQ-5D-5L-proxy.

We also identified a moderate correlation between the informal-carer-completed EQ-5D-5L-proxy and DEMQOL-Proxy-U, with the DEMQOL-Proxy-U providing higher utility scores especially at lower levels. These results closely resemble those of the 2 comparable Australian studies[4,9] (eg, we reported rho = 0.386, versus 0.389 in 1 of the Australian studies, and a mean difference of 0.214 versus 0.202). These results might have arisen because the EQ-5D-5L has a much lower bound than the DEMQOL-Proxy-U (−0.285 vs 0.363, respectively).

Comparison With Existing Studies

Three existing (repeated) cross-sectional studies compared various combinations of resident-completed and proxy-completed versions of the EQ-5D and the DEMQOL in Britain[19] or Australia.[4,9] In contrast to our study, these included people without dementia[4] or people who were restricted to unrepresentative subgroups (eg, people recently hospitalized[9] or with significant depressive symptoms).[19] None of these studies compared formal-carer with informal-carer proxies or DEMQOL-Proxy-U with other quality of life measures (QOL-AD-NH and QUALID), and 2 of the studies did not compare self- and proxy-responses for the same residents.

Our study uniquely collected data from 2 dementia-specific scales of cognitive and functional impairment (FAST and CDR), rather than using more general cognition or mental health measures (eg, the Mini-Mental State Exam [MMSE] or the PAS-Cog), and included residents with the full range of dementia severity. Larger sample sizes, repeated measures, and/or longer follow-up periods also enabled analyses of responsiveness that likely support more robust causal inference than can be done with cross-sectional comparisons.

Limitations

Some caution is necessary when comparing person-specific utility scores measured at the same time point because each assessor completed the PBM questionnaires at different times within a 2-week window and because by design the EQ-5D-5L (which measures "health today") and the DEMQOL (which measures health "the past week") assess HRQOL over different time periods (Table 1).

Although this study included multiple utility measures for each resident, it did not include the (resident-reported) DEMQOL-U (owing to concerns about research burden) or broader capability measures (eg, ICECAP-O [ICEpop CAPability measure for Older people][28]).

Implications and Future Research

This study can inform decisions about how to measure utility in studies of people with dementia depending, for example, on the functional or cognitive aspects of the HRQOL that researchers wish to capture and on the severity of dementia in trial participants. A future study could collect the quantitative PBM data alongside immediate qualitative interviewing to better understand reasons for observed discrepancies between resident- and proxy-rated utility. Given the difficulties of collecting utility data from residents (when they were unable) and proxies (because of recruitment challenges) in this study, as well as discrepancies between utility measures, researchers would be well advised to try to collect PBM data from both individuals and their proxies. Future work could establish how to improve the participation of proxies in trials and develop more specific guidance on when particular utility measures should be used, based, for instance, on dementia severity. This requires more information on the degree to which discrepancies between proxy- and resident-reported outcomes arise because residents are no longer able to judge their own situation reliably, for instance, and more information on which particular improvements in functional or cognitive aspects of HRQOL are most valued by payors. A potential focus of future research could be to combine different utility scores to provide a more complete utility measure that utilizes the insights provided by residents as well as their proxies, perhaps by "adjusting" the scores of one using the score provided by another.

The observed relationships between the utility scores and (non-PBM) cognitive measures could also be used to inform health state utility parameters associated with several different clinical measures in economic models, by indicating the utility decrement associated with different (worsening) dementia severity, and to inform imputation techniques in trials where some utility measures are missing.

Conclusion

There are multiple strategies to capture health state utility in people with dementia. Choice of strategy has a nontrivial impact on the utility values obtained, and these will potentially have an impact on economic evaluation results. The results presented here can inform choice of utility capture strategy. Future research should explore how to combine utility data from multiple different sources to prevent data loss and to retain the patient's perspective.

Acknowledgments

We would like to thank all the care homes, individuals with dementia, the family members of these individuals, and care home staff for taking part in this study and giving freely of their time. We would like to thank the following people who contributed to trial design, delivery and analysis: Clive Ballard, Natasha Burnley, Lynn Chenoweth, Byron Creese, Bryony Dawkins, Murna Downs, Amanda J Farrin, Jane Fossey, Lucy Garrod, Elizabeth H Graham, Ivana Holloway, Rachael Kelley, Amanda Lilley-Kelley, Joanne McDermid, Vicki McLellan, Holly Millard, Devon Perfect, Louise Robinson, Olivia Robinson, Emily Shoesmith, Najma Siddiqi, Graham Stokes, Daphne Wallace, Rebecca EA Walwyn and Edward C F Wilson. We would also like to thank Chris Albertyn, Marie Crabbe, Cara Gates, Stephanie Jones, Baber Malik, Harriet Maunsell, Kirsty Nash, Sahdia Parveen, Luisa Rabanal, Bina Sharma, Emily Standell, Miguel Vasconcelos Da Silva, and other researchers who collected the data.

Source of financial support

This project was funded by the National Institute for Health Research Health Technology Assessment programme (project number 11/15/13). The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HTA, NIHR, NHS, or the Department of Health and Social Care.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2019.07.002.

REFERENCES

1. National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal 2013*. London: National Institute for Health and Care Excellence; 2013. Process and Methods Guides No. 9. https://www.ncbi.nlm.nih.gov/books/NBK395867/pdf/Bookshelf_NBK395867.pdf. Accessed August 16, 2019.
2. Diaz-Redondo A, Rodriguez-Blazquez C, Ayala A, et al. EQ-5D rated by proxy in institutionalized older adults with dementia: psychometric pros and cons. *Geriatr Gerontol Int*. 2014;14:346–353.
3. Coucill W, Bryan S, Bentham P, Buckley A, Laight A. EQ-5D in patients with dementia: an investigation of inter-rater agreement. *Med Care*. 2001;39:760–771.
4. Easton T, Milte R, Crotty M, Ratcliffe J. An empirical comparison of the measurement properties of the EQ-5D-5L, DEMQOL-U and DEMQOL-Proxy-U for older people in residential care. *Qual Life Res*. 2018;27(5):1283–1294.
5. Rowen D, Mulhern B, Banerjee S, et al. Estimating preference-based single index measures for dementia using DEMQOL and DEMQOL-Proxy. *Value Health*. 2012;15:346–356.
6. Comans TA, Nguyen K-H, Mulhern B, et al. Developing a dementia-specific preference–based quality of life measure (AD-5D) in Australia: a valuation study protocol. *BMJ Open*. 2018;8:e018996.
7. Connolly S. Economics of dementia: a review of methods. *Dementia*. 2018. https://doi.org/10.1177/1471301218800639.
8. Surr CA, Holloway I, Walwyn REA, et al. Dementia Care Mapping^{TM} to reduce agitation in care home residents with dementia: The DCM^{TM} EPIC cluster randomised controlled trial. Health Technology Assessment. *2019. ISSN 1366-5278*.

9. Ratcliffe J, Flint T, Easton T, et al. An empirical comparison of the EQ-5D-5L, DEMQOL-U and DEMQOL-Proxy-U in a post-hospitalisation population of frail older people living in residential aged care. *Appl Health Econ Health Policy*. 2017;15:399–412.
10. Surr CA, Griffiths AW, Kelley R, et al. The Implementation of Dementia Care Mapping in a Randomized Controlled Trial in Long-Term Care: Results of a Process Evaluation. *Am J Alzheimers Dis Other Demen*. 2019;34(6):390–398.
11. https://euroqol.org/eq-5d-instruments/eq-5d-5l-available-modes-of-administration/proxy/. Accessed August 16, 2019.
12. Kind P, Dolan P, Gudex C, Williams A. 1998. Variations in population health status: results from a United Kingdom national questionnaire survey. *BMJ*. 1998;316(7133):736–741.
13. Reisberg B. Functional assessment staging (FAST). *Psychopharmacol Bull*. 1988;24:653–659.
14. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*. 1993;43:2412–2414.
15. Cohen-Mansfield J. *Instruction Manual for the Cohen-Mansfield Agitation Inventory (CMAI)*. Rockville, MD: Research Institute of the Hebrew Home of Greater Washington; 1991.
16. Weiner MF, Martin-Cook K, Svetlik DA, Saine K, Foster B, Fontaine CS. The Quality of Life in Late-Stage Dementia (QUALID) Scale. *J Am Med Dir Assoc*. 2000;1:114–116.
17. Rand S, Caiels J. *Using proxies to assess quality of life: a review of the issues and challenges*. Quality and Outcomes of Person-Centered Care Policy Research Unit. 2015. https://core.ac.uk/download/pdf/42411586.pdf. Accessed August 16, 2019.
18. Kunz S. Psychometric properties of the EQ-5D in a study of people with mild to moderate dementia. *Qual Life Res*. 2010;19:425–434.
19. Mulhern B, Rowen D, Brazier J, et al. Development of DEMQOL-U and DEMQOL-PROXY-U: generation of preference-based indices from DEMQOL and DEMQOL-PROXY for use in economic evaluation. *Health Technol Assess*. 2013;17:1–140.
20. Naglie G, Tomlinson G, Tansey C, et al. Utility-based quality of life measures in Alzheimer's disease. *Qual Life Res*. 2006;15:631–643.
21. Devine A, Taylor SJC, Spencer A, Diaz-Ordaz K, Eldridge S, Underwood M. The agreement between proxy and self-completed EQ-5D for care home residents was better for index scores than individual domains. *J Clin Epidemiol*. 2014;67:1035–1043.
22. Moyle W, Murfield JE, Griffiths SG, Venturato L. Assessing quality of life of older people with dementia: a comparison of quantitative self-report and proxy accounts. *J Adv Nurs*. 2012;68:2237–2246.
23. Jonsson L, Andreasen N, Kilander L, et al. Patient- and proxy-reported utility in Alzheimer disease using the EuroQoL. *Alzheimer Dis Assoc Disord*. 2006;20:49–55.
24. Karlawish JH, Zbrozek A, Kinosian B, Gregory A, Ferguson A, Glick HA. Preference-based quality of life in patients with Alzheimer's disease. *Alzheimers Dement*. 2008;4:193–202.
25. Thorgrimsen L, Selwood A, Spector A, et al. Whose quality of life is it anyway? The validity and reliability of the Quality of Life-Alzheimer's Disease (QOL-AD) dscale. *Alzheimer Dis Assoc Disord*. 2003;17:201–208.
26. Hounsome N, Orrell M, Edwards RT. EQ-5D as a quality of life measure in people with dementia and their carers: evidence and key issues. *Value Health*. 2011;14:390–399.
27. McCartney K, Robinson K, Horne JC, Logan PA. FinCH study team. DEMQOL: fit for purpose in care homes? *Age Ageing*. 2018;47(suppl 3):iii9–iii12.
28. van Leeuwen KM, Bosmans JE, Jansen APD, et al. Comparing measurement properties of the EQ-5D-3L, ICECAP-O, and ASCOT in frail older adults. *Value Health*. 2015;18:35–43.