

RESEARCH PAPER

Inter-rater reliability of care home staff's proxy judgements with residents' assessments of their own health-related quality of life: an analysis of the PATCH trial EQ-5D data

CHARLOTTE KELLY¹, CLAIRE HULME², LIZ GRAHAM^{3,4}, ALISON ELLWOOD⁵, ISMAIL PATEL³, BONNIE CUNDILL⁶, AMANDA FARRIN⁶, MADELINE GOODWIN⁶, KAREN HULL^{3,7}, JILL FISHER⁷, ANNE FORSTER^{3,4}

¹Hull York Medical School, University of Hull, Hull, UK

²University of Exeter Medical School, University of Exeter, Exeter, UK

³Academic Unit for Ageing and Stroke Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK

⁴University of Leeds, Leeds, UK

⁵University of Bradford, Bradford, UK

⁶Clinical Trials Research Unit, University of Leeds, Leeds, UK

⁷Leeds Neurophysiotherapy, Leeds, UK

Address correspondence to: Charlotte Kelly, Tel: 01482 463107 Email: Charlotte.kelly@hyms.ac.uk

Abstract

Objectives: to compare care staff proxies with care home residents' self-assessment of their health-related quality of life (HRQoL).

Methods: we assessed the degree of inter-rater reliability between residents and care staff proxies for the EQ-5D-5L index, domains and EQ Visual Analogue Scale at baseline, 3 months and 6 months, collected as part of the PATCH trial. We calculated kappa scores. Interpreted as <0 no agreement, 0–0.2 slight, 0.21–0.60 fair to moderate and >0.6 substantial to almost perfect agreement. Qualitative interviews with care staff and researchers explored the challenges of completing these questions.

Results: over 50% of the HRQoL data from residents was missing at baseline compared with a 100% completion rate by care staff proxies. A fair-to-moderate level of agreement was found for the EQ-5D-5L index. A higher level of agreement was achieved for the EQ-5D-5L domains of mobility and pain. Resident 'non-completers' were more likely to: be older, have stayed a longer duration in the care home, have lower Barthel Index and Physical Activity and Mobility in Residential Care (PAM-RC) scores, a greater number of co-morbidities and have joined the trial through consultee agreement. Interviews with staff and researchers indicated that it was easier to rate residents' mobility levels than other domains, but in general it was difficult to obtain data from residents or to make an accurate proxy judgement for those with dementia.

Conclusions: whilst assessing HRQoL by care staff proxy completion provides a more complete dataset, uncertainty remains as to how representative these values are for different groups of residents within care homes.

Keywords: care homes, care staff, cost effectiveness, health-related quality of life, older people, proxy measures

Key Points

- Evaluating the cost effectiveness of interventions in a care home setting requires alternative methods for collecting health-related quality of life outcomes where residents are unable to provide this information themselves.

- Care home staff proxy reporting is one way of collecting this information, but there are concerns regarding the level of agreement between care staff proxies and residents.
- This analysis has shown that there is a higher level of agreement between care home staff proxies and residents for some of the domains (e.g. mobility and pain). The non-completers (residents who did not complete the EQ-5D-5L) were more likely to be older, have lower Barthel and PAM-RC scores, a greater number of co-morbidities and have dementia. Alternative ways of ascertaining HRQoL for residents who lack the capacity to complete measures requires investigation.

Introduction

In the United Kingdom, the majority of long-term care is provided by care staff in care homes. Over 400,000 people in the United Kingdom live in care homes (with and without nursing care), with demand for this type of care forecast to rise with an increasingly ageing population [1]. This is a frail population, with many residents having complex healthcare needs, including dementia [2]. A growing number of trials of interventions to improve the quality of life of care home residents are being undertaken, which require the evaluation of these interventions' cost effectiveness.

Since 2008 the National Institute for Health and Care Excellence (NICE) in England has stated that economic evaluations of interventions should be presented using an incremental cost per quality-adjusted life year (QALY) and that the estimates of the health-related quality of life (HRQoL) used in the QALY for an adult population should use the EuroQol five-dimension (EQ-5D) questionnaire to ensure consistency [3]. The EQ-5D focuses on five dimensions of health: mobility, ability to self-care, ability to undertake usual activities, pain and discomfort and anxiety and depression—assessed using either the EQ-5D-3L (where each domain has three levels of severity) or the EQ-5D-5L (where each domain has five levels of severity). The domains are converted to an index using a set of weights [4]. The EQ Visual Analogue Scale (VAS)—which is a scale of 0–100, rated from the worst imaginable health state to the best imaginable health state—is completed alongside the EQ-5D and records a person's assessment of their health state, as they are experiencing it on that day.

Previous studies have identified low completion rates of HRQoL measures such as the EQ-5D within care home settings [5]. It is argued that, due to cognitive impairment, some care home residents struggle to answer the HRQoL questions reliably [5, 6], and there is evidence of a significant correlation with residents' difficulty answering questions and their having reduced capacity [5]. An alternative, or complementary, approach is to ask care home staff to complete the questions on their behalf. The OPERA study compared 556 care home residents' EQ-5D-3L ratings with staff ratings and found that, on average, staff proxies rated residents as having a lower HRQoL than the residents' self-assessment [7]. This comparatively lower rating of HRQoL was particularly the case for residents with depression, pain and cognitive impairment; whilst the opposite effect—overstating HRQoL compared with the residents—was seen for those with no

cognitive impairment [7]. The PEACH study compared care staff proxies and care home residents' EQ-5D-5L and EQ VAS data across three consecutive months achieving 117, 109 and 104 pairs, respectively [8]. The researchers reported that care staff proxies' values, on average, were lower than the residents' for the EQ-5D-5L index scores at baseline (0.43 versus 0.57), but the EQ-VAS results were the opposite (68 versus 65). Levels of agreement [using the Intra-Class Correlation Coefficient (ICC)] were 'fair' for the EQ VAS and 'moderate' for the EQ-5D-5L index, raising questions about the validity of this HRQoL measure in this population [8]. Other studies have identified an under-reporting of levels of discomfort and distress compared with the residents' own assessment [6].

This paper reports the inter-rater reliability of care home residents and care staff proxies who participated in the PATCH trial for the EQ-5D-5L domains, index and EQ VAS. The study adds to the literature by employing a mixed methods approach to assess the inter-rater reliability and examine the challenges experienced in collecting these measures of HRQoL in a care home population.

Methods

The PATCH trial design is reported in the protocol paper [9] and results paper [10]. It was a parallel-group, cluster-randomised controlled feasibility trial with embedded process evaluation. Ten care homes in Yorkshire were randomised (1:1) to the Skilful Care Training Package (SCTP) delivered by physiotherapists to care staff plus usual care, or to usual care. Consent for participation in data collection was sought from residents who required assistance when mobilising, regardless of capacity. Where residents lacked capacity to consent, an appropriate Consultee was approached to provide agreement to their involvement in the trial [11]. Between May 2017 and February 2018 146 residents were registered as trial participants across 10 care homes. One SCTP home closed shortly after the 3-month follow-up.

Data were collected at baseline (pre-randomisation), 3- and 6-months post-randomisation from residents, where they were able to provide information, and from care staff proxies for all participating residents. All participants included in the PATCH trial were included in this study. Questionnaires were administered face-to-face by a blinded researcher who recorded, at the end of each questionnaire, reasons for partial or non-completion. Proxy data were

Table 1. Trial participants available to complete the EQ-5D-5L Index at each time point

		Baseline (n = 146)		3 months (n = 124)		6 months (n = 104)	
		Completed	Non- or partially completed	Completed	Non- or partially completed	Completed	Non- or partially completed
		71 (48.6%)	75 (51.4%)	43 (34.7%)	81 (65.3%)	36 (34.6%)	68 (65.4%)
Age (years)	Mean (SD)	85.7 (7.47)	86.2 (8.32)	85.3 (8.05)	86.7 (8.04)	84.3 (7.71)	86.2 (7.94)
Gender	Female	51 (71.8%)	55 (73.3%)	33 (76.7%)	56 (69.1%)	24 (66.7%)	51 (75.0%)
	Male	20 (28.2%)	20 (26.7%)	10 (23.3%)	25 (30.9%)	12 (33.3%)	17 (25.0%)
Length of stay in care home (months)	Median (range)	25.0 (1, 232)	25.0 (1, 144)	22.0 (1, 232)	29.0 (1, 196)	20.5 (1, 232)	27.5 (1, 196)
PAM-RC Total Score 0–21 (higher score greater physical ability and activity)	Mean (SD)	5.8 (4.34)	3.1 (3.25)	6.1 (4.71)	3.5 (3.59)	4.4 (3.43)	4.1 (4.35)
Barthel Total Score^a 0–20 (higher score greater self-care ability)	Mean (SD)	5.8 (4.75)	2.3 (2.87)	6.7 (5.09)	2.6 (3.07)	5.1 (4.95)	3.3 (3.56)
Co-morbidities^b	0 co-morbidities	1 (1.4%)	2 (2.7%)	2 (4.7%)	1 (1.2%)	0 (0.0%)	3 (4.4%)
	1–2 co-morbidities	24 (33.8%)	13 (17.3%)	13 (30.2%)	19 (23.5%)	12 (33.3%)	17 (25.0%)
	3+ co-morbidities	46 (64.8%)	60 (80.0%)	28 (65.1%)	61 (75.3%)	24 (66.7%)	48 (70.6%)
Capacity to consent	Yes	24 (33.8%)	0 (0.0%)	19 (44.2%)	1 (1.2%)	15 (41.7%)	0 (0.0%)
	No	47 (66.2%)	75 (100.0%)	24 (55.8%)	80 (98.8%)	21 (58.3%)	68 (100.0%)

^aMissing for one participant who did not complete or partially completed the EQ-5D questionnaire at 3 and 6 months. ^bNumber (percentage) of registered residents with a confirmed diagnosis of the condition is reported.

completed via (blinded) researcher interview with staff members who knew the resident well. Staff members were asked to rate from the perspective of the resident. In 8 out of 10 care homes the same staff member was the proxy for all time points and all residents. In one care home one staff member completed all baseline and 6-month time points and another the 3-month time point and in the final care home two care staff covered the residents at each time point.

Residents' characteristics collected included: gender, age, co-morbidities, months resident at the care home and levels of mobility and independence—including the Barthel Index of Activities of Daily Living [12] and Physical Activity and Mobility in Residential Care Scale (PAM-RC) [13]. Where all EQ-5D-5L domains had been completed they were converted to index values on the scale of -0.285 (extreme problems on all the EQ-5D-5L domains) to 1 (no problems on any of the EQ-5D-5L domains) using the England EQ-5D-5L value set published by Devlin *et al.* [4].

Analysis of inter-rater agreement was cross sectional at each time point (baseline, 3 and 6 months) and performed using STATA 15.0. The weighted kappa statistic was used to assess levels of agreement for the EQ-5D-5L domains, using linear weights for the five levels and confidence intervals (CIs) calculated using bootstrapping (1,000 repetitions). The ICC was calculated for the EQ-5D-5L index value and EQ VAS using a one-way random effects model. These methods allow the measurement of agreement between two or more raters [14]. The kappa scores range from -1 to $+1$ and can be interpreted as <0 no agreement, $0-0.2$ slight, $0.21-0.40$ fair,

$0.41-0.60$ moderate, $0.61-0.80$ substantial ≥ 0.80 almost perfect agreement [15].

Semi-structured interviews were undertaken by author L.G. at the end of the trial with a sample of researchers ($N = 2$) and care home staff ($N = 7$ from five care homes) who had collected data from or provided data about participating residents, respectively. All care homes were given the opportunity to participate in these interviews, and all those who responded were included. Interviews explored, using a topic guide, the acceptability of data collection tools and processes, including researchers' and staffs' views on completion of the HRQoL measures—for example how easy or difficult they found it to provide answers and their perceptions of the relevance of questions to residents. The thematic analysis method described by Braun and Clarke was used to analyse the interview data [16].

Results

Residents' characteristics are summarised in Table 1, by those who did and did not complete all the EQ-5D-5L domains at each time point. The denominator represents all residents still participating in the trial at each time point (excluding residents who died or moved from the home at time points beyond baseline). Less than 50% of the residents fully completed the EQ-5D-5L at baseline, declining to 34% of those available at 6 months. Completers were more likely to be younger, have higher levels of self-care ability and

Table 2. Levels of agreement for the EQ-5D-5L domains

Domain	Time point	No. with resident and proxy completed (% of total)	% Exact agreement	Kappa (95% CI)	No. of care homes
Mobility	Baseline	83 (56.85%)	40.96	0.408 (0.27, 0.54)	10
	3 months	56 (45.53%)	53.57	0.434 (0.19, 0.68)	10
	6 months	42 (40.38%)	57.14	0.444 (0.13, 0.76)	9
Self-care	Baseline	80 (54.79%)	28.75	0.068 (−0.04, 0.18)	10
	3 months	47 (38.21%)	34.04	0.153 (0.08, 0.23)	10
	6 months	42 (40.38%)	34.48	0.30 (−0.02, 0.62)	9
Usual activities	Baseline	74 (50.68%)	27.03	0.084 (−0.04, 0.21)	10
	3 months	50 (40.65%)	20.00	0.135 (−0.04, 0.31)	10
	6 months	47 (46.08%)	37.50	0.228 (−0.05, 0.50)	9
Pain/discomfort	Baseline	82 (56.16%)	42.68	0.178 (0.08, 0.27)	10
	3 months	58 (47.15%)	44.68	0.199 (0.09, 0.38)	10
	6 months	47 (45.19%)	37.50	0.234 (0.09, 0.38)	9
Anxiety/depression	Baseline	80 (54.79%)	25.00	−0.045 (−0.11, 0.02)	10
	3 months	52 (42.28%)	38.46	0.115 (−0.08, 0.31)	10
	6 months	46 (44.23%)	23.91	−0.054 (−0.17, 0.07)	9

There are differing numbers completed across the domains, as some residents did not complete all questions.

physical activity, have fewer co-morbidities and have the capacity to consent.

The resident and care staff proxy questionnaires (where both were complete) and inter-rater agreements for the EQ-5D-5L domains are summarised in Table 2. Care staff proxy data were provided for all available residents at each time point, but there were varying amounts of missing data for the residents' self-assessment. Residents were most likely to complete the mobility domain, which also had the highest level of agreement between the matched pairs with kappa scores of >0.4 (moderate agreement) at each time point. There was less agreement between residents and proxies for the usual activities and anxiety and depression domains. The anxiety and depression domain resulted in a negative Kappa score indicating that there was 'no agreement'.

The EQ-5D-5L index and EQ-5D VAS pairings for residents and care staff proxies are presented in Table 3. The EQ-5D-5L index requires all five domains to be completed, resulting in fewer pairs to compare than in Table 2. For the EQ-5D-5L index there were 71 pairings at baseline, 42 at 3 months and 36 at 6 months. On average care staff assessed residents to be in a poorer state of health than the residents reported themselves (at baseline 0.44 versus 0.56), although the 95% CIs are overlapping indicating this difference is not statistically different. The ICC values for the EQ-5D-5L index were all >0.3 and at 6 months >0.45, so in the realms of fair to moderate levels of agreement. For the EQ VAS there were 57 pairings at baseline, 25 at 3 months and 27 at 6 months. The EQ VAS scores had similar levels of agreement, but much lower rates of completion than the EQ-5D-5L index, with only 20% of residents completing the EQ-5D VAS at 3 months and 26% at 6 months.

Researchers being unable to engage with residents due to communication or cognitive difficulty were the most common reasons for non-completion of part or all of the questions. Other reasons included that the resident could not

understand or hear the question, was asleep, was unwell or was distressed by the questions. Reasons for the low response to the EQ VAS included that some residents couldn't see the scale or were unable to understand what they needed to do—despite the researcher being there to guide them through it.

A key theme identified from the interviews was that some staff seemed to conceptualise quality of life for those with cognitive impairment as being quite different from those who did not have dementia.

Alzheimer's they just get things sometimes back to front and it's just a total different scenario. They're not anxious, they're not depressed, they just really don't know what comes next.

[quote from member of care staff]

Researchers found the anxiety and depression question difficult to ask residents, as some residents would become upset. Care staff often reported that anxiety was contingent upon personal care activities (e.g. toileting or transfers) rather than being an ongoing state, which made it more difficult to answer the question.

Researchers and staff suggested that some of the EQ-5D-5L domains were not appropriate for care home residents, particularly 'usual activities'—where the examples given are often irrelevant to residents (e.g. housework). The mobility domain was seen as more straightforward, as staff had a good understanding of residents' dependency levels (as part of their care plans), and researchers found it easier to ascertain mobility levels from residents through conversation.

... things like mobility you can sort of have a conversation, and get them to pin that down.

[quote from researcher]

Where residents did engage, researchers were not always confident that they were providing meaningful answers due

Table 3. Levels of agreement for the EQ-5D-5L index and EQ-5D VAS

	Time point	No. with resident and proxy completed (% of total)	Resident mean (95% CI)	Proxy mean (95% CI)	ICC (95% CI)	No. of care homes
Index Score	Baseline	71 (48.6%)	0.56 (0.48, 0.64)	0.44 (0.38, 0.50)	0.368 (0.184, 0.577)	10
	3 months	43 (35.0%)	0.54 (0.44, 0.64)	0.46 (0.37, 0.54)	0.329 (0.041, 0.570)	10
	6 months	36 (34.0%)	0.49 (0.38, 0.64)	0.34 (0.24, 0.44)	0.453 (0.148, 0.679)	9
EQ VAS	Baseline	57 (39.04)	61.47 (55.18, 67.77)	61.88 (56.23, 67.53)	0.239 (−0.024, 0.471)	10
	3 months	25 (20.3%)	65.76 (59.09, 72.43)	64.32 (54.27, 74.37)	0.478 (0.103, 0.732)	10
	6 months	27 (26.0%)	58.70 (48.94, 68.46)	65.93 (56.83, 75.02)	0.376 (0.017, 0.653)	9

to their cognitive impairment—meaning they were either unable to answer or did not provide an accurate account of their abilities, as confirmed by care staff. Staff commented that: it was difficult to obtain a ‘true’ account from residents with cognitive impairment; residents are often unable to tell staff members how they are feeling and that some residents say they are fine when staff know they are not or vice versa. This made it difficult to provide a proxy rating from the perspective of the residents, as shown by these staff quotes:

... you don't know how a person's feeling if they can't tell you how they're feeling, so that can be quite difficult.

... where I were putting somebody they were completely put their self somewhere different, you know, it's, it is quite hard that one.

The EQ VAS was seen as particularly difficult to complete, as residents were confused by how to assign a number to their health, and staff found it difficult to make a subjective judgment on a resident's behalf.

Discussion

To assess the cost effectiveness of interventions in a care home population, where not all residents are able to complete the questions, proxy-reported HRQoL outcomes are often collected. Using proxies for HRQoL where people are unable to make this assessment themselves is not a new phenomenon [17]. The key consideration is whether there is confidence that the proxy-reported outcomes are comparable with self-complete. The PATCH trial identified a greater level of agreement between care staff proxies and residents for some of the domains (e.g. mobility and pain), which was supported by the interviews with staff and researchers. The mobility domain was seen to be easier to complete than other domains that were more subjective and so less straightforward to answer (e.g. anxiety and depression). These differences in the

ease of completion of the domains reinforce findings from other studies where this has been identified [18].

A tendency for proxy values to be lower than self-reported values was identified in other studies [7, 18, 19]. This was particularly evident for certain groups of care home residents, such as those with dementia [20, 21] and older people [22]. Lower levels of agreement have been reported for residents with severe Parkinson's disease and depression [23], but higher levels of resident and care worker proxy agreement where the individuals were in better health [17]. Inaccurate HRQoL data risks reporting gains (or losses) from interventions that are not representative of those experienced by care home residents and calls into question the impact of combining self-complete and proxy in one analysis.

In the PATCH trial we were able to achieve a 100% completion rate for questionnaires completed by the care staff proxies. In comparison, <50% of residents completed the five domains at baseline, declining to 34% at 6 months. The most common reason recorded for non-completion was a lack of cognitive ability, indicating that some of the missing data may include particular groups of care home residents excluded from the paired analysis who genuinely need proxy assessment of their HRQoL (as they are unable to understand or communicate their answers) [24] and for whom we have been unable to ascertain whether the proxy ratings provided reflected their HRQoL. At the present time there is no formal screening or cut off suggested for use of the proxy version of the EQ-5D.

The DEMQOL-U, a measure of HRQoL for people with dementia (proxy and self-report), has been used as alternative preference-based measures within care home research [25, 26]. In the UK, given NICE guidance [3] these are typically used alongside the EQ-5D, often as sensitivity analyses [27]. Divergence in scoring between the EQ-5D and DEMQOL-U and their proxies suggest use of the DEMQOL-U for a subgroup of participants with cognitive impairment or dementia is not an appropriate substitute for the EQ-5D-5L [27–29].

If HRQoL measures such as the EQ-5D are to be used to assess cost effectiveness of interventions in care home settings, to focus on only those residents that could self-complete limits the generalisability of results. As a number of studies have identified, this depends on the care home population that is being studied, as there is likely to be better agreement for those whose health is either very good or very poor, but currently less evidence for those in-between [17].

Within our study, we used the tariff developed by Devlin *et al.* [4] for the EQ-5D-5L. Concerns have been raised over the quality and reliability of this tariff [30]. Whilst NICE recommends use of a mapping function [31] instead of the tariff [32], without a gold standard is not possible to conclude that the tariff necessarily mis-represents health state preferences [33]. The ‘interim’ solution of a mapping function is also thought to be limited [34]. Given the controversy we used the Devlin tariff as proposed in our original protocol. Use of this tariff rather than the mapping function is likely to impact on the index score and ICCs but not on the VAS or domain responses. Whilst we believe use of the tariff is unlikely to change our conclusions, we recommend future research uses the new EQ-5D-5L tariff once available.

Strengths and limitations

A mixed methods approach was employed to provide a greater understanding of the difficulties in collecting data from a population group that in many cases were unable to complete the questions—often due to cognitive impairment. One of the key strengths of the study was that in the majority of care homes the same staff proxy filled in the questions at all time points, minimizing the influence on agreement from multiple raters. The interviews were conducted by the trial manager who was familiar with all aspects of the trial, so had a good understanding of the staff member experiences. The study did include a comparison of the characteristics of those participants included (with complete EQ-5D-5L Index scores) versus not included in the paired analysis but did not focus on how participant characteristics might affect inter-rater agreement of those that were paired.

Conclusions

This trial and others have identified a high level of missing data in the responses of care home residents, which advocates the use of care staff proxies in evaluating HRQoL—to allow the inclusion of all residents’ data in evaluations. However, we have shown that, where it was possible to compare staff proxies with the residents’ responses, there was only a fair to moderate level of agreement, providing a degree of uncertainty regarding the representativeness of proxy assessments of residents’ HRQoL. Alternative methods for assessing residents’ HRQoL should be explored, particularly for those with cognitive impairment.

Acknowledgements: The authors wish to thank and acknowledge the support of: all participating care homes’ staff, residents and their relatives; the Clinical Research Network staff who supported the running of the study in North Yorkshire; members of the Trial Steering Committee and the Lay Advisory Group—for their important and insightful contributions to the trial.

Declaration of Conflicts of Interest: Prof. Hulme was a member of the Chartered Society of Physiotherapy Scientific Panel between 2011 and 2016.

Declaration of Sources of Funding: This work was supported by a grant from the Chartered Society of Physiotherapy Charitable Trust (grant number OPA/14/03). The funding agency played no role in the design, execution, analysis and interpretation of data or the writing of this manuscript.

References

1. LaingBuisson. Care Homes For Older People UK Market Report Market Report, 2019.
2. Lievesley N, Crosby G, Bowman C. The Changing Role of Care Homes. London: Centre for Policy on Ageing, 2011.
3. EuroQol. A new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16: 199–208.
4. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-5L value set for England. *Health Econ* 2018; 27: 7–22.
5. Gordon AL, Franklin M, Bradshaw L, Logan P, Elliott R, Gladman JRF. Health status of UK care home residents: a cohort study. *Age Ageing* 2014; 43: 97–103.
6. Benson T, Bowman C. Health status of care home residents: practicality and construct validity of data collection by staff at scale. *BMJ Open Quality* 2019; 8: 1–8.
7. Parker B, Petrou S, Underwood M, Madan J. Can care staff accurately assess health-related quality of life of care home residents? A secondary analysis of data from the OPERA trial. *BMJ Open* 2017; 7: e012779.
8. Usman A, Lewis S, Hinsliff-Smith K *et al.* Measuring health-related quality of life of care home residents, comparison of self-report with staff proxy responses for EQ-5D-5L and HowRu: protocol for assessing proxy reliability in care home outcome testing. *BMJ Open* 2018; 8: e022127.
9. Graham L, Cicero R, Clarke D *et al.* PATCH: posture and mobility training for care staff versus usual care in care homes: study protocol for a randomised controlled trial. *Trials* 2018; 19: 521.
10. Graham L, Ellwood A, Hull K *et al.* A posture and mobility training package for care home staff: results of a cluster randomised controlled feasibility trial (the PATCH trial). *Age Ageing* 2020; 49: 821–28.
11. The Department for Constitutional Affairs. The Mental Capacity Act 2005 Code Of Practice. United Kingdom: Crown Copyright, 2007.
12. Collin C, Wade D, Davies S, Horne V. The Barthel ADL index: a reliability study. *Int Disabil Stud* 1988; 10: 61–3.

13. Whitney J, Close JC, Lord SR, Jackson SHD. The physical activity and mobility in residential care (PAM-RC) scale. Pre-publication data.
14. Mandrekar JN. Measures of Interrater agreement. *J Thorac Oncol* 2011; 6: 6–7.
15. Hartwell M, Rubinstein E, Hayes W. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33.
16. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006; 3: 77–101.
17. Rand S, Caiels J. Using proxies to assess quality of life: a review of the issues and challenges D.P. 2899, Editor. Canterbury: University of Kent, 2015.
18. Usman A, Lewis S, Hinsliff-Smith K *et al.* Measuring health-related quality of life of care home residents: comparison of self-report with staff proxy responses. *Age Ageing* 2019; 48: 407–13.
19. 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: 760–71.
20. Arons AMM, Krabbe PFM, Schölzel-Dorenbos CJM, van der Wilt GJ, Rikkert MGMO. Quality of life in dementia: a study on proxy bias. *BMC Med Res Methodol* 2013; 13: 110.
21. 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.
22. Scocco P, Fantoni G, Caon F. Role of depressive and cognitive status in self-reported evaluation of quality of life in older people: comparing proxy and physician perspectives. *Age Ageing* 2005; 35: 166–71.
23. Martínez-Martín P, Benito-León J, Alonso F, Catalán MJ, Pondal M, Zamarbide I. Health-related quality of life evaluation by proxy in Parkinson's disease: approach using PDQ-8 and EuroQoL-5D. *Mov Disord* 2004; 19: 312–8.
24. Graham C. Incidence and impact of proxy response in measuring patient experience: secondary analysis of a large postal survey using propensity score matching. *International J Qual Health Care* 2016; 28: 246–52.
25. Smith SC, Lamping DL, Banerjee S *et al.* Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technol Assess* 2005; 9: 1–93 iii-iv.
26. 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: 346–56.
27. Meads DM, Martin A, Griffiths A *et al.* Cost-effectiveness of dementia care mapping in care-home settings: evaluation of a randomised controlled trial. *Appl Health Econ Health Policy* 2020; 18: 237–47.
28. 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: 1283–94.
29. 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.
30. Hernandez Alava M, Wailoo A, Pudney S. Quality review of a proposed EQ-5D-5L value set for England EEPUR Report, 2018.
31. van Hout B, Janssen MF, Feng YS *et al.* Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 2012; 15: 708–15.
32. NICE. Position statement on use of the EQ-5D-5L value set for England (updated October 2019). National Institute for Health and Care Excellence 2019 <https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l> (accessed 29 November 2020).
33. NICE. The EQ-5D-5L valuation set for England A view from the side. 2019 <https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisal/Brouwer-advice-EQ-5D-5L-valuation.pdf> (accessed 29 November 2020).
34. Norman R, Olsen JA. Competing views on the English EQ-5D-5L valuation set. *Value Health* 2020; 23: 574–5.

Received 19 August 2020; editorial decision 4 December 2020