Nikolova S, **Hulme C**, West R, Pendleton N, Heaven A, Bower P, Humphrey S, Farrin A, Cundill B, Lawton R, Clegg A. Normative estimates and agreement between two measures of health-related quality of life in older people with frailty: findings from the Community Ageing Research 75+ cohort.

Background

Previous studies have summarised evidence on health-related quality of life (HRQOL) for older people, identifying a range of measures that have been validated, but have not sought to present results by degree of frailty. Furthermore, previous studies did not typically use quality of life measures that generate an overall health utility score. Health utility scores are a necessary component of Quality Adjusted Life Year calculations used to estimate cost-effectiveness of interventions.

Methods

We calculated normative estimates in terms of mean and standard deviation for EQ-5D-5L, SF-36 and SF-6D for a range of established frailty models. We compared response distributions across dimensions of the measures and investigated agreement using Bland-Altman and Interclass Correlation techniques.

Results

EQ-5D-5L, SF-36 and SF-6D scores decrease and their variability increases with advancing frailty. There is strong agreement between EQ-5D-5L and SF-6D across the spectrum of frailty. Agreement is lower for people who are most frail, indicating that different components of the two instruments may have greater relevance for people with advancing

frailty in later life. There is a greater risk of ceiling effects using EQ-5D-5L rather than SF-6D.

Conclusions.

We recommend SF-36/SF-6D as an appropriate measure of HRQOL for clinical trials if fit older people are the planned target. In trials of interventions involving older people with increasing frailty we recommend that both EQ-5D-5L and SF36/SF6D are included, and are used in sensitivity analyses as part of cost-effectiveness evaluation.

Highlights

1. International guidelines identify development, evaluation and implementation of new interventions to improve QOL for older people with frailty as a key priority.

2. Evidence on normative estimates and agreement for different measures of HRQOL across the spectrum of frailty is critical for designing interventions and cost-effectiveness evaluation.

3. Researchers should consider using SF-36/SF-6D if fit older people are the planned target. In interventions involving older people with increasing frailty, both ED-5D and SF-6D should be included.

Introduction

Background

The World Health Organisation (WHO) defines quality of life (QOL) as an 'An individual's perceptions of their position in life, in the context of the culture and value systems in which they live, and in relation to their goals, expectations, standards and concerns'[1]. International guidelines identify development, evaluation and implementation of new interventions and services to improve QOL for older people with frailty as a key priority[2],[3]. A 2012 consensus report by the US Institute of Medicine has recommended a focus on QOL outcome measures for research and programme evaluation of interventions for people living with long-term health conditions[4].

Previous reviews have summarised evidence on health-related quality of life (HRQOL) measures for older people, identifying a wide range of measures that have been validated, but have not sought to present results stratified by degree of frailty[5]. More recently, reviews have focused on QOL for people living with frailty, but have identified limitations in the evidence base[6]. For example, although frailty is best understood as a graded condition, with evidence for the existence of mild frailty, or 'pre-frailty', that typically precedes development of more advanced frailty studies typically dichotomised frailty into 'not frail' and 'frail' categories[6]. Furthermore, included studies did not typically use QOL measures that enable generation of an overall health utility score, whereby individual health profiles are converted into single utility scores by applying pre-existing weights based on preferences of the general population. Health utility scores are a necessary component of Quality Adjusted Life Year (QALY) calculations used in health economic evaluations to estimate value for money of interventions.

Notably, no studies have evaluated the EuroQol 5-dimension health questionnaire (EQ-5D) in frailty as a well-established measure of HRQOL that enables generation of a health utility score. EQ-5D is the preferred UK National Institute for Health and Care Excellence (NICE) measure of HRQOL in adults [7] and is also the most evaluated HRQOL measure internationally[8],[9]. Furthermore, although previous studies have evaluated the short-form 36 item health questionnaire in frailty (SF-36)[6], none evaluated the short-form 6 dimension (SF-6D) health utility score that can be derived from the SF-36 for health economic modelling. Although the EQ-5D and SF-6D both enable derivation of a health utility score, and have been demonstrated to converge at the aggregate level, there is ongoing uncertainty regarding differences across patient groups and illness severity[10]. The two measures differ in terms of their dimensions, items, and preference weights and, therefore, can potentially assign different utility scores to the same individual [10,11]. Furthermore, SF-6D has the potential to tap into broader aspects of HRQOL through its role and social functioning dimensions. These are particularly salient for the population of older people living with frailty as they often have complex health and social care needs and, thus, the social value of an intervention may be more important than health improvement.

The absence of evidence on health utility scores for older people living with frailty is problematic, as normative estimates are needed for design of clinical trials to evaluate new interventions. Furthermore, HRQOL estimates for people living with different grades of frailty inform development of robust economic models, for example decision analytic cost effectiveness models that incorporate transition between frailty categories. Also, investigation of agreement between different HRQOL measures across different frailty categories would help inform selection of instruments for both observational research studies and clinical trials.

Objectives

To report normative estimates for EQ-5D and SF-6D for a range of established frailty measures, compare response distribution across dimensions of the two HRQOL measures, and investigate agreement between EQ-5D and SF-6D using Bland-Altman and Intraclass Correlation techniques.

Methods

Study design

Secondary analysis of prospective cohort data from the Community Ageing Research 75+ (CARE75+) study, collected between December 2014 and November 2018.

Setting

Multi-site, community-based cohort study, recruiting from UK general practices across a range of urban and rural areas, with wide sociodemographic representation[12].

Participants

People aged 75 years and over and living at home were eligible. Care home residents, people living at home and bedbound, and people in the terminal stage of life were excluded.

Variables

Frailty measures

1) Phenotype model

The phenotype model of frailty, based on the five physical characteristics as reported in the original Cardiovascular Health Study (slow walking speed, weight loss, exhaustion, weak grip strength, low energy expenditure), uses standardised cut points[13]. Those with no characteristics were identified as fit, one or two characteristics as pre-frail and three to five characteristics as frail.

2) Cumulative deficit model

The research-standard 60 item frailty index (FI) is based on the cumulative deficit model of frailty and previously validated as part of the English Longitudinal Study of Ageing (ELSA)[14]. The FI score is calculated as an equally weighted proportion of the number of deficits present in an individual relative to the total possible. The FI groups individuals into four categories: very fit (FI score of 0-0.10); well (>0.10-0.14), vulnerable (>0.14-0.24), and frail (>0.24)[15].

3) Electronic frailty index (eFI)

The eFI score is based on the cumulative deficit model of frailty, including 36 variables recorded in the primary care electronic health record (EHR) as part of routine care. The score is calculated as an equally weighted proportion of the number of deficits present in an individual relative to the total possible. The eFI enables identification of frailty categories: fit (0-0.12), mild frailty (0.12-0.24), moderate frailty (0.24-0.36), severe frailty (>0.36)[16].

HRQOL measures

1) EuroQol 5-dimension health questionnaire, 5-level version (EQ-5D-5L)

The EQ-5D-5L five dimensions are: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 levels of severity: no problems, slight problems, moderate problems, severe problems and extreme problems. The scores for each of the five dimensions are combined in a five digit number representing 3125 different health states that can be converted into a utility index ranging from -0.29 to 1 (0 for dead, 1 for perfect health and negative values for states worse than death) for use in economic evaluation[17].

2) SF-36

The RAND short-form 36-Item Health Survey (SF-36) questionnaire includes 36 questions spanning eight health domains: physical functioning; bodily pain; role limitations due to physical health problems; role limitations due to personal or emotional problems; general mental health; social functioning; energy/fatigue; and general health perceptions. It also includes a single item that provides an indication of perceived change in health. The SF-36 enables calculation of Physical Component Summary (PCS) and Mental Component Summary (MCS) scores. SF-36 domain scores, PCS and MCS are on a 0-100 scale with higher scores indicating better health.

3) SF-6D

The SF-6D is a health utility score derived from 11 items of the SF-36 questionnaire. The items are converted into a six-dimension health state classification system, the SF-6D, with four to six levels, allowing for a total of 18,000 unique health states. Dimensions of SF-6D include physical functioning, role limitations, social functioning, pain, mental health, and vitality. The SF-6D index score has values ranging from 0.29 to 1, with lower values representing worse health-related quality of life[18].

Methods of assessment

All measures, except eFI scores, were obtained during face-to-face assessments in the participant's own home. eFI scores were obtained directly from primary care Electronic Health Records (EHR).

Bias

All measures were collected using an electronic data capture system, and researchers were unable to review previous scores at follow-up time points for the same individual, limiting potential for assessment bias.

Statistical methods

We analysed the two HRQOL measures (EQ-5D-5L and SF-6D), the eight individual dimensions of SF-36, and the SF-36 PCS and MCS scores. We generated summary statistics for the entire sample and for frailty subsamples. We used qualitative, Bland Altman (BA), and quantitative, Intraclass Correlation (ICC), techniques to examine agreement between the two HRQOL measures.

The ICC method formally tests significance of agreement in the sample under study. We used consistency of agreement ICC (CA-ICC) to account for the fact that the two utility scores are measured on different scales using a two-way mixed-effect ICC whereby the two HRQOL measures were modelled as fixed effects. The ICC method is dependent on the range of the measurement rather than the actual scale of measurement. Further, a high ICC is based on the assumption that discrepancies in measuring health utility are the same across the possible range of outcomes. This might be considered too restrictive when we compare two health utility measures that, under an ordinality assumption, only need to preserve ranking to be equivalent and thus allow for non-constant biases across the different values of the indices. We therefore used an additional, qualitative measure of agreement, the Bland-Altman plot[19 20] which shows variation in agreement over the entire range of values. We performed two data transformations for the Bland-Altman analysis. First, to omit individual-specific clustering, we excluded all but the last observation for each individual. Second, we collapsed the data to averages of EQ-5D for each value of SF-6D and averages of SF-6D for each value of EQ-5D and retained one observation per individual for the analysis.

We analysed the distributions of self-reported responses for individual dimensions of EQ-5D-5L and SF-6D questionnaires for the whole sample and for subsamples based on frailty categories. We also examined correlation between dimensions of these two questionnaires using Spearman's rank correlation coefficient.

Missing data

Data are assumed missing at random throughout the analysis.

Results

Participants

Data from 2472 assessments of 1038 individual CARE75+ participants are included, with 75% of the study population aged 75–84 and slightly more women (52.7%). Based on the phenotype model, 20.2% of the sample were classified as fit, 51.4% pre-frail and 28.4% frail.

According to the cumulative deficit model 28.3% were classified as fit, 15.5% well, 30% vulnerable and 26.2% frail. The eFI distribution of frailty suggests that 22.4% were fit, 32.8% had mild frailty, 32.3% moderate frailty and 12.5% severe frailty (Table 1).

Main results

Quality of Life scores

Table 2 presents normative data, in the form of means and standard deviations, for EQ-5D-5L, SF-6D, PCS, MCS and the eight dimensions of SF-36 for the sample as a whole and by frailty categories.

EQ-5D-5L and SF-6D utility scores consistently decrease with frailty. The EQ-5D-5L mean exceeds the SF-6D mean for all frailty categories. In general, the utility score mean difference decreases with advancing frailty across all three indices.

Mean SF-36 scores decrease across all eight dimensions with increasing frailty. Similarly, SF-36 PCS scores decrease with increasing frailty. Although SF-36 MCS scores decrease with frailty, the differences between two consecutive groups for the Phenotype and eFI model are small (less than 2 points, or 4%) and somewhat larger (within 4.76 points, or 9.2%) for the Frailty Index.

Variance estimates increase with frailty for most indices. This pattern is different on dimensions related to physical functioning. In particular, standard deviation is lower for the most frail category on PCS and two of its components (physical function and role limitations due to physical problems).

Normative estimates were further stratified by age group and sex (Tables 4-7 in online appendix). Quality of life scores consistently decrease with age. Although men, in general, report higher quality of life, women with advancing frailty and aged 85 and older report higher or similar EQ-5D, SF-6D, composite MCS and mental health and emotional role SF-36 dimensions.

Variance estimates increase with age and are higher for women. For individuals with advancing frailty aged 85 and older, the standard deviation is consistently lower on the composite PCS and physical functioning and physical role SF-36 dimensions across all frailty models.

Agreement

CA-ICC results indicate stronger agreement between individual EQ-5D-5L and SF-6D across the entire sample (CA-ICC 0.61), compared with the generally lower CA-ICC estimates for individual frailty subgroups. (Table 3). The CA-ICC estimates are lower for the fit and frail categories of the phenotype model, but close in magnitude to the CA-ICC estimate for the entire sample for the pre-frail category. The FI CA-ICC estimates for different frailty subgroups are low in magnitude. In the case of eFI, the CA-ICC is of similar magnitude across different frailty categories and closer in magnitude to the estimate for the entire sample. Precision of CA-ICC estimates (confidence interval (CI) width 0.05) is higher for the entire sample compared to frailty subsamples.

The Bland-Altman plot (Figure 1) constructed using averages of one utility measure given the value of the other shows that there is systematic variation in EQ-5D-5L and SF-6D scores. Frailer individuals have lower average value of the two utility measurements, compared with fit individuals. Within the frail group, SF-6D scores are typically higher, compared to the EQ-5D-5L. Conversely, within the fitter group, EQ-5D-5L scores are typically higher. Results indicate increased greater variation in estimates with advancing frailty.

Given the differences in range and valuation of the EQ-5D and SF-6D utility measures we checked whether ranks of utility scores are better suited for the Bland-Altman analysis. The plot for ranks (in online appendix) shows higher variability in scores at better levels of health. This, at least partially, can be explained by the fact that the number of individuals with similar EQ-5D and SF-6D is significantly higher at the healthier end of the utility spectrum. As a result, the same error in measurement between EQ-5D and SF-6D will lead to larger discrepancy in terms of rank. The Bland Altman plot for ranks is also symmetric around 0 suggesting there is no bias in predicting rank of one utility score using rank of the other.

Correlation between EQ-5D-5L and SF-6D dimensions

The correlations between similar dimensions [21] of the EQ-5D-5L and SF-6D utility indices are either high or moderate. The lowest correlations are observed between three pairs: (1) EQ-5D-5L pain/discomfort dimension and SF-6D mental health ($\rho = 0.189$); (2) EQ-5D-5L anxiety/depression and SF-6D pain ($\rho = 0.190$); and (3) EQ-5D-5L anxiety/depression and SF-6D vitality ($\rho = 0.190$).[1] For each dimension of EQ-5D-5L a large proportion of the responses is concentrated in the top level. 37% of the overall study population reported scoring optimal HRQOL (EQ-5D score of 11111), but this was not observed with SF-6D with only 1% scoring the highest possible score. The predominant response for SF-6D physical functioning is level 2, while responses on pain and vitality have two equally probable levels (levels 1 and 2 for pain and 2 and 3 for vitality).[2]

[1] For Spearman's correlation estimates check online appendix.

[2] For more details on distributions of self-responses check online appendix

Discussion

Key results

Quality of life scores

Our results indicate that mean EQ-5D-5L and SF-6D scores decrease, but overall variability of scores increases, with advancing frailty. Compared with EQ-5D-5L, SF-6D utility index value is higher for people with advancing frailty, consistent with previous research that has compared the two measures in more severe illness states[10,11].

Mean scores for the eight dimensions of SF-36 decrease with frailty. The decrease in mean scores is more notable for physical components of SF-36 and the overall physical component summary score, compared with the mental component summary score. Differences between means for the MCS are small with larger differences between groups of the FI model. As opposed to the EQ-5D-5L and SF-6D scores, variability estimates for the eight SF-36 dimensions typically decrease with advancing frailty. Notable exceptions are the PCS and its physical function and physical-role components, where variability increases with advancing frailty. The variability of physical characteristics appears to decrease as individuals become very old, potentially reflecting greater similarity in physical capabilities in very advanced old age.

Agreement

The CA-ICC for the entire sample is larger than the frailty subgroup estimates as it is affected by the variability across a population. We found that variability within eFI categories and pre-frail category of the phenotype index is similar to the variability in the entire sample as indicated by CA-ICC estimates. The CA-ICC precision for the entire sample is larger as the CI width is directly related to sample size.

We also observed stronger agreement between utility values at higher health levels on the Bland-Altman plot. Due to higher concentration of individuals at this end of the utility spectrum, the Bland-Altman plot based on ranks demonstrates higher variability. We conclude that value, not rank, is the appropriate measure of analysis.

Response distribution

The five level EQ–5D-5L has been shown to reduce the ceiling effects of the earlier three level version (EQ-5D-3L)[22]. This study however has identified that more than one in three older individuals (37%) scored in the top level on all five dimensions, indicating optimal HRQOL[23], even though they had frailty classed as advanced by frailty models (8% by the Frailty Index and 15% by the Phenotype and eFI models), raising ongoing concern for ceiling effects with the five-level version in some groups of older people[24].

Limitations

The findings in this study are based on a sample of individuals who are older than 75, live at home, with a relatively low prevalence of dementia. As a result, normative estimates and additional findings from this study cannot necessarily be extrapolated to older people living with dementia or care home residents[25].

In this study we assessed agreement between EQ-5D-5L and SF-6D. EQ-5D and SF-6D measures are used in healthcare decision making by NICE in the UK and health technology assessment agencies in other countries including Brazil, China, Norway, South Korea, and Spain[9], with the EQ-5D being the preferred measure of HRQOL in adults. NICE currently does not recommend using the 5L valuation set[26]. Existing evidence[27], however, suggests that the 5L has superior measurement properties than the 3L and is preferable in population with multimorbidities, likely to have some similar characteristics to the population of older people living with frailty. As the cost-effectiveness results obtained from the two measures will likely differ, a sensitivity analysis using SF-6D to explore uncertainty in estimates for the population of individuals with increasing frailty is thus needed.

Interpretation

This study provides important information on normative estimates and agreement for different measures of health-related quality of life across the spectrum of frailty. These normative estimates can be used for robust sample size calculations by trialists investigating novel interventions for older people with frailty where health-related quality of life is the primary outcome of interest.

Our findings indicate that health related quality of life decreases with advancing frailty when either EQ-5D-5L or SF-6D are used as the measure. There appears to be greater impact on physical health related quality of life than mental health-related quality of life, which may in part be explained by the greater emphasis on physical characteristics within frailty models.

Overall health utility scores are more consistent for people who are fit, with greater variability in scores for people with increasing frailty, while physical components of the SF-36 and composite PCS demonstrate consistent decline with frailty. Findings are consistent across different frailty measures and constructs. We have identified the possibility of a greater risk of ceiling effects using the EQ-5D-5L compared with the SF-6D.

Findings indicate good agreement between EQ-5D-5L and SF-6D across the spectrum of frailty, lending support for the two measures identifying a common construct of health-related quality of life in frailty. Agreement is lower for those who are most frail, indicating that different components of the two instruments may have greater relevance for people with advancing frailty in later life.

We recommend that researchers consider using SF-36/SF-6D as an appropriate measure of health-related quality of life for clinical trials involving older people if there are concerns about the impact of ceiling effects for outcome measurement, for example if fit older people are the planned target. In trials involving older people with increasing frailty, where the social value of an intervention may be more relevant and ceiling effects less of a concern, we recommend that both EQ-5D-5L and SF-36/SF-6D are included as measures of health-related quality of life, and are used in sensitivity analysis as part of planned cost-effectiveness evaluation. Further research to clarify individual priorities for older people living with different degrees of frailty will help guide the future selection of appropriate tools for measurement of health-related quality of life in this population.

Bibliography

1. Study protocol for the World Health Organization project to develop a Quality of Life assessment instrument (WHOQOL). *Qual Life Res*. 1993; 2(2):153-159. doi:10.1007/bf00435734 ;

2. Fit for frailty. British Geriatrics Society. https://www.bgs.org.uk/resources/resource-series/fit-for-frailty. Published 2014. Accessed September 25, 2019.

3. Multimorbidity: clinical assessment and management. NICE. https://www.nice.org.uk/guidance/ng56. Published 2016. Accessed September 25, 2019..

4. Living Well With Chronic Illness: A Call for Public Health Action. *Mil Med*. 2015;180(5):485-487. doi:10.7205/milmed-d-15-00034

5. Haywood K, Garratt A, Fitzpatrick R. Quality of life in older people: a structured review of self-assessed health instruments.*Qual Life Res*. 2006;6(2):181-194. doi:10.1586/14737167.6.2.181

6. Crocker T, Brown L, Clegg A et al. Quality of life is substantially worse for community-dwelling older people living with frailty: systematic review and meta-analysis. *Qual Life Res*. 2019;28(8):2041-2056. doi:10.1007/s11136-019-02149-1 .

7. Guide to the methods of technology appraisalNice.org.uk. NICE. https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781. Published 2013. Accessed September 25, 2019.

8. Rowen D, Azzabi Zouraq I, Chevrou-Severac H, van Hout B. International Regulations and Recommendations for Utility Data for Health Technology Assessment. *Pharmacoeconomics*. 2017;35(S1):11-19. doi:10.1007/s40273-017-0544-y

9. Summary of guidance on health-utility measures by selected health technology assessment agencies. RTIHS. https://www.rtihs.org/sites/default/files/HTA_document_4_March_2019.pdf. Published 2019. Accessed September 25, 2019.

10. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ*. 2004;13(9):873-884. doi:10.1002/hec.866

11. Barton G, Sach T, Avery A et al. A comparison of the performance of the EQ-5D and SF-6D for individuals aged ≥ 45 years. *Health Econ*. 2008;17(7):815-832. doi:10.1002/hec.1298.

12. Heaven A, Brown L, Young J et al. Community ageing research 75+ study (CARE75+): an experimental ageing and frailty research cohort. *BMJ Open*. 2019;9(3):e026744. doi:10.1136/bmjopen-2018-026744

13. Fried LP, Tangen CM, Walston J, et al Frailty in older adults: evidence for a phenotype. *J Gerantol A Biol Sci Med Sci*. 2001, 56(3) doi:10.1093/gerona/56.3.m146.

14. Marshall A, Nazroo J, Tampubolon G, Vanhoutte B. Cohort differences in the levels and trajectories of frailty among older people in England. *J Epidemiol Community Health*. 2015, 69(4): 316-321. doi: 10.1136/jech-2014-204655.

15. Hubbard RE, Goodwin VA, Llewellyn DJ, Warmoth K, Lang IA. Frailty, financial resources and subjective well-being in later life. *Arch Gerontol Geriatr*. 2014; 58(3):364-369. doi: 10.1016/j.archger.2013.12.008.

16. Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing*. 2017. doi: 10.1093/ageing/afx001

17. Devlin NJ, Shah KK, Feng Y, Mulhern B, Hout BV. Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Econ*. 2017;27(1):7-22. doi: 10.1002/hec.3564.

18. Janssen MF, Bonsei GJ, Luo N. Is EQ-5D-5L Better Than EQ-5D-3L? A Head-to-Head Comparison of Descriptive Systems and Value Sets from Seven Countries. *Pharmacoeconomics*. 2018;36(6):675-697. doi: 10.1007/s40273-018-0623-8.

19. Altman JD, Bland JM**.** Measurement in Medicine: the Analysis of Method Comparison Studies. *The Statistician*. 1983;32(3):307. doi: 10.2307/2987937

20. Bland JM, Altman JD. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.*1986;327(8746): 307-310. doi:10.1016/S0140-6736(86)90837-8.

21. Whitehurst DG, Bryan S. Another Study Showing that Two Preference-Based Measures of Health-Related Quality of Life (EQ-5D and SF-6D) are Not Interchangeable. But Why Should We Expect Them to Be? *Value Health*. 2011;14(4): 531-538; doi: 10.1016/j.jval.2010.09.002.

22. Janssen MF, Pickard AS, Golicki D, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-£L across eight patient groups: a multi-country study. *Qual Life Res*. 2012;22(7):1717-1727. doi:10.1007/s11136-012-0322-4.

23. Lutomski JE, Krabbe PFM, Bleijenberg N, et al. Measurement properties of the EQ-5D across four major geriatric conditions: Findings from TOPICS-MDS. *Health Qual Life Outcomes*. 2017;15(1). doi:10.1186/s12955-017-0616-x

24. Huber M, Vogelman M, Leidl R. Valuing health-related quality of life: systematic variation in health perception. *Health Qual Life Outcome*s, 2018;16(1). doi: 10.1186/s12955-018-0986-8.

25. Usman A, Lewis S, Nihsliff-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 outcomes testing. *BMJ Open*. 2018;8(8) doi: 10.1136/bmjopen-2018-022127.

26. Position statement on use of the EQ-5D-5L valuation set for England (updated November 2018). NICE. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l. Accessed September 25, 2019

27. Thompson AJ, Turner AJ. A Comparison of the EQ-5D-3L and EQ-5D-5L. *PharmacoEconomics*. 2020. doi:10.1007/s40273-020-00893-8

28. Hout BV, Janssen M, Feng Y-S, et al. Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets. *Value Health*. 2012;15(5):708-715. doi:10.1016/j.jval.2012.02.008

Table 1: Summary statistics for the sample of older people with frailty

Table 2 Means and stds for EQ-5D, SF-6D, PCS, MCS and eight dimensions of SF-36 for the whole sample and by frailty categories

.

Table 3: CA-ICC between EQ-5D-5L and SF-6D for the whole sample and by frailty categories of Phenotype, Frailty Index and eFI models.

Figure 1: Bland Altman limits of agreement between EQ5D and SF6D values

Figure1 shows the relationship between EQ-5D and SF-6D sample means difference for different values of the average of the two variables.