
Article:
Penton, H., Dayson, C., Hulme, C. et al. (2022) A qualitative investigation of older adults' conceptualization of quality of life and a think-aloud content validation of the EQ-5D-5L, SF-12v2, Warwick Edinburgh mental well-being scale, and Office of National Statistics-4. Value in Health, 25 (12). pp. 2017-2027. ISSN: 1098-3015

https://doi.org/10.1016/j.jval.2022.04.1735


ScienceDirect

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

Patient-Reported Outcomes

A Qualitative Investigation of Older Adults' Conceptualization of Quality of Life and a Think-Aloud Content Validation of the EQ-5D-5L, SF-12v2, Warwick Edinburgh Mental Well-Being Scale, and Office of National Statistics-4.

Hannah Penton, PhD, Christopher Dayson, MA, Claire Hulme, PhD, Tracey Young, PhD

ABSTRACT

Objectives: Old age is characterized by declining health, comorbidities, and increasing health and social care service use. Traditionally, patient-reported outcome measures (PROMs) including the EQ-5D-5L and SF-12v2 have focused on health. Nevertheless, aged care often aims to improve broader elements of quality of life (QoL), captured by well-being measures, such as the Warwick Edinburgh Mental Wellbeing Scale (WEMWBS) and Office of National Statistics-4 (ONS-4). This study investigates older adults' conceptualization of QoL and the content validity of the EQ-5D-5L, SF-12v2, WEMWBS, and ONS-4 in measuring their QoL.

Methods: Qualitative cognitive think-aloud interviews were undertaken with older adults aged 75+, exploring their views on what was important to QoL and, for each measure, the relevance, acceptability, and interpretation of items; suitability of response options; and the comprehensiveness of the measure. Conceptualization of QoL was analyzed thematically and content validity using framework analysis.

Results: Twenty interviews were undertaken. Older adults' conceptualization of QoL centered on health, ability to perform usual activities, social contact, and emotional functioning. Possible response shift was observed, as older adults assessed their health relative to lower health expectations at their age or to people in worse states. Participants questioned the relevance of negatively phrased mental items and often preferred the functioning-focused EQ-5D-5L to more subjective ONS-4 and WEMWBS items. Domains suggested to improve comprehensiveness included social contact, coping, security, dignity, and control.

Conclusions: These findings are useful to researchers developing new PROMs for older adults or for the developers of included PROMs considering permanently adapting or bolting-on domains to improve content validity in older adults.

Keywords: SF-12v2, EQ-5D-5L, cognitive interview, content validity, ONS-4, older adults, quality of life, WEMWBS.

VALUE HEALTH. 2022; 25(12):2017–2027

Introduction

The quality-adjusted life-year, a unit that combines both length and quality of life (QoL), is a widely used outcome measure to inform decision making in health technology assessment.[1] QoL is measured using patient-reported outcome measures (PROMs), which historically have focused strongly on health.[1-3] For example, the UK's National Institute for Health and Care Excellence and the Dutch Zorginstituut require QoL be measured using the EQ-5D measure of health, whereas the Canadian Agency for Drugs and Technologies in Health suggests either the EQ-5D or Short-Form 6-Dimension, the preference-based measure in the popular SF instrument family, including the 36-Item Short Form Survey and SF-12.[4-9]

There is increasing recognition that this health focused approach may be inappropriate when evaluating health and social care interventions for older adults.[2,10,11] The World Health Organization defines healthy aging as the process of developing and maintaining the functional ability that enables well-being in older age.[12] Therefore, services for older adults often have broader benefits outside health, including increased social participation and independence.[2,10,11,13] These broader benefits may be missed by commonly used health measures and therefore not accounted for in economic evaluation, resulting in these services being undervalued.[1] This could result in inefficient service provision for an intensive group of service users, which could be detrimental to the QoL of older adults and the economic efficiency of health and social care budgets.

The need to include wider benefits has been recognized by health technology assessment agencies in the evaluation of social care services, where a large proportion of spending goes to older populations.[4,5] The National Institute for Health and Care Excellence social care guidance manual suggests using well-being measures to capture broader QoL.[14] Two potential well-being measures are the Warwick Edinburgh Mental Wellbeing Scale (WEMWBS) and the Office of National Statistics personal well-being questions (ONS-4).[15,16] These measures have been included in several large population health and well-being surveys and have also been used to evaluate health and social care interventions aimed at older adults in the United Kingdom.[17-20]

Despite such PROMs being widely used in older adults, there is little evidence of their validity in this important group of service users, because they are often overlooked in measuring development and psychometric testing.[2,10] Content validity is argued to be the most important element of measurement performance,[21] given that the validity of the data received from questionnaires is dependent on whether the questions and response options are understood by the respondent, relevant to the concept being measured, and whether the important aspects of that concept are comprehensively captured. Nevertheless, a systematic review investigating the psychometric performance of the EQ-5D, SF-12, WEMWBS, and ONS-4 in older adults failed to identify any studies examining the content validity of these measures in older adults, except one study on the Dutch 3-level version of EQ-5D.[13,22]

In-depth qualitative interviews are often used in measure development to understand what is important to the concept being measured in the target population, as a basis for selecting dimensions.[23,24] Cognitive interviews are commonly used to examine the content validity of existing or experimental versions of PROMs.[13,25] Cognitive techniques including think-aloud and verbal probing allow in-depth investigation of participants' response process when completing a PROM.[25] This can identify response issues that threaten content validity, enabling the maximization of validity and reliability and the minimization of measurement error in PROM data.[26-28]

There are various points in the response process where response issues may threaten validity.[25] Respondents may not understand the question or response options or interpret them differently than intended. This may bias conclusions drawn from responses and invalidate comparisons among individuals. Respondents may feel the questions are inappropriate or irrelevant to them, or there may be a mismatch between the response options provided and the individual's desired response, meaning they may not engage fully with the question potentially leaving it blank or providing an invalid answer. Finally, the respondent may edit their response, for example, responding more positively than their true state, again leading to bias.

The objective of this study was to conduct a qualitative investigation of older adults' conceptualization of QoL and to investigate the content validity of the EQ-5D-5L, SF-12v2, WEMWBS, and ONS-4 in measuring the QoL of older adults.

Methods

Sample and Recruitment

Participants were recruited from the Community Aging Research 75+ (CARE75+) cohort study, which investigates frailty transitions and health over time in adults aged 75+ years.[29] Participants for this study were recruited from the Bradford and Leeds area. Patient and public involvement and engagement on study design and recruitment were sought through the CARE75+ Frailty Oversight Group, an independent older lay reference group.[30]

Eligible participants were aged 75+ years and had a recent (assessed in CARE75+) Montreal cognitive assessment score ≥26, the cutoff for normal mental capacity, indicating sufficient mental capacity to consent and comprehend the interview tasks.[31]

Participants were recruited using convenience sampling. Contact details of eligible CARE75+ participants who consented to be contacted about future research projects were sent to the researcher (H.P.) via encrypted email. Invitation letters including an information sheet, consent form, and response card were sent to potential participants inviting them to participate in one-to-one face-to-face interviews with the researcher. Participants could respond by email, telephone, or returning the response card, using the included stamped and addressed envelope. This study was approved by the Health Research Authority and South West Frenchay NHS Research Ethics Committee in December 2017.

Measures

EQ-5D-5L

The EQ-5D-5L measure of health contains 5 items: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.[32] Each question has 5 response options covering no problems, slight problems, moderate problems, severe problems, and extreme problems/unable to. Utilities are generated using the accompanying country specific value set. The EQ-5D-5L also contains a visual analog scale (VAS) where respondents rate their health today between 0, worst imaginable health, and 100, best imaginable health.

SF-12v2

The SF-12v2 measure of health status contains 12 items, each with 3 to 5 response options, covering 8 domains including general health, physical functioning, physical role, pain, emotional role, mental health, vitality, and social functioning.[7] This study used the English standard 4-week recall version.

WEMWBS

The WEMWBS is a measure of positive mental well-being containing 14 items about feeling optimistic about the future, useful, relaxed, interested in others, good about oneself, close to others, loved, interested in new things, confident, cheerful, able to deal with problems, think clearly, and make up one's mind about things and having energy. Each item has 5 response options: none of the time, rarely, some of the time, often, and all of the time.[16]

ONS-4

The ONS-4 was developed by the Office of National Statistics to measure personal well-being.[15] It contains 4 items covering life satisfaction, the extent to which people feel the things they do in life are worthwhile, happiness, and anxiety. Respondents can respond between 0 ("not at all") and 10 ("completely").

Interview Procedure

All interviews were conducted by a single interviewer (H.P.) with a background in health economics and, more specifically, patient outcomes research and psychometric testing of PROMs. The interviewer was trained in qualitative and mixed methods research methods and conducting interviews and focus groups.

Cognitive interviewing is mentally demanding, because participants are required to elaborate on their thought processes while completing PROMs and recall additional details after completion in verbal probing. The WEMWBS and SF-12v2 are relatively long, increasing the burden for participants completing

these measures. To reduce the cognitive demand on frail older participants, each interview only covered 2 PROMs, in varying combinations and orders (Table 1). Two pilot interviews were conducted to check that the interview design was appropriate and feasible.

The interviewer began each interview by explaining what would be covered in the interview, explaining the think-aloud process and reminding participants they could stop the interview at any time. Participants were then asked to sign the consent form and complete some demographic questions (age, ethnicity, education background, living situation, and long-term conditions).

The topic guide began by discussing the participant's definition of QoL and what they needed in life to achieve a good QoL. The first PROM was then provided. Participants were asked to think aloud, saying whatever they were thinking, while completing the measure and were prompted to "continue thinking aloud" if they became silent and stopped explaining how they were arriving at answers. Once participants completed the PROM, they were asked for their initial impressions, including whether it was clear, easy to understand, and of acceptable length.

Retrospective think-aloud techniques, asking respondents to explain how they arrived at responses that were not verbalized during the concurrent think-aloud, were used in combination with retrospective verbal probing, to further explore participants' interpretation and understanding of terms in each question and whether the questions were relevant, important to their QoL, and acceptable to ask to someone like themselves. Once all items had been discussed, participants were asked whether the measure missed anything important to their QoL. A break was offered, before the process was repeated for the second PROM. Finally, respondents were asked which of the 2 measures they preferred.

Interviews were audio recorded and transcribed verbatim by the interviewer. The interviewer made field notes during the interviews, noting down nonverbal cues to provide relevant context to participant responses and analysis.

Analysis

Data analysis was performed iteratively by the interviewer during the data collection period to monitor data saturation. Saturation was determined once interviews produced no further response issues or codes that added substantial understanding of the data. No new codes or response issues were identified beyond the eighth interview for any of the PROMs included.

Transcripts were read repeatedly to ensure familiarity and entered into NVIVO 11 for data management.[33] To explore participants' understanding of QoL and what was important to their QoL, interviews were initially analyzed using thematic analysis, following the procedure outlined by Braun and Clarke.[34] During repeated readings of transcripts, data were coded iteratively and then codes grouped into themes.

Response issues that threatened the content validity of the measures were analyzed using framework analysis. The framework was adapted from an existing framework of response issues from a Dutch content validation study of the EQ-5D-3L, ICECAP-O, and ASCOT in older people.[13] This framework, based on Tourangeau's model of survey response, examined response issues related to understanding, interpretation, the suitability of response options, and response editing.[35] Adaptation of this framework for the current study was considered appropriate given the similarity in the aims, populations, and measures being assessed in the 2 studies and the use of Tourangeau's model of survey response, which provided a strong theoretical foundation to the framework. The framework was adapted during data collection, by adding categories related to the format of the measures, the relevance and acceptability of items, and comprehensiveness of measures. When analyzing the content validity of each measure, a report was created per item, listing the verbatim transcription of each respondents' comments on that item, as recommended by Knafl et al.[28] Comments on the measure as a whole were also grouped. Comments about each item and each measure in general were then reviewed using the framework.

To minimize subjective bias arising from the fact that the data analysis was performed solely by the interviewer, reflective practice was used to ensure that the researcher reflected on and accounted for personal biases. The researcher also continually discussed the codes that were being developed and response issues identified with the research team who were involved in the project. This formed part of the reflective practice because they could question and provide feedback on the analysis as it was presented.

Results

Respondents

A total of 122 people were invited to participate, and 40 responses were received. Twenty-two agreed, but one was uncontactable and one consented, but during the interview the researcher could not be sure they had sufficient capacity to consent because of a decline in mental capacity. The interview was stopped, and data were excluded from analysis.

Interviews occurred between February and May 2018, mostly in participants' homes. Twenty participants were interviewed and included in the analysis. Saturation was reached and recruitment terminated after the 20th interview, when each measure had been discussed by 10 participants. All interviews were one to one except 2 participants who wanted to be interviewed together and 1 participant who requested her daughter be present. All participants completed both measures in a single interview.

Most participants were female (65%), lived alone (55%), and had a long-term condition (75%) (Table 1). Four lived in assisted living facilities (20%), whereas the rest lived at home. Participants were spread across the age and Fried Frailty categories, with a score of 0 = fit, 1 to 2 = prefrail, and 3 to 5 = frail (assessed in CARE75+).[29,36]

Conceptualization of QoL

To fully understand content validity issues, it is first important to understand participants' definition of QoL and what is important to their QoL. What was important to participants' QoL centered around 4 themes: health, ability to perform usual activities, social participation, and emotional functioning.

Health was important to all participants' QoL as health affected their ability to undertake activities they valued or enjoyed and participate in regular social interaction. Mrs Eighteen said, "If you've got your health at my age, you don't need a lot more in life because you can get out and about and do stuff."

The way health was discussed provided important insights into the way older adults assess their QoL, which will affect their responses to PROMs. Participants commonly expressed that their expectations of their health had declined with old age. Mr Two said, "I expect to go down a bit. You don't expect to stay the same active as you were 10 years ago."

Participants often assessed their health relative to people they knew of a similar age who were in worse health. These people were used as examples of how lucky participants were to be in a relatively better state. This often led to them describing problems

Table 1. Participant characteristics and PROMs completed.

PROMs completed	Participant	Age category	Fried frailty	Long-term condition	Lives alone
EQ-5D-5L – WEMWBS	Mr One	75-79	1	Yes	Yes
	Mrs Eight	85-89	3	Yes	Yes (assisted living)
WEMWBS – EQ-5D-5L	Mrs Seven	85-89	5	Yes	Yes
EQ-5D-5L – SF-12v2	Mr Two	90-94	4	Yes	Yes
	Mrs Twenty	75-79	4	Yes	Yes
SF-12v2 – EQ-5D-5L	Mrs Fifteen	90-94	3	No	Yes
EQ-5D-5L – ONS-4	Mrs Eleven	80-84	4	Yes	Yes (assisted living)
	Mr Fourteen	75-79	2	Yes	Yes
ONS-4 – EQ-5D-5L	Mrs Four	80-84	1	Yes	With husband
	Mr Five	75-79	2	Yes	With Wife
WEMWBS – SF-12v2	Mrs Nine	80-84	3	Yes	Yes
	Mrs Eighteen	80-84	1	No	Yes (assisted living)
SF-12v2 – WEMWBS	Mrs Sixteen	90-94	2	No	With child's family
	Mr Twelve	75-79	3	No	With Wife
WEMWBS – ONS-4	Mrs Ten	90-94	2	Yes	Yes
	Mr Nineteen	85-89	4	Yes	Yes
ONS-4 – WEMWBS	Mrs Thirteen	90-94	3	Yes	Yes (assisted living)
SF-12v2 – ONS-4	Mrs Three	85-89	2	No	Yes
ONS-4 – SF-12v2	Mrs Six	75-79	3	Yes	Yes
	Mr Seventeen	75-79	2	Yes	With child's family

ONS-4 indicates Office of National Statistics-4; PROM, patient-reported outcome measure; WEMWBS, Warwick Edinburgh Mental Wellbeing Scale.

with their own health but then going on to view their own state very positively, given that they were not as badly off as others.

Peoples' ability to perform their usual activities inside and outside the home was central to their QoL. Ability to perform daily activities within the home was closely related to independence and pride, shown by comments such as "I don't want folk molly-coddling me. I want to do it myself. I know I struggle[…], but I get there in the end" by Mrs Eleven. Their ability to engage in activities outside the home was related to their sense of control and their ability to engage in social contact.

Regular social contact was central to QoL, with everyone mentioning some form of social contact, including family, partner/spousal relationships, friends, or carers. Loneliness was often discussed as a big problem in older people. Mr Two said, "People don't realise how much difference somebody calling in makes to a person on their own[…] It's the most important thing, loneliness."

Participants often discussed the emotional impact of aging and strategies for coping with this. Stoicism was commonly conveyed, with comments about enduring problems and hardship without complaint or displays of emotion, including "I'm not into emotion no. […] you have to get on with it" (Mrs Nine) and "I'm not one that dwells on things. […] I try to look on the bright side" (Mrs Sixteen). Participants often expressed that it was not good for them to dwell on things they could not control and it was better to carry on and maintain a positive outlook.

Response Issues

The final response issue framework is displayed and explained in Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.04.1735. Tables 2 to 5 summarize the response issues identified for the EQ-5D-5L, SF-12v2, WEMWBS, and ONS-4, respectively. The most commonly identified issues and those that present the biggest threats to content validity are summarized below. More detailed results and quotes can be found in the full report of the study available elsewhere.[22]

EQ-5D-5L

Narrow interpretation

All respondents considered being able to do their usual activities important to their QoL. Nevertheless, usual activities were often narrowly interpreted as asking only about household chores. For example, Mrs Seven said, "Usual activities… do you mean cleaning and that?" Participants more rarely interpreted it to include activities outside of the home, despite mentioning them elsewhere.

Respondents often focused on one aspect of the double-barreled items, only mentioning pain for pain/discomfort or only mentioning either anxiety or depression. This may be more problematic for anxiety/depression because these are more distinct concepts.

Recall period

Some respondents struggled to stick to the recall period of "today," as their state varied over time, which sometimes led to response option selection issues. For example, Mrs Seven said, "moderate problems walking about, but sometimes when my back's bad, severe problems. Can I tick 2? No uhhh well, at the moment, I haven't severe problems, so I'll put moderate problems." Phrases such as "at the moment" and "normally I'm alright" (Mrs Eight) suggested that some respondents were averaging over longer time periods.

Table 2. EQ-5D-5L response issues (n = 10).

	EQ-5D-5L						
	Measure whole	Mobility	Self-care	Usual activities	Pain/ discomfort	Anxiety/ depression	VAS
Response issue							
Practical completion							
Length of measure							
Layout of measure							
Understanding							
Odd wording							
Difficult wording							
Recall							
Wrong time period		1					
Interpretation							
Difficult interpretation							
Wrong interpretation			1				
Narrow interpretation				7	7	3	
Response option selection							
Format difficult							
Different answers for different aspects of item		2					
Response options partly applicable							
Irrelevant response options							
Missing intermediate response options							
Similar response options							
Order of options							
Inconsistent response							
Positive responding		3	1	1			5
Acceptability							
Item inappropriate							
Relevance/comprehensiveness							
Similar question							
Item irrelevant						1	
Important aspects missing	3						

VAS indicates visual analog scale.

Positive responding

Respondents answered more positively than expected for several items. This was particularly noticeable for mobility and the VAS. For example, Mrs Four who was confined to a wheelchair selected "severe problems in walking about" for mobility rather than "unable" and selected 90 on the VAS. Mr Two chose "slight problems in walking" for mobility despite having said "my legs are buggered. I've got an electric kart yeah. I can walk, but not far." Further probing of positive responding gave the impression that such respondents interpreted perfect health on the VAS or no problems on the items as the best possible given their age and situation and made a relative assessment based on this, feeling that their health truly was good considering their peers. Mr Fourteen directly discussed the impact of this relativity on his answer, by saying:

> *This scale thing is quite hard because, for example, I pretty well thought 75 but if I was 50 (years old), I wondered if I would put it quite a bit lower. So, it can be very misleading […] everything is relative.*

SF-12v2

Layout

The SF-12v2 questions are lengthy, and participants often had to read them several times to fully understand what was asked. Questions are often presented together in clusters, which caused confusion with some participants reading the long introduction to the cluster and beginning to answer, without realizing that the question was finished below. One participant did not understand that multiple questions were asked within a cluster and only answered one.

Odd wording

The examples of moderate activities created response option selection issues. Participants felt these activities required very different levels of physical ability, as they could mostly move a table or vacuum, but could not (or would not) bowl or play golf. Mrs Six said:

Table 3. SF-12v2 response issues (n = 10).

Response issue	SF-12v2											
	Measure whole	General health	Moderate activities	Stairs	PR accomplish	PR limited	ER accomplish	ER careful	Pain	Calm/ peaceful	Energy	Downhearted/ low	Social activities
Practical completion													
Length of measure													
Layout of measure	5												
Understanding													
Odd wording			3							2	1		
Difficult wording													
Recall													
Wrong time period		1									1		
Interpretation													
Difficult Interpretation													
Wrong interpretation					1	1							
Narrow interpretation					4	4	2	2					
Response option selection													
Format difficult													
Different answers for different aspects			4								1		1
Options partly applicable													
Irrelevant options													
Missing options													
Similar options													
Order of options													
Inconsistent response							1	1			1		
Positive responding		1	1	1	1	1					1	1	
Acceptability													
Item inappropriate													
Relevance/comprehensiveness													
Similar question					2	2	1	1				1	
Item irrelevant				1			5	5			2		
Important aspects missing	2												

ER indicates emotional role; PR, physical role.

Bowling or playing golf I just wouldn't do it… so is that being limited then? It just doesn't come up in my life. No not limited at all[…] But I've missed that bit out altogether – bowling and golf.

People were unsure whether to ignore the examples and think about activities they considered moderate, ignore the 2 more vigorous activities that did not apply to their life, imagine how limited they would be in these activities and use this response, or provide a middle response over all the activities suggested. The chosen strategy affects their response and comparability across respondents.

Narrow interpretation

In the role items, regular daily activities were often interpreted as housework rather than activities outside the home.

Table 4. WEMWBS response issues (n = 10).

Response issue	WEMWBS														
	Measure whole	Optimistic future	Useful	Relaxed	Interested people	Energy	Deal problems	Think clearly	Feel good self	Close to People	Confident	Make up own mind	Loved	Interest New Things	Cheerful
Practical completion															
Length of measure															
Layout of measure															
Understanding															
Odd wording															
Difficult wording															
Recall															
Wrong time period															
Interpretation															
Difficult interpret									1						
Wrong interpret					3				1						
Narrow interpret					1					1				2	
Response option selection															
Format difficult															
Different answers different aspects						1									
Options partly applicable															
Irrelevant options															
Missing options															
Similar options															
Order of options															
Inconsistent response						3	2								
Positive responding															
Acceptability															
Item inappropriate		2	2			2		1							
Relevance/ comprehensiveness															
Similar question															
Item irrelevant		4	2	1	1				1	2			4		1
Important aspects missing	1														

WEMWBS indicates Warwick Edinburgh Mental Wellbeing Scale.

Table 5. ONS-4 response issues (n = 10).

Response issue	ONS-4				
	Measure whole	Life satisfaction	Worthwhile	Happy	Anxious
Practical completion					
Length of measure					
Layout of measure	5				
Understanding					
Odd wording					
Difficult wording					
Recall					
Wrong time period			1		
Interpretation					
Difficult interpretation					1
Wrong interpretation			5		
Narrow interpretation					
Response option selection					
Format difficult	1				
Different answers for different aspects		1			
Options partly applicable					
Irrelevant options					
Missing options					
Similar options					
Order of options					
Inconsistent response					3
Positive responding		1			
Acceptability					
Item inappropriate					
Relevance/comprehensiveness					
Similar question				1	1
Item irrelevant		1	5	1	1
Important aspects missing	1				

ONS-4 indicates Office of National Statistics-4.

Item irrelevant

Some respondents considered the emotional role questions irrelevant to their QoL. Most respondents said anxiety and depression were not issues that they thought about or affected them. Mrs Nine responded, "I'm not into emotion no[…] you have to get on with it." Most people referred to a "carry on" attitude, stating that they did not dwell on things they could not control.

WEMWBS

Interpretation

Several participants questioned whether feeling interested in other people meant being nosy. Mr Twelve said, "Is that being a peeping tom, nosing out the window? (laughs)"

Item inappropriate/irrelevant

Several participants considered asking whether older adults felt optimistic about the future inappropriate or irrelevant as their future was uncertain. Mrs Nine responded, "Feeling optimistic about the future (laughs) what future? Future is tomorrow or today; it's not beyond." Feeling useful was considered inappropriate by frailer respondents who could no longer perform traditional roles, including Mrs Thirteen who said "I don't expect to be useful (laughs)… I've done my job!" Participants also considered having energy "to spare" unrealistic at their age.

ONS-4

Layout

The response scale layout caused confusion. The first 3 items are positively worded so higher numbered responses, toward "completely," indicate better well-being, but anxiety is negatively worded with higher responses suggesting worse well-being. Some participants failed to notice the scale reversal and provided invalid responses, suggesting much higher anxiety levels than intended.

Interpretation

There was a clear spilled in how respondents interpreted the worthwhile item, with some considering whether the things they did were worthwhile to the community and others considering

whether things were worthwhile to themselves. "They're worthwhile to me yes. I don't know if they're worthwhile to society… but to me, yes my life is worthwhile." (Mr Nineteen). Healthier respondents tended to consider the community, whereas frailer respondents were more likely to interpret this as whether they could do basic tasks for themselves, which were "worth doing."

Item irrelevant

Some respondents considered the worthwhile question irrelevant, either because they did not think about whether the things they did were "worthwhile" or because it was no longer relevant, because of their age limiting their functional ability.

Comprehensiveness

Most participants felt the measures were comprehensive; nevertheless, a few suggested additional domains for each PROM (Table 6). Although it is noted for which measure additional domains were mentioned, these additional domains could be relevant to any of the measures included. Relationships, social contact, and loneliness were broadly considered central to older adults' QoL. For example, Mr Two stated, "I think loneliness is the most important question."

One participant noted that the way older adults were treated was very important to their QoL as many older people felt they were a burden on family and society. Other suggestions included coping and support, financial security and feeling secure about future service requirements and living arrangements, control over daily life, and independence.

Measure Preference

Eleven participants preferred 1 measure over the other (Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.04.1735). Several participants who received a health and a well-being measure found the health measure easier to answer as questions focused on ability rather than subjective feelings and concepts that they did not think about in daily life. Several participants preferred other measures over the SF-12v2 because they found the SF-12v2 questions confusing.

Of the 9 participants who stated no preference between the measures they received, 4 thought both measures should be used together, given that they covered different areas of QoL. These 4 participants all received one health and one well-being measure, suggesting they recognized the different coverage of these measures and considered both important.

Discussion

This study found that what was important to the QoL of older adults centered around 4 key themes: health, ability to perform usual activities, social contact, and emotional functioning. These themes closely reflect the World Health Organization definition of healthy aging, which centers on maintaining functional ability to enable the well-being of older adults[12] and key themes identified in the literature.[2,3,37-39] This provides support that these themes are generalizable to older adults more broadly. Developers of PROMs for use in older adults should consider these core domains.

The way health was discussed showed that participants often assessed their health relative to other people of their age who were worse off or relative to expectations of their health given their age or situation. By lowering their benchmark for good health, participants were able to rate their state more positively on the EQ-5D-5L and SF-12v2. Positive responding, because of adaptation, lowering benchmarks, or comparison to others in worse situations, has also been observed by in previous content validation of the EQ-5D in older adults.[13,40] This could indicate response shift because participants may have recalibrated their benchmark for good health as they aged or their health declined. This was seen in participants' responses to various items, including global assessments of health, life satisfaction, physical functioning, and usual activities/roles. Response shift has been widely observed in older adults' responses to PROMs.[39,41-43] Response shift affects the validity of responses and comparisons across participants given that, if different individuals use different strategies or reference points to assess their health, scores are not comparable.

All PROMs had interpretation issues. Usual activities were narrowly interpreted to mean solely housework on the EQ-5D-5L and SF-12v2, respondents sometimes struggled to stick to the recall period on these measures, and only one aspect of double-barreled EQ-5D-5L items were considered. These issues have been noted in previous content validation studies of the EQ-5D-3L in older populations.[13,39] The ONS-4 worthwhile item was interpreted in different ways, to mean worthwhile either to themselves or to society, consistent with findings from a previous content validation.[44] Response option selection issues were also seen for the SF-12v2 moderate activities item. These issues will affect responses and the validity of comparing responses across individuals, leading to biased scores and decision making.

The SF-12v2 and ONS-4 layouts caused confusion, affecting the validity of responses, in line with previous evidence.[44] The relevance of at least one item was questioned on every measure; nevertheless, this was more widespread on the WEMWBS, ONS-4, and SF-12v2 than the EQ-5D-5L. Subjective well-being and negatively worded mental health items were more commonly considered irrelevant to older participants because they adopted a stoic attitude. Although emotional functioning was often discussed as important to QoL, discussion centered on maintaining a positive outlook, stoically refusing to dwell on problems, and carrying on. Negatively worded mental health items went

Table 6. Additional dimensions suggested for each measure.

Dimension	EQ-5D-5L	SF-12v2	WEMWBS	ONS-4
Relationships/social contact/loneliness	✔	✔		
Control over daily life	✔			
Coping/support		✔		✔
Future security			✔	✔
The way people are treated	✔			

Note. For interest, it is noted for which measure additional domains were mentioned. Nevertheless, these additional domains could be relevant to any of the measures included.
ONS-4 indicates Office of National Statistics-4; WEMWBS, Warwick Edinburgh Mental Wellbeing Scale.

against this positive approach and were therefore not considered relevant by some participants. Frailer participants found few of the subjective WEMWBS and ONS-4 items relevant to their life anymore, given that increasing frailty meant they focused on basic functioning rather than broader elements of QoL connected to having a role. These issues could negatively affect respondent engagement, resulting in higher levels of missing or invalid responses.

Participants often found the concise, functioning-focused EQ-5D-5L items easier to answer and more relevant to daily life than other measures, suggesting that the EQ-5D-5L may be the best starting point for measuring the effectiveness of health and social care interventions for older adults. Nevertheless, participants did not feel the EQ-5D-5L comprehensively covered their QoL. The ease of answering the EQ-5D, combined with a lack of comprehensiveness are also reflective of previous findings.[13,39] Therefore, it may require adaptation, either through permanent adaptation or the additional of bolt-on dimensions for use in older respondents.[45-47]

Additional dimensions suggested to improve the comprehensiveness of each measure for older adults included: social contact, coping/support, future security, the way people were treated, and control/independence. These domains could be appropriate bolt-ons to an existing measure of QoL or included in a new measure aimed at older adults. These suggested domains closely align with domains of other PROMs developed for older adults or with their qualitative input, including ICECAP-O domains of attachment, control, and security and ASCOT domains of social participation, control, occupation, and dignity.[3,48]

Limitations

The sample did not include older people living in care homes or nonwhite ethnicities. The CARE75+ cohort from which this sample was recruited did not recruit individuals living in care homes. Given the time available for this project and the difficulties associated with recruiting older adults, particularly those requiring the level of care provided by care homes, a pragmatic choice to recruit through the CARE75+ was made, acknowledging that care home residents would be missed. Although nonwhite ethnicities were invited to participate, none responded. Therefore, findings may be broadly generalizable to an older community-dwelling white British population, but content validity in older adults from nonwhite cultural backgrounds and those living in care homes should be investigated further.

Participants had previous experience in CARE75+ of being asked questions about their health and QoL, including the EQ-5D-3L and 36-Item Short Form Survey. This may have affected this study's findings because participants may have been more accepting of such PROMs than older adults who have never experienced them. Although their prior experience of thinking about their health and QoL in CARE75+ may also mean they found it easier to form opinions about these concepts, which would be beneficial to the study, this focus on health in the CARE75+ and experience in answering health-related PROMs may have also influenced the preference for functioning-focused items and measures. Nevertheless, similarities with previous studies in which older respondents also found the EQ-5D easy to answer provide some confidence in results.[13,39]

Identification and classification of response issues are subjective, particularly when recognizing positive responding/response shift. It was easier to identify where participants had rated themselves substantially more positively than expected for functional ability items than subjective well-being questions. This may have biased the identification of response shift toward health measures.

Conclusions

Response issues were found for each measure, which may bias scores obtained and any allocation decisions based on them. The impact of response shift should also be considered, and methods for controlling for this should be examined. These findings could be useful to the owners of the included measures. Alterations to wording and layout could be made where confusion or interpretation issues were observed and the comprehensiveness of the measures could be improved for older adults, either through permanent additions or by using bolt-ons in older populations. Findings are also useful to guide the development of new measures aimed at or relevant to older adults.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2022.04.1735.

Article and Author Information

Accepted for Publication: April 24, 2022

Published Online: June 24, 2022

doi: https://doi.org/10.1016/j.jval.2022.04.1735

Author Affiliations: School of Health and Related Research, University of Sheffield, Sheffield, England, UK (Penton, Young); Centre for Regional Economic and Social Research, Advanced Wellbeing Research Centre, Sheffield, England, UK (Dayson); Institute of Health Research, South Cloisters, University of Exeter, Exeter, England, UK (Hulme).

Correspondence: Hannah Penton, PhD, Open Health Group, Marten Meesweg 107, Rotterdam, 3068 AV, The Netherlands. Email: hannahpenton@openhealthgroup.com

Author Contributions: *Concept and design:* Penton, Dayson, Hulme, Young
Acquisition of data: Penton
Analysis and interpretation of data: Penton, Dayson, Hulme, Young
Drafting of manuscript: Penton, Dayson, Hulme, Young
Critical review of paper for intellectual content: Penton, Dayson, Hulme, Young
Obtaining funding: Dayson, Hulme, Young
Supervision: Dayson, Hulme, Young

Conflict of Interest Disclosures: The authors reported no conflicts of interest. The views expressed in this publication are those of the authors and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

Funding/Support: This study was funded by the National Institute for Health Research Yorkshire and Humber Applied Research Collaboration.

Role of the Funder/Sponsor: The funder had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the articles; or in the decision to submit it for publication.

REFERENCES

1. Brazier J, Tsuchiya A. Improving cross-sector comparisons: going beyond the health-related QALY. *Appl Health Econ Health Policy*. 2015;13(6):557–565.
2. Milte CM, Walker R, Luszcz MA, Lancsar E, Kaambwa B, Ratcliffe J. How important is health status in defining quality of life for older people? An exploratory study of the views of older South Australians. *Appl Health Econ Health Policy*. 2014;12(1):73–84.

3. Grewal I, Lewis J, Flynn T, Brown J, Bond J, Coast J. Developing attributes for a generic quality of life measure for older people: preferences or capabilities? *Soc Sci Med*. 2006;62(8):1891–1901.
4. Guide to the methods of technology appraisal 2013:1-93. National Institute for Health and Care Excellence. https://www.nice.org.uk/process/pmg9/chapter/foreword. Accessed June 20, 2022.
5. Guideline for economic evaluations in healthcare. Zorginstituut Nederlands. https://tools.ispor.org/PEguidelines/source/Netherlands_Guideline_for_economic_evaluations_in_healthcare.pdf. Accessed September 5, 2021.
6. Ware J, Kosinski MA, Dewey J. *How to Score Version 2 of the SF-36® Health Survey*. Lincoln, RI: Quality Metric Incorporated; 2000.
7. Ware Jr JE, Kosinski M, Turner-Bowker DM, Gandek B. *How to Score Version 2 of the SF-12 Health Survey (With a Supplement Documenting Version 1)*. Lincoln, RI: Quality Metric Incorporated; 2002.
8. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care*. 2004;42(9):851–859.
9. Guidelines for the economic evaluation of health technologies: Canada. CADTH. https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-0. Accessed September 5, 2021.
10. Ratcliffe J, Lancsar E, Flint T, et al. Does one size fit all? Assessing the preferences of older and younger people for attributes of quality of life. *Qual Life Res*. 2017;26(2):299–309.
11. Makai P, Brouwer WB, Koopmanschap MA, Stolk EA, Nieboer AP. Quality of life instruments for economic evaluations in health and social care for older people: a systematic review. *Soc Sci Med*. 2014;102:83–93.
12. Beard JR, Officer A, de Carvalho IA, et al. The World report on ageing and health: a policy framework for healthy ageing. *Lancet*. 2016;387(10033):2145–2154.
13. van Leeuwen KM, Jansen AP, Muntinga ME, et al. Exploration of the content validity and feasibility of the EQ-5D-3L, ICECAP-O and ASCOT in older adults. *BMC Health Serv Res*. 2015;15(1):201.
14. The social care guidance manual. National Institute for Health and Care Excellence. https://pubmed.ncbi.nlm.nih.gov/27905705/. Accessed June 16, 2018.
15. Hicks S, Tinkler L, Allin P. Measuring subjective well-being and its potential role in policy: perspectives for the UK Office for National Statistics. *Soc Indic Res*. 2013;114:73–86.
16. Tennant R, Hiller L, Fishwick R, et al. The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): development and UK validation. *Health Qual Life Outcomes*. 2007;5:63.
17. Health survey for England. Health and Social Care Information Centre, Department of Health. https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england; 2014. Accessed April 4, 2019.
18. People keeping well in their community. Sheffield Clinical Commissioning Group. https://democracy.sheffield.gov.uk/documents/s38335/People%2520Keeping%2520Well%2520-%2520Briefing%2520Note.pdf; 2015. Accessed April 14, 2019.
19. *Care home resident interviews: phase one implementation - care homes not in Vanguard*. Healthwatch Wakefield; 2016.
20. Black SV, Cooper R, Martin KR, Brage S, Kuh D, Stafford M. Physical activity and mental well-being in a cohort aged 60-64 years. *Am J Prev Med*. 2015;49(2):172–180.
21. Cosmin taxonomy of measurement properties. COSMIN Group. https://www.cosmin.nl/tools/cosmin-taxonomy-measurement-properties/. Accessed December 27, 2018.
22. Penton H. *An Investigation Into the Psychometric Performance of Existing Measures of Health, Quality of Life and Wellbeing in Older Adults*. Sheffield, United Kingdom: University of Sheffield; 2019.
23. Rothrock NE, Kaiser KA, Cella D. Developing a valid patient-reported outcome measure. *Clin Pharmacol Ther*. 2011;90(5):737–742.
24. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity–establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1–eliciting concepts for a new PRO instrument. *Value Health*. 2011;14(8):967–977.
25. Collins D. Pretesting survey instruments: an overview of cognitive methods. *Qual Life Res*. 2003;12(3):229–238.
26. Jobe J. The application of cognitive methods to the design of health survey questionnaires. *Am J Epidemiol*. 1990;132(4):824.
27. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res*. 2009;18(9):1263–1278.
28. Knafl K, Deatrick J, Gallo A, et al. The analysis and interpretation of cognitive interviews for instrument development. *Res Nurs Health*. 2007;30(2):224–234.
29. 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.
30. Heaven A, Brown L, Foster M, Clegg A. Keeping it credible in cohort multiple Randomised Controlled Trials: the Community Ageing Research 75+ (CARE 75+) study model of patient and public involvement and engagement. *Res Involv Engagem*. 2016;2:30.
31. Davis DHJ, Creavin ST, Yip JLY, Noel-Storr AH, Brayne C, Cullum S. Montreal cognitive assessment for the diagnosis of Alzheimer's disease and other dementias. *Cochrane Database Syst Rev*. 2015;2015(10):CD010775.
32. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–1736.
33. *NVivo qualitative data analysis software*. QSR International Pty Ltd. Version 10; 2012. Version 11, 2013. https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home. Accessed June 20, 2022.
34. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77–101.
35. Tourangeau R, Rips L, Rasinki K. *The Psychology of Survey Response*. Cambridge, United Kingdom: The Cambridge University Press; 2000.
36. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):M146–M156.
37. Bulamu NB, Kaambwa B, Ratcliffe J. A systematic review of instruments for measuring outcomes in economic evaluation within aged care. *Health Qual Life Outcomes*. 2015;13:179.
38. Moser DK, Heo S, Lee KS, et al. 'It could be worse... lot's worse!' Why health-related quality of life is better in older compared with younger individuals with heart failure. *Age Ageing*. 2013;42(5):626–632.
39. Engel L, Bucholc J, Mihalopoulos C, et al. A qualitative exploration of the content and face validity of preference-based measures within the context of dementia. *Health Qual Life Outcomes*. 2020;18(1):178.
40. Ernstsson O, Burström K, Heintz E, Mølsted Alvesson H. Reporting and valuing one's own health: a think aloud study using EQ-5D-5L, EQ VAS and a time trade-off question among patients with a chronic condition. *Health Qual Life Outcomes*. 2020;18(1):388.
41. Mallinson S. Listening to respondents: a qualitative assessment of the Short-Form 36 health status questionnaire. *Soc Sci Med*. 2002;54(1):11–21.
42. Hulme C, Long AF, Kneafsey R, Reid G. Using the EQ-5D to assess health-related quality of life in older people. *Age Ageing*. 2004;33(5):504–507.
43. Ubel P, Jankovic A, Smith D, Langa K, Fagerlin A. What is perfect health to an 85-year-old?: evidence for scale recalibration in subjective health ratings. *Med Care*. 2005;43(10):1054–1057.
44. Ralph K, Palmer K, Olney J. Subjective well-being: a qualitative investigation of subjective well-being questions. office of national statistics: a working paper for the Technical Advisory Group. https://gss.civilservice.gov.uk/wp-content/uploads/2018/10/Subjective-well-being-a-qualitative-investigation-of-subjective-well-being-questions.pdf. Accessed March 28, 2021.
45. Finch AP, Brazier JE, Mukuria C, Bjorner JB. An exploratory study on using principal-component analysis and confirmatory factor analysis to identify bolt-on dimensions: the EQ-5D case study. *Value Health*. 2017;20(10):1362–1375.
46. Finch AP, Brazier JE, Mukuria C. Selecting bolt-on dimensions for the EQ-5D: examining their contribution to health-related quality of life. *Value Health*. 2019;22(1):50–61.
47. Finch A, Brazier J, Mukuria C. Selecting bolt-on dimensions for the EQ-5D: testing the impact of hearing, sleep, cognition, energy and relationships on preferences using pairwise choices. *Med Decis Making*. 2021;41(1):89–99.
48. Netten A, Burge P, Malley J, et al. Outcomes of social care for adults: developing a preference-weighted measure. *Health Technol Assess*. 2012;16(16):1–166.