



ELSEVIER

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Well-being of Older People (WOOP): Quantitative validation of a new outcome measure for use in economic evaluations

Mariska Q.N. Hackert^{a,*}, Job van Exel^{a,b}, Werner B.F. Brouwer^{a,b}

^a Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Burgemeester Oudlaan 50, P.O. Box 1738, 3000 DR, Rotterdam, the Netherlands

^b Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands

ARTICLE INFO

Keywords:

Well-being
Measurement
Psychometric properties
Older people
The Netherlands

ABSTRACT

There is a need for comprehensive measures to evaluate the benefits of health and social care services for older people. The newly developed Well-being of Older People measure (WOOP) aims to capture all aspects that older people find important to their well-being. This study explores the validity and test-retest reliability of the WOOP. Between December 2017 and January 2018, an online survey was used to retrieve data from 1113 people aged 65 years and older in the Netherlands. Regression analyses on Cantril's Ladder scores were conducted to explore the relative importance of the items of the WOOP. Dimensionality was checked using exploratory factor analysis. Convergent and discriminant validity were investigated by relating the WOOP to several measures of health and well-being. Test-retest reliability was examined using data from 269 respondents that participated in a second online survey, distributed one week after the first. The items of the WOOP were significantly associated with Cantril's Ladder scores. When regressed simultaneously this was still true for all but the 'social contacts' item and one level of the 'acceptance and resilience' item. The dimensionality analysis revealed three factors, of which two included items of the WOOP and the EQ-5D-5L and the third only items of the WOOP. The WOOP correlated moderately to highly with physical health, and (very) highly with (mental) health and well-being measures. The test-retest reliability in terms of ICC was high, whereas the kappa for the items was fair to good, except for two items. Overall, the WOOP seems to capture aspects relevant to the well-being of older people adequately, and the results of first validity and reliability tests were satisfactory. Before the WOOP can be used in economic evaluations, further validation in a variety of health and social care settings is recommended, and utility weights need to be determined.

1. Introduction

The efficient and fair allocation of scarce health and social care resources is an important challenge in many countries. Economic evaluations can inform allocation decisions by comparing the costs and benefits of alternative treatment options. That way, resources can be directed towards those services that provide the most value for money. Such decisions are increasingly required, due to technological advances and also because of the ageing of populations across the globe, and the expected associated increase in demand of health and social care services (WHO, 2015). Traditionally, economic evaluations in the form of a cost-utility analysis (CUA) use quality-adjusted life-years (QALYs) to assess care services' benefits. QALYs comprise gains in life years and health-related quality of life (HRQoL). Using a multi-attribute utility measure such as the EuroQol five-dimensional questionnaire (EQ-5D), HRQoL is typically measured and valued on a utility scale that is

anchored on the state perfect health (with value 1) and the state dead (with value 0), with health states worse than dead (with a value < 0) also being possible (Drummond et al., 2015; Makai et al., 2014). However, QALYs may fall short in capturing all benefits of health and social care services for older people, as these typically aim to improve well-being beyond health, in particular in long-term care and social care. Hence, it has been argued that well-being measures may be more appropriate measures of benefit than QALYs in this context (Bulamu et al., 2015; Drummond et al., 2009; Makai et al., 2014). One could view this as an expansion of the evaluative scope of economic evaluations, moving beyond HRQoL to general quality of life. Extensive overviews and theoretical debates exist on the exact nature and definition of general quality of life, or well-being (Makai et al., 2014; Van Leeuwen et al., 2019). Here, we draw on the findings of a recent study that investigated what (older) people themselves consider to be important for their well-being (Hackert et al., 2019b). In recent years,

* Corresponding author.

E-mail addresses: hackert@eshpm.eur.nl (M.Q.N. Hackert), vanexel@eshpm.eur.nl (J. van Exel), brouwer@eshpm.eur.nl (W.B.F. Brouwer).

<https://doi.org/10.1016/j.socscimed.2020.113109>

Received in revised form 30 April 2020; Accepted 3 June 2020

Available online 13 June 2020

0277-9536/ © 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

several and diverse well-being measures were developed to enable a more comprehensive benefit assessment in evaluations of care services for older people, aiming to capture benefits more broadly than only health (for an overview, see Bulamu et al., 2015; Cleland et al., 2019; Makai et al., 2014).

Most of the existing well-being measures may, however, not be directly suitable for use in economic evaluations. For example, some lack content validity, as they are based on expert opinion regarding relevant well-being domains, instead of older people's own views on what is important to their well-being (Bowling et al., 2013; Bowling and Stenner, 2011). Moreover, several measures contain too many items to reasonably allow self-completion by the target population (Bowling et al., 2013), and many do not have utility scores to reflect the relative importance of each of their items to older people's overall well-being (Makai et al., 2014). Two measures that do not have these shortcomings and therefore could be useful in the context of economic evaluations, are the Adult Social Care Outcomes Toolkit (ASCOT) (Netten et al., 2012) and the ICEpop CAPability measure for Older people (ICECAP-O) (Coast et al., 2008; Grewal et al., 2006). Both measures have demonstrated to have good psychometric properties (Hackert et al., 2017, 2019a; Proud et al., 2019; Van Leeuwen et al., 2015). However, the evaluative scope of the measures is different. The ICECAP-O aims to measure capability well-being (i.e., what people *can do or be*), while the ASCOT aims to measure social care-related outcomes and therefore appears to have a narrower, more focused scope of benefit assessment and application.

Although the ASCOT and the ICECAP-O seem to be the most promising well-being measures currently available for use in economic evaluations of health and social care services for older people (Bulamu et al., 2015; Makai et al., 2014), it is still an open question whether these measures adequately cover all domains that are relevant to older people's well-being. Most importantly, domains such as health, finances and resilience are not directly captured in these measures (Coast et al., 2008; Grewal et al., 2006; Netten et al., 2012). It has been suggested that the ICECAP-O may capture these domains indirectly through the items that are directly measured (Grewal et al., 2006), but it is unclear whether this ensures a full capturing and valuing of these domains. For example, some studies showed that physical health may not be fully captured in both the ICECAP-O and the ASCOT (Davis et al., 2013; Hackert et al., 2017, 2019a; Van Leeuwen et al., 2015). Because less attention for well-being domains relevant to older people in the evaluation of health and social care services and the subsequent decision-making process about these services seems undesirable, the development and exploration of more comprehensive well-being measures remains warranted.

In this context, recently the Well-being of Older People measure (WOOP) was developed that aims to capture and value a comprehensive set of domains relevant to the well-being of older people. To meet this objective, the WOOP was grounded in the outcomes of a Q-methodology study examining the heterogeneity in views of older people in the Netherlands on what constitutes well-being (Hackert et al., 2019b). Although all individuals may be assumed to aim to optimise their well-being, they may nonetheless differ in their views on what constitutes well-being. Integrating this heterogeneity, the development resulted in a nine-item draft version of the WOOP (developed in Dutch; for the English translation, see Appendix A). Five of the nine WOOP items (i.e., 'physical health', 'mental health', 'social contacts', 'receive support' and 'acceptance and resilience') emerged as important domains in two or more of the five identified views in the Q-methodology study, while the other four items (i.e., 'feeling useful', 'independence', 'making ends meet' and 'living situation') were each considered important in a specific view on what constitutes well-being. A first qualitative validation study, conducted in parallel to this study, indicated that the WOOP indeed captures the domains relevant to the well-being of older people, as identified by themselves, satisfactorily. The current quantitative validation study presents the outcomes of diverse validity and test-

retest reliability tests of the draft version of the WOOP, hence testing the psychometric properties of this newly developed well-being measure for older people.

2. Data and methodology

2.1. Sampling strategy

Using an online survey, data were gathered in two stages between December 2017 and January 2018. A sampling agency recruited 1113 respondents aged 65 years and older in the Netherlands, approximately representative of this population in terms of age, sex and level of education. In the first stage (t0), respondents completed a questionnaire asking about their well-being, health and several background characteristics. One week after sending out the first questionnaire (t1), all respondents were approached with an invitation to participate in a second online survey. After a reminder, 269 respondents agreed to participate in this second stage and completed the second questionnaire. This sub-sample was used to check the test-retest reliability of the WOOP. No missing data were reported as all questions were mandatory. Participation in both stages of the study was voluntary, based on informed consent and could be terminated at any point. Information about the non-response to the first questionnaire and the number of people who started but did not finish the first and second questionnaire was not made available by the sampling agency.

2.2. Well-being of Older People measure (WOOP)

The WOOP aims to capture and value the well-being of older people through nine items, each covering a relevant well-being domain: 'physical health', 'mental health', 'social contacts', 'receive support', 'acceptance and resilience', 'feeling useful', 'independence', 'making ends meet' and 'living situation'. The WOOP focusses on the functionings of older people (i.e., what people *do or are*) rather than their capabilities, as the ICECAP-O does, because functionings are more likely interpreted in a uniform way by respondents (Al-Janabi, 2018; Al-Janabi et al., 2013). This is also in line with commonly applied HRQoL-measures, that also focus on functionings (Makai et al., 2014). For all items, response options were defined on a 5-point scale, representing an excellent (5), good (4), fair (3), poor (2) or bad (1) level of functioning on the specific item. Summing the item scores therefore leads to a well-being score with a theoretical range from 9 to 45, with higher scores indicating higher well-being.

Respondents completed the draft version of the WOOP at t0 and t1 (see Appendix A). In the second questionnaire, they were also asked to report whether between t0 and t1 an important life event had occurred that affected their well-being (and, hence, WOOP score) and, if so, to describe this event.

2.3. Background characteristics

Information was gathered on respondents' age, sex and marital status. The highest level of education attained was categorised as low (primary, secondary or lower vocational education), middle (middle vocational education), or high education (higher vocational or academic education). Material deprivation was defined as not being able to pay for at least one of the following four expenses: (i) heating the house, (ii) membership of a (sport)club, (iii) visiting family and friends, and (iv) paying a €1000 on unforeseen expenses without being in debt or taking a loan (SCP, 2014). The number of health problems was measured using the thirteen item Comorbidity Index (CI) (Sangha et al., 2003). Respondents' health care utilisation was split into informal and formal care. Informal care comprised family or friends providing care and support, whereas formal care concerned the use of at least one of the following services: (i) home help, (ii) home care, (iii) day-centre care, (iv) living in supported housing, or (v) living in a nursing home.

2.4. Health

Respondents' health was measured using the five-level EQ-5D (EQ-5D-5L), EuroQol Visual Analogue Scale (EQ-VAS), Barthel Index, and the Physical health and Mental health Component Score of the 12-item Short-Form Health Survey (SF-12 PCS and SF-12 MCS). The EQ-5D-5L (Herdman et al., 2011; Janssen et al., 2013) measures general HRQoL using five items: 'mobility', 'self-care', 'usual activities', 'pain and discomfort', and 'anxiety and depression'. On each item, respondents can indicate the severity of their health problems using five response levels. A utility score, ranging between -0.446 and 1 , was calculated using utility weights for the Netherlands (Versteegh et al., 2016). The EQ-VAS (Herdman et al., 2011) comprises a vertical, visual analogue scale on which people can indicate how they rate their health today. The scale ranges from 0 (worst imaginable health) to 100 (best imaginable health). The Barthel Index (Post et al., 1995) examines respondents' ability to perform activities of daily living (ADL), using seven items referring to specific activities (e.g., eating, bathing and dressing) and three items referring to mobility. Applying the scoring system, a score ranging between 0 and 20 was computed, with higher scores indicating better physical health. The SF-12 PCS and SF-12 MCS (Ware et al., 1995, 1996) comprise twelve items to measure the impact of respectively physical and mental health limitations on respondents' functioning. Using a standard scoring algorithm, scores were generated with a mean of 50 and an SD of 10 , with higher scores indicating better functioning.

2.5. Well-being

Respondents' well-being was measured using the ICECAP-O, the Satisfaction with Life Scale (SWLS) and Cantril's Ladder. The ICECAP-O (Grewal et al., 2006) uses five items with each four response levels to capture capability well-being among older people. A 0 to 1 scale was created using preference-based weights (Coast et al., 2008), with higher scores indicating higher well-being. The SWLS (Pavot and Diener, 1993) captures people's judgement of their life satisfaction. It consists of five items with a 7-point response scale, leading to a score ranging between 5 and 35 , with higher scores indicating greater life satisfaction. Cantril's Ladder (Cantril, 1965) comprises a vertical, visual analogue scale on which people can rate their life. An 11-rung ladder was used ranging from 0 (worst possible life) to 10 (best possible life).

2.6. Ethics

The study protocol was approved by the Medical Ethics Review Board of the Erasmus Medical Centre, case number: MEC-2017-571. Participation in the study was voluntary and could be terminated at any point; incomplete questionnaires were discarded. By signing up to the panel of the sampling agency and submitting their data after completing the full questionnaire, respondents gave informed consent to the use of their responses for the purpose of this study.

2.7. Analytic strategy

Data were analysed using R (R Foundation for Statistical Computing, Vienna, Austria). To assess the association between the WOOP items, Spearman correlations were calculated. The following ranges were used to classify correlation strength: less than 0.10 , trivial; 0.10 – 0.29 , small; 0.30 – 0.49 , moderate; 0.50 – 0.69 , high; 0.70 – 0.89 , very high; and 0.90 or higher, (nearly) perfect (Hopkins, 2002). To examine the extent to which higher scores on the WOOP items are related to better overall well-being, and check for potential overlap between the items, we consecutively used the WOOP items separately and then simultaneously to predict respondents' scores on Cantril's Ladder, while controlling for age, sex and level of education. Because lower levels of well-being were rarely reported, response levels were merged

to ensure that, for all items, all (merged) levels represented at least 10% of the observations. To ensure comparability between items, having a bad (1) to fair (3) level of well-being was set as the reference category, except for the item 'mental health'; because of the skewed response to this item, the reference category was having a bad (1) to good (4) level of well-being.

Using exploratory factor analysis (EFA) in the package "psych" (Revelle, 2015), dimensionality was tested by examining the potential overlap in underlying constructs (i.e., factors) between the items of the WOOP and the EQ-5D-5L. The Bartlett test of sphericity, the Kaiser-Meyer Olkin measure of sampling adequacy and the multicollinearity test from the package "usdm" (Naimi, 2015) confirmed the suitability of this method. The Kaiser criterium, the scree plot, and the interpretability of the models were used to select the appropriate number of factors. Maximum likelihood extraction was used, next to principal axis factoring, to control for the possible violation of multivariate normality. Using the package "GPArotation" (Bernaards and Jennrich, 2015), oblimin rotation was applied to allow factors to be correlated. For each item, only the highest factor loading was reported, which needed to be $\geq |0.40|$ to be reliable for interpretation (Hair et al., 1998). Each factor was defined by determining the common denominator among the items that loaded on it.

Convergent validity was tested by relating the WOOP and its items to measures of health (EQ-5D-5L, EQ-VAS, Barthel Index, SF-12 PCS and SF-12 MCS) and well-being (ICECAP-O, SWLS and Cantril's Ladder). Discriminant validity was investigated by assessing the ability of the WOOP to differentiate between older people based on their background characteristics, and levels of health and well-being (i.e., above versus below or equal to the study sample average). For this purpose, t-tests and One-Way Analyses of Variance were executed. When unequal variances were detected using the Levene test from the package "car" (Fox et al., 2019), the Welch test was applied.

The test-retest reliability of the WOOP was assessed in a subgroup of our sample ($N = 269$) with Spearman correlations, percentages of complete agreement, κ coefficients and intraclass correlation coefficients (ICCs). Percentages of complete agreement concern the share of respondents with exactly the same score on the WOOP at t_0 and t_1 . Considering that these similarities may have emerged from chance alone, chance-corrected agreements in the form of κ coefficients were calculated using the package "psych" (Revelle, 2015). The guidelines of Fleiss (Streiner and Norman, 2008) were used to classify their strength: less than or equal to 0.40 , poor; 0.41 – 0.75 , fair to good; 0.76 – 1.00 , excellent. In addition, ICCs were computed using the package "irr" (Gamer et al., 2019), as partial agreement can also be considered relevant for a measure based on scaled responses (e.g., the minor difference between a score of 40 at t_0 and 41 at t_1). Because the focus was on the difference in scores for single respondents on t_0 and t_1 , ICC(1,1) was applied using the One-Way Random model, with values ≥ 0.70 considered as being acceptable (Hoefman et al., 2011; Shrout and Fleiss, 1979; Streiner and Norman, 2008).

3. Results

3.1. Descriptive characteristics

The descriptive statistics of the study samples at t_0 and t_1 are displayed in Table 1. At t_0 , respondents on average were 74 years old, with 19% being 80 years or older. Both sexes were equally represented, and 65% of the respondents were married or living together. Respectively, 40%, 33% and 27% had attained a low, middle and high education. Almost all were retired (89%), and 28% indicated material deprivation in at least one of the indicated expense categories. 61% reported two or more health problems on the Comorbidity Index, 10% received informal care, and 29% received at least one type of formal care. Respondents who also participated in the second survey at t_1 ($N = 269$) were significantly younger, more often male, retired and less often

Table 1
Descriptive statistics of the study sample (N = 1113) and test-retest reliability sample (N = 269); and the discriminatory power of the WOOP (N = 1113).

Variable		Descriptive statistics										Discriminatory power (t0)	
		Study sample (t0)					Test-retest reliability sample (t1)						
		%	Mean	SD	Min.	Max.	%	Mean	SD	Min.	Max.		Mean WOOP score
Age ^a	< 80 years old	81.31	73.77	5.88	65.00	100.00	83.64	72.86	5.79	65.00	94.00	36.63	
	≥80 years old	18.69					16.36					36.32	
Sex	Female	50.31					41.26					36.11	***
	Male	49.69					58.74					37.05	
Marital status	Married or living together	64.51					66.91					37.01	***
	Never married	4.22					2.23					36.47	
	Divorced	11.41					9.29					35.55	
	Widowed	19.86					21.56					35.78	
Education	Low	40.16					39.03					36.35	
	Middle	32.79					34.20					36.51	
	High	27.04					26.77					36.97	
Material deprivation ^a	No	72.42	0.41	0.77	0.00	4.00	75.09	0.38	0.77	0.00	4.00	37.58	***
	Yes, in at least one expense category	27.58					24.91					33.94	
Comorbidity ^a	No	16.62	2.29	1.78	0.00	10.00	15.24	2.49	1.86	0.00	10.00	39.04	***
	One	22.10					17.10					38.14	
	Two or more	61.28					67.66					35.34	
Informal care	No	90.30					89.96					37.05	***
	Yes	9.70					10.04					32.13	
Formal care ^a	No	71.34	0.37	0.66	0.00	5.00	73.98	0.36	0.67	0.00	3.00	37.07	***
	Yes, use at least one type	28.66					26.02					35.33	
EQ-5D-5L	≤0.81	35.67	0.81	0.19	-0.22	1.00						33.10	***
	> 0.81	64.33										38.50	
EQ-VAS	≤73.73	42.59	73.73	18.85	3.00	100.00						33.92	***
	> 73.73	57.41										38.54	
Barthel Index	≤19.32	27.31	19.32	1.59	4.00	20.00						34.18	***
	> 19.32	72.69										37.47	
SF-12 PCS	≤46.17	38.90	46.17	10.38	13.85	63.07						33.88	***
	> 46.17	61.10										38.29	
SF-12 MCS	≤52.81	36.48	52.81	7.95	20.53	67.18						33.67	***
	> 52.81	63.52										38.24	
ICECAP-O	≤0.85	34.05	0.85	0.13	0.08	1.00						32.45	***
	> 0.85	65.95										38.70	
SWLS	≤27.05	42.59	27.05	5.96	7.00	35.00						33.82	***
	> 27.05	57.41										38.62	
Cantril's Ladder	≤7.12	50.31	7.12	1.46	0.00	10.00						34.17	***
	> 7.12	49.69										39.01	
WOOP			36.57	4.66	20.00	45.00		36.53	4.91	22.00	45.00		

WOOP, Well-being of Older People measure; EQ-5D-5L, five-levels EuroQol five-dimensional questionnaire; EQ-VAS, EuroQol Visual Analogue Scale; SF-12 PCS, 12-item Short-Form Health Survey Physical health Component Score; SF-12 MCS, 12-item Short-Form Health Survey Mental health Component Score; ICECAP-O, ICEpop CAPability measure for Older people; SWLS, Satisfaction with Life Scale.

*** = p < .001, ** = p < .005, * = p < .05.

^a Both the index score and the categorised scores are displayed.

never married, and had more health problems than those who only participated in the first survey.

At t0, respondents had a fairly good health, with mean scores of 0.81 on the EQ-5D-5L and 73.7 on the EQ-VAS. Moreover, they reported high independence in ADL, with a mean Barthel Index score of 19. On average, they scored 46 and 53 on the SF-12 PCS and SF-12 MCS, respectively, indicating fair levels of physical and mental health.

Also, respondents' well-being, measured by the ICECAP-O, was high, with a mean score of 0.85. Average values on other well-being measures were slightly lower, namely 27 on the SWLS and 7.1 on Cantril's Ladder. On a theoretical range between 9 and 45, respondents' mean score on the WOOP was 37. For each item a majority of respondents scored good or excellent levels of well-being (see Fig. 1). Remarkably, 89% of respondents reported no 'mental health' problems. In contrast, fair to bad levels of well-being were most often mentioned on the items 'physical health', 'feeling useful' and 'making ends meet'.

Table 2 presents the Spearman correlations of the items of the WOOP. Overall, the correlations show a diverse picture, most being moderate or small. Only 'social contacts' correlated highly (0.53) with 'receive support'. 'Acceptance and resilience' had moderate correlations with all but one of the other items ('mental health'). 'Mental health' and 'making ends meet' had small to even trivial correlations with most

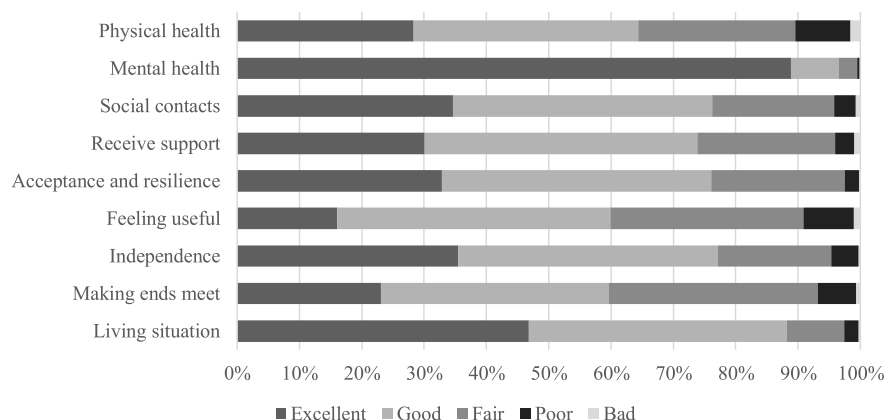
items.

Table 3 displays the results of the regression analyses of Cantril's Ladder on each WOOP item separately (models I – IX) and all simultaneously (model X). For all items, a higher score was associated with a significantly higher score on Cantril's Ladder. When all items were added simultaneously (model X), both levels of the item 'social contacts' became insignificant, as well as the middle level of the 'acceptance and resilience' item.

We repeated these analyses with SWLS as dependent variable. The findings for models I to IX were very similar; all item levels were significantly related to SWLS, and higher-level scores were associated with higher SWLS scores. Except for some levels no longer being significantly different from each other, the most important difference in model X with SWLS as dependent variable was that the excellent level of 'social contacts' and the good level of 'acceptance and resilience' were now statistically significant, and (ii) the excellent level of 'independence' had a higher coefficient than the good level, as one would expect.

3.2. Dimensionality

Table 4 presents the results of the EFA to test the overlap in underlying constructs between the items of the EQ-5D-5L and the WOOP.



WOOP, Well-being of Older People measure

Fig. 1. Response distribution on the items of the WOOP (N = 1113).

In the three-factor model, the first and the third factor contained items of both measures. The first factor included the WOOP item ‘physical health’ and the EQ-5D-5L items ‘mobility’, ‘self-care’, ‘usual activities’ and ‘pain and discomfort’. The third factor encompassed the WOOP item ‘mental health’ and the EQ-5D-5L item ‘anxiety and depression’. The second factor contained only items of the WOOP, including ‘social contacts’, ‘receive support’, ‘acceptance and resilience’, ‘feeling useful’ and ‘living situation’. The three factors explained 23%, 14% and 8% of the total variance, respectively. The WOOP items ‘independence’ and ‘making ends meet’ both had a high uniqueness (0.58 and 0.80) and did not display factor loadings above the threshold value of 0.40 on any of the three factors.

3.3. Convergent validity

Table 5 presents the Spearman correlations of (the items of) the WOOP with several measures of health and well-being. The WOOP correlated highly with the general health measures EQ-5D-5L and EQ-VAS, and the mental health measure SF-12 MCS. In contrast, mixed evidence was found in relation to physical health, as our measure and its item ‘physical health’ correlated (very) highly with the SF-12 PCS but only moderately with the Barthel Index. The item ‘mental health’ had mostly small correlations with the other health measures, with the correlation with the SF-12 MCS being moderate.

In addition, the WOOP correlated (very) highly with the well-being measures ICECAP-O, SWLS and Cantril’s Ladder. Except for ‘mental health’, all items of the WOOP correlated at least moderately with these measures as well.

Table 2 Spearman correlations of the items of the WOOP (N = 1113).

	Mental health	Social contacts	Receive support	Acceptance and resilience	Feeling useful	Independence	Making ends meet	Living situation
Physical health	0.21***	0.25***	0.14***	0.36***	0.36***	0.39***	0.24***	0.13***
Mental health		0.25***	0.16***	0.28***	0.21***	0.23***	0.15***	0.07*
Social contacts			0.53***	0.47***	0.45***	0.39***	0.23***	0.35***
Receive support				0.38***	0.36***	0.27***	0.25***	0.34***
Acceptance and resilience					0.47***	0.49***	0.32***	0.30***
Feeling useful						0.47***	0.28***	0.26***
Independence							0.31***	0.30***
Making ends meet								0.35***

WOOP, Well-being of Older People measure.

*** = p < .001, ** = p < .005, * = p < .05.

3.4. Discriminant validity

Table 1 displays the discriminatory power of the WOOP. The WOOP differentiated between respondents based on their background characteristics (sex, marital status, material deprivation, comorbidity, and utilisation of informal and formal care) in the expected direction. Also, a higher score on the WOOP was found for those who reported an above average level of health (measured by the EQ-5D-5L, EQ-VAS, Barthel Index, SF-12 PCS and MCS) or well-being (measured by the ICECAP-O, SWLS and Cantril’s Ladder).

3.5. Test-retest reliability

Table 6 shows the test-retest reliability statistics of (the items of) the WOOP. The partial agreement measured by the ICC was high (0.89), but complete agreement (16.7%) and chance-corrected agreement were low (κ = 0.11). The WOOP items’ complete agreement was between 55% and 87%. ‘Physical health’ and ‘making ends meet’ displayed good κ and ICC values, whereas ‘mental health’, ‘social contacts’, ‘feeling useful’, ‘independence’ and ‘living situation’ showed fair to good κ values that approximated the ICC threshold value. ‘Receive support’ and ‘acceptance and resilience’ had poor κ values (0.34 and 0.39) and the lowest ICC scores (0.60 and 0.63).

Because 30 respondents reported that a life event occurred that had influenced their well-being score between t0 and t1, test-retest reliability was also analysed disregarding these respondents. The results of this analysis were very similar. Of the 30 respondents that reported a significant life event between t0 and t1, 18 reported to have experienced a negative life event related to their physical health, and 11 related to their social life. Regression analyses indicated that experiencing

Table 3
Linear regression analyses of Cantril's Ladder on the items of the WOOP (N = 1113).^a

	Model I	Model II	Model III	Model IV	Model V	Model VI	Model VII	Model VIII	Model IX	Model X
Physical health										
Bad - fair	Ref.									Ref.
Good	0.89***									0.35***
Excellent	1.16***									0.36***
Mental health										
Bad - good		Ref.								Ref.
Excellent		1.16***								0.41***
Social contacts										
Bad - fair			Ref.							Ref.
Good			0.95***							0.16
Excellent			1.46***							0.17
Receive support										
Bad - fair				Ref.						Ref.
Good				1.02***						0.26*
Excellent				1.27***						0.30*
Acceptance and resilience										
Bad - fair					Ref.					Ref.
Good					1.07***					0.16
Excellent					1.64***					0.36**
Feeling useful										
Bad - fair						Ref.				Ref.
Good						1.16***				0.44***
Excellent						1.55***				0.52***
Independence										
Bad - fair							Ref.			Ref.
Good							1.23***			0.48***
Excellent							1.59***			0.38**
Making ends meet										
Bad - fair								Ref.		Ref.
Good								1.03***		0.42***
Excellent								1.27***		0.50***
Living situation										
Bad - fair									Ref.	Ref.
Good									1.11***	0.58***
Excellent									1.76***	0.84***
Constant	6.20***	5.86***	5.97***	6.05***	5.96***	6.02***	5.79***	6.33***	5.67***	4.36***
Adjusted R-Squared	0.12	0.08	0.15	0.13	0.19	0.20	0.18	0.15	0.16	0.38

WOOP, Well-being of Older People measure.

*** = p < .001, ** = p < .005, * = p < .05.

^a B reported controlled for age (ref. = 65 years old), sex (ref. = female) and education (ref. = low).

Table 4
Factor loadings for the items of the WOOP and the EQ-5D-5L (N = 1113).^a

Measure	Item	Uniqueness	Factor		
			1	2	3
WOOP	Physical health	0.35	0.77		
	Mental health	0.61			0.66
	Social contacts	0.45		0.70	
	Receive support	0.51		0.77	
	Acceptance and resilience	0.54		0.46	
	Feeling useful	0.55		0.46	
	Independence	0.58			
	Making ends meet	0.80			
	Living situation	0.75		0.51	
	EQ-5D-5L	Mobility	0.28	0.90	
	Self-care	0.64	0.58		
	Usual activities	0.30	0.81		
	Pain and discomfort	0.39	0.77		
	Anxiety and depression	0.44			0.73
Total variance explained			0.23	0.14	0.08
Correlation with Factor 2			0.37		
Correlation with Factor 3			-0.43	-0.49	

WOOP, Well-being of Older People measure; EQ-5D-5L, five-levels EuroQol five-dimensional questionnaire.

^a Only presented is the highest factor loading per item, which is also ≥ 0.40.

a negative physical health event did affect the 'physical health' item score at t1, but not respondents' overall well-being score. Experiencing a negative social life event did not affect respondents' score on the 'social contacts' item or the WOOP at t1.

4. Discussion and conclusion

4.1. Main findings

In a sample of 1113 Dutch adults aged 65 years and older, we tested the validity of the draft version of the WOOP measure, a new multi-dimensional well-being measure for older people. The correlation analysis showed that correlations between the items of the WOOP were all trivial to moderate, except for the correlation between 'social contacts' and 'receive support'. Regression analyses indicated that a higher level of well-being on all items was associated with a significantly higher score on Cantril's Ladder and SWLS, both general measures of overall well-being. However, when regressed simultaneously, not all levels of all items were significantly related to Cantril's Ladder and SWLS, in particular for the item 'social contacts'. Furthermore, the dimensionality analysis showed that two of the items of the WOOP and the five dimensions of the EQ-5D-5L related to two shared factors, whereas five items of the WOOP defined an additional factor. Convergent and discriminant validity analyses indicated (very) high correlations of the WOOP with measures of well-being (ICECAP-O, SWLS and Cantril's Ladder) and health (EQ-5D-5L, EQ-VAS). Regarding

Table 5
Spearman correlations of the WOOP with diverse measures of health (EQ-5D-5L – SF-12 MCS) and well-being (ICECAP-O – Cantril's Ladder) (N = 1113).

WOOP		EQ-5D-5L	EQ-VAS	Barthel Index	SF-12 PCS	SF-12 MCS	ICECAP-O	SWLS	Cantril's Ladder
Overall		0.61***	0.54***	0.32***	0.51***	0.52***	0.72***	0.59***	0.62***
Items	Physical health	0.75***	0.62***	0.40***	0.73***	0.29***	0.43***	0.32***	0.35***
	Mental health	0.30***	0.19***	0.15***	0.15***	0.37***	0.27***	0.26***	0.24***
	Social contacts	0.30***	0.27***	0.13***	0.22***	0.37***	0.48***	0.41***	0.37***
	Receive support	0.18***	0.19***	0.04	0.13***	0.29***	0.41***	0.34***	0.33***
	Acceptance and resilience	0.45***	0.40***	0.22***	0.35***	0.47***	0.52***	0.42***	0.44***
	Feeling useful	0.43***	0.38***	0.23***	0.36***	0.36***	0.51***	0.39***	0.46***
	Independence	0.46***	0.39***	0.30***	0.39***	0.35***	0.51***	0.37***	0.39***
	Making ends meet	0.29***	0.27***	0.17***	0.25***	0.26***	0.46***	0.42***	0.43***
	Living situation	0.17***	0.23***	0.07*	0.14***	0.23***	0.38***	0.39***	0.39***

WOOP, Well-being of Older People measure; EQ-5D-5L, five-levels EuroQol five-dimensional questionnaire; EQ-VAS, EuroQol Visual Analogue Scale; SF-12 PCS, 12-item Short-Form Health Survey Physical health Component Score; SF-12 MCS, 12-item Short-Form Health Survey Mental health Component Score; ICECAP-O, ICEpop CAPability measure for Older people; SWLS, Satisfaction with Life Scale.

*** = p < .001, ** = p < .005, * = p < .05.

physical health, the WOOP and the ‘physical health’ item of the WOOP correlated (very) highly with the SF-12 PCS, but only moderately with the Barthel Index. This mixed result may be because the SF-12 PCS and the Barthel Index differ in their measurement of physical health, as their mutual correlation was only moderate (0.43). It is also noteworthy that in our sample the Barthel Index hardly showed any variance, with almost all respondents having a very high score. Regarding mental health, the WOOP correlated highly and the ‘mental health’ item of the WOOP correlated moderately with the SF-12 MCS. This may be an artefact of the lack of variation in scores on the WOOP item in this sample, as 89% reported an excellent level of mental health, but it may also be that aspects of mental health are not fully captured through this single item (and perhaps indirectly through some of the other items that correlated moderately with SF-12 MCS, such as ‘acceptance and resilience’). Finally, the WOOP differentiated significantly in expected directions between subgroups of older people based on their background characteristics (sex, marital status, material deprivation, comorbidity, and formal and informal care) and the above-mentioned well-being and health measures.

The test-retest reliability analysis of the WOOP in a subsample of 269 respondents showed that the complete and (chance-corrected) absolute agreement (κ) was fair to good for most items, but poor for the WOOP and the items ‘receive support’ and ‘acceptance and resilience’. Partial agreement of the WOOP in terms of ICC was high, which seems a more relevant indicator as the WOOP score ranges from 9 to 45.

4.2. Comparability with previous studies

The correlations of the WOOP with measures of physical, mental and overall health (Barthel Index, SF-12 MCS, EQ-5D-5L, EQ-VAS) and

Table 6
Test-retest reliability of (the items of) the WOOP in Spearman correlations, percentages of complete agreement, kappa statistics (κ) and intraclass correlation coefficients (ICC) at t0 and t1 (N = 269).

WOOP	Spearman correlation	% complete agreement	k	95% CI κ		ICC	95% CI ICC		
				Lower bound	Upper bound		Lower bound	Upper bound	
Overall	0.87	16.73	0.11	0.07	0.16	0.89	0.86	0.91	
Items	Physical health	0.80	73.23	0.63	0.55	0.70	0.82	0.78	0.86
	Mental health	0.66	86.99	0.48	0.35	0.60	0.63	0.56	0.70
	Social contacts	0.64	62.08	0.43	0.34	0.52	0.68	0.61	0.74
	Receive support	0.59	55.39	0.34	0.25	0.43	0.60	0.52	0.67
	Acceptance and resilience	0.62	59.48	0.39	0.30	0.48	0.63	0.55	0.69
	Feeling useful	0.69	64.68	0.49	0.40	0.57	0.72	0.65	0.77
	Independence	0.69	63.20	0.45	0.36	0.53	0.68	0.62	0.74
	Making ends meet	0.84	76.21	0.66	0.58	0.73	0.84	0.81	0.88
	Living situation	0.64	67.66	0.46	0.37	0.55	0.70	0.63	0.76

WOOP, Well-being of Older People measure.

general well-being (SWLS, Cantril's Ladder) are largely similar to those of the ICECAP-O and the ASCOT reported in previous studies (Hackert et al., 2017, 2019a; Van Leeuwen et al., 2015). However, the correlation of the WOOP with the physical health measure SF-12 PCS was high, whereas the correlation of the SF-12 PCS with the ASCOT was small and with the ICECAP-O was moderate in the study by Van Leeuwen et al. (2015).

The three-factor model demonstrated that the WOOP and the EQ-5D-5L tapped into two shared factors. The factor defined by the WOOP item ‘physical health’ and the EQ-5D-5L domains ‘mobility’, ‘self-care’, ‘usual activities’ and ‘pain and discomfort’ may be interpreted to cover physical functioning. The factor defined by the WOOP item ‘mental health’ and the EQ-5D-5L domain ‘anxiety and depression’ appears to cover mental functioning. The WOOP items ‘physical health’ and ‘mental health’ displayed high loadings on the respective factors, suggesting that they have a strong relation with the underlying construct. This is partly in contrast with previously performed dimensionality analyses on the ICECAP-O and the EQ-5D-5L, which showed lower correlations for the ICECAP-O item ‘control’ (0.44, 0.58) with a factor associated with the EQ-5D-5L domains describing physical health (Davis et al., 2013; Hackert et al., 2019a). This may signal that the WOOP captures the life domain physical functioning better. However, remarkably, the WOOP item ‘independence’, which may be expected to capture a similar construct as the ICECAP-O item ‘control’, did not display a loading greater than the predefined threshold value. This raises the question whether these two items indeed measure the same underlying construct. Five items of the WOOP loaded on an additional factor that did not include items of the EQ-5D-5L, which demonstrates that the WOOP also captures outcomes beyond health.

4.3. Study limitations and strengths

Our findings should be interpreted with some caution, as generalisability is limited by the use of data from one country gathered online in a sample of relatively healthy members of a panel. The WOOP seemed sufficiently valid and reliable within this group and this context, but, obviously, our results may not be generalisable to different populations or contexts (Streiner and Norman, 2008). In this study, older people in the highest age groups, and those with low levels of well-being and (physical and mental) health were underrepresented. This likely holds even stronger for those more dependent on health and social care (e.g., living in a nursing home), who may even be missing from our sample. Hence, the validity of the WOOP in people in poor well-being states, scoring low on the WOOP items, remains under-examined. Note that harder to reach groups (e.g. those living in nursing homes, older persons with a non-western background, and those in the highest age groups) were included in the Q-study underlying the WOOP (Hackert et al., 2019b).

In addition, insight in possible selection effects was limited, as we had no access to numbers of, or data on respondents who did not accept the invitation to participate, or started but did not finish the online questionnaire. This may also relate to the burden of completing an (online) questionnaire. Also, using the internet to collect our data has consequences for our sample and data. Older people with low computer literacy may not have been reached. Moreover, those who did participate in the online questionnaire may have been less engaged than for instance when interviewed in person. This may have influenced the reliability of the study outcomes. We also note that another selection of instruments, to compare the WOOP to, could have been chosen in the context of its validation, which may have provided other insights. While this obviously relates to currently existing and already validated measures, it also needs noting that this field is currently developing fast. Future studies could also consider inclusion of measures currently under development, like the E-QALY (Mukuria et al., 2018), to compare the WOOP to (although the E-QALY is not specifically aimed at older persons). Further validation, especially in these above-mentioned groups, using other modes of administration of the WOOP, and using other instruments to compare the WOOP to, is therefore highly encouraged.

An important issue for instruments designed for self-completion, like the WOOP, is their feasibility for all members of the target population. The fact that the WOOP measure, and its items and their descriptions and response levels, were based on a Q-methodology study (Hackert et al., 2019b), in which the wording of different items was discussed with older people themselves, may have contributed to the comprehensibility of the measure. However, for people with cognitive impairments, like people with dementia, it may still be challenging to complete such measures. When and how to switch to proxy versions, as well as the validity of proxy completion of the WOOP, remain questions for future research to answer.

Regarding potential overlap between the WOOP items, the results from the regression analyses may suggest that the content of 'social contacts' is also captured indirectly through other WOOP items. In particular, 'social contacts' and 'receive support' were highly correlated. This warrants further investigation. Even though we tried to correct for events potentially affecting respondents' well-being between t0 and t1, the test-retest reliability scores may be influenced by events or experiences during the holiday period in which the data were collected (Christmas and New Year). Moreover, the time between completing the t0 and t1 questionnaire ranged from 1 to 12 days because response time to the invitations varied, which may also have affected the test-retest reliability results.

Finally, parallel to the study presented in this paper, a qualitative validation study was conducted, in the sample of 269 respondents at t1, focussing on the feasibility and content validity of the draft version of the WOOP. The yet unpublished results indicate that all aspects

relevant to the well-being of older people seemed to be captured in the WOOP: the items cover the central aspects of well-being that respondents mentioned when describing what well-being meant to them, and their understanding of the items corresponded closely to the descriptions that were formulated for the items (see Appendix A). Based on the feedback respondents provided, some minor adjustments were made to the wording of the WOOP in order to improve clarity and a common understanding of the items. The modified, final version of the WOOP, translated into English by a certified translator, is included in Appendix B (and in Dutch in Appendix C). Considering that the changes made to the WOOP were all very minor and did not affect the structure and nature of the measure, we do not expect the psychometric properties of the final version of the WOOP to differ from what we presented here for the draft version.

4.4. Research and policy recommendations

The study outcomes indicate that the WOOP is a comprehensive and valid alternative measure of well-being for usage in economic evaluations of health and social care services for older people. In addition, the WOOP clearly captures effects beyond health as assessed through HRQoL-measures like the EQ-5D. Just as the ICECAP-O, the WOOP is a generic measure that allows for comparisons across the health and social care sectors. Unlike the ICECAP-O, the WOOP directly asks people about their physical and mental health, implying that it, in principle, does not need to be used in combination with a HRQoL-measure in order to capture the full effects of an intervention for older people. Using a single outcome measure to evaluate interventions can make decision-making more straightforward.

Further research is required before the WOOP can be used in economic evaluations. Even though the draft version of the WOOP showed to have a good validity, and a fair to good test-retest reliability, further validation of the final version of the WOOP is recommended in different contexts (e.g., specific disease areas, higher age groups or older people living in nursing homes). In addition, greater insight is needed in the performance of the WOOP in older people with more severe physical and mental health problems, which were underrepresented in the current sample. Other psychometric properties should also be examined, including responsiveness, inter-rater reliability and sensitivity-to-change. In particular the latter is important, if the WOOP is used to measure improvements in well-being that are considered relevant by older people. Lastly, utility weights should be developed to allow computing a single utility score that accounts for the relative importance of each item to overall well-being.

Credit author statement

Mariska Q.N. Hackert: Conceptualization, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Job van Exel: Conceptualization, Investigation, Writing - review & editing, Supervision. Werner B.F. Brouwer: Conceptualization, Investigation, Writing - review & editing, Supervision.

Declaration of competing interest

None.

No financial support was received for this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.socscimed.2020.113109>.

References

- Al-Janabi, H., 2018. Do capability and functioning differ? A study of U.K. survey responses. *Health Econ.* 27 (3), 465–479. <https://doi.org/10.1002/hec.3586>.
- Al-Janabi, H., Keeley, T., Mitchell, P., Coast, J., 2013. Can capabilities be self-reported? A think aloud study. *Soc. Sci. Med.* 87, 116–122. <https://doi.org/10.1016/j.socscimed.2013.03.035>.
- Bernaards, C., Jennrich, R., 2015. Package “GPArotation”. Available from: <https://cran.r-project.org/web/packages/GPArotation/GPArotation.pdf> and for a description of the package see: <https://CRAN.R-project.org/package=GPArotation>.
- Bowling, A., Hankins, M., Windle, G., Bilotta, C., Grant, R., 2013. A short measure of quality of life in older age: the performance of the brief Older People's Quality of Life questionnaire (OPQOL-brief). *Arch. Gerontol. Geriatr.* 56 (1), 181–187. <https://doi.org/10.1016/j.archger.2012.08.012>.
- Bowling, A., Stenner, P., 2011. Which measure of quality of life performs best in older age? A comparison of the OPQOL, CASP-19 and WHOQOL-OLD. *J. Epidemiol. Community Health* 65, 273–280. <https://doi.org/10.1136/jech.2009.087668>.
- Bulamu, N.B., Kaambwa, B., Ratcliffe, J., 2015. A systematic review of instruments for measuring outcomes in economic evaluation within aged care. *Health Qual. Life Outcome* 13, 179. <https://doi.org/10.1186/s12955-015-0372-8>.
- Cantril, H., 1965. *The Pattern of Human Concerns*. Rutgers University Press, New Brunswick.
- Cleland, J., Hutchinson, C., Khadka, J., Milte, R., Ratcliffe, J., 2019. A review of the development and application of generic preference-based instruments with the older population. *Appl. Health Econ. Health Pol.* 17, 781–801. <https://doi.org/10.1007/s40258-019-00512-4>.
- Coast, J., Flynn, T.N., Natarajan, L., Sproston, K., Lewis, J., Louviere, J.J., Peters, T.J., 2008. Valuing the ICECAP capability index for older people. *Soc. Sci. Med.* 67 (5), 874–882. <https://doi.org/10.1016/j.socscimed.2008.05.015>.
- Davis, J.C., Liu-Ambrose, T., Richardson, C.G., Bryan, S., 2013. A comparison of the ICECAP-O with EQ-5D in a falls prevention clinical setting: are they complements or substitutes? *Qual. Life Res.* 22, 969–977. <https://doi.org/10.1007/s11136-012-0225-4>.
- Drummond, M., Brixner, D., Gold, M., Kind, P., McGuire, A., Nord, E., 2009. Toward a consensus on the QALY. *Value Health* 12 (Suppl. 1), S31–S35. <https://doi.org/10.1111/j.1524-4733.2009.00522.x>.
- Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L., Torrance, G.W., 2015. *Methods for the Economic Evaluation of Health Care Programmes*, fourth ed. Oxford University Press, New York.
- Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., et al., 2019. Package “car”. Available from: <https://cran.r-project.org/web/packages/car/car.pdf> and for a description of the package see: <https://CRAN.R-project.org/package=car>.
- Gamer, M., Lemon, J., Fellows, I., Singh, P., 2019. Package “irr”. Retrieved January 30, 2019 from <https://cran.r-project.org/web/packages/irr/irr.pdf> and for a description of the package see: <https://CRAN.R-project.org/package=irr>.
- Grewal, I., Lewis, J., Flynn, T., Brown, J., Bond, J., Coast, J., 2006. Developing attributes for a generic quality of life measure for older people: preferences or capabilities? *Soc. Sci. Med.* 62 (8), 1891–1901. <https://doi.org/10.1016/j.socscimed.2005.08.023>.
- Hackert, M.Q.N., van Exel, J., Brouwer, W.B.F., 2017. Valid outcome measures in care for older people: comparing the ASCOT and the ICECAP-O. *Value Health* 20 (7), 936–944. <https://doi.org/10.1016/j.jval.2017.03.012>.
- Hackert, M.Q.N., van Exel, J., Brouwer, W.B.F., 2019a. Does the ICECAP-O cover the physical, mental and social functioning of older people in the UK? *Qual. Life Res.* 28, 761–770. <https://doi.org/10.1007/s11136-018-2042-x>.
- Hackert, M.Q.N., Brouwer, W.B.F., Hoefman, R.J., van Exel, J., 2019b. Views of older people in The Netherlands on wellbeing: a Q-methodology study. *Soc. Sci. Med.* 240, 112535. <https://doi.org/10.1016/j.socscimed.2019.112535>.
- Hair, J.F., Tatham, R.L., Anderson, R.E., Black, W., 1998. *Multivariate Data Analysis*, fifth ed. Prentice-Hall, London.
- Herdman, M., Gudex, C., Lloyd, A., Janssen, M.F., Kind, P., Parkin, D., Bonsel, G., Badia, X., 2011. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual. Life Res.* 20, 1727–1736. <https://doi.org/10.1007/s11136-011-9903-x>.
- Hoefman, R.J., van Exel, N.J.A., Foets, M., Brouwer, W.B.F., 2011. Sustained informal care: the feasibility, construct validity and test-retest reliability of the CarerQol instrument to measure the impact of informal care in long-term care. *Aging Ment. Health* 15 (8), 1018–1027. <https://doi.org/10.1080/13607863.2011.575351>.
- Hopkins, W.G., 2002. A New View of Statistics: a Scale of Magnitudes for Effect Statistics. Retrieved from. <https://www.sportssci.org/resource/stats/effectmag.html>.
- Janssen, M.F., Pickard, A.S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., Swinburn, P., Busschbach, J., 2013. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual. Life Res.* 22, 1717–1727. <https://doi.org/10.1007/s11136-012-0322-4>.
- Makai, P., Brouwer, W.B.F., Koopmanschap, M.A., Stolk, E.A., Nieboer, A.P., 2014. Quality of life instruments for economic evaluations in health and social care for older people: a systematic review. *Soc. Sci. Med.* 102, 83–93. <https://doi.org/10.1016/j.socscimed.2013.11.050>.
- Mukuria, C., Conell, J., Carlton, J., Peasgood, T., Brazier, J., Scope, A., et al., 2018. Developing content for a new generic QALY measure: results from a qualitative literature review (E-QALY Project). *Value Health ABSTRACT ONLY* 21 (Suppl. 1), S110. <https://doi.org/10.1016/j.jval.2018.04.743>.
- Naimi, B., 2015. Package “usdm”. Available from: <http://cran.r-project.org/web/packages/usdm/usdm.pdf> and for a description of the package see: <https://CRAN.R-project.org/package=usdm>.
- Netten, A., Burge, P., Malley, J., Potoglou, D., Towers, A.-M., Brazier, J., Flynn, T., Forder, J., Wall, B., 2012. Outcomes of social care for adults: developing a preference-weighted measure. *Health Technol. Assess.* 16 (16). <https://doi.org/10.3310/hta16160>.
- Pavot, W., Diener, E., 1993. Review of the satisfaction with life scale. *Psychol. Assess.* 5 (2), 164–172.
- Post, M.W.M., van Asbeck, F.W.A., van Dijk, A.J., Schrijvers, A.J.P., 1995. Nederlandse interviewversie van de Barthel-index onderzocht bij dwarslaesiepatiënten. *Ned. Tijdschr. Geneesk.* 139 (27), 1376–1380.
- Proud, L., McLoughlin, C., Kinghorn, P., 2019. ICECAP-O, the current state of play: a systematic review of studies reporting the psychometric properties and use of the instrument over the decade since its publication. *Qual. Life Res.* 28, 1429–1439. <https://doi.org/10.1007/s11136-019-02114-y>.
- Revelle, W., 2015. Package “psych”. Available from: <https://personality-project.org/r/psych-manual.pdf> and for a description of the package see: <https://CRAN.R-project.org/package=psych>.
- Sangha, O., Stucki, G., Liang, M.H., Fossel, A.H., Katz, J.N., 2003. The self-administered comorbidity questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Care Res.* 49 (2), 156–163. <https://doi.org/10.1002/art.10993>.
- SCP, 2014. *Replicatie Van Het Meetinstrument Voor Sociale Uitsluiting*. SCP, Den Haag, the Netherlands.
- Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86 (2), 420–428.
- Streiner, D.L., Norman, G.R., 2008. *Health Measurement Scales: A Practical Guide to Their Development and Use*, fourth ed. Oxford University Press, New York.
- Van Leeuwen, K.M., Bosmans, J.E., Jansen, A.P.D., Hoogendijk, E.O., van Tulder, M.W., van der Horst, H.E., Ostelo, R.W., 2015. Comparing measurement properties of the EQ-5D-3L, ICECAP-O, and ASCOT in frail older adults. *Value Health* 18 (1), 35–43. <https://doi.org/10.1016/j.jval.2014.09.006>.
- Van Leeuwen, K.M., van Loon, M.S., van Nes, F.A., Bosmans, J.E., de Vet, H.C.W., Ket, J.C.F., Widdershoven, G.A.M., Ostelo, R.W.J.G., 2019. What does quality of life mean to older adults? A thematic synthesis. *PLoS One* 14 (3), e0213263. <https://doi.org/10.1371/journal.pone.0213263>.
- Versteegh, M.M., Vermeulen, K.M., Evers, S.M.A.A., de Wit, G.A., Prenger, R., Stolk, E.A., 2016. Dutch tariff for the five-level version of EQ-5D. *Value Health* 19 (4), 343–352. <https://doi.org/10.1016/j.jval.2016.01.003>.
- Ware, J.E., Kosinski, M., Keller, S.D., 1995. *SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales*. The Health Institute, New England Medical Centre, Boston, MA.
- Ware, J.E., Kosinski, M., Keller, S.D., 1996. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med. Care* 34 (3), 220–233.
- WHO, 2015. *World Report on Ageing and Health*. World Health Organization, Geneva, Switzerland.