Social Science & Medicine 301 (2022) 114901

Contents lists available at ScienceDirect

Social Science & Medicine

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

Estimating an anchored utility tariff for the well-being of older people measure (WOOP) for the Netherlands

Sebastian Himmler[a,b,c,*], Marcel Jonker[a,b,c], Frédérique van Krugten[a,b,c], Mariska Hackert[a], Job van Exel[a,b,c], Werner Brouwer[a,b]

[a] Erasmus School of Health Policy & Management, Erasmus University Rotterdam (ESHPM), Rotterdam, the Netherlands
[b] Erasmus Centre for Health Economics Rotterdam (EsCHER), Erasmus University Rotterdam, Rotterdam, the Netherlands
[c] Erasmus Choice Modelling Centre (ECMC), Erasmus University Rotterdam, Rotterdam, the Netherlands

ARTICLE INFO

Keywords:
Utility measurement
Discrete choice experiment
Well-being
Older people
WOOP

ABSTRACT

Objective: Health economic evaluations using common health-related quality of life measures may fall short in adequately incorporating all relevant benefits of health and social care interventions targeted at older people. The Well-being of Older People measure (WOOP) is a broader well-being measure that comprises nine well-being domains. The objective of this study was to estimate a utility tariff for the WOOP, to facilitate its application in cost-utility analyses.

Methods: A discrete choice experiment (DCE) with duration approach was set up and fielded among 2,012 individuals from the Netherlands aged 65 years and above. Matched pairwise choice tasks, colour-coding and level overlap were used to reduce the cognitive burden of the DCE. The choice tasks were created using a Bayesian heterogeneous D-efficient design. The estimation procedure accommodated for nonlinear time preferences via an exponential discounting function.

Results: The estimation results showed that 'physical health', 'mental health', and 'making ends meet' were the most important well-being domains for older people, followed by 'independence' and 'living situation'. Of somewhat lesser importance were domains like 'social life', 'receiving support' and 'feeling useful'. The generated utility tariffs can be used to translate well-being states described with the WOOP to a utility score between −0.616 and 1.

Conclusions: This study established a tariff for the WOOP, which will facilitate its use in economic evaluations of health and social care interventions targeted at older people, first of all in the Netherlands.

1. Introduction

Health care, social care and long-term care spending is increasing worldwide (Lorenzoni et al., 2019), propelled by the interaction of ageing populations, increased public expectations, and advances in medical technology (de Meijer et al., 2013). In high income countries, health care spending in the age group above 65 years is already two to three times higher compared to spending in all other age groups combined (Papanicolas et al., 2020). Therefore, the efficient use of scarce care resources, especially within this age group, is crucial. Health economic evaluations, like cost-utility analyses, are established tools to assess whether care services are offering value for money and, therefore, are worthwhile investing in. The results of such analyses guide policy makers in their endeavour to provide the best possible care from the available budget. So far, cost-utility analyses predominantly use quality-adjusted life-years (QALYs) as outcome measure, which combine health-related quality of life (HRQoL) with length of life (Neumann et al., 2016).

Especially in long-term care, social care and end-of-life care, which often aim to improve (or preserve) quality of life domains beyond health, generic HRQoL measures may fall short of measuring the full benefits of these services (Makai et al., 2014). As a result, different well-being measures have been developed that aim to capture these quality of life domains beyond health (Bulamu et al., 2015; Cleland et al., 2019; Helter et al., 2019; Makai et al., 2014). However, in developing these measures, lay perspectives on what is important for the well-being of older people have often been overlooked (Bowling and Stenner, 2011), as well as the heterogeneity in older people's views on

* Corresponding author. Erasmus School of Health Policy & Management (ESHPM), P.O. Box 1738, Rotterdam, 3000DR, the Netherlands.
E-mail address: himmler@eshpm.eur.nl (S. Himmler).

https://doi.org/10.1016/j.socscimed.2022.114901
Received 11 January 2022; Received in revised form 1 March 2022; Accepted 10 March 2022
Available online 17 March 2022

what constitutes well-being (Hackert et al., 2019). Moreover, some of the existing well-being measures are very lengthy and, therefore, not well-suited for self-completion. Most also lack a utility tariff to reflect the relative importance of their domains to overall well-being (Makai et al., 2014). While measures like Adult Social Care Outcomes Toolkit (ASCOT) and the ICEpop CAPability measure for Older people (ICE-CAP-O) do not seem to have these shortcomings (Grewal et al., 2006; Netten et al., 2012), questions remain about their evaluative scope (Makai et al., 2014). For instance, these measures do not directly measure the quality of life domain 'health' (Grewal et al., 2006; Netten et al., 2012), even though older people consider this to be (very) important for their well-being (Hackert et al., 2017, 2019). While health supposedly is captured indirectly in the ICECAP-O, research suggests that this may not be sufficiently the case, in particular physical health (Davis et al., 2013; Hackert et al., 2020; Van Leeuwen et al., 2015).

To overcome some of the shortcomings of existing well-being measures, an alternative measure was developed: the Well-being of Older People measure (WOOP) (Hackert et al., 2021). Its domains are directly based on the views of older people in the Netherlands themselves on what constitutes well-being (Hackert et al., 2019). and covers a comprehensive set of nine well-being domains: 'physical health', 'mental health', 'social life', 'receiving support', 'acceptance and resilience', 'feeling useful', 'independence', 'making ends meet', and 'living situation'. For each of the domains, respondents can indicate their level of functioning by selecting one of five response categories (see Appendix). Qualitative research confirmed the content validity and feasibility of the WOOP as it demonstrated that it captured the important domains of well-being for older people and was considered clear and suitable to self-report their level of well-being (Hackert et al., 2021). Quantitative research showed satisfactory to good results for construct, convergent and discriminant validity, as well as test-retest reliability (Hackert et al., 2020).

Utility tariffs for the WOOP are currently lacking, which clearly hampers its application in (economic) evaluations of health and social care services for older people (Neumann et al., 2016). Therefore, the objective of this study is to estimate a Dutch WOOP utility tariff. The structure of this paper is as follows: the next paragraph specifies the methods, with an emphasis on the design of the choice experiment and the data collection; subsequently, the results are presented, including the WOOP utility tariff; finally, we discuss our findings and their implications.

2. Methods

To estimate utility tariffs for the WOOP for the Netherlands a discrete choice experiment was designed. More specifically, a 'DCE with duration' approach was employed, entailing including duration of life as an additional attribute in the choice tasks. This allows anchoring of utilities on a scale from 0 (dead) to 1 (perfect well-being) (Rowen et al., 2015). This method was preferred over standard gamble and time-trade-off approaches due to concerns relating to the cognitive burden of these iterative procedures, the size of the WOOP instrument, and due to the possibility of administering DCE tasks online (Mulhern et al., 2014).

The traditional estimation approach for DCE with duration data assumes linear time preferences. This implies that the general public is willing to give up a constant proportion of remaining life years for a certain health improvement, without consideration of the number of life years that remain (Dolan and Stalmeier, 2003). Previous work provided evidence that this assumption does not hold in DCE with duration data and that it would introduce biased parameters, as health state preferences would be contaminated by time preferences (Craig et al., 2018; Jonker et al., 2018a). As such, we did not want to presume linear time preferences from the outset and selected an approach that can accommodate non-linear time preferences with a more flexible approach (Jonker et al., 2018a). How this was achieved is outlined below under 'conceptual framework'.

2.1. Attributes, levels, and matched choice task

Attributes and levels used in the choice experiment were defined by the descriptive system of the WOOP (see Appendix) (Hackert et al., 2020). Each of the nine domains of the WOOP is represented by one item with five response levels, generally ranging from excellent (level 1) to bad (level 5). Physical health level 1, for instance, represents being very satisfied with one's physical health. In addition to the nine WOOP domains, a duration attribute was included to enable trade-offs between quality and duration of life. Duration was specified in years using 17 values (0.25, 0.5, 1, 2, …, 15). The values and the range thereof were selected to provide realistic quantities of remaining life years in our target population (smallest and highest values were designed to appear less frequently than the more commonly occurring and hence more realistic middle values). To further increase realism in the choice tasks, we ruled out that the following attribute levels could appear together: Level 1 of independence together with either level 5 of physical or mental health, as well as level 1 of social life and level 5 of support. In a previous data collection with 1,113 respondents, the first two combinations did not occur in the data, while the latter occurred just once (Hackert et al., 2020).

To reduce the cognitive burden of the ten-attribute choice task for the target population, we undertook several steps. First, descriptions of domains and levels were carefully simplified by the researchers involved in the development and qualitative work of the WOOP instrument. Full domain descriptions were still accessible to respondents in the choice task upon moving the cursor over the abbreviated versions. Second, a previously used matched pairs choice task format, which was found to reduce the cognitive burden of choice tasks, was applied (Fig. 1) (Jonker et al., 2017, 2018a). This entailed a first choice between two well-being states A and B, both with equal duration, followed by a matched second choice between the same well-being state B and perfect well-being. This format already simplified the choice tasks by avoiding simultaneous comparisons between the quantity and quality of well-being. This feature of this choice task format additionally helps respondents to treat health and duration multiplicatively. This is theoretically required, but not the case for most respondents when using a traditional, single choice, DCE with duration format (Jonker and Norman, 2022). To further reduce the complexity of the choice tasks, five out of the nine domains were constrained to be overlapped (i.e., well-being states differed in only four domains). To highlight the differences, the level descriptions were colour-coded using shades of purple (with darker shades representing worse levels). This combination of level overlap and colour-coding successfully reduced drop-out rates and attribute non-attendance in earlier studies (Jonker et al., 2018b, 2019, 2018b). The second choice was between the same well-being state B and perfect well-being, but with a shorter duration. We confirmed the feasibility of the final choice tasks in think-aloud interviews among individuals aged 65 years and above. In the executed think-aloud protocol, users were asked to verbalize their thoughts as they completed the full concept online survey, in which the DCE was embedded (for the elements of the survey see "Data collection and survey design"). Data saturation was reached after four think-aloud interviews. Obtained information was summarised into three meaningful categories: instructions choice tasks, instructions other tasks, overall layout. Based on the corresponding insights, minor changes were made to the layout of the survey and to the instructions accompanying the warm-up choice tasks.

2.2. Experimental design

Optimizing the statistical efficiency of the DCE design was crucial due to the large descriptive system of the WOOP (a total of 1,953,125 possible well-being states) and the imposed level overlap constraints. Therefore, an efficient design was implemented and optimized using the TPC-QALY software package (Jonker and Bliemer, 2019). More specifically, a Bayesian heterogeneous D-efficient design with ten sub-designs

Fig. 1. Visual presentation of the pairwise choice task (translated from Dutch).

was used. This implied a simultaneous optimisation of the efficiency of ten separate designs, as well as the efficiency of their aggregate. To give more detail, the D-efficiency criterion was calculated with 100 Bayesian draws based on the weighted average of the overall (i.e., combined) D-error (0.25) and D-errors of the individual blocks (0.75). An exponential discount function was assumed, which appeared to be the most efficient discount function tested with the TPC-QALY software (Jonker and Bliemer, 2019). The design was optimized for the above-described matched choice task format (see also Fig. 1). The number of matched choice tasks per respondent was set at 15, resulting in ten versions and a total set of 300 paired comparisons between two well-being states. Priors for optimizing the initial design were informed by logit model estimates of WOOP best-worst scaling data (N = 310) from a previous study (Himmler et al., 2020a). The experimental design was updated twice after calculating priors based on 201 and a total of 514 completes to further increase the efficiency of the design.

2.3. Data collection and survey design

The DCE was embedded in an online questionnaire and administered to citizens in the Netherlands aged 65 years and above recruited from the panel of the market research company Dynata. We aimed to sample around 2,000 respondents, representative in terms of age and gender, using stratified sampling. After completion, respondents could make a small donation to a charity of their choice. Data collection took place between December 2020 and March 2021.

The survey started with a description of its purpose and a consent form. Next, respondents had to rate their well-being using the WOOP. The DCE training procedure started with a two-alternative choice task with three (randomly selected) WOOP domains. Subsequently, the complexity of the introductory choice task was increased step by step. First, colour-coding was introduced. Second, the duration attribute was added. Third, alternative C and, therefore, the second of the pairwise choice tasks was included. Fourth, all nine WOOP domains were included. Colour-coding, level overlap, and duration were explicitly described. Respondents were randomised to one of the ten blocks of 15 choice tasks between two well-being states. To avoid ordering biases, further randomisation took place regarding the order of choice tasks, the order of the well-being states within choice tasks (A and B), and the order of WOOP domains across respondents (constant order per respondent). The 15 choice tasks were split in three blocks of five tasks, interrupted by two sets of standard socio-demographic questions to reduce response fatigue with respect to the choice tasks. The questionnaire ended with cognitive debriefing questions, an inquiry into whether COVID-19 changed the importance of the WOOP domains for respondents' well-being, and measures for health (EQ-VAS) and life satisfaction (Cantril's ladder).

2.4. Conceptual framework and statistical analysis

In line with the conceptual framework of time-preference corrected QALY tariffs (see Jonker et al., 2018a), the utility derived by individual i for well-being state j in choice task t was defined as the product of the quality of life of the well-being state and the net present value (NPV) of the number of years lived in that well-being state, or:

$$U_{ijt} = quality_{ijt} * NPV(years)_{ijt} + \varepsilon_{ijt} \tag{1}$$

An exponential discount function was used, which has a single discount rate parameter (r) that controls the degree of discounting and results in the following specification of the NPV:

$$NPV(years)_{ijt} = \left(1 - \exp\left(-r*years_{ijt}\right)\right) \Big/ (\exp(r) - 1) \tag{2}$$

The quality of life component in equation (1) was defined as follows:

Step 1.

$$quality_{ijt} = \beta_{i1} + \sum_{d=1}^{9} \beta_{i(d+1)} * WOOPdomain_{ijtd} \tag{3a}$$

in which β_i denotes a respondent-specific parameter vector that captures the importance of the nine WOOP domains ($i.e.\ \beta_{i(2-10)}$) relative to each other and to perfect well-being (i.e., excellent levels in all domains, captured by the β_{i1} intercept), and.

Step 2.

$$WOOPdomain_{ijtd} = \sum_{l=1}^{5} \gamma_{dl} * X_{ijtdl}. \tag{3b}$$

in which γ_d denotes a WOOP domain-specific parameter vector that measures the relative importance of levels 2, 3 and 4 relative to levels 1 and 5 of each WOOP domain, subject to the constraints that $\gamma_{d1} \equiv 0$ and $\gamma_{d5} \equiv 1$ for identification, and where X_{ijtd} denotes a dummy-coded vector that equals 1 for the level at which each WOOP domain was presented to the respondent in the specific choice task, and 0 otherwise.

This specification was programmed in the BUGS language and fitted with OpenBUGS using Markov Chain Monte Carlo (MCMC) techniques. A technical appendix provides details about the statistical modelling. Worthy to note here is that the used approach implies by construction that the QALY decrements for levels 2 to 4 are monotonically increasing proportional to the γ_d WOOP domain-specific level importance parameters.

3. Results

A total of 2,660 respondents provided informed consent to participate in the study, of which 2,169 (82%) started with the DCE valuation tasks after the warm-up tasks. 2,012 respondents completed the full survey, which constitutes 93% of those who started with the DCE valuation tasks. The average age was 73 years, with 57% of respondents being male. The gender distribution of respondents above 75 years did not reflect the targeted sample quota, with females in this age category being underrepresented and males overrepresented (Table 1). Respondents generally reported high levels of well-being in the nine well-being domains of the WOOP (Fig. 2). Lower levels were most frequently reported for the domains 'physical health', 'social life', 'feeling useful', and 'making ends meet'.

The average survey completion time was 34 min (median 24 min). Speeding, defined as a completion time of less than one-third of the median, occurred in 2% of responses (speeders were not excluded from the analysis). The cognitive debriefing questions in general provided favourable results, for instance, 78% of individuals at least partially agreed to the statement that the choice tasks were 'clear' to them (details in suppl. material, Table A1). We did not find large or significant differences in the response patterns to the cognitive debriefing questions between the three different experimental designs used (suppl. material, Table A2). This alleviates concerns about sacrificing (too much) respondent efficiency at the gain of statistical efficiency, which has been discussed before (Flynn et al., 2016; Mulhern et al., 2017).

Table 1
Study sample characteristics (N = 2,012).

	Sample	Sampling quota (census data)[a]
Maledsl	57.3%	
Age in years (SD)[b]	73.3 (5.6)	
Age and gender distribution		
65–74 male	27%	26%
75+ male	30%	18%
65–74 female	28%	27%
75+ female	14%	29%
Finished tertiary education	35.4%	
Married	64.7%	
Employment		
Retired	84.9%	
Gainfully employed	6.0%	
Informal work and volunteering	5.6%	
Other	3.5%	
Country of birth		
Netherlands	94.2%	
Other	5.8%	
Cantril's Ladder (SD)	7.6 (1.2)	
EQ-VAS (SD)	73.4 (18.6)	

Note: SD, Standard deviation.
[a] Data from Statistics Netherlands (Centraal Bureau voor de Statistiek) 2020.
[b] Age ranged from 65 to 101.

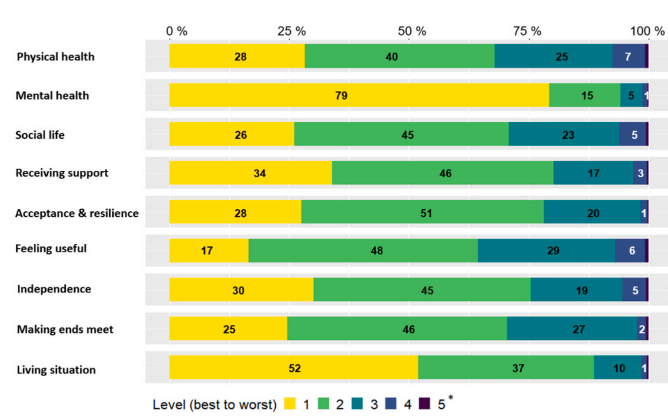

Fig. 2. Distribution of responses to the nine well-being domains of the WOOP (N = 2,012).
Note: * The worst level was selected by less than 1% of respondents in all WOOP dimensions.

3.1. Utility estimates

The calculated domain importance coefficients (equation (3a)) show that 'physical health' and 'mental health', and to a lesser degree 'making ends meet', were the most important well-being domains among the older people in our sample (Table 2). Similarly, when summarising and plotting the terms used by respondents for describing well-being in their own words, physical and mental health were most frequently mentioned (see suppl. material, Figure A1).

The anchored domain level utility weights are presented in Fig. 3 (suppl. material, Table A4 shows the 95% CI). By construction, the estimated domain level weights are logically consistent within all nine well-being domains and non-positive. Two levels failed to reach statistical significance (i.e., the second-best levels of 'acceptance & resilience'

Table 2
Domain importance on latent utility scale.

Domain	Mean	Lower 95% CI	Upper 95% CI	SD
Physical health	−1.381	−1.482	−1.283	0.051
Mental health	−1.507	−1.615	−1.401	0.055
Social life	−0.556	−0.606	−0.507	0.025
Receiving support	−0.457	−0.506	−0.409	0.025
Acceptance and resilience	−0.543	−0.596	−0.493	0.026
Feeling useful	−0.426	−0.475	−0.380	0.025
Independence	−0.718	−0.781	−0.657	0.032
Making ends meet	−1.136	−1.218	−1.054	0.042
Living situation	−0.674	−0.735	−0.615	0.031

CI = Credible Interval.

and ‘making ends meet’). The strongest decrements were found for ‘mental health’ (−0.329), ‘physical health’ (−0.302) and ‘making ends meet’ (−0.248), followed by ‘independence’ (−0.157) and ‘living environment’ (−0.147). ‘Social life’, ‘receiving support’, ‘acceptance and resilience’, and ‘feeling useful’ were generally perceived as less important for well-being. The theoretical spread of the WOOP utility ranges from −0.616 (worst possible state) to 1 (best possible state). The estimated discount rate was 0.173, considerably larger than has been found in a related general population study (0.057) (Jonker et al., 2018a). A higher discount rate may relate to a lower remaining life expectancy among older people. Jonker et al. (2018a) also found that people with more severe health problems had a higher discount rate, thus finding this higher discount rate for older people, who tend to have more, and more severe health problems, is not completely unexpected.

Applying the utility tariffs to the WOOP responses in the sample produced a mean WOOP utility of 0.856 (SD 0.120). A utility value of 1 was observed for 34 respondents (1.7%) and a utility value below 0 (−0.067) for one respondent. The 25%, 50%, and 75% quantiles were 0.831, 0.889, and 0.929, respectively. When plotting utilities against EQ-VAS and Cantril’s ladder (Fig. 4), a strongly positive correlation was observed (r = 0.59 and r = 0.54, respectively) with similar trends for males and females.

3.2. COVID-19 impact

Results for the question about whether the importance of the different well-being domains had changed due to COVID-19 were the following: The domains ‘physical health’, ‘mental health’,

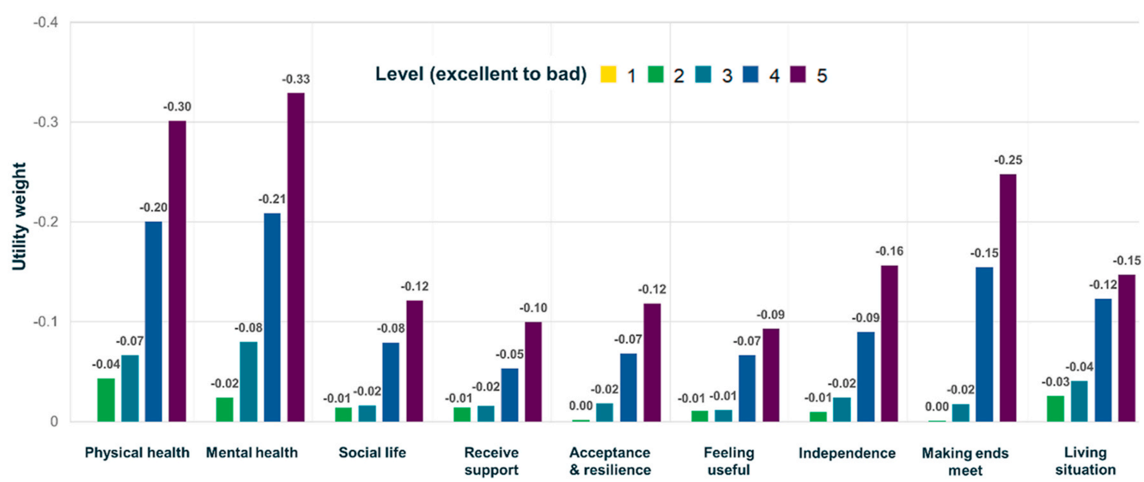

Fig. 3. Utilities weights of the WOOP domain levels, with level 1 (excellent) as reference category.

Fig. 4. WOOP utility values plotted against health (EQ-VAS) and life satisfaction.
Note: For illustration purposes, jitter was added to the EQ-VAS and life satisfaction values, which are bounded on 0 to 100 and 0 to 10 range, respectively.

'independence', and 'social life' generally appear to have become more important (suppl. material, Figure A2). Depending on the dimension, between 61% and 74% of respondents indicated that each respective domain had remained equally important for their well-being, with the lowest value observed for 'social life'.

4. Discussion

Given the increasing relevance of health and social care services for older people, and the fact that these services usually aim to improve well-being rather than health (alone), adequate instruments for measuring the well-being of older people are required. The WOOP was recently developed for this purpose. To be useful as outcome measure in economic evaluations, such a measure ideally is accompanied by utility tariffs. Hence, in this study we present the results of a discrete choice experiment fielded among 2,012 individuals in the Netherlands aged 65 years and above to obtain preference-based utility tariffs for the WOOP. The resulting tariffs enable transformation of well-being states described with the WOOP into a utility score anchored on perfect well-being (1) and dead (0), and hence the use of the WOOP as outcome measure in cost-utility analyses of interventions in health and social care aimed at older people.

We elicited preferences from individuals aged 65 years and above, hence in the group of older people themselves and not in the general adult population as is commonly done for other outcome measures. Therefore, the utility tariffs for the WOOP reflect the relative importance for well-being of the different domains and functioning levels therein in the target population of the WOOP. This approach was deemed most relevant in informing the allocation of resources intended to improve the well-being of older people, and especially to evaluate optimal allocation *within* the budget for health and social care services for older people according to their preferences. Therefore, in contrast to measures like the EQ-5D, the WOOP is specifically targeted at one age group and not intended for comparisons across all adult age groups.

Given the large descriptive system and the target population of the WOOP, the optimal elicitation method was selected based on a preceding study (Himmler et al., 2020a). Furthermore, we undertook several steps to reduce the cognitive burden of the choice tasks. Based on the responses to the cognitive debriefing statements presented to respondents after the choice tasks, it seems that the combination of a stepwise introduction to the experiment, colour-coding, level overlap and the separation of the trade-offs between well-being domains and duration was successful in reducing the cognitive burden to a manageable amount for this sample of older people.

In line with the Q-methodology study conducted to identify the domains of the WOOP (Hackert et al., 2019) and with previous research (Douma et al., 2017; Van Leeuwen et al., 2019), we found that 'physical health' and 'mental health' were the most important domains for the well-being of older people, followed by 'making ends meet'. Domains like 'independence', 'social life', 'receiving support' and 'feeling useful' seem to be of somewhat lesser importance to their well-being. The relatively low importance of the domain 'social life' was somewhat surprising given the results of previous research (Douma et al., 2017; Hackert et al., 2019; Van Leeuwen et al., 2019).

4.1. Strengths and limitations

Previous studies estimating utility tariffs for well-being measures primarily applied best-worst scaling (BWS) approaches (Coast et al., 2008; Flynn et al., 2015; Netten et al., 2012) (or intend to do so (Ratcliffe et al., 2019)). Therefore, a noteworthy strength of the applied methodology is that it provides a feasible alternative approach, which was also shown to be preferable for older people in terms of the cognitive burden of choice tasks in a previous study (Himmler et al., 2020a). Moreover, the DCE design with a duration attribute allowed anchoring the utility weights of the WOOP on a QALY-like scale, facilitating a more straightforward combination of length and quality of life in computing the benefits of interventions. The applied approach furthermore accounts for non-linear time preferences, which otherwise would bias estimates in DCE with duration approaches (Jonker et al., 2018a). The estimated discount rate of 0.173 implies that parameter estimates would have been severely biased by time preferences if we would have assumed linear time preferences. A more general implication of this is that for older people, estimated/empirical discount rates are much higher than the discount rates used in traditional HTA calculations, which mostly range between 1.5% and 5% (Attema et al., 2018). The use of exponential (as opposed to, for example, hyperbolic) discounting in our analysis is also consistent with the common approach to discounting of health effects in health technology assessment.

More particularly, the implemented modelling approach has the advantage that it reduces the number of respondent-specific parameters, allows for correlated preferences between the WOOP domains, and produces readily available estimates of the relative importance of the WOOP domains, while ensuring a logically consistent utility tariff. The modelling approach used here was more structured than the one used by Jonker et al. (2018a), but a more parsimonious model structure was crucial considering the large descriptive system of the WOOP and the limited number of respondents relative to the number of utility decrements.

While the pilot tests indicated that we reduced the complexity of the DCE choice tasks to a manageable cognitive burden for most respondents, decision heuristics could still have played an important role. For instance, while the colour-coding helped in identifying the differences between the two well-being states in a choice task, it may have stimulated respondents to focus on the colour intensity when making their choices. To what extent respondents used decision heuristics in general is unknown, although 89% of respondents reported to have compared all different aspects before making their choices (suppl. material, Table A1). At the same time, colour-coding combined with level overlap have previously been established as effective strategies to reduce the use of (other) decision heuristics (Jonker et al., 2017).

Reducing the complexity also entailed simplifying the attribute and level descriptions in the choice tasks—as is common in health state valuation (see e.g. (Devlin et al., 2018). We do not know whether and how this might have impacted the interpretation of attributes and levels, as we did not formally test the equivalence of abbreviated and full descriptions. Nevertheless, this is not expected to have had a substantial impact on the valuation results. First, as much as possible, all domains and levels were abbreviated in the same manner. Second, prior to the valuation tasks, individuals already were introduced to the full WOOP instrument with the full descriptions. Third, respondents had the full attribute descriptions available as mouse-over elements in the choice task to ease interpretation.

A clear limitation of the analysis relates to the representativeness of the sample, which is hampered by two factors. First, females aged 75 years and above were underrepresented in the sample (see Table 1). The market research company was unable to reach the desired number of completes in this group even after considerable effort.

Second, and more importantly, people above 65 years of age, who are part of online survey panels, and perhaps especially those above 75 years, likely will not be fully representative of this age group in terms of functioning, living situation, digital skills, and cognitive ability. Unfortunately, we did not collect data about these characteristics, but we could use EQ-VAS values as an indicator. When we compare age-stratified EQ-VAS values in our sample with data from a previous large-scale study among community dwelling Dutch elderly (suppl. Materials, Table A3), we find that individuals in our sample have a lower level of health (Mangen et al., 2017). This might hint towards capturing a wider range of respondents than just community dwelling individuals, but this cannot be confirmed. At the same time, it is very likely that people in poor health and well-being states are underrepresented in our study. For instance, another study found that among residents of nursing homes in the Netherlands, the mean EQ-5D VAS score was 64.8 (SD 21.7) (Nijsten et al., 2019), which is clearly lower than in our sample. The underrepresentation of individuals in poorer states, including those in nursing homes may explain the high levels of well-being in most domains of the WOOP observed in our sample.

Worth mentioning in this context is that 18% of survey participants dropped out during the introduction and warm-up tasks. It is likely that this drop-out is related to the cognitive capabilities of participants, which may have further contributed to an analysis sample of relatively capable, healthy, and happy respondents.

The preferences of older people in poorer states, including those that who are frail, dependent, or living in nursing homes, may thus differ from what we observed in our sample. These groups are, however, difficult to reach and experience more difficulty with participating in (this type of) research. As such, the utility tariffs presented here may not fully reflect the preferences of the older population in its entirety. Given the aim of the WOOP, assessing whether preferences regarding the WOOP states differ in the subgroup of the oldest old and frailest individuals is important, but appears to require a different study design. This might include purposive sampling, also within nursing homes, and interviewer-assisted survey techniques, with an adjusted and simplified choice experiment.

Finally, we emphasize that the data for this study was collected during the COVID-19 pandemic, an extraordinary context with special relevance to older peoples' well-being. Our attempts to assess the impact of this on the estimated preferences showed that 'physical health', 'mental health', and 'independence' domains may have especially increased in importance (see suppl. material, Figure A1). It is not clear whether possible effects of the pandemic on preferences for the WOOP domains are temporary or may last after the pandemic is over. After all, a possible effect of the current crisis may be that people became more aware of what they consider most important for their well-being and, hence, the preferences we measured in this study may even be closer to their true preferences.

4.2. Application and future research

The estimated utility tariffs enable the use of the WOOP in economic evaluations of health and social care interventions targeted at older people, first of all in the Netherlands. Using the WOOP may provide a more comprehensive overview of the benefits of such interventions as compared to health-related quality of life measures (e.g. EQ-5D), but also as compared to the ICECAP-O and the ASCOT (Makai et al., 2014). Moreover, the WOOP has the advantage over other well-being measures that its utility tariff is anchored on dead and perfect wellbeing, facilitating a more straightforward combination of length and quality of life in computing the benefits of interventions. Consequently, the WOOP may also be useful when evaluation cross-sectoral interventions, for instance health and social care services combined with housing or income support. However, until further research confirms the (psychometric) validity of the WOOP and assessing interventions in health and social care in terms of their full benefits to older people becomes more established, we would advocate the use of the WOOP next to standard measures of health-related quality of life. This is also in line with the current recommendation of the Dutch health care institute (Zorginstituut Nederlands) for the use of the ICECAP-O. We do note that since the WOOP captures broader wellbeing *including* health, the measure cannot be readily added to results obtained using generic health-related quality of life measures, as this would imply double-counting.

Decision makers and analysts need to be aware that using an outcome measure like the WOOP, which focuses on broader outcomes than health and is conceptually targeted at a specific age group, makes it difficult to compare the results of evaluation studies with those using other outcome measures. Hence, the comprehensiveness and relevance of the WOOP in the specific context of health and social care for older people comes at the price of reducing the comparability of findings with those from economic evaluations in other populations or focused on health as outcome. Furthermore, for economic evaluations using the WOOP to be truly informative for decision-making about whether or not to implement particular health and social care services, a threshold value representing the monetary value of a well-being adjusted life year (WALY) is required. Considering that the scope of benefits is broader, it is likely to be higher than that the threshold for a QALY. While different methods may be used to estimate such a threshold value (Himmler et al., 2020b; Kinghorn and Afentou, 2021; Ryen and Svensson, 2015) an important conceptual question will be whether this valuation should be done within the target population, as the beneficiaries of health and social care interventions, or within the general public, as the payer of such interventions in a collective system (like in the Netherlands). A last noteworthy aspect of the use of broader outcome measures in general is that by extending the scope of the benefit dimension, one needs to consider also extending the cost dimension beyond health care to stay within a consistent framework.

5. Conclusion

By generating utility weights, the WOOP can now be used in economic evaluations of health and social care services targeted at older people, first of all in the Netherlands. Furthermore, the methodological approach used in this study may be helpful for future studies valuing newly developed measures with similarly large descriptive systems.

Author statement

Sebastian Himmler: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Marcel Jonker**: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing. **Frédérique van Krugten**: Formal analysis, Resources, Writing – original draft, Writing – review & editing. **Mariska Hackert**: Conceptualization, Writing – original draft, Writing – review & editing. **Job van Exel**: Conceptualization, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Werner Brouwer**: Conceptualization, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition

Financial disclosure

Sebastian Himmler received funding from a Marie Sklodowska-Curie fellowship financed by the European Commission (Grant agreement No. 721402). The data collection was also funded through that grant. The sponsor had no in study design; in the collection, analysis and interpretation of data; in the writing of the articles; and in the decision to submit it for publication.

Ethics approval

The study was reviewed and approved by the Research Ethics Committee of the Erasmus School of Health Policy & Management.

Availability of data and material

The full survey (in Dutch), raw survey data, Stata data preparation code, and OpenBUGS model code for the computation of the QALY decrements can be accessed on the Open Science Framework (anonymous link): https://osf.io/ysajr/.

Declaration of competing interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Acknowledgements

The authors would like to thank members of the Erasmus Choice Modelling Center for ideas and feedback on the study at different stages.

Appendix

Well-being of Older People measure (WOOP)

For each section, select the description that is most appropriate for you today.

Physical health

Consider physical conditions or ailments and other physical impairments that affect your daily functioning.

☐ I have no problems with my physical health
☐ I have slight problems with my physical health
☐ I have moderate problems with my physical health
☐ I have severe problems with my physical health
☐ I have very severe problems with my physical health

Mental health

Consider problems with your ability to think, anxiety, depression and other mental impairments that affect your daily functioning.

☐ I have no problems with my mental health
☐ I have slight problems with my mental health
☐ I have moderate problems with my mental health
☐ I have severe problems with my mental health
☐ I have very severe problems with my mental health

Social life

Consider your relationship with your partner, family or other people who are important to you. This concerns the amount and quality of the contact you have.

☐ I'm very satisfied with my social life
☐ I'm satisfied with my social life
☐ I'm reasonably satisfied with my social life
☐ I'm dissatisfied with my social life
☐ I'm very dissatisfied with my social life

Receive support

Everyone needs help or support sometimes. Consider practical or emotional support, for example from your partner, family, friends, neighbours, volunteers or professionals. This concerns being able to count on support when you need it, as well as the quality of the support.

☐ I'm very satisfied with the support I get, when needed
☐ I'm satisfied with the support I get, when needed
☐ I'm reasonably satisfied with the support I get, when needed
☐ I'm dissatisfied with the support I get, when needed
☐ I'm very dissatisfied with the support I get, when needed

Acceptance and resilience

Consider your acceptance of your current circumstances and your ability to adapt to changes to these, whether or not with support of your religion or belief.

☐ I'm more than able to deal with my circumstances and changes to these
☐ I'm able to deal with my circumstances and changes to these
☐ I'm reasonably able to deal with my circumstances and changes to these
☐ I'm not able to deal with my circumstances and changes to these
☐ I'm not at all able to deal with my circumstances and changes to these

Feeling useful

Consider meaning something to others, your environment or a good cause.

☐ I feel very useful
☐ I feel useful
☐ I feel reasonably useful

☐ I do not feel useful
☐ I do not feel at all useful

Independence

Consider being able to make your own choices or doing the activities that you find important.

☐ I feel very independent
☐ I feel independent
☐ I feel reasonably independent
☐ I feel dependent
☐ I feel very dependent

Making ends meet

Consider having enough money to meet your daily needs and having no money worries.

☐ I'm more than able to make ends meet
☐ I'm able to make ends meet
☐ I'm reasonably able to make ends meet
☐ I'm not able to make ends meet
☐ I'm not at all able to make ends meet

Living situation

Consider living in a house or neighbourhood you like.

☐ I'm very satisfied with my living arrangements
☐ I'm satisfied with my living arrangements
☐ I'm reasonably satisfied with my living arrangements
☐ I'm dissatisfied with my living arrangements
☐ I'm very dissatisfied with my living arrangements

Technical appendix

The specification described by equations (1)–(3b) was programmed in the BUGS language and fitted with OpenBUGS using Markov Chain Monte Carlo (MCMC) techniques. This involved the selection of prior densities for the model parameters and updating these densities with the likelihood of the observed data. A multivariate normal prior was placed on the β_i parameters, i.e., $\beta_i \sim MVN(\mu,\mathrm{T})$. Uninformative normal priors (i.e., with means of 0 and standard deviations of 10) were assigned to μ and a Wishart prior with an identity scale matrix and 10 degrees of freedom to the precision matrix T. A uniform (0,1) prior was placed on r, and Dirichlet priors with concentration parameters equal to 1.0 were assigned to a set of latent $\gamma^*_{d(1:4)}$ parameters that were subsequently transformed into $\gamma_{d(1:5)}$ by setting $\gamma_{d1} \equiv 0$ and defining $\gamma_{dl} = \sum_{m=1}^{l} \gamma^*_{dm}$ for $l \in 2-5$. This ensured, by construction, monotonically increasing $\gamma_{d(1:5)}$ parameters that automatically adhered to the required $\gamma_{d1} \equiv 0$ and $\gamma_{d5} \equiv 1$ constraints, leading to monotonically increasing level-importance parameter estimates within each of the WOOP domains.

Standard Gibbs updates were used to update μ and Σ, antithetic Metropolis-within-Gibbs update steps to update β, slice sampling update steps to update r, and non-conjugate random-walk Dirichlet update steps were used to update the γ^* parameters. In addition, the implied decrements for the WOOP attribute levels on the QALY scale were calculated by first dividing all elements in the mean vector (μ) by the first element (μ_1), which ensured that the value of full well-being was equal to 1.0, and then, for each WOOP domain, multiplying the scaled average domain importance (i.e. μ_{2-10}/μ_1) with the corresponding domain-specific γ_d parameter. Given the constraints on γ_d, this implies, by definition, that QALY decrements for level 1 are 0 and that the QALY decrements for levels 5 are equal to $\mu_{(d+1)}/\mu_1$ and thereby equal to the relative importance of the respective WOOP domain. Furthermore, this implies that the QALY decrements for levels 2 to 4 are monotonically increasing proportional to the γ_d WOOP domain-specific level importance parameters.

Appendix A. Online supplementary materials

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2022.114901.

References

Attema, A.E., Brouwer, W.B.F., Claxton, K., 2018. Discounting in economic evaluations. Pharmacoeconomics 36, 745–758. https://doi.org/10.1007/s40273-018-0672-z.
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, 1–23. https://doi.org/10.1186/s12955-015-0372-8.
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, 874–882. https://doi.org/10.1016/j.socscimed.2008.05.015.
Craig, B.M., Rand, K., Bailey, H., Stalmeier, P.F.M., 2018. Quality-adjusted life-years without constant proportionality. Value Health 21, 1124–1131. https://doi.org/10.1016/j.jval.2018.02.004.
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.
de Meijer, C., Wouterse, B., Polder, J., Koopmanschap, M., 2013. The effect of population aging on health expenditure growth: a critical review. Eur. J. Ageing 10, 353–361. https://doi.org/10.1007/s10433-013-0280-x.

Devlin, N.J., Shah, K.K., Feng, Y., Mulhern, B., van Hout, B., 2018. Valuing health-related quality of life: an EQ-5D-5L value set for England. Health Econ 27, 7–22. https://doi.org/10.1002/hec.3564.

Dolan, P., Stalmeier, P., 2003. The validity of time trade-off values in calculating QALYs: constant proportional time trade-off versus the proportional heuristic. J. Health Econ. 22, 445–458. https://doi.org/10.1016/S0167-6296(02)00120-0.

Douma, L., Steverink, N., Hutter, I., Meijering, L., Bowers, B.J., 2017. Exploring subjective well-being in older age by using participant-generated word clouds. Gerontol. 57, 229–239. https://doi.org/10.1093/geront/gnv119.

Flynn, T.N., Bilger, M., Malhotra, C., Finkelstein, E.A., 2016. Are efficient designs used in discrete choice experiments too difficult for some respondents? A case study eliciting preferences for end-of-life care. Pharmacoeconomics 34, 273–284. https://doi.org/10.1007/s40273-015-0338-z.

Flynn, T.N., Huynh, E., Peters, T.J., Al-Janabi, H., Clemens, S., Moody, A., Coast, J., 2015. Scoring the icecap-a capability instrument. Estimation of a UK general population tariff. Health Econ 24, 258–269. https://doi.org/10.1002/hec.3014.

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, 1891–1901. https://doi.org/10.1016/j.socscimed.2005.08.023.

Hackert, M.Q.N., Brouwer, W.B.F., Hoefman, R.J., van Exel, J., 2019. 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.

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, 936–944. https://doi.org/10.1016/j.jval.2017.03.012.

Hackert, M.Q.N., van Exel, J., Brouwer, W.B.F., 2021. Content validation of the well-being of older people measure (WOOP). Health Qual. Life Outcome 19, 200. https://doi.org/10.1186/s12955-021-01834-5.

Hackert, M.Q.N., van Exel, J., Brouwer, W.B.F., 2020. Well-being of older people (WOOP):Quantitative validation of a new outcome measure for use in economic evaluations. Soc. Sci. Med. 259, 113109. https://doi.org/10.1016/j.socscimed.2020.113109.

Helter, T.M., Coast, J., Łaszewska, A., Stamm, T., Simon, J., 2019. Capability Instruments in Economic Evaluations of Health-Related Interventions: a Comparative Review of the Literature, Quality of Life Research. Springer International Publishing. https://doi.org/10.1007/s11136-019-02393-5.

Himmler, S., Soekhai, V., van Exel, J., Brouwer, W., 2020a. What works better for preference elicitation among older people? Cognitive burden of discrete choice experiment and case 2 best-worst scaling in an online setting. J. Choice Model. 38, 100265. https://doi.org/10.1016/j.jocm.2020.100265.

Himmler, S., van Exel, J., Brouwer, W., 2020b. Estimating the monetary value of health and capability well-being applying the well-being valuation approach. Eur. J. Health Econ. 21, 1235–1244. https://doi.org/10.1007/s10198-020-01231-7.

Jonker, M.F., Attema, A.E., Donkers, B., Stolk, E.A., Versteegh, M.M., 2017. Are health state valuations from the general public biased? A test of health state reference dependency using self-assessed health and an efficient discrete choice experiment. Heal. Econ. (United Kingdom) 26, 1534–1547. https://doi.org/10.1002/hec.3445.

Jonker, M.F., Bliemer, M.C.J., 2019. On the optimization of bayesian D-efficient discrete choice experiment designs for the estimation of QALY tariffs that are corrected for nonlinear time preferences. Value Health 22, 1162–1169. https://doi.org/10.1016/j.jval.2019.05.014.

Jonker, M.F., Donkers, B., de Bekker-Grob, E., Stolk, E.A., 2019. Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. Heal. Econ. (United Kingdom) 28, 350–363. https://doi.org/10.1002/hec.3846.

Jonker, M.F., Donkers, B., de Bekker-Grob, E.W., Stolk, E.A., 2018a. Advocating a paradigm shift in health-state valuations: the estimation of time-preference corrected QALY tariffs. Value Health 21, 993–1001. https://doi.org/10.1016/j.jval.2018.01.016.

Jonker, M.F., Donkers, B., de Bekker-Grob, E.W., Stolk, E.A., 2018b. Effect of level overlap and color coding on attribute non-attendance in discrete choice experiments. Value Health 21, 767–771. https://doi.org/10.1016/j.jval.2017.10.002.

Jonker, M.F., Norman, R., 2022. Not all respondents use a multiplicative utility function in choice experiments for health state valuations, which should be reflected in the elicitation format (or statistical analysis). Heal. Econ. (United Kingdom) 31, 431–439. https://doi.org/10.1002/hec.4457.

Kinghorn, P., Afentou, N., 2021. Eliciting a monetary threshold for a year of sufficient capability to inform resource allocation decisions in public health and social care. Soc. Sci. Med. 279, 113977. https://doi.org/10.1016/j.socscimed.2021.113977.

Lorenzoni, L., Marino, A., Morgan, D., James, C., 2019. Health Spending Projections to 2030: New Results Based on a Revised OECD Methodology. OECD Heal. Work. Pap. https://doi.org/10.1787/5667f23d-en.

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.

Mangen, M.J.J., Bolkenbaas, M., Huijts, S.M., van Werkhoven, C.H., Bonten, M.J.M., de Wit, G.A., 2017. Quality of life in community-dwelling Dutch elderly measured by EQ-5D-3L. Health Qual. Life Outcome 15, 1–6. https://doi.org/10.1186/s12955-016-0577-5.

Mulhern, B., Bansback, N., Brazier, J., Buckingham, K., Cairns, J., Devlin, N., Dolan, P., Hole, A.R., Kavetsos, G., Longworth, L., Rowen, D., Tsuchiya, A., 2014. Preparatory study for the revaluation of the EQ-5D tariff: methodology report. Health Technol. Assess. 18 https://doi.org/10.3310/hta18120.

Mulhern, B., Bansback, N., Hole, A.R., Tsuchiya, A., 2017. Using discrete choice experiments with duration to model EQ-5D-5L health state preferences: testing experimental design strategies. Med. Decis. Mak. 37, 285–297. https://doi.org/10.1177/0272989X16670616.

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 https://doi.org/10.3310/hta16160.

Neumann, P.J., Sanders, G.D., Russell, L.B., Siegel, J.E., Ganiats, T.G., 2016. Cost Effectiveness in Health and Medicine. Oxford University Press, New York.

Nijsten, J.M.H., Leontjevas, R., Smalbrugge, M., Koopmans, R.T.C.M., Gerritsen, D.L., 2019. Apathy and health-related quality of life in nursing home residents. Qual. Life Res. 28, 751–759. https://doi.org/10.1007/s11136-018-2041-y.

Papanicolas, I., Marino, A., Lorenzoni, L., Jha, A., 2020. Comparison of health care spending by age in 8 high-income countries. JAMA Netw. Open 3, e2014688. https://doi.org/10.1001/jamanetworkopen.2020.14688.

Ratcliffe, J., Cameron, I., Lancsar, E., Walker, R., Milte, R., Hutchinson, C.L., Swaffer, K., Parker, S., 2019. Developing a new quality of life instrument with older people for economic evaluation in aged care: study protocol. BMJ Open 9, e028647. https://doi.org/10.1136/bmjopen-2018-028647.

Rowen, D., Brazier, J., Van Hout, B., 2015. A comparison of methods for converting DCE values onto the full health-dead QALY scale. Med. Decis. Mak. 35, 328–340. https://doi.org/10.1177/0272989X14559542.

Ryen, L., Svensson, M., 2015. The willingness to pay for a quality adjusted life year: a review of the empirical literature. Health Econ 24, 1289–1301. https://doi.org/10.1002/hec.3085.

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, 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 Veti, 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.