Mariska Hackert

Care for Well-being
Development of the
Well-being of Older People measure (WOOP)

Care for Well-being

Development of the

Well-being of Older People measure (WOOP)

Zorg voor welzijn

Ontwikkeling van de

Well-being of Older People uitkomstmaat (WOOP)

Mariska Quirina Nikita Hackert

Care for Well-being

Development of the

Well-being of Older People measure (WOOP)

Zorg voor welzijn

Ontwikkeling van de

Well-being of Older People uitkomstmaat (WOOP)

Proefschrift

ter verkrijging van de graad van doctor aan de

Erasmus Universiteit Rotterdam

op gezag van de

rector magnificus

Prof.dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.
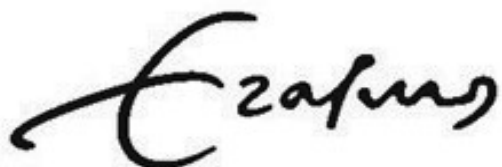
De openbare verdediging zal plaatsvinden op

donderdag 17 september 2020 om 13:30 uur

door

Mariska Quirina Nikita Hackert

geboren te Venlo

Erasmus University Rotterdam

Promotiecommissie

Promotoren

Prof.dr. N.J.A. van Exel

Prof.dr. W.B.F. Brouwer

Overige leden

Prof.dr. M.P.M.H. Rutten – van Mölken

Prof.dr. C.D. Dirksen

Prof.dr. E. Buskens

Contents

Chapter 1

Introduction

1.1 | Population ageing: A challenge for public health

Worldwide the sustainability of health care systems is an issue of concern, importantly fuelled by the continuing ageing of populations. In OECD countries (Organisation for Economic Co-operation and Development), the percentage of older people, in this thesis defined as those aged 65 years and older, has grown from less than 9% in 1960 to 17% in 2015, and is expected to reach 28% in 2050 (OECD, 2017). This development is due to a combination of declining fertility rates and higher life expectancies. Declining fertility rates are the result of less child mortality, greater access to contraception and changing gender norms, while higher life expectancies are related to greater access to care and better quality of care, as well as improved lifestyles and living conditions (OECD, 2017; WHO, 2015).

The mere observation that the number of older people continues to increase, in absolute and relative terms, potentially conceals the challenge that arises from the expectation that people will become increasingly old. In the Netherlands, for instance, people at the age of 65 are now predicted to have a remaining life expectancy of almost 20 years, and this figure is expected to increase further. However, many of these life years at older ages are not spent in perfect health. In the Netherlands, only half of the additional 20 years people aged 65 are projected to live, are expected to be healthy life years (OECD, 2017). With ageing, people are more likely to develop multi-morbidity, geriatric syndromes, and other disabilities (OECD, 2017; WHO, 2015). To illustrate this, the prevalence of dementia is especially high among people aged 90 years and over (e.g. 40% in OECD countries), which suggests that this disease may become more common in the future (OECD, 2017). Jointly, these developments and projections indicate that the need for care and the complexity of care needs will significantly increase over the coming years, in particular in relation to long-term care (OECD/EU, 2018).

At the same time, policy makers face the problem that health care resources are scarce and typically insufficient to meet all demands for health and social care in the population. If they choose to devote more funds from the available health care budget to meet the increasing care needs of older people, this means that they have to reduce spending on care for other groups. Alternatively, increasing government spending on health as a response either comes at the expense of spending on other policy areas, such as education and infrastructure, or of raising taxes or premiums. Financing potential additional expenditures on care for older people is even more troublesome in view of the declining proportion of the population that

is economically active. Across OECD countries, the number of individuals of working-age (15 – 64 years) per older person (≥ 65 year) is predicted to drop from four in 2012 to two in 40 years' time (OECD, 2017). Such figures raise questions about the sustainability of health care financing and the limits to intergenerational solidarity.

Consequently, policy makers are required to make choices about which interventions to fund and which not to fund from the limited health care budget. In this context, economic evaluations are of increasing interest as a transparent and systematic tool to help in the optimal allocation of finite health care resources over competing alternatives. A prerequisite of such evaluations is that all relevant costs and benefits are adequately captured (Brazier et al., 2017; Drummond et al., 2015). Especially in the context of care for older persons, which may aim to improve their health or, more broadly, their well-being, this is challenging. This dissertation focusses on this issue, by specifically addressing the issue of how to adequately measure the benefits of the wide range of health and social care interventions for older people in economic evaluations.

1.2 | Economic evaluations to support informed decisions in health care

Economic evaluations provide information on the costs and benefits of health and social care interventions, to aid policy makers in financing those that provide the most value for money. Various types of economic evaluations exist that especially differ in their unit of measuring benefits. Cost-benefit analysis (CBA) expresses all costs and benefits of an intervention in monetary terms. In this context, health outcomes thus need to be captured in monetary terms, which is difficult and contentious. Cost-effectiveness analysis (CEA) expresses the benefits of interventions in natural units that are relevant to the specific intervention under study, such as life years gained or event-free life years. As interventions may differ, CEA outcomes from different settings may be incomparable, which complicates decision-making. Cost-utility analysis (CUA) overcomes this comparability issue by capturing health benefits of interventions in terms of quality-adjusted life-years (QALYs). QALYs comprise gains in both length and health-related quality of life (HRQoL). HRQoL is typically measured using preference-based measures, such as the EuroQol five-dimensional questionnaire (EQ-5D) (The EuroQol Group, 1990) or the Short-Form Six-Dimension health index (SF-6D) (Brazier et al., 2002). These measures have descriptive systems that cover various health domains, such as mobility, self-care and pain, on which respondents can indicate their level of

functioning. Considering that certain health problems may have a greater impact on people's overall health utility than others, these health states can be transformed into a utility score using a weighting system (or tariff). This weighting system typically reflects population preferences for the health states that are distinguished with the descriptive system. The utility scores are usually anchored on a standardised scale, ranging from 1, being in perfect health, to 0, being dead, with negative values accounting for health states considered to be worse than being dead. CUA allows comparisons of interventions affecting health in different ways, as QALYs generically combine changes in important health domains and longevity in a single metric. In this way, interventions for the same condition, or even for different conditions, that affect different domains of health, can be compared relatively straightforwardly, also in relation to their costs (Brazier et al., 2017; Drummond et al., 2015).

1.3 | Challenges in economic evaluations of health and social care services for older people

Nevertheless, the use of CUA in its traditional form may not be completely appropriate in the evaluation of health and social care services for older people. Health improvements need not be the only, or even the primary goal of services for older people. Rather, such services may also be targeted at other life domains than health (captured in QALYs) that are also important to older people's well-being. For example, an intervention facilitating ageing in place needs not necessarily improve older people's health, but it presumably does promote a sense of connection, security and autonomy (WHO, 2015). As these outcomes are valued by older people (Douma et al., 2017; Gabriel and Bowling, 2004), and affect their well-being beyond health, using QALYs in this case is likely to result in policy makers being misinformed about the full benefits of such an intervention to the target population. Consequently, this may lead to inefficient spending of scarce resources. It has been recognised that, compared to conventional HRQoL-measures like the EQ-5D, broader outcome measures are required to establish a more complete evaluation of the benefits of care services for older people (Bulamu et al., 2015; Drummond et al., 2009; Makai et al., 2014a).

Several well-being measures have been developed in recent years (for an overview, see Bulamu et al., 2015; Cleland et al., 2019; Makai et al., 2014a), but most of them are considered unsuitable for application in CUA. These measures often contain too many items

for older people to self-report (Bowling et al., 2013) and do not have utility scores to reflect the relative importance of the included items to overall well-being (Makai et al., 2014a). Another important shortcoming is that the well-being domains they include are often based on expert views instead of older people's own views regarding what is relevant for their well-being (Bowling et al., 2013; Bowling and Stenner, 2011).

The ICEpop CAPability measure for Older people (ICECAP-O) (Coast et al., 2008; Grewal et al., 2006) and the Adult Social Care Outcomes Toolkit (ASCOT) (Netten et al., 2012), both developed in the United Kingdom (UK), do not have these shortcomings. The ICECAP-O measures capability well-being in five domains: 'attachment', 'security', 'role', 'enjoyment' and 'control', each with four response levels ('none', 'a little', 'a lot' and 'all'). Compared to functionings (i.e. what people *do* or *are*), capabilities capture what people *can do* or *be*, irrespectively of whether or not they choose to do or be this way (Sen, 1982). To illustrate this, two older people may have the same low level of social functioning, where one feels lonely but has no capability to have more social interactions, while the other prefers to be alone but has the capability to have more social interactions. The ASCOT covers eight domains that presumably are most affected by social care: 'personal safety', 'control over daily life', 'occupation', 'social participation and involvement', 'personal cleanliness and comfort', 'food and drink', 'accommodation cleanliness and comfort' and 'dignity'. Four response levels aim to capture both capabilities ('ideal state') and functionings ('no needs', 'some needs' and 'high needs'). Both the ASCOT and the ICECAP-O can be transformed into a utility scale ranging from 0 to 1, with higher scores reflecting greater well-being, and the ASCOT having negative values to account for well-being states worse than being dead.

1.4 | Psychometric properties of possible outcome measures

Developing broader outcome measures for use in economic evaluations of health and social care services for older people is necessary but not sufficient. These measures of course also need to be fit for purpose in terms of their psychometric properties. To date, limited evidence suggests that the ASCOT and the ICECAP-O have acceptable levels of (test-retest) reliability, and that the ICECAP-O is sufficiently responsive to change (Van Leeuwen et al., 2015; Netten et al., 2012; Proud et al., 2019; Rand et al., 2017). Compared to the ASCOT, the ICECAP-O has been widely validated (Makai et al., 2014a; Proud et al., 2019), with head-to-head comparisons of both measures being rare (Rand et al., 2017; Van Leeuwen et al.,

2015). The ASCOT and the ICECAP-O demonstrated good construct validity through their comparison with socio-demographic characteristics, HRQoL and existing well-being measures (Makai et al., 2014b; Proud et al., 2019; Rand et al., 2017; Van Leeuwen et al., 2015). Nevertheless, some issues relating to content validity have been noted (Netten et al., 2012; Proud et al., 2019). While the ICECAP-O is more generic, the ASCOT was specifically designed for use in economic evaluations of social care interventions (Netten et al., 2012). This may limit the suitability of the ASCOT in evaluating the benefits of other types of services, while using different outcome measures as a solution reduces the comparability of outcomes across interventions and further complicates decision-making.

Furthermore, even though health is a crucial element in older people's well-being (Douma et al., 2017; Gabriel and Bowling, 2004), the ASCOT and the ICECAP-O do not directly ask people about their health. Consequently, several studies examined whether both measures indirectly cover health through their items. The ASCOT and the ICECAP-O were found to correlate highly with mental functioning, but moderately with physical functioning (Netten et al., 2012; Van Leeuwen et al., 2015). For the ICECAP-O, these findings are in line with an exploratory factor analysis (EFA) demonstrating the items of the ICECAP-O and the EQ-5D to capture separate factors (Davis et al., 2013). Also, they are in concordance with EFA findings on a sister measure of the ICECAP-O (i.e. the ICECAP-A aimed at the adult population) (Keeley et al., 2016). However, these findings contradict previous studies that displayed a high correlation of the ICECAP-O with physical functioning (Makai et al., 2014b), or fairly similar correlations of the ICECAP-O with physical and mental functioning (Makai et al., 2013). Due to these mixed findings, the relation of both measures with health needs further investigation. As economic evaluations require the full benefit of an intervention to be captured, it is crucial to know which aspects of health, if any, are not adequately covered by the ASCOT and the ICECAP-O. If some aspects are not adequately covered, these could be captured using complementary measures of health. However, this solution would then need to be traded off against the potential problem of double-counting other health effects and the challenge of combining different measures into a single unit of benefit.

Whether existing broader outcome measures sufficiently recognise and reflect the heterogeneity in older people's views on well-being also deserves further investigation. What constitutes well-being for older people likely depends on their personal and contextual

circumstances and preferences. Only a few studies considered this heterogeneity directly (Bowling, 2006; Douma et al., 2017; Farquhar, 1995; Jopp et al., 2015; Wilhelmson et al., 2005), while most other studies only reported sample averages and overall rankings of aspects of well-being. Similar to health, indirect or partial coverage of any important well-being domain (to some older people) may lead to inadequate estimations of the benefits of particular interventions and, as a consequence, potentially to a suboptimal allocation of resources (Brown, 2006). Hence, it seems worthwhile to further explore outcome measures that sufficiently capture all well-being domains that are relevant to older people.

1.5 | **Research objectives**

This thesis focuses on methodological challenges related to economic evaluations of health and social care services for older people, with a special emphasis on testing and developing adequate outcome measures capable of capturing well-being. Several issues are addressed in this context. First, this thesis examines the validity of the ASCOT and the ICECAP-O, by comparing them to each other and to various measures of physical and mental health. Secondly, this thesis examines the possible diversity in older people's views on what constitutes their well-being. Finally, using the insights derived from answering the first two research questions, this thesis reports the development of a new well-being measure for use in economic evaluations of health and social care services for older people, named the Well-being of Older People measure (WOOP). Using mixed methods, the validity, feasibility and reliability of the WOOP are tested.

Summarising, this thesis deals with four research questions:

1. To what extent are the ASCOT and the ICECAP-O valid measures of the well-being of older people?
2. To what extent do views on what constitutes their well-being differ between older people?
3. How can this heterogeneity in views be integrated in a comprehensive measure of the well-being of older people?
4. How valid, feasible and reliable is this new measure of the well-being of older people?

The outline of this thesis is as follows:

Chapter 2 presents a head-to-head comparison of the ASCOT and the ICECAP-O in a sample of social care users in the UK. In chapter 3, the ICECAP-O is tested in terms of its validity, with a special focus on its relation with the EQ-5D, again using a UK sample of older people. Chapter 4 reports the findings of a Q-methodology study performed in the Netherlands to investigate the heterogeneity in views of people aged 65 and older regarding what constitutes their well-being. Chapters 5 describes the development of the Well-being of Older People measure (WOOP), based on the findings from chapter 4. Chapter 5 and 6 present a qualitative and quantitative exploration of the feasibility, reliability and validity of this new measure. Chapter 7 discusses the findings of the previous chapters in relation to the research questions and the findings of previous studies, and addresses the strengths and weaknesses of the studies presented in this thesis. Chapter 7 ends with some research and policy recommendations.

Chapter 2

Valid outcome measures in care for older people: Comparing the ASCOT and the ICECAP-O
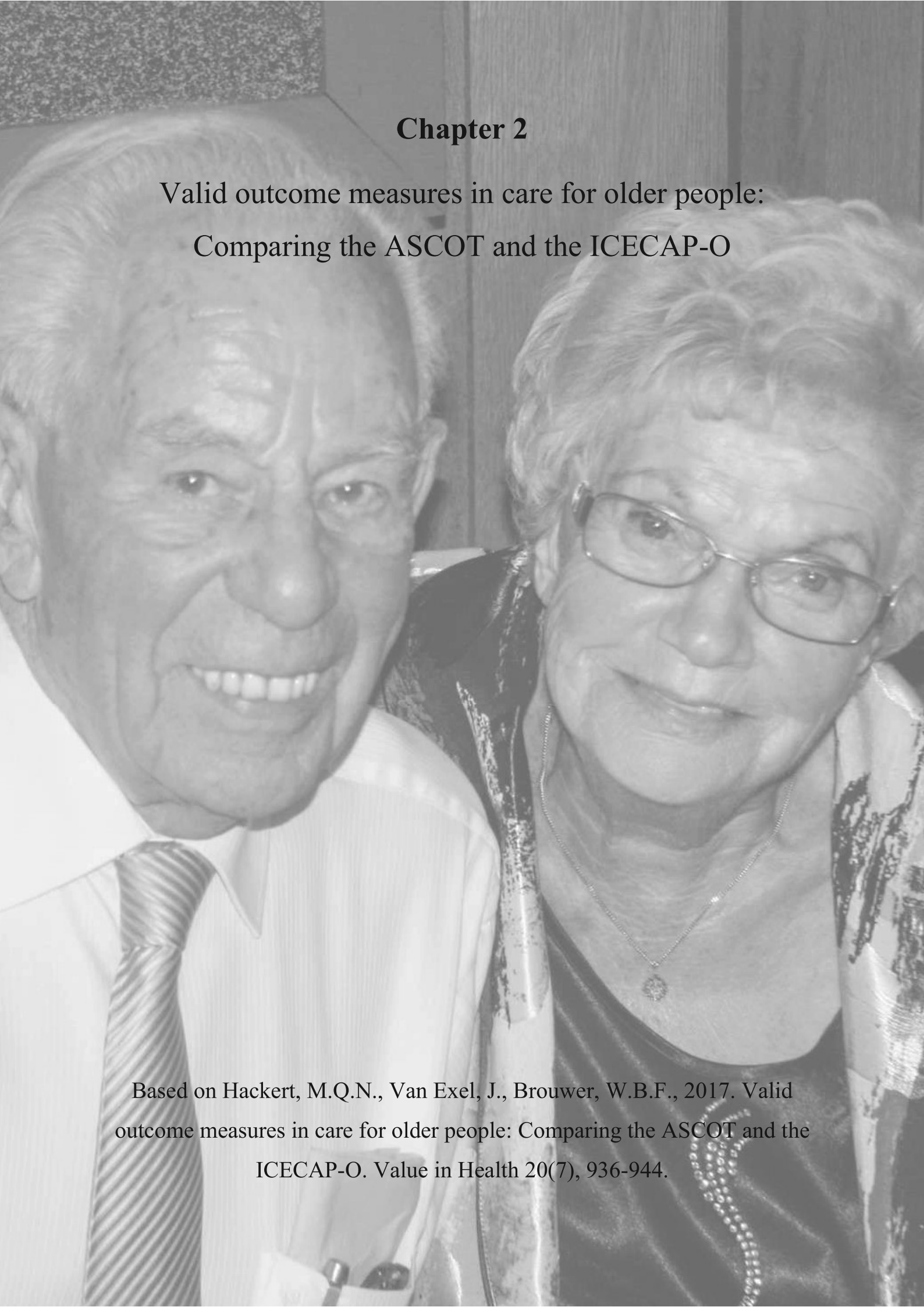
Abstract

Background

Limited health care budgets and population ageing result in a need to compare care services on their benefits and costs. Because services for older people often aim to improve multiple life aspects, valid measures are needed to examine their benefits on individuals' health and well-being simultaneously. Two measures may meet this end: the Adult Social Care Outcomes Toolkit (ASCOT) and the ICEpop CAPability measure for Older people (ICECAP-O). This chapter compares the validity of both measures.

Methods

A sampling agency gathered cross-sectional data in May 2015. Using exploratory factor analysis, the underlying factor structure of the ASCOT and the ICECAP-O was examined for the first time. Convergent and discriminant validity in relation to health measures (five-level EuroQol five-dimensional questionnaire, EuroQol Visual Analogue Scale, Barthel Index and Geriatric Depression Scale-15) and well-being measures (Older People's Quality of Life questionnaire-13, Satisfaction with Life Scale and Cantril's Ladder) were tested using Spearman rank correlations and variance analysis.

Results

The ASCOT and the ICECAP-O tapped into a shared factor, whereas both measures also loaded on two separate factors. The ASCOT and the ICECAP-O correlated highly with the health and well-being measures, but the correlation with the physical health measure Barthel Index was moderate. Both measures discriminated between subgroups of respondents.

Conclusions

The ASCOT and the ICECAP-O seem promising measures to evaluate well-being among older people, whereby the ASCOT seems more specific to social care-related outcomes. The performance of both measures in other respondent groups and countries, and their relation to physical health, need to be further examined before their use in economic evaluations can be recommended.

2

2.1 | **Introduction**

Economic evaluations are increasingly used to advise decision makers about the optimal allocation of scarce resources. Also, in care for older people, information about the costs and benefits of competing interventions helps to fund those services that improve the lives of old adults most efficiently. Benefits are generally assessed using quality-adjusted life-years (QALYs), the arithmetic product of the life expectancy and health status of the individual in these remaining life years. On the basis of individuals' functional abilities in various health domains (e.g. mobility, pain, and anxiety), a multi-attribute utility instrument such as the EuroQol five-dimensional questionnaire (EQ-5D) is typically used to value individuals' health status on a scale from 0 (being dead) to 1 (being in perfect health), with negative values accounting for health states worse than being dead (Drummond et al., 2015; Makai et al., 2014a).

However, using QALYs is likely to result in the underestimation of valuable aspects of care services for older people, because these often aim to improve not only health but also well-being (Bulamu et al., 2015; Drummond et al., 2009; Harwood, 2008; Makai et al., 2014a). Two measures were developed in the United Kingdom (UK) that may allow for a more complete evaluation of services' welfare impact: 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 seem to have a broader evaluative space than do common QALY measures such as the EQ-5D and to date their validity seems promising. The ASCOT (Kaambwa et al., 2015; Malley et al., 2012; Netten et al., 2012; Van Leeuwen et al., 2015) and the ICECAP-O (Couzner et al., 2012; Flynn et al., 2011; Makai et al., 2013, 2014b; Van Leeuwen et al., 2015) generally display moderate to high correlations with well-being and health measures, and sufficient power to differentiate between subgroups of older people.

Nevertheless, important aspects of the validity of the ASCOT and the ICECAP-O remain underexamined. In particular, head-to-head comparisons of both measures are rare, with only one such study published so far (Van Leeuwen et al., 2015). Although the ASCOT and the ICECAP-O lack direct measures of health, the study by Van Leeuwen et al. (2015), performed in the Netherlands, indicated that both measures correlate highly with mental health, whereas they correlate only moderately with physical health. For the ASCOT, these

findings are in line with previous validity tests (Netten et al., 2012), whereas they contradict the expectations of Makai et al. (2014a), who predicted that mental health is likely not fully captured. For the ICECAP-O, these findings contradict the results of some earlier studies (Makai et al., 2013, 2014b), but they are in line with the findings of other previous research (Davis et al., 2013; Makai et al., 2014a). Hence, these mixed results leave a question as to whether the ASCOT and the ICECAP-O should be used alongside health measures such as the EQ-5D because they may complement each other, or whether this would lead to double counting of health effects (Makai et al., 2014a) and the ASCOT and the ICECAP-O should therefore be used in isolation. Using separate health and well-being measures in combination has some disadvantages. First, disentangling these two benefits may be difficult because the relation between well-being and health is complex and bi-directional (Habibov and Afandi, 2016; Sabatini, 2014). Second, having two rather than a single all-encompassing measure of benefit makes decision-making less straightforward. Therefore, further exploration is necessary to investigate whether both measures sufficiently include physical and mental health in their measurement of well-being.

This chapter adds to the validation study of Van Leeuwen et al. (2015) by including a sample of older social care users in the UK, and by using a larger set of measures and a greater variety of validity tests. To explore whether the ASCOT and the ICECAP-O capture distinct or similar aspects of well-being, this study is the first to investigate the dimensionality of both measures through exploratory factor analysis (EFA). Moreover, the relation of both measures with health is explored by correlating and testing their discriminatory power with health and well-being measures, such as the five-level EQ-5D (EQ-5D-5L) and Cantril's Ladder.

2.2 | **Methods**

2.2.1 | **Sampling strategy**

Data were gathered in May 2015 using a Web-based questionnaire. A sampling agency was employed to contact a small, representative group of British social care users of the age of 70 years and older from an online panel. All respondents were selected on the basis of the fact that they had used home help / home care (77%) or day care / day-centre activities (26%) or care home services (including residential care and nursing homes) (17%) in the past two weeks. This criterion needed to be met for the responses on the social care measure ASCOT to be valid. Participation was voluntary, on the basis of informed consent, and could be

terminated at any point. 205 respondents completed the questionnaire. Representativeness in terms of age, sex, and educational level proved difficult in relation to the selection criteria and the online recruitment strategy. On the basis of a pilot study in which respondents needed to properly fill out the questionnaire as quickly as possible, the minimal completion time was set to five minutes. Because no questions could be skipped and the response time of all respondents was more than this threshold, no missing data were reported.

2.2.2 | Dependent variables

The ASCOT (Netten et al., 2012) aims to measure social care-related quality of life. On the basis of the Older People's Utility Scale (OPUS) and mixed methods research, the measure includes eight items to describe well-being: 'personal safety', 'control over daily life', 'occupation', 'social participation and involvement', 'personal cleanliness and comfort', 'food and drink', 'accommodation cleanliness and comfort', and 'dignity'. Four response levels are defined for each item, on which respondents can describe their well-being level: 'ideal state', 'no needs', 'some needs', and 'high needs'. According to Netten et al. (2012), the 'ideal state' was added to the response levels to indicate individuals' capabilities instead of functionings as reflected in the other three levels. Functionings cover what individuals do to achieve well-being, whereas capabilities indicate individuals' ability to achieve certain well-being states, irrespectively of doing so. Developed by a combination of time trade-off (TTO) and best-worse scaling (BWS), UK population values were used to weigh the items in accordance to their relative importance, which led to a final scale ranging from -0.171 to 1. Higher scores reflect greater well-being, with negative values accounting for care states considered worse than being dead.

The ICECAP-O (Grewal et al., 2006) intends to capture capability well-being among older people. After extensive qualitative research, five items were selected to measure well-being: 'attachment', 'security', 'role', 'enjoyment', and 'control'. Four response levels were defined for respondents to describe their level of ability to achieve these well-being states: 'all', 'a lot', 'a little' and 'none'. Population values were calculated by BWS in the UK (Coast et al., 2008) and scores on the items were converted to a 0 to 1 scale, with higher scores indicating greater well-being.

2.2.3 | **Independent variables**

Information was obtained on individuals' age, sex, marital status, educational level, and income. Furthermore, various health and well-being measures were included in the questionnaire. Health was assessed by the valid and reliable EQ-5D-5L (Van Reenen and Janssen, 2015), EuroQol Visual Analogue Scale (EQ-VAS) (Van Reenen and Janssen, 2015), Barthel Index (Gompertz et al., 1994; Hartigan, 2007) and Geriatric Depression Scale-15 (GDS-15) (Smarr and Keefer, 2011; Yesavage and Sheikh, 1986). The EQ-5D-5L (Van Reenen and Janssen, 2015) examines general health using five items: 'mobility', 'self-care', 'usual activities', 'pain and discomfort', and 'anxiety and depression'. Five levels, indicating no problems to extreme problems, are used by individuals to describe their own health status on these items. By summing the scores, a 0 to 20 misery index (MI) was calculated, with higher scores indicating greater health problems. Also, a utility score, ranging from -0.281 to 1, was calculated on the basis of population values retrieved in England (Devlin et al., 2016). Higher scores indicate a higher health-related quality of life and negative values account for health states considered worse than being dead. Because results proved to be robust for both scales, only the analyses with the utility scores are presented. The EQ-VAS (Van Reenen and Janssen, 2015) comprises a vertical, visual analogue scale on which respondents can indicate their health. The scale ranged from 0 (worst imaginable health) to 100 (best imaginable health). The self-completion postal version of the Barthel Index (Gompertz et al., 1994) assesses physical health by determining individuals' ability to perform activities of daily living. Summing the scores on ten items, the scale ranged from 0 to 20, with higher scores indicating greater physical health. The GDS-15 (Yesavage and Sheikh, 1986) examines mental health using 15 items on which respondents indicate depressive symptoms (yes or no) during the past week. Final scores ranged between 0 and 15, with a score of 15 indicating bad mental health.

Well-being was measured by the valid and reliable Older People's Quality of Life questionnaire-13 (OPQOL-13) (Bowling et al., 2013), Satisfaction with Life Scale (SWLS) (Pavot and Diener, 1993) and Cantril's Ladder (Cantril, 1965). The OPQOL-13 (Bowling et al., 2013) includes 13 statements to measure well-being among older people. A five-point response scale ranging from 'strongly disagree' to 'strongly agree' was used to sum the scores on the items to a scale ranging from 13 to 65. The SWLS (Pavot and Diener, 1993) examines respondents' global cognitive judgment of one's life satisfaction. A seven-point response

scale on five items was used to sum the scores on the items to a scale ranging from 5 to 35. Cantril's Ladder (Cantril, 1965) uses a vertical, visual analogue scale in which respondents can indicate to what extent they had the best possible life. An 11-rung ladder was used ranging from 0 (worst possible life) to 10 (best possible life).

To facilitate comparison between measures within the context of this study, the scores of all variables except the ASCOT, ICECAP-O, and EQ-5D-5L were linearly rescaled to a range of 0 to 1. For comparison of scores with the literature, the original scores are also presented within parentheses in Table 2.1.

2.2.4 | Analytic strategy

Data were analysed using the R package (R Foundation for Statistical Computing, Vienna, Austria). The association between the items of the ASCOT and the ICECAP-O was examined using Spearman rank correlations. The guidelines of Hopkins (2002) were applied to evaluate their strength: < 0.10, trivial; 0.10 – 0.29, small; 0.30 – 0.49, moderate, 0.50 – 0.69, high; 0.70 – 0.89, very high; ≥ 0.90, (nearly) perfect. To determine whether the items of the ASCOT and the ICECAP-O could be reduced to the same underlying constructs, EFA (Beaujean, 2013) was used to investigate both measures' dimensionality. EFA is a multidimensional scaling technique that uses the correlation matrix of the items of the ASCOT and the ICECAP-O to reduce them to a smaller set of underlying theoretical constructs, which are also called factors. Each construct can be interpreted by defining the common denominator among those items that were scaled to it. In contrast to the sister method, principal component analysis, EFA realistically assumes that a part of the variance of each item is unique, which holds that the variance of each item cannot be fully explained by the other items. The suitability of EFA was tested using the Bartlett's test of sphericity (Williams et al., 2010), the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (Williams et al., 2010), and the multicollinearity test from the package 'usdm' (Naimi, 2015). On the basis of conventional Pearson correlations and maximum likelihood extraction (ML) (Basto and Pereira, 2012; Flora and Curran, 2004), a scree plot, parallel analysis, and very simple structure (VSS) (Beaujean, 2013; Williams et al., 2010) were considered for extracting the appropriate number of factors. Using the package 'psych' (Revelle, 2015), a number of robustness checks were performed (for more information, see Appendix 2.A). The effect of the possible violation of the assumption of multivariate normality was assessed using principal axing factoring (PAF) (Baglin, 2014). The influence of the possible violation of the

Pearson assumption of linearity was examined using polychoric correlations (Basto and Pereira, 2012; Flora and Curran, 2004). During interpretation of the models, Oblimin rotation was applied using the package 'GPArotation' (Bernaards and Jenrich, 2015) to allow factors to be correlated. Only the highest factor loading for each item was reported, which needed to be greater or equal to 0.40 to be regarded as reliable for interpretation (Hair et al., 1998). All estimated models demonstrated marginal differences in their main outcomes. Therefore, the most commonly used EFA model based on Pearson correlations and maximum likelihood extraction is reported, because this model gives insight into the uniqueness of all items.

In addition, the convergent validity of the ASCOT and the ICECAP-O was determined by calculating Spearman rank correlations to test their association with the health (EQ-5D-5L, EQ-VAS, Barthel Index and GDS-15) and well-being (OPQOL-13, SWLS and Cantril's Ladder) measures. Discriminant validity was examined by calculating the ability of the ASCOT and the ICECAP-O to differentiate between older people with diverse background characteristics, and different levels of health and well-being (i.e. above-average as compared with equal or below-average scores in the sample). One-way analyses of variance were performed in combination with the Levene's test on the violation of the homogeneity of variances. Although one-way analyses of variance are proven to be robust against violations of the normal distribution assumption (Schmider et al., 2010), severe deviations on the dependent variables were checked using quantile-quantile plots.

2.3 | **Results**

2.3.1 | **Sample characteristics**

The descriptive statistics of the study sample are presented in Table 2.1. On average, the respondents were 76 years old. Half of them were women and 56% lived with a partner. About 55% completed at least further education college. More than 44% had a gross monthly income of at least £1600, and 68% reported to make ends meet (fairly) easily.

On average, participants were in fair health. They had mean scores of 0.66 on the EQ-5D-5L and 0.60 on the EQ-VAS. More specifically, they reported a high level of independency with a mean Barthel Index score of 0.84, and a fair level of depressive symptoms with a GDS-15 score of 0.34.

Well-being, measured by the ASCOT and the ICECAP-O, was high. On average, the older people scored 0.81 on the ASCOT and 0.79 on the ICECAP-O. These mean values are similar to those obtained by Van Leeuwen et al. (2015). More in particular, most respondents reported high levels of well-being on the items of both measures (see Figure 2.1). The lowest levels were most frequently mentioned on the ASCOT items 'occupation', 'social participation and involvement', and 'control over daily life', and the ICECAP-O items 'security', 'enjoyment', and 'role'. Overall, higher levels of well-being were mentioned on the ASCOT. In contrast, well-being calculated by measures that were not based on population preference weights was moderate, with mean values of 0.72 on the OPQOL-13, 0.59 on the SWLS and 0.65 on Cantril's Ladder.

2
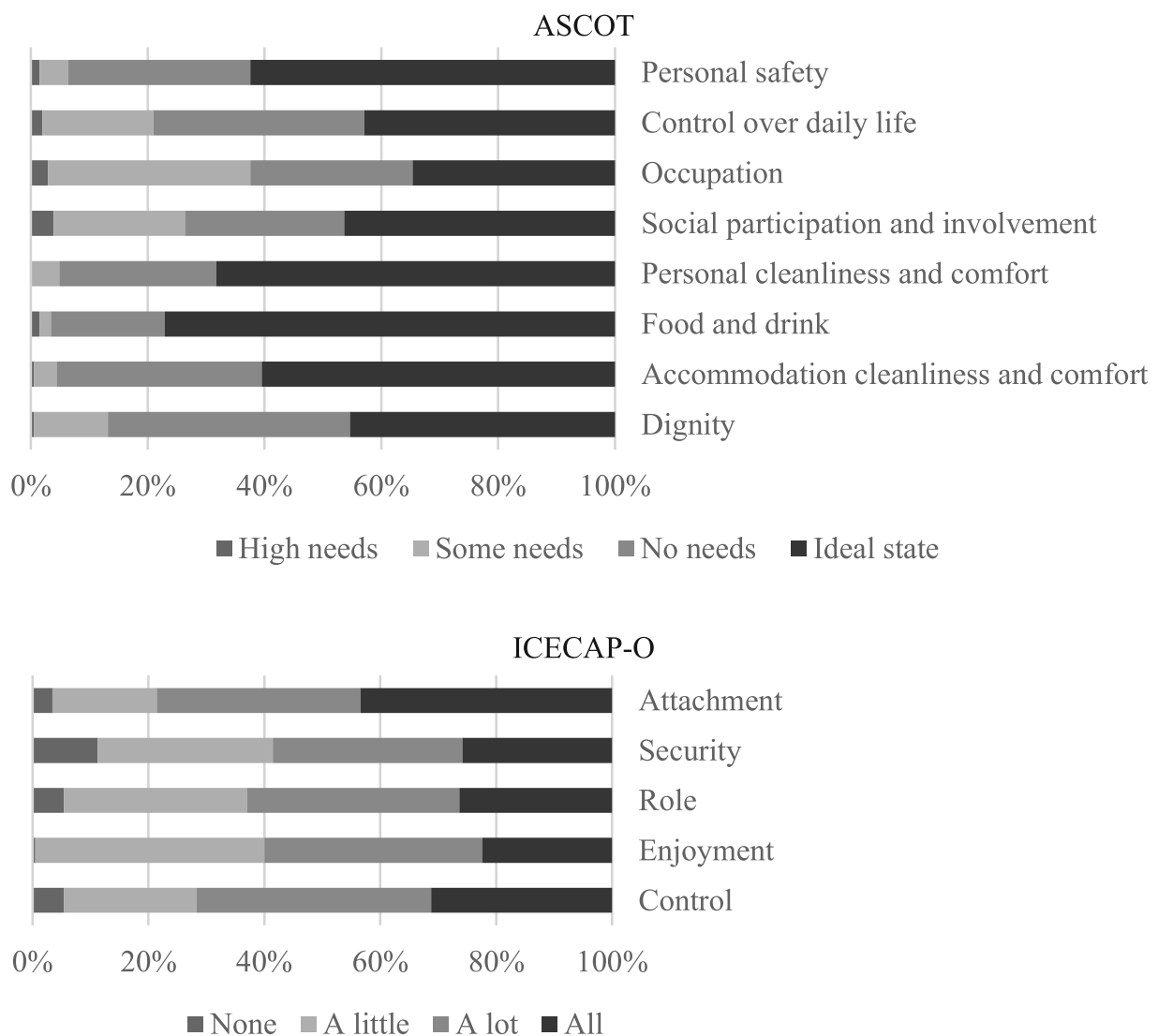
Figure 2.1 | Response distribution on the items of the ASCOT and the ICECAP-O (N=205)

ASCOT, Adult Social Care Outcomes Toolkit; ICECAP-O, ICEpop CAPability measure for Older people

Table 2.1 | Descriptive statistics of the study sample and discriminatory power of the ASCOT and the ICECAP-O (N=205)

					Descriptive statistics		Discriminatory power	
	Mean	SD	Min.	Max.	Categories	%	Mean ASCOT	Mean ICECAP-O
ASCOT	0.81	0.18	0.03	1.00	≤ 0.81	44.88		0.65***
					> 0.81	55.12		0.90
ICECAP-O	0.79	0.17	0.25	1.00	≤ 0.79	45.37	0.67***	
					> 0.79	54.63	0.92	
Age	76.02	5.47	70.00	96.00	≤ 76.02	59.51	0.79	0.78
					> 76.02	40.49	0.83	0.80
Sex					Female	49.76	0.82	0.80
					Male	50.24	0.80	0.78
Marital status					Married or cohabiting as married	55.61	0.82	0.81
					Never married	4.39	0.93	0.83
					Separated	14.15	0.78	0.74
					Widowed	25.85	0.78	0.76
Education					None or primary	9.27	0.75	0.72
					Secondary school	35.61	0.83	0.80
					Further education college or university	55.12	0.81	0.79
Household income					≤ £1599, gross per month	55.61	0.80	0.76*
					> £1599, gross per month	44.39	0.82	0.82
Making ends meet					With great difficulty	5.85	0.73	0.70*
					With some difficulty	25.85	0.78	0.75
					Fairly easy	42.44	0.82	0.79
					Easily	25.85	0.84	0.84

Table 2.1 | Continued

	Descriptive Statistics						Discriminatory power	
	Mean	SD	Min.	Max.	Categories	%	Mean ASCOT	Mean ICECAP-O
EQ-5D-5L	0.66	0.28	-0.23	1.00	≤ 0.66	38.54	0.68***	0.66***
					> 0.66	61.46	0.89	0.87
EQ-VAS [a]	0.60	0.23	0.00	1.00	≤ 0.60	47.80	0.71***	0.69***
	(60.07)	(22.63)	(0.00)	(100.00)	> 0.60	52.20	0.90	0.87
Barthel Index [a]	0.84	0.20	0.05	1.00	≤ 0.84	31.71	0.68***	0.67***
	(16.88)	(3.97)	(1.00)	(20.00)	> 0.84	68.29	0.87	0.84
GDS-15 [a]	0.34	0.27	0.00	1.00	≤ 0.34	60.00	0.89***	0.88***
	(5.16)	(4.02)	(0.00)	(15.00)	> 0.34	40.00	0.69	0.65
OPQOL-13 [a]	0.72	0.16	0.31	1.00	≤ 0.72	47.80	0.68***	0.66***
	(50.47)	(8.21)	(29.00)	(65.00)	> 0.72	52.20	0.92	0.90
SWLS [a]	0.59	0.25	0.00	1.00	≤ 0.59	45.37	0.68***	0.66***
	(22.63)	(7.50)	(5.00)	(35.00)	> 0.59	54.63	0.91	0.89
Cantril's Ladder [a]	0.65	0.22	0.00	1.00	≤ 0.65	46.34	0.70***	0.67***
	(6.45)	(2.16)	(0.00)	(10.00)	> 0.65	53.66	0.90	0.89

[a] Linearly rescaled scores (of 0 - 1); original scores of measures within parentheses.

ASCOT, Adult Social Care Outcomes Toolkit; ICECAP-O, ICEpop CAPabilty measure for Older people; EQ-5D-5L, five-levels EuroQol five-dimensional questionnaire; EQ-VAS, EuroQol Visual Analogue Scale; GDS-15, Geriatric Depression Scale-15; OPQOL-13, Older People's Quality of Life questionnaire-13; SWLS, Satisfaction with Life Scale

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

2.3.2 | Dimensionality

To test the association between the ASCOT and the ICECAP-O, a scatterplot of the overall score of both measures was constructed (see Figure 2.2). In addition, Spearman rank correlations were computed (see Table 2.2). The overall score of the ASCOT correlated very highly with the overall score of the ICECAP-O (0.81).
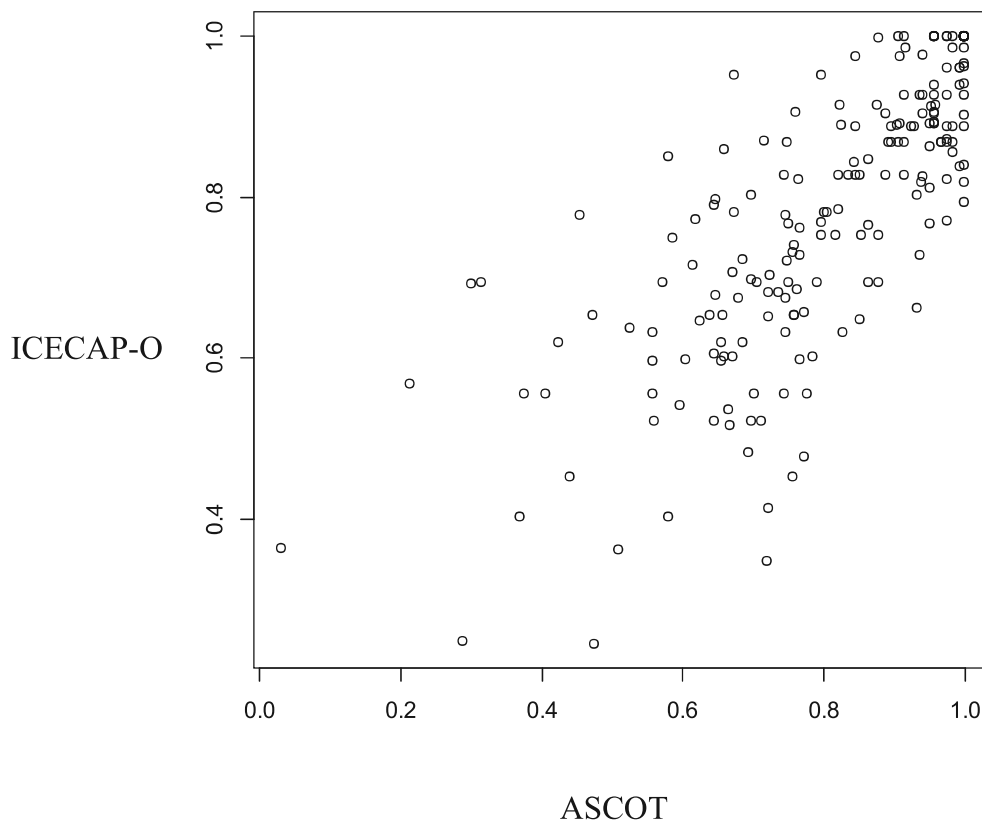
ASCOT

Figure 2.2 | Scatterplot of the overall scores of the ASCOT and the ICECAP-O (N=205)

ASCOT, Adult Social Care Outcomes Toolkit; ICECAP-O, ICEpop CAPability measure for Older people

More in particular, the ASCOT items 'personal safety', 'control over daily life', 'occupation', and 'social participation' correlated (very) highly with the overall score of the ICECAP-O ($\geq$ 0.55), whereas all of the ICECAP-O items correlated highly with the overall score of the ASCOT ($\geq$ 0.56). In addition, many of the items of both measures correlated highly with each other. The only exceptions were the ASCOT items 'personal cleanliness and comfort', 'food and drink', 'accommodation cleanliness and comfort', and 'dignity', which correlated only moderately with the overall score of the ICECAP-O ($\leq$ 0.49) as well as with its items.

Table 2.2 | Spearman correlation matrix of the items and overall scores of the ASCOT and the ICECAP-O (N=205)

ASCOT	ICECAP-O					
	Attach-ment	Security	Role	Enjoy-ment	Control	Overall score
Personal safety	0.44***	0.42***	0.43***	0.37***	0.45***	0.55***
Control over daily life	0.31***	0.50***	0.58***	0.55***	0.59***	0.64***
Occupation	0.46***	0.58***	0.66***	0.68***	0.57***	0.73***
Social participation and involvement	0.52***	0.49***	0.53***	0.55***	0.54***	0.66***
Personal cleanliness and comfort	0.28***	0.31***	0.48***	0.36***	0.50***	0.49***
Food and drink	0.28***	0.26***	0.28***	0.24***	0.25***	0.31***
Accommodation cleanliness and comfort	0.41***	0.37***	0.30***	0.37***	0.29***	0.40***
Dignity	0.36***	0.39***	0.21**	0.25***	0.20**	0.36***
Overall score	0.56***	0.65***	0.66***	0.66***	0.64***	0.81***

ASCOT, Adult Social Care Outcomes Toolkit; ICECAP-O, ICEpop CAPability measure for Older people

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

Table 2.3 presents the results of the EFA. Taken together, the items of the ASCOT and the ICECAP-O led to a three-factor model. The first factor included the ASCOT items 'control over daily life', 'occupation', and 'social participation and involvement', and the ICECAP-O items 'security', 'role', 'enjoyment', and 'control'. All these items had a low uniqueness (≤ 0.48) and together they explained 27% of the variance in the ASCOT and the ICECAP-O items. The second factor contained the ASCOT items 'personal cleanliness and comfort', and 'food and drink', whereas the third factor comprised the ASCOT item 'accommodation cleanliness and comfort' and the ICECAP-O item 'attachment'. Nevertheless, in the principal axis factoring (PAF) model with polychoric correlations, the ASCOT item 'accommodation cleanliness and comfort' did not load on the third factor, but it loaded on the second factor (for more information, see Appendix 2.A). Both the second and third factors explained about 10% of the total variance. The ASCOT items 'food and drink' and 'accommodation cleanliness and comfort' displayed a high uniqueness (respectively, 0.67 and 0.63). The same holds for the ASCOT items 'personal safety' and 'dignity' (≥ 0.64), which displayed no factor loadings higher than the threshold value of 0.40 on any of the three factors (in all but one combination, see Appendix 2.A).

Table 2.3 | Oblimin-rotated factor loadings for the items of the ASCOT and the ICECAP-O (N=205)

		Uniqueness	Factors [a]		
			1	**2**	**3**
ASCOT	Personal safety	0.64			
	Control over daily life	0.46	0.61		
	Occupation	0.35	0.65		
	Social participation and involvement	0.48	0.48		
	Personal cleanliness and comfort	0.27		0.81	
	Food and drink	0.67		0.55	
	Accommodation cleanliness and comfort	0.63			0.42
	Dignity	0.75			
ICECAP-O	Attachment	0.38			0.72
	Security	0.39	0.60		
	Role	0.30	0.86		
	Enjoyment	0.33	0.78		
	Control	0.30	0.82		
Proportion of total variance explained			0.27	0.10	0.10
Correlation with factor 1				0.54	0.43
Correlation with factor 2					0.20

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

ASCOT, Adult Social Care Outcomes Toolkit; ICECAP-O, ICEpop CAPability measure for Older people

2.3.3 | **Convergent validity**

Table 2.4 demonstrates the Spearman rank correlations of the ASCOT and the ICECAP-O with the health and well-being measures. In all cases, the measures demonstrated slightly higher correlations with the ICECAP-O than with the ASCOT.

The ASCOT and the ICECAP-O correlated (very) highly with the general health measures EQ-5D-5L and EQ-VAS and the mental health measure GDS-15. In contrast, both measures correlated only moderately with the physical health measure Barthel Index. Their correlation with the well-being measures OPQOL-13, SWLS and Cantril's Ladder were (very) high. Interestingly, the correlations of the ASCOT and the ICECAP-O items with these measures showed a more diverse picture. For instance, the ASCOT items 'personal cleanliness and comfort', 'food and drink', 'accommodation cleanliness and comfort', and 'dignity' displayed only small to moderate (or even no) correlations with the health and well-being measures.

Table 2.4 | Spearman correlation matrix of the overall scores of the ASCOT and the ICECAP-O with diverse health (EQ-5D-5L to GDS-15) and well-being (OPQOL-13 to Cantril's Ladder) measures (N=205)

	EQ-5D-5L	EQ-VAS	Barthel Index	GDS-15	OPQOL-13	SWLS	Cantril's Ladder
ASCOT							
Personal safety	0.45***	0.45***	0.31***	-0.50***	0.53***	0.48***	0.43***
Control over daily life	0.57***	0.58***	0.45***	-0.54***	0.58***	0.56***	0.51***
Occupation	0.55***	0.51***	0.30***	-0.62***	0.66***	0.65***	0.58***
Social participation and involvement	0.50***	0.48***	0.35***	-0.56***	0.61***	0.59***	0.52***
Personal cleanliness and comfort	0.44***	0.43***	0.44***	-0.45***	0.49***	0.42***	0.41***
Food and drink	0.20**	0.27***	0.27***	-0.23***	0.31***	0.29***	0.26***
Accommodation cleanliness and comfort	0.22**	0.32***	0.16*	-0.34***	0.41***	0.38***	0.40***
Dignity	0.20**	0.33***	0.11	-0.30***	0.35***	0.43***	0.36***
Overall score	0.63***	0.64***	0.45***	-0.69***	0.76***	0.74***	0.66***
ICECAP-O							
Attachment	0.37***	0.39***	0.16*	-0.47***	0.53***	0.60***	0.53***
Security	0.50***	0.53***	0.32***	-0.58***	0.63***	0.71***	0.62***
Role	0.58***	0.57***	0.44***	-0.60***	0.70***	0.63***	0.58***
Enjoyment	0.54***	0.52***	0.33***	-0.62***	0.68***	0.67***	0.62***
Control	0.64***	0.62***	0.53***	-0.54***	0.65***	0.58***	0.53***
Overall score	0.68***	0.66***	0.49***	-0.73***	0.80***	0.82***	0.74***

ASCOT, Adult Social Care Outcomes Toolkit; ICECAP-O, ICEpop CAPabilty measure for Older people; EQ-5D-5L, five-levels EuroQol five-dimensional questionnaire; EQ-VAS, EuroQol Visual Analogue Scale; GDS-15, Geriatric Depression Scale-15; OPQOL-13, Older People's Quality of Life questionnaire-13; SWLS, Satisfaction with Life Scale

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

2.3.4 | Discriminant validity

The discriminatory power of the ASCOT and the ICECAP-O is presented in Table 2.1. On average, a higher score on the ASCOT and the ICECAP-O was found for older people with an above-average level of health (measured by the EQ-5D-5L, EQ-VAS, Barthel Index, and GDS-15) and well-being (measured by the OPQOL-13, Cantril's Ladder, and SWLS). The ICECAP-O also differentiated between respondents on both income variables. Those who had a higher income or were better able to make ends meet reported a higher ICECAP-O score.

2.4 | Discussion

2.4.1 | Main findings

This study was the first to explore the underlying factor structure of the ASCOT and the ICECAP-O. Moreover, through extended measures and validity tests, the measurement properties of the ASCOT and the ICECAP-O, which were found to be promising in previous studies (Couzner et al., 2012; Flynn et al., 2011; Kaambwa et al., 2015; Makai et al., 2013, 2014b; Malley et al., 2012; Netten et al., 2012; Van Leeuwen et al., 2015), were tested further. Using a sampling agency, data were gathered through a Web-based questionnaire distributed among 205 social care users of the age of 70 years and older in the UK. EFA, Spearman rank correlations and variance analysis were applied.

In summary, EFA showed that the ASCOT and the ICECAP-O partly measured the same construct. Both measures, however, also loaded on separate factors. Spearman rank correlations demonstrated that both measures correlated highly with various health (EQ-5D-5L, EQ-VAS, and GDS) and well-being (OPQOL-13, SWLS, and Cantril's Ladder) measures, with the exception of the physical health measure Barthel Index. Variance analysis displayed that the ASCOT and the ICECAP-O discriminated between subgroups of older people with diverse health and well-being levels.

2.4.2 | Comparability with previous studies

The factor structure revealed that most items of the ASCOT and the ICECAP-O loaded primarily on the first factor (see Table 2.3), which supports the assumption that the ASCOT and the ICECAP-O overlap in their measurement of well-being. In addition, the ASCOT item 'accommodation cleanliness and comfort' and the ICECAP-O item 'attachment' loaded on

factor three. On the basis of the content of both items, this results seems counterintuitive, especially because their correlation was only moderate (see Table 2.2). It seemed logical that the ICECAP-O item 'attachment' would load on the same factor as the ASCOT item 'social participation and involvement', in particular because these items were highly correlated. However, when we loosened the assumptions under which the EFA was performed, the ASCOT item 'accommodation cleanliness and comfort' shifted to factor two, on which the ASCOT items personal 'cleanliness and comfort' and 'food and drink' also loaded (see Appendix 2.A). The ASCOT items 'personal safety' and 'dignity' were not associated with any of the factors at all. Overall, it seems that the ICECAP-O item 'attachment', referring to love and friendship, has a different meaning to respondents than the ASCOT item 'social participation and involvement'. Moreover, the ASCOT measures specific social care-related outcomes that are not included in the ICECAP-O, which is in line with the aim and composition of the measure (Netten et al., 2012).

The correlations of the ASCOT and the ICECAP-O with various health (EQ-5D-5L, EQ-VAS, Barthel Index, and GDS-15) and well-being (OPQOL-13, SWLS, and Cantril's Ladder) measures (see Table 2.4) support the convergent validity outcomes of previous studies (Couzner et al., 2012; Kaambwa et al., 2015; Makai et al., 2013, 2014b; Netten et al., 2012; Van Leeuwen et al., 2015). Although the strength of the correlations varied, the results are comparable to those obtained in the Netherlands (Makai et al., 2013; Van Leeuwen et al., 2015), Germany (Makai et al., 2014b) and Australia (Couzner et al., 2012; Kaambwa et al., 2015). Similar to previous studies (Courzner et al., 2012; Makai et al., 2013), the strength of the correlations with the well-being measures was fairly equal to those with the health measures. In line with Netten et al. (2012), Van Leeuwen et al. (2015) and Davis et al. (2013), both the ASCOT and the ICECAP-O displayed a stronger correlation with the mental health measure GDS-15 than with the physical health measure Barthel Index. These findings for the ICECAP-O contradict those of Makai et al. (2013), who found comparable correlations of the ICECAP-O with physical and mental health, and of Makai et al. (2014b), who found a high correlation of the ICECAP-O with physical health.

The power of the ASCOT and the ICECAP-O to differentiate between subgroups of older people (see Table 2.1) supports the main discriminant validity outcomes of previous studies (Couzner et al., 2012; Flynn et al., 2011; Kaambwa et al., 2015; Makai et al., 2013, 2014b; Malley et al., 2012; Netten et al., 2012). Both measures discriminated between respondents

with diverse health and well-being levels (on the basis of the EQ-5D-5L, EQ-VAS, Barthel Index, GDS-15, OPQOL-13, SWLS, and Cantril's Ladder). Also, the ICECAP-O differentiated between participants according to their income, which is interesting in relation to the likely impact of income on capability.

2.4.3 | Study limitations and strengths

Some study limitations need to be considered when interpreting the results. The validity of the study is influenced by the use of data from one country and the specific sample of modest size. The aim was to retrieve, as much as possible, a sample of social care users in the UK representative in terms of age, sex, and educational level. Nevertheless, because of the use of an online panel, people older than 75 years and those with a lower educational level or income were difficult to reach and therefore under-represented in our study. Also, using the Internet as medium could have lowered the level of engagement of respondents, which may have influenced the reliability of the results. Because of the use of a sampling agency, we did not have access to the response rate and data of respondents who did not complete the questionnaire. We acknowledge that this limits our insight into possible issues regarding the data collection and completion of the measures.

Notwithstanding these limitations, this first head-to-head comparison within the UK supports the validity of the ASCOT and the ICECAP-O. A broad selection of health and well-being measures was included to test the convergent and discriminant validity of both measures. For the first time, the factor structure of both measures was examined. The convergence in findings from the diverse estimated models shows the results to be robust, and the comparability with the findings of other studies indicates that they are externally valid. Excluding respondents with a response time of less than 10 or more than 60 minutes did not affect the main results presented in this chapter.

2.4.4 | Research and policy recommendations

The results of this study indicate that the ASCOT and the ICECAP-O are promising measures to evaluate care services for older people by broadening the evaluative space beyond the health domain to well-being. Therefore, both can potentially be applied in a large number of sectors such as social care and institutionalised care.

Which measure to choose may depend on the objective of the study. If the aim is to evaluate social care services, the use of the ASCOT is recommended. In contrast to the ICECAP-O,

the ASCOT covers an additional construct that aims to measure specific social care-related outcomes. However, if the aim is to evaluate well-being among diverse subgroups of older people, using the ICECAP-O may be the better choice, because it measures well-being without including specific care-related or disease-related outcomes. The latter also implies that if comparability across studies and interventions is aimed for, the ICECAP-O may be considered more appropriate.

Nevertheless, further research is required before the use of the ASCOT and the ICECAP-O can be recommended in economic evaluations. Even though both measures demonstrate to have sufficient reliability, responsiveness and validity in our study and the one performed by Van Leeuwen et al. (2015), further research is needed to support these promising results. The EFA should be reproduced in different contexts (e.g. specific disease areas, particular age groups, or other cultural settings). In particular, the limited coverage of the ICECAP-O item 'attachment' by the ASCOT should be re-examined. Beyond the scope of this research, efforts should be made to investigate the measures' sensitivity to change. For the ASCOT and the ICECAP-O to be useful for evaluating competing services in care for older people, they should be able to detect relevant improvements in the health and well-being from such services.

Also, insight is needed in evaluating physical health in services for older people. Especially when we aim to evaluate the benefits of care interventions that affect physical health, among other life aspects, we need an evaluative measure that is capable of detecting these improvements. Although mental health and well-being seem well covered by the ASCOT and the ICECAP-O, the empirical evidence of the limited coverage of physical health is expanding (Davis et al., 2013; Van Leeuwen et al., 2015). Even though the ASCOT was originally intended to be used alongside the EQ-5D whereas the ICECAP-O was not, using the ASCOT or the ICECAP-O in combination with a physical health measure in evaluative studies is not likely to solve this shortcoming, because of the chance of double counting of health effects and the increased difficulty of the decision-making process. The current status quo requires further attention, especially because physical health contributes greatly to old adults' sense of well-being (Gabriel and Bowling, 2004; Nimrod and Ben-Shem, 2015; Wilhelmson et al., 2005).

2.5 | **Conclusion**

This study was the first to explore the underlying factor structure of the ASCOT and the ICECAP-O. Both measures tapped into the same factor, whereas both also loaded on two separate factors. Specifically, the ASCOT covered a construct that aims to measure social care-related outcomes that was not included in the ICECAP-O. The convergent and discriminant validity tests supported the promising validity outcomes of previous studies (Couzner et al., 2012; Flynn et al., 2011; Kaambwa et al., 2015; Makai et al., 2013, 2014b; Malley et al., 2012; Netten et al., 2012; Van Leeuwen et al., 2015). The ASCOT and the ICECAP-O were closely related to diverse health and well-being measures, and both measures differentiated between subgroups of older people. Before the ASCOT and the ICECAP-O should be recommended to evaluate the benefits of care services for old adults, their promising performance should be confirmed in multiple countries and other groups of older people. Moreover, their relation with physical health and their sensitivity to change need to be further examined. In particular, further research should investigate the potential for combining the content of either the ASCOT or the ICECAP-O with health measures, such as the EQ-5D, in a single measure that is sensitive to changes in well-being.

Appendix 2.A | Robustness checks for the oblimin-rotated factor loadings for the items of the ASCOT and the ICECAP-O (N=205)

	Pearson, ML Factors [a,b]			Pearson, PAF Factors [a,c]			Polychoric, ML Factors [a,d]			Polychoric, PAF Factors [a,e]		
	1	2	3	1	2	3	1	2	3	1	2	3
ASCOT												
Personal safety												
Control over daily life	0.61			0.59			0.62			0.59		
Occupation	0.65			0.66			0.66			0.69		
Social participation and involvement	0.48			0.49			0.50			0.55		
Personal cleanliness and comfort		0.81			0.82			0.90			0.86	
Food and drink		0.55			0.52			0.68			0.71	
Accommodation cleanliness and comfort			0.42			0.44		0.43	0.44		0.53	
Dignity						0.40						
ICECAP-O Attachment			0.72			0.70			0.73			0.68
Security	0.60			0.57			0.68			0.66		
Role	0.86			0.86			0.90			0.91		
Enjoyment	0.78			0.78			0.75			0.76		
Control	0.82			0.82			0.85			0.84		
Proportion of total variance explained	0.27	0.10	0.10	0.30	0.12	0.12	0.33	0.17	0.13	0.34	0.18	0.12
Correlation with factor 1		0.54	0.43		0.53	0.44		0.59	0.43		0.60	0.36
Correlation with factor 2			0.20			0.22			0.26			0.25

[a] Only presented are the highest factor loadings per item, which are also ≥ 0.40. [b] Replicated from Table 2.3. [c] Model checks effect of loosening assumption of multivariate normality. [d] Model checks effect of loosening linearity assumption. [e] Model checks effect of loosening both assumptions simultaneously.

ASCOT, Adult Social Care Outcomes Toolkit; ICECAP-O, ICEpop CAPability measure for Older people

2

Chapter 3

Does the ICECAP-O cover the physical, mental and social functioning of older people in the UK?

Abstract

Background

The ICEpop CAPability measure for Older people (ICECAP-O) is intended for use in economic evaluations of care services for older people. Although studies support the validity of the ICECAP-O, it does not directly ask older people about their health. This raises questions about its ability to capture health indirectly. Previous studies found mixed results in this aspect, especially for physical health. This chapter further investigates whether the ICECAP-O indirectly includes health.

Methods

Using a cross-sectional design, a sampling agency retrieved data from 516 people aged 70 years and older from the United Kingdom through an online questionnaire. The overlap in underlying constructs of the ICECAP-O and five-levels EuroQol five-dimensional questionnaire (EQ-5D-5L) was assessed using exploratory factor analysis. Spearman correlations and variance analysis were conducted by relating the ICECAP-O to measures of physical, mental and social functioning.

Results

The ICECAP-O and EQ-5D-5L items loaded on two factors. Their overlap was limited, as four out of five EQ-5D-5L items loaded on the first factor, while four out of five ICECAP-O items loaded on the second. The ICECAP-O correlated highly with (mental and social functioning) health measures, and was able to differentiate between individuals with different scores on these measures. However, the correlation with the Barthel Index, a measure of physical functioning, was moderate.

Conclusions

The ICECAP-O may not fully cover all aspects of health. Therefore, a complementary health measure should be used in addition to the ICECAP-O to capture the full benefits of care interventions for older people in economic evaluations.

3.1 | **Introduction**

Economic evaluations of care services for older people are indispensable. Due to population ageing, services need to be compared in terms of their costs and benefits to ensure an efficient allocation of finite resources. In economic evaluation studies, benefits of care interventions are often assessed using quality-adjusted life-years (QALYs). These comprise individuals' life duration corrected for their health during those years. Health is typically measured by multi-attribute utility instruments such as the EuroQol five-dimensional questionnaire (EQ-5D). Based on individuals' functional abilities in various health domains (e.g. mobility, self-care and anxiety), this measure values individuals' health on a scale from 0, being dead, to 1, being in perfect health, and negative values accounting for health states worse than being dead. By determining the QALY gain and the incremental costs of an intervention relative to a relevant comparator, those care services can be detected that yield the most health per invested monetary unit (Drummond et al., 2015; Makai et al., 2014a).

However, using QALYs to calculate benefits of care services may not always be appropriate nor lead to an efficient spending of limited care resources. This holds especially when health measures like the EQ-5D do not adequately capture all relevant outcomes of the intervention under study. Services for older people often do not only intend to improve health, but also, or perhaps especially, aim to affect broader well-being. This may include gains in self-management, social relations or enjoyment, which are valued by older people and hence should be taken into account when assessing the benefits of an intervention (Gabriel and Bowling, 2004; Wilhelmson et al., 2005). Failing to do so may lead decision makers to be misinformed about the full consequences of care services. Interventions may then potentially be under- or overvalued. Consequently, this may lead to the suboptimal allocation of scarce care resources (Bulamu et al., 2015; Drummond et al., 2009; Harwood, 2008; Makai et al., 2014a).

Compared to conventional health measures such as the EQ-5D, a more complete evaluation of the benefits of services for older people may be established by the ICEpop CAPability measure for Older people (ICECAP-O) (Grewal et al., 2006). Developed by extensive qualitative research in the United Kingdom (UK), the ICECAP-O focuses more broadly on quality of life rather than solely health to capture capability well-being among older people. Capability well-being focuses on individuals' ability to achieve certain well-being states,

irrespectively of actually doing so. The ICECAP-O measures well-being in five domains (using one item per domain), which are weighted to reflect their relative importance (Coast et al., 2008). To date, validation studies show that the ICECAP-O in general correlates moderately to highly with other health and well-being measures and has sufficient power to differentiate between subgroups of older people (Couzner et al., 2012; Flynn et al., 2011; Makai et al., 2013, 2014b; Van Leeuwen et al., 2015; and chapter 2).

Notwithstanding these promising validity outcomes, some caution is warranted when using the ICECAP-O in economic evaluations. Even though the ICECAP-O is considered a measure of capability well-being, it would be expected to cover health as one of the main pillars of overall well-being (Bulamu et al., 2015; Drummond et al., 2009; Makai et al., 2014a). However, the ICECAP-O does not directly ask older people about their health (Grewal et al., 2006). A number of studies examined if the ICECAP-O is able to indirectly capture health through its items. Using exploratory factor analysis (EFA), Davis et al. (2013) demonstrated that the items of the ICECAP-O and the EQ-5D mainly capture separate factors. Based on this finding, they concluded that the items of the ICECAP-O do not adequately cover physical functioning. Their conclusion is in line with the study outcomes of Keeley et al. (2016), who performed EFA on a sister measure of the ICECAP-O (i.e. the ICECAP-A aimed at the adult population). Similar findings were reported in chapter 2 and by Van Leeuwen et al. (2015), who found that the ICECAP-O correlates highly with mental functioning, but only moderately with physical functioning. Contrary to these findings, Makai et al. (2014b) reported a high correlation of the ICECAP-O with physical functioning. Moreover, Makai et al. (2013) displayed fairly similar correlations of the ICECAP-O with mental and physical functioning. Due to these mixed findings, the relationship between the ICECAP-O and health remains unclear and requires further study. After all, if the aim of an economic evaluation is to capture the full benefit of an intervention for older people, it is important to know which aspects of health are not adequately covered by the ICECAP-O and should be captured using complementary measures of health; or, when a general measure of health like the EQ-5D is used alongside the ICECAP-O, whether this would potentially lead to the double counting of some health effects. Therefore, the extent to which the ICECAP-O includes (physical) health can have consequences both for research as well as for subsequent decision-making and may make both the separate use of the ICECAP-O as well as its combined use with measures like the EQ-5D less straightforward.

Hence, this study aims to further disentangle the relation of the ICECAP-O with health. Using cross-sectional data from the UK, EFA was performed to assess whether the ICECAP-O and the EQ-5D cover similar or distinct theoretical constructs and to determine whether the ICECAP-O may be used as a single comprehensive outcome measure in economic evaluations of care services for older people. Using the broad definition of health by the World Health Organization (WHO, 1948), convergent and discriminant validity tests were conducted to examine the relation of the ICECAP-O with the health aspects physical, mental and social functioning. Spearman rank correlations and variance analysis were used to investigate the ability of the ICECAP-O to differentiate between individuals based on e.g. their health status.

3.2 | Data and methodology

3.2.1 | Sampling strategy

In April – May 2015, data were obtained from 516 British people aged 70 and above using a Web-based questionnaire. The age threshold of 70 years was set based on the increasing age which qualifies for senior status, as the life expectancy and the retirement age keep rising. A sampling agency was instructed to gather a representative sample in terms of age, sex and education, but representativeness proved to be difficult in relation to the selection criteria and the online recruitment strategy. Informed consent was obtained from all respondents included in the study. Participation could be terminated at any point. The minimum response time was set to five minutes based on a pilot study in which individuals were asked to properly fill out the questionnaire as quickly as possible. Because no respondent had a completion time below this threshold and all questions were mandatory, no missing data were reported.

3.2.2 | Measures

The ICECAP-O (Grewal et al., 2006) includes five items of well-being: 'attachment' (love and friendship), 'security' (thinking about the future without concern), 'role' (doing things that make you feel valued), 'enjoyment' (enjoyment and pleasure) and 'control' (independence). Older people can indicate on each item to what extent they can achieve these well-being states using four response levels: 'all', 'a lot', 'a little' and 'none'. Population values were applied to weigh all items to their relative importance (Coast et al., 2008) and obtain an overall score ranging from 0 to 1, with higher scores indicating greater well-being.

Information was collected on respondents' age, sex, education and income. Also, comorbidity was measured by the Charlson Comorbidity Index (Chaudhry et al., 2005). Well-being was measured using the reliable and valid Older People's Quality of Life questionnaire-13 (OPQOL-13) (Bowling et al., 2013), the Satisfaction with Life Scale (SWLS) (Pavot and Diener, 1993) and Cantril's Ladder (Cantril, 1965). The OPQOL-13 (Bowling et al., 2013) consists of 13 health and broader quality of life statements on which respondents can indicate their level of agreement using a five-point response scale. Summing the responses to the statements leads to a score ranging from 13 to 65, with higher scores indicating greater well-being. The SWLS (Pavot and Diener, 1993) uses five items to measure individuals' level of life satisfaction. Respondents use a seven-point response scale to indicate their level of agreement with each item. A score ranging from 5 to 35 can be calculated by summing the item scores, with higher scores indicating greater satisfaction with life as a whole. Cantril's Ladder (Cantril, 1965) comprises a visual analogue scale (VAS) in the shape of a ladder ranging from 0 (worst possible life) at the bottom to 10 (best possible life) at the top, on which participants can indicate how they perceive their life.

Happiness was assessed using the Subjective Happiness Scale (SHS) (Lyubomirsky and Lepper, 1999) and Happiness Visual Analogue Scale (Happiness VAS). The SHS (Lyubomirsky and Lepper, 1999) is a valid and reliable measure that consists of four items on which individuals need to rate their happiness relative to others using seven response levels. By averaging responses to the four items a score ranging from 1 to 7 was retrieved, with higher scores indicating greater happiness. The Happiness VAS comprised a horizontal bar on which individuals could indicate their level of happiness. The bar ranged from 0 to 10, with a 10 indicating the highest level of happiness.

Health was measured using the reliable and valid five-levels EuroQol five-dimensional questionnaire (EQ-5D-5L) (Van Reenen and Janssen, 2015) , EuroQol Visual Analogue Scale (EQ-VAS) (Van Reenen and Janssen, 2015), Barthel Index (Gompertz et al., 1994), Geriatric Depression Scale-15 (GDS-15) (Smarr and Keefer, 2011; Yesavage and Sheikh, 1986) and Brief Loneliness Scale (BLS) (Hughes et al., 2004). The EQ-5D-5L (Van Reenen and Janssen, 2015) assesses generic health using five items: 'mobility', 'self-care', 'usual activities', 'pain and discomfort' and 'anxiety and depression'. On each item, respondents can choose between five response options to indicate how many problems they experience. A utility score was derived by weighting the relative importance of the items of the EQ-5D-

5L using the value-set for England (Devlin et al., 2016). This score ranges from -0.281 to 1, with higher scores indicating greater health and negative scores accounting for health states worse than being dead. We also performed our analyses with an unweighted index, that contained the summed responses to the five EQ-5D-5L items, ranging from 0 (i.e. perfect health) to 20 (i.e. extreme health problems on each item). This removed the influence of the value set, which has recently been debated (NICE, 2017). As our results proved to be robust for both scenarios, only the analyses with the utility score are reported. The EQ-VAS (Van Reenen and Janssen, 2015) comprises a VAS on which individuals can indicate their level of health. The scale ranges from 0 (worst imaginable health) to 100 (best imaginable health). To capture the diverse elements included in the broad definition of health by the WHO (1948), the Barthel Index (Gompertz et al., 1994) was used to examine respondents' physical functioning. The measure contains ten items on which people can indicate their ability to perform activities of daily living. Applying a scoring system leads to a score ranging from 0 to 20, with higher scores indicating greater physical functioning. The GDS-15 (Yesavage and Sheikh, 1986) was used to assess participants' mental functioning. The measure consists of 15 items on which individuals can indicate whether they experienced depressive symptoms in the past week. By summing the item scores a 0 to 15 scale was calculated, with higher scores indicating less mental functioning. The BLS (Hughes et al., 2004) was used to measure individuals' social functioning. The measure contains three items on which individuals can indicate their perception of social isolation using three response options. By summing the item scores a score between 3 and 9 is obtained, with higher scores indicating less social functioning.

To facilitate comparison between the measures within the context of this study, the scores of all variables, except the ICECAP-O and the EQ-5D-5L, were linearly rescaled to a range between 0 and 1. For comparison with the literature, the original scores are also presented in Table 3.1 within parentheses.

3.2.3 | Analytic strategy

Data were analysed using the R-package (R Foundation for Statistical Computing, Vienna, Austria). Spearman rank correlations were calculated to examine the relation between the items of the ICECAP-O and the EQ-5D-5L. The guidelines of Hopkins (2002) were applied to evaluate their strength: < 0.10, trivial; 0.10 – 0.29, small; 0.30 – 0.49, moderate; 0.50 – 0.69, high; 0.70 – 0.89, very high; ≥ 0.90, (nearly) perfect. To determine whether both

measures partly capture the same underlying constructs, as proposed in Davis et al. (2013), EFA (Beaujean, 2013) was performed. EFA is a multidimensional scaling technique that uses the correlation matrix of the items of the ICECAP-O and the EQ-5D-5L to reduce them to a smaller set of constructs to explore their underlying theoretical structure. Those items that belong to the same construct, also called a factor, can be interpreted to give meaning to that construct. In contrast to the sister technique principal component analysis, EFA realistically assumes that a part of the variance of each item is unique, which holds that the variance of each item cannot be fully explained by the other items. The use of the method in this study was approved based on the Bartlett's test of sphericity (Williams et al., 2010), Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (Williams et al., 2010), and a multicollinearity test conducted using the package 'usdm' (Naimi, 2015). Based on conventional maximum likelihood extraction and Pearson correlations (Basto and Pereira, 2012; Flora and Curran, 2004), diverse methods were used to extract the appropriate number of factors, including a scree plot, parallel analysis and a very simple structure (Basto and Pereira, 2012; Beaujean, 2013; Williams et al., 2010). Using the package 'psych' (Revelle, 2015), multiple robustness checks were performed. Polychoric correlations were used to test the influence of the possible violation of linearity (Basto and Pereira, 2012; Flora and Curran, 2004). Also, principal axis factoring was used to check the impact of the possible violation of multivariate normality (Baglin, 2014). While interpreting the models, oblimin rotation was applied using the package 'GPArotation' (Bernaards and Jennrich, 2015) to allow factors to be correlated. Because the various models displayed minor differences, only the EFA based on maximum likelihood and Pearson correlations is presented, as this model gives insight in the uniqueness of all items. Only the highest factor loading for each item was reported, which needed to be equal or greater than 0.40 to be considered reliable for interpretation (Hair et al., 1998). Next, convergent validity was tested by calculating Spearman rank correlations. The ICECAP-O was related to various well-being (OPQOL-13, SWLS and Cantril's Ladder), happiness (SHS and Happiness VAS) and health (EQ-5D-5L, EQ-VAS, Barthel Index, GDS-15 and BLS) measures. Using the package 'ppcor' (Kim, 2015) Spearman rank correlations were checked, controlling for individuals' age, sex, education and income (all included as interval or ratio measures). Discriminant validity was tested by the ability of the ICECAP-O to differentiate between subgroups of older people, which were created using the background characteristics and previously mentioned measures. In combination with the Levene's test on the violation

of the homogeneity of variances, T-tests were performed for two group comparisons and one-way ANOVA's for multiple group comparisons.

3.3 | Results

3.3.1 | Sample characteristics

The descriptive statistics of the study sample are displayed in Table 3.1. The respondents were on average 75 years old. A slight majority was male and 49% completed at least further education college. About 46% had an income of at least £1600 per month, whereas 70% reported to make ends meet (fairly) easily. The mean comorbidity score was 1.46 with a range between 0 and 7 diseases mentioned. 23% reported no diseases, whereas 37% reported one disease and 40% reported two diseases or more.

The mean overall score on the ICECAP-O was high (0.81). Most of the older people reported a great level of well-being on all ICECAP-O items (see Figure 3.1). The lowest levels were mentioned on the items 'security' and 'enjoyment'. In contrast, well-being measured by the OPQOL-13, SWLS and Cantril's Ladder was moderate to high, with mean values of 0.76, 0.63 and 0.70.
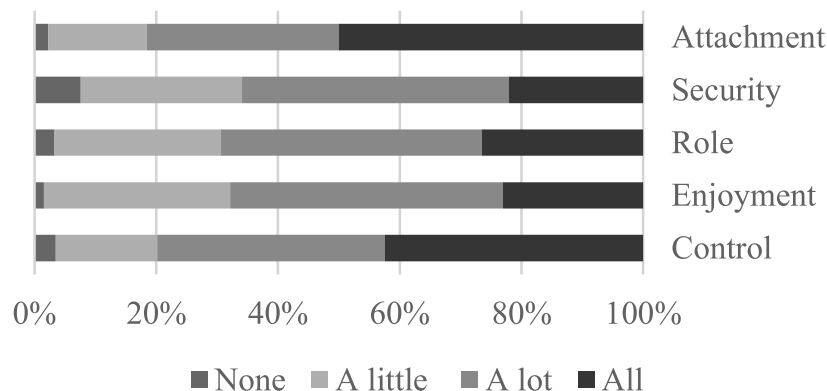
Figure 3.1 | Response distribution on the items of the ICECAP-O (N=516)

ICECAP-O, ICEpop CAPability measure for Older people

Furthermore, older people reported a high level of happiness, with a mean score of 0.70 on the SHS and 0.74 on the Happiness VAS. Comparable levels were derived for health, measured by the EQ-5D-5L (0.74) and EQ-VAS (0.66). On average, the respondents were able to function independently and had only minor signs of depression and social isolation, as evident from the mean values of the Barthel Index (0.90), GDS-15 (0.26) and BLS (0.28).

Table 3.1 | Descriptive statistics of the study sample and discriminatory power of the ICECAP-O (N=516)

	Descriptive statistics						Discriminatory power
	Mean	**SD**	**Min.**	**Max.**	**Categories**	**%**	**Mean ICECAP-O**
ICECAP-O	0.81	0.15	0.25	1.00			
Age	75.08	4.97	70.00	96.00	$\leq$ 75.08	62.02	0.81
					> 75.08	37.98	0.81
Sex					Female	46.32	0.81
					Male	53.68	0.82
Education					None or primary	11.05	0.78
					Secondary school	39.53	0.82
					Further education college or university	49.42	0.81
Household income					$\leq$ £1599, gross per month	54.26	0.79***
					> £1599, gross per month	45.74	0.84
Making ends meet					With great difficulty	4.26	0.72***
					With some difficulty	26.16	0.77
					Fairly easy	42.25	0.83
					Easily	27.33	0.84
Comorbidity	1.46	1.24	0.00	7.00	$\leq$ 1.46	59.69	0.84***
					> 1.46	40.31	0.77
OPQOL-13 [a]	0.76 (52.32)	0.14 (7.35)	0.31 (29.00)	1.00 (65.00)	$\leq$ 0.76	55.23	0.75***
					> 0.76	44.77	0.90
SWLS [a]	0.63 (23.97)	0.22 (6.68)	0.00 (5.00)	1.00 (35.00)	$\leq$ 0.63	41.28	0.70***
					> 0.63	58.72	0.89
Cantril's Ladder [a]	0.70 (6.95)	0.19 (1.91)	0.00 (0.00)	1.00 (10.00)	$\leq$ 0.70	36.24	0.70***
					> 0.70	63.76	0.88
SHS [a]	0.70 (5.20)	0.21 (1.24)	0.00 (1.00)	1.00 (7.00)	$\leq$ 0.70	44.38	0.74***
					> 0.70	55.62	0.87
Happiness VAS [a]	0.74 (7.42)	0.18 (1.82)	0.00 (0.00)	1.00 (10.00)	$\leq$ 0.74	38.95	0.71***
					> 0.74	61.05	0.88
EQ-5D-5L	0.74	0.24	-0.23	1.00	$\leq$ 0.74	32.75	0.69***
					> 0.74	67.25	0.87
EQ-VAS [a]	0.66 (66.44)	0.21 (20.78)	0.00 (0.00)	1.00 (100.00)	$\leq$ 0.66	38.37	0.72***
					> 0.66	61.63	0.87
Barthel Index [a]	0.90 (17.99)	0.15 (3.08)	0.05 (1.00)	1.00 (20.00)	$\leq$ 0.90	27.33	0.71***
					> 0.90	72.67	0.85
GDS-15 [a]	0.26 (3.94)	0.24 (3.62)	0.00 (0.00)	1.00 (15.00)	$\leq$ 0.26	58.33	0.89***
					> 0.26	41.67	0.71

Table 3.1 | Continued

| | Descriptive statistics | | | | | | Discrimina-tory power |
	Mean	SD	Min.	Max.	Categories	%	Mean ICECAP-O
BLS [a]	0.28 (4.66)	0.29 (1.72)	0.00 (3.00)	1.00 (9.00)	≤ 0.28 > 0.28	54.65 45.35	0.88*** 0.74

[a] Linearly rescaled scores (of $0 - 1$); original scores of measures within parentheses.

ICECAP-O, ICEpop CAPability measure for Older people; OPQOL-13, Older People's Quality of Life questionnaire-13; SWLS, Satisfaction with Life Scale; SHS, Subjective Happiness Scale; Happiness VAS, Happiness Visual Analogue Scale; EQ-5D-5L, five-levels EuroQol five-dimensional questionnaire; EQ-VAS, EuroQol Visual Analogue Scale; GDS-15, Geriatric Depression Scale-15; BLS, Brief Loneliness Scale

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

3.3.2 | Dimensionality

In Figure 3.2, the association between the ICECAP-O and the EQ-5D-5L is displayed in a scatterplot. Spearman rank correlations between the overall scores and items of both measures are displayed in Table 3.2, all controlled for age, sex, education and income, and original bivariate correlations are presented within parentheses.
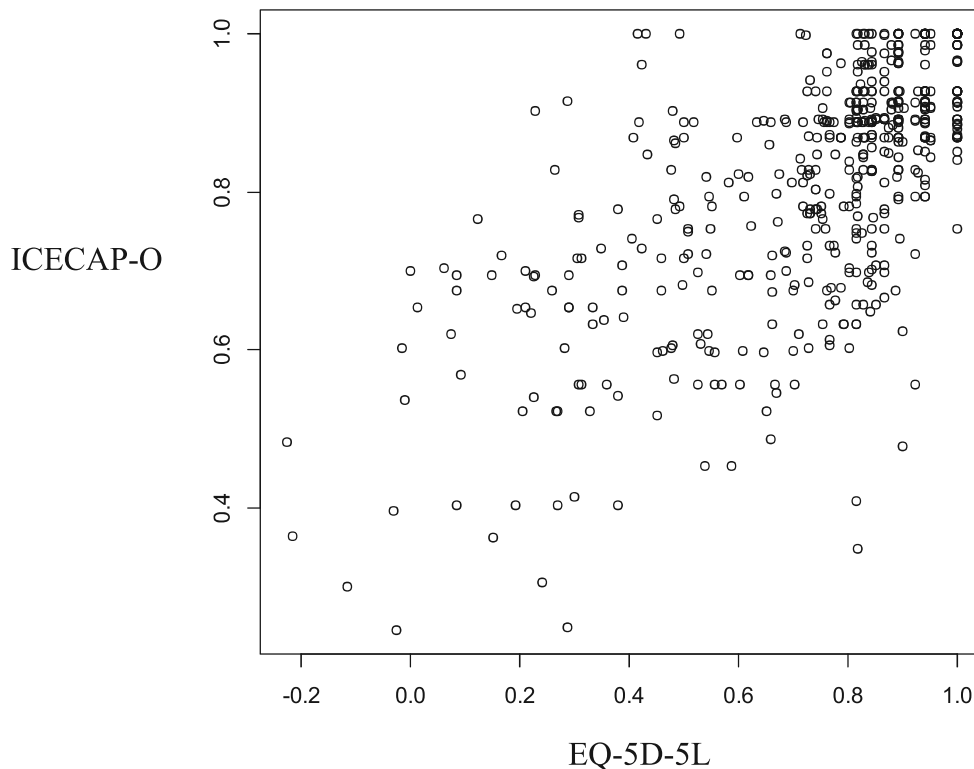
Figure 3.2 | Scatterplot of the overall scores of the ICECAP-O and the EQ-5D-5L (N=516)

ICECAP-O, ICEpop CAPability measure for Older people; EQ-5D-5L, five-levels EuroQol five-dimensional questionnaire

Table 3.2 | Spearman correlation matrix of the items and overall scores of the ICECAP-O and the EQ-5D-5L (N=516). [a]

ICECAP-O	EQ-5D-5L					
	Mobility	Self-care	Usual activities	Pain/ discomfort	Anxiety / depression	Overall score
Attachment	0.15*** (0.15***)	0.15*** (0.15***)	0.15*** (0.16***)	0.10* (0.13**)	0.30*** (0.32***)	0.25*** (0.26***)
Security	0.27*** (0.30***)	0.25*** (0.27***)	0.35*** (0.38***)	0.26*** (0.30***)	0.45*** (0.47***)	0.43*** (0.46***)
Role	0.47*** (0.49***)	0.40*** (0.42***)	0.51*** (0.53***)	0.33*** (0.37***)	0.40*** (0.42***)	0.56*** (0.58***)
Enjoyment	0.40*** (0.43***)	0.31*** (0.34***)	0.43*** (0.46***)	0.32*** (0.36***)	0.44*** (0.46***)	0.52*** (0.55***)
Control	0.58*** (0.59***)	0.54*** (0.55***)	0.63*** (0.64***)	0.44*** (0.47***)	0.35*** (0.37***)	0.62*** (0.65***)
Overall score	0.49*** (0.50***)	0.43*** (0.45***)	0.55*** (0.57***)	0.38*** (0.42***)	0.53*** (0.55***)	0.63*** (0.65***)

[a] Controlled for age, sex, educational level and income; original correlation coefficients within parentheses.

ICECAP-O, ICEpop CAPability measure for Older people; EQ-5D-5L, five-levels EuroQol five-dimensional questionnaire

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

The overall score of the ICECAP-O correlated highly with the overall score of the EQ-5D-5L (0.63). More specifically, the ICECAP-O items 'role', 'enjoyment' and 'control' correlated highly with the overall score of the EQ-5D-5L, while the items 'attachment' and 'security' showed only small to moderate correlations. In addition, the ICECAP-O items correlated at least moderately with the EQ-5D-5L items, except for the items 'attachment' and 'security'.

The EFA results are presented in Table 3.3. When the ICECAP-O items were combined with the items of the EQ-5D-5L, the scree plot, parallel analysis and very simple structure indicated either a two or three factor structure. As the literature supports the two factor structure, we decided to display those results.

The first factor included the ICECAP-O item 'control' and the EQ-5D-5L items 'mobility', 'self-care', 'usual activities' and 'pain and discomfort'. Davis et al. (2013) labelled this factor as 'physical functioning'. All items had a low uniqueness ($\leq$ 0.49) and together, they explained 31% of the total variance of all items. Factor two comprised the ICECAP-O items 'attachment', 'security', 'role' and 'enjoyment' and the EQ-5D-5L item 'anxiety and

depression'. Davis et al. (2013) labelled this factor as 'psychosocial well-being'. In total, they explained 23% of the variance of all items. In particular, the items 'attachment', 'security', 'anxiety and depression' showed high unique variances ($\geq$ 0.54). The overlap between the items of the ICECAP-O and the EQ-5D-5L was even less in the three factor structure, where the ICECAP-O item 'control' loaded on a third factor.

Table 3.3 | Oblimin-rotated factor loadings for the items of the ICECAP-O and the EQ-5D-5L (N=516)

| | | Uniqueness | Factors [a] | |
			1	2
ICECAP-O	Attachment	0.72		0.63
	Security	0.54		0.68
	Role	0.37		0.64
	Enjoyment	0.36		0.78
	Control	0.40	0.58	
EQ-5D-5L	Mobility	0.24	0.90	
	Self-care	0.42	0.76	
	Usual activities	0.19	0.89	
	Pain / discomfort	0.49	0.71	
	Anxiety / depression	0.65		0.55
Proportion of total variance			0.31	0.23
Correlation with factor 1				0.59

[a] Only presented is the highest factor loading per item, which is also $\geq$ 0.40.

ICECAP-O, ICEpop CAPability measure for Older people; EQ-5D-5L, five-levels EuroQol five-dimensional questionnaire

3.3.3 | **Convergent validity**

Table 3.4 demonstrates the Spearman rank correlations of the overall score of the ICECAP-O with the well-being, happiness and health measures. The ICECAP-O correlated (very) highly with the well-being measures OPQOL-13, SWLS and Cantril's Ladder. Comparable correlations were found with the happiness (SHS and Happiness VAS) and health (EQ-5D-5L and EQ-VAS) measures. The ICECAP-O correlated highly with the mental and social functioning measure GDS-15 and BLS. In contrast, its correlation with the physical health measure Barthel Index was only moderate. The correlations of the items of the ICECAP-O with previous measures showed a similar picture, although the correlations were generally lower.

Table 3.4 | Spearman correlation matrix of the items and overall score of the ICECAP-O with diverse well-being (OPQOL-13 to Cantril's Ladder), happiness (SHS to Happiness VAS) and health (EQ-5D-5L to BLS) measures (N=516) [a]

ICECAP-O	OPQOL-13	SWLS	Cantril's Ladder	SHS	Happiness VAS	EQ-5D-5L	EQ-VAS	Barthel Index	GDS-15	BLS
Attachment	0.37***	0.47***	0.40***	0.38***	0.37***	0.25***	0.25***	0.07	-0.32***	-0.51***
	(0.37***)	(0.47***)	(0.40***)	(0.37***)	(0.38***)	(0.26***)	(0.26***)	(0.07)	(-0.32***)	(-0.51***)
Security	0.47***	0.57***	0.48***	0.42***	0.47***	0.43***	0.40***	0.21***	-0.47***	-0.36***
	(0.50***)	(0.59***)	(0.50***)	(0.43***)	(0.48***)	(0.46***)	(0.42***)	(0.23***)	(-0.49***)	(-0.38***)
Role	0.56***	0.55***	0.55***	0.45***	0.51***	0.56***	0.51***	0.38***	-0.52***	-0.44***
	(0.59***)	(0.58***)	(0.58***)	(0.47***)	(0.53***)	(0.58***)	(0.54***)	(0.40***)	(-0.55***)	(-0.46***)
Enjoyment	0.58***	0.61***	0.60***	0.50***	0.56***	0.52***	0.49***	0.30***	-0.55***	-0.50***
	(0.61***)	(0.64***)	(0.63***)	(0.52***)	(0.58***)	(0.55***)	(0.53***)	(0.34***)	(-0.58***)	(-0.52***)
Control	0.52***	0.48***	0.50***	0.37***	0.41***	0.62***	0.55***	0.49***	-0.49***	-0.31***
	(0.55***)	(0.51***)	(0.52***)	(0.39***)	(0.42***)	(0.65***)	(0.58***)	(0.50***)	(-0.52***)	(-0.34***)
Overall score	0.66***	0.72***	0.67***	0.56***	0.61***	0.63***	0.58***	0.39***	-0.64***	-0.55***
	(0.68***)	(0.74***)	(0.69***)	(0.57***)	(0.63***)	(0.65***)	(0.60***)	(0.42***)	(-0.66***)	(-0.57***)

[a] Controlled for age, sex, educational level and income; original correlation coefficients within parentheses.

ICECAP-O, ICEpop CAPability measure for Older people; OPQOL-13, Older People's Quality of Life questionnaire-13; SWLS, Satisfaction with Life Scale; SHS, Subjective Happiness Scale; Happiness VAS, Happiness Visual Analogue Scale; EQ-5D-5L, five-levels EuroQol five-dimensional questionnaire; EQ-VAS, EuroQol Visual Analogue Scale; GDS-15, Geriatric Depression Scale-15; BLS, Brief Loneliness Scale

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

3.3.4 | **Discriminant validity**

In Table 3.1, the discriminatory power of the ICECAP-O is presented. The overall score of the ICECAP-O differentiated between older people based on their income. The group with a higher income or who made ends meet (fairly) easily had a greater mean ICECAP-O score than those who did not reach these levels. Also, the ICECAP-O discriminated between respondents based on their score on the well-being (OPQOL-13, SWLS and Cantril's Ladder), happiness (SHS and Happiness VAS) and health (EQ-5D-5L, EQ-VAS, Barthel Index, GDS-15 and BLS) measures. Those who scored equal or below the average value differentiated from those who scored above this threshold.

3.4 | **Discussion**

3.4.1 | **Main findings**

In a sample of 516 adults aged 70 years and older from the UK, EFA showed that the ICECAP-O and the health measure EQ-5D-5L tap into two shared underlying constructs. Nevertheless, the overlap between both measures was limited, as four out of five EQ-5D-5L items loaded on the first factor, while four out of five ICECAP-O items loaded on the second. Using Spearman rank correlations and variance analysis in combination with a broad range of measures covering the definition of health by the WHO (1948), the convergent and discriminant validity of the ICECAP-O was examined. The ICECAP-O correlated highly with mental (GDS-15) and social (BLS) functioning and overall measures of health (EQ-5D-5L, EQ-VAS), but only a moderate correlation was found with the Barthel Index, a measure of physical functioning. Finally, the ICECAP-O correlated highly with well-being (OPQOL-13, SWLS and Cantril's Ladder) and happiness (SHS and Happiness VAS) measures, and was able to differentiate between subgroups of older people based on their income and their scores on the above mentioned measures.

3.4.2 | **Comparability with previous studies**

This study confirmed that the ICECAP-O largely covers health with the exception of physical functioning. The factor structure of the items of the ICECAP-O and the EQ-5D-5L was similar to that obtained by Davis et al. (2013) and Keeley et al. (2016), although the ICECAP-O item 'control' and the EQ-5D-5L item 'pain and discomfort' loaded on a unique factor in this study. The Spearman rank correlations seem to support this conclusion. In contrast to the

studies of Makai et al. (2013; 2014b), but in line with chapter 2 and Van Leeuwen et al. (2015), the ICECAP-O correlated to a greater extent with the GDS-15, a measure of mental functioning, than the Barthel Index, a measure of physical functioning. Also, in contrast to Makai et al. (2013), the ICECAP-O correlated highly with the social functioning measure BLS. Moreover, the ICECAP-O correlated highly with the generic health measures EQ-5D-5L and EQ-VAS, and diverse well-being (OPQOL-13, SWLS and Cantril's Ladder) and happiness (SHS and Happiness VAS) measures. Although the strength of these correlations varied to some extent, this study broadly confirmed the validity outcomes of previous studies performed in the Netherlands (Makai et al., 2013; Van Leeuwen et al., 2015), UK (chapter 2), Germany (Makai et al., 2014b) and Australia (Couzner et al., 2012). Finally, the research findings supported the discriminatory validity outcomes obtained in previous studies (Flynn et al., 2011; Makai et al., 2013, 2014b; and chapter 2). The ICECAP-O differentiated between older people based on their income, and their score on the measures used to check the convergent validity.

3.4.3 | Study limitations and strengths

Some study limitations are worth mentioning. The results of this study are limited by its cross-sectional design, as we were unable to examine the extent to which the ICECAP-O is sensitive to the impact of various care services and changes in (self-perceived) well-being and health. The generalisability of the study outcomes may be hampered by the use of an online panel. Due to the limited online participation of the eldest and the lower educated, they were underrepresented in this study. We could not access data of respondents who did not complete the questionnaire. We acknowledge that this limits our insight in possible issues regarding the response rate and completion of the measures. Nevertheless, respondents showed considerable heterogeneity, and the study included a wide range of measures to test the validity of the ICECAP-O. Repeating all analyses in a dataset excluding respondents with a response time below 10 or above 60 minutes did not affect our main results.

3.4.4 | Research and policy recommendations

The ICECAP-O may be a promising candidate to replace health measures in evaluation studies of care for older people. The ICECAP-O not only includes specific items on well-being, but is also able to capture aspects of health through them. However, further research is required on several aspects of the measure. So far, analyses on the coverage of similar

underlying factors as the EQ-5D were conducted in Canada (Davis et al., 2013) and the UK. Research in other countries and in diverse subgroups of older people should be stimulated to re-examine the uptake of physical functioning by the ICECAP-O. Whether or not the ICECAP-O and physical functioning are related to each other may depend on the sample under study. Limited overlap in the factor structure and moderate correlations (0.30 – 0.49) between the ICECAP-O and physical functioning were observed in samples of frail older people and social care users (Davis et al., 2013; Van Leeuwen et al., 2015; and chapter 2). On the other hand, high correlations were reported in a sample of post-hospitalised older people (-0.51) (Makai et al., 2013) and a sample of older people diagnosed with dementia (0.72) (Makai et al., 2014b). In particular in the latter study the correlation between the ICECAP-O and physical functioning was very high, but it needs noting that nursing professionals were used as proxy respondents in this study, which may have affected this outcome. More knowledge on the relationship between the ICECAP-O and physical functioning in different samples remains warranted. To strengthen the evidence in this area, future studies could relate the ICECAP-O to clinical measures of physical and mental health in addition to the self-reported questionnaires used here and in other studies. If future studies support the results discussed here, efforts should be made to integrate this aspect of health into the measure, as using the ICECAP-O and a generic measure of health like the EQ-5D simultaneously may lead to the double counting of certain health effects. Also, population values should be developed in other countries to determine the relative importance of the ICECAP-O items there. Finally, the ICECAP-O should be examined on its sensitivity to change. The minimal clinical important difference (MCID) should be calculated to derive the smallest change in the ICECAP-O that is meaningful for the elderly. Using the one-half standard deviation benchmark, in this study the MCID would be 0.08.

3.5 | Conclusion

This study supported the validity outcomes of the ICECAP-O found in previous studies (Couzner et al., 2012; Davis et al., 2013; Flynn et al., 2011; Makai et al., 2013, 2014b; Van Leeuwen et al., 2015; and chapter 2). The ICECAP-O displayed convergent and discriminant validity with diverse well-being, happiness and health measures. However, physical functioning did not appear to be fully captured, as most of the ICECAP-O and the EQ-5D-5L items loaded on different factors in the EFA, and the ICECAP-O correlated moderately

with the Barthel Index. As the ICECAP-O apparently does not fully cover the health effects of interventions, the measure should be used with caution and perhaps in addition to a complementary measure of health to evaluate the full benefits of interventions for older people. How both measures can be combined should be investigated further.

3

Chapter 4

Views of older people in the Netherlands on well-being:
A Q-methodology study

Abstract

Background

Population ageing and restricted budgets result in the need for an efficient allocation of scarce resources in care services for older people. As these services tend to address more than only health, diverse well-being measures have been developed to assess their benefits in economic evaluations. These measures are grounded in research on well-being of older people and its determinants. Little is known about possible heterogeneity in this context and the extent to which well-being measures cover the aspects of well-being that are most important to older people with different views on well-being.

Methods

We conducted a Q-methodology study between December 2016 and October 2017 to investigate the variety in views among people aged 65 and older in the Netherlands on what is important to their well-being. A purposive sample of 53 respondents ranked 34 opinion statements according to importance to their well-being and explained their ranking during a follow-up interview. Data were analysed using by-person factor analysis to identify common patterns in the rankings of the statements.

Results

Five distinct views were extracted in which different aspects were considered important: (i) health, financial security and a life partner; (ii) family, support and physical functioning; (iii) autonomy, mental health and helping others; (iv) social contacts, support, mental health and religion; and (v) a life partner, social contacts, living environment and adaptation.

Conclusions

This heterogeneity in views of older people on what constitutes well-being supports the use of person-centred approaches in care services for older people. Arguably, (evaluations of) policies and services for older people should take this plurality into consideration.

4.1 | Introduction

The ageing of the world population has led to an increasing need for health and social care services. In particular, long-term care expenditures are expected to at least double by 2060 (OECD, 2017). Considering that public resources are finite, priority setting regarding which care services to fund is inevitable. Increasingly, economic evaluations are used to inform such decisions by comparing the benefits of care services to their costs. In this context, benefits are typically assessed in terms of quality-adjusted life-years (QALYs), which comprise both length and health-related quality of life (HRQoL). HRQoL is usually measured and valued using a generic instrument, such as the Euroqol five-dimensional questionnaire (EQ-5D). By considering costs and benefits of care services, those services can be funded that provide the most HRQoL in the population from the available budget (Drummond et al., 2015; Makai et al., 2014a).

However, using this approach may lead to a suboptimal allocation of scarce resources in the context of care services for older people. These services commonly aim to improve the well-being of older people more broadly than only via their HRQoL. These broader benefits should therefore also be considered when such services are evaluated (Bulamu et al., 2015; Drummond et al., 2009; Harwood, 2008; Makai et al., 2014a). To illustrate this, take providing care through day-care centres as an example: this care allows older people to live independently and in the neighbourhood of their social circle for a longer period of time. Even though these benefits may not be part of the calculation of QALYs, they do represent value to the target population (Gabriel and Bowling, 2004; Wilhelmson et al., 2005). Such benefits could be captured by using broader measures of outcome than HRQoL.

In recent years, several well-being measures have been developed for this purpose (for an overview, see Bulamu et al., 2015; Makai et al., 2014a). Important examples include the Adult Social Care Outcomes Toolkit (ASCOT) (Netten et al., 2012) and ICEpop CAPability measure for Older people (ICECAP-O) (Coast et al., 2008; Grewal et al., 2006). The ASCOT aims to capture outcomes in social care settings, including eight domains: 'personal safety', 'control over daily life', 'occupation', 'social participation and involvement', 'personal cleanliness and comfort', 'food and drink', 'accommodation cleanliness and comfort', and 'dignity'. The ICECAP-O was designed as a more generic measure of well-being, covering the broad domains: 'attachment' (love and friendship), 'security' (thinking about the future

without concern), 'role' (doing things that make you feel valued), 'enjoyment' (enjoyment and pleasure) and 'control' (independence). Several studies have shown that these measures appear to be valid and have sufficient power to differentiate between relevant subgroups of older people (Van Leeuwen et al., 2015; and chapters 2 and 3), suggesting they could be useful to assess the broader well-being benefits of policies and services for older people.

However, some studies have also indicated that the ASCOT and ICECAP-O may not cover all potential benefits of care services for older people. Prior to the development of well-being measures, explorations of aspects of life relevant to older people's well-being have been carried out which show some common denominators: physical health, mental health, emotional well-being, social contacts, autonomy, activities, finances and environment (Bowling et al., 2013; Bulamu et al., 2015; Ferrans and Powers, 1985; Grewal et al., 2006; Makai et al., 2014a; Netten et al., 2012; Power et al., 2005). In this context it is important to note that measures like the ASCOT and ICECAP-O do not ask older people directly about their health (Grewal et al., 2006; Netten et al., 2012), while at the same time they also do not seem to fully cover physical health through the items they do include (Davis et al., 2013; Van Leeuwen et al., 2015; and chapters 2 and 3). Furthermore, the developers of the ICECAP-O acknowledge the influence of finances and environment on the well-being of older people, but the instrument measures their impact indirectly as well (Grewal et al., 2006). Consequently, the current guideline in the Netherlands is to use a HRQoL-measure like the EQ-5D alongside the ASCOT or ICECAP-O in economic evaluations of care services for older people (Makai et al., 2014a; Versteegh et al., 2016a). Obviously, using two partially overlapping measures side-by-side instead of a single all-encompassing measure of outcome makes decision-making more complex (chapter 2).

Another important point to consider is that the views of older people on what constitutes well-being are likely to be heterogeneous and adjusted to personal and contextual circumstances. Only a few studies considered this heterogeneity directly (Bowling, 2006; Douma et al., 2017; Farquhar, 1995; Jopp et al., 2015; Wilhelmson et al., 2005), as previous studies mostly focused on 'averages' and overall rankings of aspects of well-being.

Therefore, this study aims to contribute to the identification of aspects of life that should be covered by a comprehensive measure of well-being of older people, taking into consideration the variety in views in this group about what is important for well-being. This study uses Q-methodology to describe the views regarding what constitutes well-being among people aged

65 and older in the Netherlands. Its results may be relevant for the (further) development of comprehensive well-being measures for use in economic evaluations.

4.2 | Materials and methods

Q-methodology combines qualitative and quantitative approaches for the systematic study of subjective phenomena, i.e. opinions, values or beliefs (McKeown and Thomas, 2013; Watts and Stenner, 2012). Generally, a purposively selected group of respondents is asked to rank a set of opinion statements, according to importance or agreement, and to explain their ranking in an interview. By-person factor analysis is used to identify distinct patterns in the rankings of the statements, the quantitative data, which are then interpreted and described as shared views on the topic of study. These interpretations are validated and can be illustrated using the qualitative data obtained from respondents. Over the past ten years, Q-methodology has increasingly been applied in health and social sciences on a variety of topics, for example on views on priority-setting in health care (e.g. Van Exel et al., 2015; McHugh et al., 2015).

4.2.1 | Development of the statement set

To be able to explore what older people consider important to their well-being, the set of opinion statements presented to them in the card sorting exercise needed to represent all potentially relevant aspects. To ensure this, a review of the literature was conducted taking two recent, related reviews on well-being measures in care for older people as a starting point (Bulamu et al., 2015; Makai et al., 2014a). They provided an overview of the aspects that were most important to older people. In addition, from their reference lists, articles were retrieved that described aspects important to older people's well-being as used in the development of the well-being measures that were indicated to be most promising in these reviews (for an overview, see Table 4.1). Articles examining possible heterogeneity in the importance of well-being aspects were also added (Bowling, 2006; Douma et al., 2017; Farquhar, 1995; Jopp et al., 2015; Wilhelmson et al., 2005). An overview was compiled and categorised by one researcher and then reviewed by two other researchers. The list resulting from this iterative process (see Table 4.1) was adopted as the theoretical structure for the development of a representative set of opinion statements. One researcher compiled a long-list of 150 opinion statements, categorised according to the theoretical structure, after which three researchers extensively discussed and refined this list into a comprehensive and manageable set of 34 statements (see the right column of Table 4.1).

Table 4.1 | Theoretical structure for the development of the opinion statements

Domain	Content	Population heterogeneity	Statements (in Dutch available upon request) [a]
Health	Health / functioning [2,3,5,8-11]. No morbidity / impairments [2,10]. Able to do activities [3,8]. Be fit / energetic [9]. No pain [7-9]. Dignity [2,4]. Good death / dying [7]. Lifestyle [11].	▪ *Health / functioning:* (-/+) eldest [8,9,11], (+) women [9], (-) high educated women [9], (+) assisted-living [11], (+) ill women [9]	1. Being physically able to perform my daily activities 2. Having no physical illness or disability 3. Having no physical pain 4. Having the prospect of many life years ahead
Physical	▪ Physical health / functioning [1,2,5,6]. Freedom / mobility [2,8-11]. Sensory abilities [7-9]. Appearance [4,6].	▪ *No pain:* (+) ill [9] ▪ *Physical health:* (+) men [9], (+) high educated men [9], (+) healthy [9], (-) ill men [9] ▪ *Appearance:* (+) ill women (i.e. hygiene) [9]	5. Living a dignified and healthy life 6. Being mentally active 7. Having no problems with my ability to think
Mental	▪ Mental / psychological / cognitive health / functioning [1,2,5,10,11]. Active mind [10]. Able to remember [9]. No depression [9].	▪ *Mental health:* (+) eldest [11], (+) high educated women [9]	8. Being able to communicate well with others 9. Not feeling anxious or depressed
Emotional well-being	Life satisfaction [5,6,10]. Happiness [2,6,8,11]. Feeling good [9]. Peace of mind / be content [2,6,9,11]. Enjoy life [9,10]. Be positive [5,7,8,10,11]. No worries / security (present / future) [1-4]. No stress [6]. Self-acceptance [6]. Coping [3]. Religion / view of life [1-3,5,6,9,11]. Self-realisation [1]. Personality (goals / values) [1,6,9,11].	▪ *Life satisfaction / happiness / enjoy life:* (+) healthy [10]	10. Not being worried about my future 11. Being able to look back on how my life went with satisfaction 12. Accepting my circumstances as they are 13. Being able to adjust to new circumstances 14. My religion or view of life

Table 4.1 | Continued

Domain	Content	Population heterogeneity	Statements (in Dutch available upon request)[a]
Social contacts	Social functioning / life [1,11]. Social relations (e.g. partner, children, (great-)grandchildren, family, friends, colleagues, neighbours, acquaintances, communities / social clubs, pets) [2,3,5,6,8-11]. Social participation [2-5,7]. Connectedness [2,3,5,7]. No loneliness [8,10,11]. Receiving support [3,5,11]. Helping others [3,9]. Health loved ones (no deaths) [8,11].	▪ *Social relations:* (-) eldest [9], (+) women [9,10], (+) high educated women [9], (+) assisted-living [11], (-/+) ill [9,10] ▪ *Partner:* (-) eldest [11], (-) women [11], (-) assisted-living [11] ▪ *Diverse relations:* (-) eldest (i.e. friends, social clubs) [11], (-) assisted-living (i.e. social clubs, colleagues, pets, communities) [11], (+) assisted-living (i.e. neighbours) [11] ▪ *Social participation:* (-) men [10], (+) high occupation [10], (+) healthy [10], (+) mobility [10] ▪ *Helping others:* (+) men [11], (+) community-dwelling [11] ▪ *Health loved ones:* (+) women [11]	15. Having a life partner 16. Having a good relationship with my close family 17. Having contact with people I care about 18. Receiving support from family, friends or neighbours when I need it 19. Feeling valued by helping others

4

Table 4.1 | Continued

Domain	Content	Population heterogeneity	Statements (in Dutch available upon request) [a]
Autonomy	Autonomy [1,7]. Independence [1-3,5,9-11]. In daily activities [2]. Freedom [1,5,7]. To make decisions [7]. Be in control [1-5,7]. To do what you want to do [7,9]. The way you use your time [7].	▪ *Independence*: (-) mobility [10]	20. Not being dependent on others for daily activities 21. Being able to go wherever I want to go 22. Having the freedom to make my own choices 23. Being able to spend my time as I want 24. Doing the things I consider important in life
Activities	Activities [3,5,7-9,11]. Activity level [1,7]. Purpose in life / achievement / self-worth [1,3,7,9]. Occupation (work / volunteering) [1,3,4,6,10,11]. Personal development (learning new things) [11]. Participate in communities [7,10]. Pleasure / leisure activities [1-3,5,6,8-11].	▪ *Activities*: (+) youngest [10,11], (+) community-living [11] ▪ *High-intensity activities*: (-) eldest [11], (-) assisted-living [11], (-) ill women (i.e. household work) [9] ▪ *Occupation*: (+) youngest [11], (+) men [11], (+) community-dwelling [11]	25. Learning new skills 26. Meeting new people 27. Having the feeling that I matter 28. Staying involved in the issues at play in my community
Finances	Adequate finances [1,3,5,8-11]. Pay bills [5,10]. Standard of living [1,6]. Wealth [2,3]. For security [3], control / independence [3,9], enjoyment [3]. No worries [9].	▪ *Finances*: (+) married [10], (+) high occupation [10], (+) home owners [10]	29. Having no financial concerns

Table 4.1 | Continued

Domain	Content	Population heterogeneity	Statements (in Dutch available upon request) [a]
Environ-ment	Environment [1,5,10,11]. Surroundings [3]. Societal criticism [11]. Political situation [11]. Community [1,2,10]. Neighbourhood [1,5,10]. Resources / facilities [5,10,11]. Accessibility [2]. Comfort [4,9,11]. Security / safety [1,3,4,11]. Home [2-5,8,9,11]. Keep own home [9]. Keep own things [11].	■ *Neighbourhood:* (-) men [10], (+) mobility [10] ■ *Home:* (+) eldest [11], (+) assisted-living [11] ■ *Keep own home:* (+) eldest [9], (-) high educated men [9] ■ *Safety:* (-) eldest [11], (+) women [11] ■ *Amenities:* (+) eldest [11], (-) assisted-living [11] ■ *Outside:* (-) eldest [11], (+) women [11]	30. Independent living 31. Living in an area in which I feel comfortable 32. Living in an area with sufficient facilities 33. Living in a stable environment with minor risk of social unrest 34. Living in a country in which my interests are heard and represented

[a] Initial long-list of opinion statements based on the ICECAP-O (Coast et al., 2008; Grewal et al., 2006), ASCOT (Netten et al., 2012), Older People's Quality of Life questionnaire (OPQOL-35,-13) (Bowling et al., 2013; Bowling and Stenner, 2011), Control Autonomy Self-realisation and Pleasure (CASP-19) (Hyde et al., 2003), Ferrans and Powers Quality of Life instrument (Ferrans and Powers, 1985), World Health Organisation Quality of Life instrument (WHOQoL-Old, -Bref) (Power et al., 2005; The WHOQOL Group, 1998).

Note: [1] Makai et al., 2014a, [2] Bulamu et al., 2015, [3] Grewal et al., 2006, [4] Netten et al., 2012, [5] Bowling et al., 2013, [6] Ferrans and Powers, 1985, [7] Power et al., 2005, [8] Farquhar, 1995, [9] Wilhelmson et al., 2005, [10] Bowling, 2006, [11] Douma et al., 2017.

Jopp et al., 2015 was not included, as their study sample to identify the most important aspects to ageing also consisted of young adults. We do acknowledge that views on well-being may differ to respondents' cultural background.
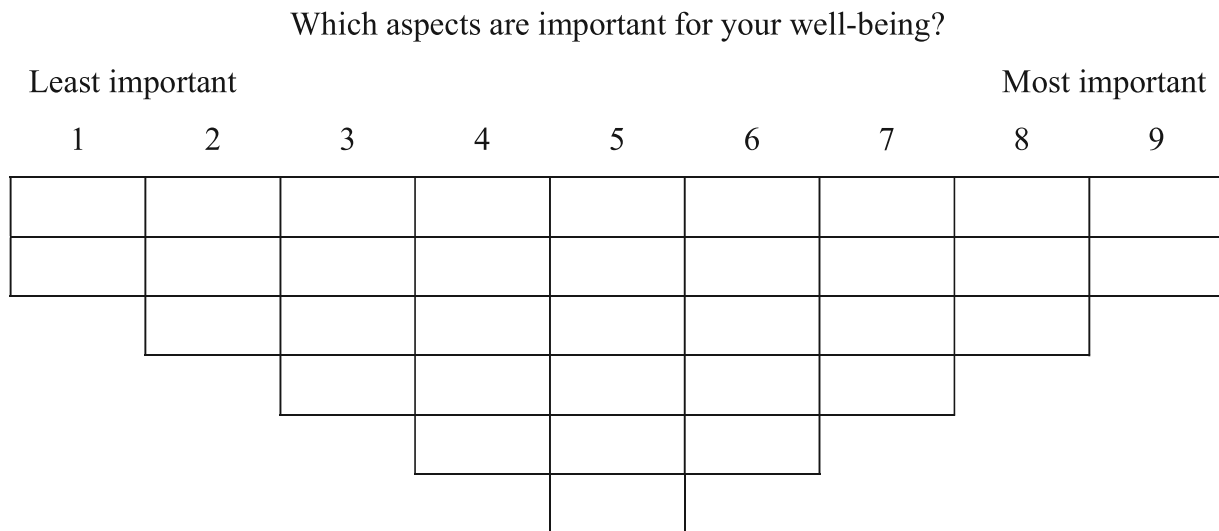
Each domain included as many statements as needed to cover the variety of aspects considered relevant within that domain. The statements were numbered and printed on equally sized and styled cards. A forced-choice sorting grid was defined, consisting of a quasi-normal and symmetrical distribution ranging from 1, the least important aspects, to 9, the most important aspects to older people's well-being (see Figure 4.1).

Which aspects are important for your well-being?

Least important   Most important

   1   2   3   4   5   6   7   8   9

Figure 4.1 | The sorting grid.

20 individual pilot interviews were conducted with people aged 65 years and older to ensure that the statements covered all relevant aspects contributing to well-being in old age and the research materials were clear and practical. Respondents first performed the ranking exercise, after which they were asked to reflect on the representativeness of the statement set and the feasibility of the sorting procedure. Based on these individual interviews no changes in content or wording of statements and instructions were needed.

4.2.2 | Sampling strategy

A total of 53 interviews were conducted by one researcher in the Netherlands between December 2016 and October 2017. Through purposive sampling, we aimed to recruit respondents that jointly would present the variety of views on well-being prevalent among older people. Recruitment was based on a set of socio-demographic characteristics that were shown to be related to well-being within this population (see Table 4.1). A sampling agency recruited 33 respondents, all living independently, in two waves. First, 20 respondents were selected based on their age and gender for inclusion in the pilot tests. They could be retained for the main study as no changes were needed in the study materials. Secondly, 13 additional

respondents were recruited based on their education level and cultural background (e.g. Surinamese and Antillean). As some relevant groups of older people (e.g. living in a nursing home, with a Turkish or Moroccan background) were underrepresented in the panel of the sampling agency, additional respondents were recruited by contacting nursing homes, day-care facilities and interest groups for the target population. In each of three nursing homes located in three different regions (Swalmen, Zeist and Zwolle) four respondents were included. To complete the study sample, eight respondents were recruited through two day-care centres in Amsterdam and through snowballing. Interviews took place at the Erasmus University in Rotterdam, at the day-care centres or at older people's own homes. A small financial incentive (i.e. a gift voucher of €20) was used to recruit respondents from underrepresented groups.

4.2.3 | **Data collection**

A similar data collection process was used in the pilot study and the main study. Specifically, data were collected in an individual interview setting. Respondents were each provided with an introduction letter, an informed consent form, the statement set on cards and the sorting grid. Before the interview started, respondents were informed about the objective of the study, the process of the ranking exercise and the intended use of the data. They were able to stop the interview at any time. After respondents signed the informed consent form, they first were asked to read the 34 statements considering well-being in order to acquaint themselves with the complete set of well-being aspects (see Table 4.1). Next, to facilitate the ranking of all the statements in accordance to their relative importance, they divided the statements into three piles: important, neutral / do not know and not important to their well-being. Then, they read the statements in the important pile again, and ranked them from most to least important on the sorting grid (see Figure 4.1), generally working from the right side to the middle. After finishing, they repeated this step for the statements in the not important pile, placing them from the left side of the sorting grid to the middle. Lastly, they placed the statements from the neutral / do not know pile in the remaining open spots in the middle of the sorting grid. Respondents then were asked to check their ranking of the statements and make adjustments until they felt satisfied that it reflected their view on what is important for well-being. Next, respondents were asked to elaborate on their ranking, always including those placed in the two extreme columns of the sorting grid. If time allowed, the interviewer asked respondents to further explain the placement of other statements, for example when statements belonging

to the same domain (see Table 4.1) were ranked remarkably differently. Permission was granted by all but one respondent to audio record this part of the interview. Finally, respondents provided information on their health and background characteristics through a questionnaire. All respondents finished the ranking exercise. However, due to lack of time, two of the 53 follow-up interviews were not completed. In addition, due to language restrictions, the interviews with the Turkish and Moroccan respondents were less elaborate and specific. While these interviews were still useful for helping with interpretation of the data, they proved less useful for extracting quotes.

4.2.4 | Ethics

The study protocol was approved by the Medical Ethics Committee Erasmus MC of Rotterdam, the Netherlands, MEC-2016-667.

4.2.5 | Analytic strategy

To identify patterns in the rankings of the statement set, by-person factor analysis was conducted using the package 'qmethod' (Zabala and Held, 2018) in R (R Foundation for Statistical Computing, Vienna, Austria). The range of the sorting grid from 1 to 9 was rescaled from -4 to +4. The data were placed in a matrix with rows and columns containing statements and respondents, respectively, with cell values representing respondents' score on the sorting grid for each statement. A correlation matrix between respondents' sorts was built, after which principal component analysis (PCA) was applied to reduce this matrix into factors, ordering them by the total variability that they explain. Oblimin rotation was administered to obtain a clearer and simpler structure of the data. Oblimin rotation permits statistically correlated factors to emerge, which seemed appropriate in the current study as, for example, we anticipated that health could be important in more than one view. The following criteria were used for selection of the number of factors: (i) scree test, (ii) Kaiser-Guttman criterion (eigenvalues > 1), (iii) at least two respondents that load significantly at $p < 0.05$ on one factor only (i.e. factor loadings are not confounded, as these best exemplify the viewpoint represented by the factor and hence are called 'exemplars' for that factor) (for details, see Table 4.3) and (iv) the most interpretable and coherent reduction of the data (Watts and Stenner, 2012; Zabala, 2014). Using a mixed methods approach, the factors were interpreted and described as views on well-being. This approach first entailed considering the characterising and distinguishing statements. Characterising statements are those with a

score of -4, -3, +3 and +4 in the factor, whereas distinguishing statements are those with a statistically significantly different score as compared to the other factors. Second, verbal explanations of respondents loading on that factor from the follow-up interviews were used to check and refine the interpretation of each factor.

4.3 | Results

4.3.1 | Study sample

The descriptive statistics of the study sample are presented in Table 4.2. The age of respondents ranged between 65 and 98 years, but most respondents were aged between 65 and 79 years old (77%). Gender and educational levels were fairly equally represented in the sample, as intended. 25% was living in a nursing home, and 28% was recruited on their non-western background.

Table 4.2 | Descriptive statistics of the study sample (N=53)

		N	%	Mean	SD	Min.	Max.
Age				74.68	9.49	65.00	98.00
Sex	Male	24	45.28				
	Female	29	54.72				
Education level	None, primary school (ULO/MULO), secondary school (VMBO)	18	33.96				
	Secondary school (HAVO/VWO), community college (MBO)	18	33.96				
	Higher education (HBO/WO)	17	32.08				
Housing	Independent living	40	75.47				
	Nursing home (incl. sheltered housing)	13	24.53				
Cultural background	Predominantly Dutch	38	71.70				
	Turkish	4	7.55				
	Moroccan	3	5.66				
	Surinamese	4	7.55				
	Antillean	4	7.55				
EQ-5D-5L [a]				0.76	0.24	0.01	1.00

[a] Versteegh et al., 2016b

4

4.3.2 | **Views on well-being**

Analysis of the 53 rankings of the statement set indicated that the data supported five factors, i.e. five distinct views on what constitutes well-being for older people. Table 4.3 presents the factor loadings: fourteen respondents were associated with factor 1, eight with factor 2, eight with factor 3, four with factor 4 and five with factor 5. Next to these thirty-nine exemplars, twelve respondents were mixed loaders (i.e. confounded) and two respondents were null loaders (i.e. no significant loadings). Table 4.4 shows the composite sorts of the statements for the five factors. In total, the factors explained 52% of the variance in respondents' rankings. Correlations between the factors were low to moderate and ranged between 0.10 and 0.46.

Table 4.3 | Factor loadings (N=53)

ID	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
1	-0.54*	0.20	0.19	0.41	0.20
2	-0.20	-0.10	0.24	-0.32	0.43
3	0.51*	0.25	0.10	-0.19	0.10
4	0.24	-0.21	0.29	-0.20	0.33
5	-0.13	0.26	0.59*	0.33	0.06
6	0.14	0.14	0.55*	-0.14	0.08
7	-0.09	-0.01	0.82*	0.00	0.01
8	0.35	0.27	0.31	0.38	0.04
9	0.08	0.69*	0.17	-0.47	-0.17
10	0.50	0.06	0.51	-0.11	-0.16
11	0.30	0.16	0.25	-0.37	0.35
12	-0.01	0.54*	0.31	0.13	-0.07
13	0.03	0.68*	0.25	0.07	0.03
14	0.44	0.40	0.04	-0.15	0.42
15	0.59*	0.38	0.02	-0.11	0.11
16	0.56*	0.40	0.06	-0.29	0.25
17	0.07	-0.02	0.36	-0.10	0.59*
18	0.59*	0.04	0.39	0.12	0.14
19	0.03	0.34	0.63*	0.07	-0.02
20	0.36	0.25	0.24	0.05	0.20
21	0.25	0.50*	0.18	0.17	-0.10
22	0.35	-0.06	0.07	-0.16	-0.54*
23	0.04	-0.09	0.59*	0.11	0.33
24	0.80*	-0.23	-0.21	-0.01	-0.03
25	-0.04	-0.27	0.47*	0.34	0.14
26	0.67*	0.22	-0.20	-0.08	0.12
27	0.56*	0.27	0.13	-0.01	0.11
28	0.16	-0.06	0.46	-0.43	0.36
29	0.48	0.50	0.12	-0.01	0.08
30	-0.10	0.66*	-0.18	0.15	0.07

Table 4.3 | Continued

ID	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
31	0.70*	-0.05	0.23	-0.07	-0.07
32	0.10	0.07	0.52*	-0.25	0.08
33	0.27	-0.19	-0.13	0.43	0.29
34	-0.27	0.76*	-0.26	0.04	0.18
35	0.35	0.28	-0.27	0.24	0.48
36	-0.01	0.68*	-0.05	0.24	-0.02
37	0.37*	-0.16	0.08	0.26	0.12
38	-0.40	0.07	0.22	0.48*	0.03
39	-0.17	-0.09	0.11	0.70*	0.26
40	0.01	0.08	-0.05	0.75*	0.16
41	0.37	-0.05	-0.44*	0.10	0.21
42	0.67*	0.00	0.21	0.14	-0.12
43	0.86*	-0.07	-0.03	0.16	-0.15
44	0.67*	-0.05	0.02	0.42	0.25
45	0.49	0.24	0.23	-0.02	0.38
46	0.15	-0.05	0.14	0.31	0.69*
47	0.09	0.03	0.19	0.24	0.60*
48	-0.06	0.26	-0.14	-0.14	-0.31
49	0.55*	0.04	0.36	-0.33	0.07
50	0.05	0.62*	0.34	0.17	-0.26
51	0.25	-0.01	0.38	0.13	-0.67*
52	0.40	0.14	-0.09	0.45	-0.43
53	0.19	0.20	0.05	0.82*	-0.18

Note: * Denotes defining Q-sort (or exemplar) for factor, which holds that Q-sort loads statistically significantly on factor ($p < 0.05$; formula to calculate significance threshold: $(-)1.96 * \left(\frac{1}{\sqrt{34}}\right)$ i.e. loading < -0.336 or > 0.336) and is not confounded between factors (i.e. square loading is larger than sum of the squares of the loadings of the Q-sort on the other four factors). For example, respondent 1 is an exemplar for factor 1 because the factor loading (-0.54) exceeds the threshold value of -0.336 (and therefore meets criterion 1) and the square loading $(0.54)^2 = 0.292$ is larger than the sum of the squares of the loadings on the other factors: $(0.20)^2 + (0.19)^2 + (0.41)^2 + (0.20)^2 = 0.288$ (and therefore meets criterion 2).

Next to the positively loading exemplars, factors 1, 3 and 5 also had negatively loading exemplars (see Table 4.3). Factors 1 and 3 each had only one negatively loading exemplar, which, as single loaders, were both insufficient to determine a separate factor. Factor 5 had two negatively loading exemplars, which had a low mutual correlation (0.26), and also did not identify a separate factor. Exploration of solutions with more than five factors also showed there was no additional factor based on these negative loaders. The negatively loading exemplars were not used in the calculation of the statement factor scores and the interpretation of the factors, as they hold different, to some extent opposite views to the other exemplars on these factors.

In the description of the factors (see Table 4.4), the positioning of important statements on the sorting grid is given in parentheses with the statement number indicated by #. Thereafter, the position on the sorting grid is presented with an indication of significant differences in relation to the other factors (in Table 4.4, followed by * / † to indicate differences at p-levels .05 / .01, respectively); e.g. (#2, $+4^{\dagger ALL}$) means that for that view, statement 2 had a factor score of +4 and that this score was distinguishing as compared to all other factors at 0.01 level of statistical significance. In addition, verbatim quotes from exemplars are presented as illustration. They are noted in "…" with respondent ID numbers.

Table 4.4 | Factor scores per statement (N=53)

		Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
1	Being physically able to perform my daily activities	$+2^{\dagger 4 *5}$	$+3^{\dagger 4,5}$	$+3^{\dagger 4,5}$	$-3^{\dagger ALL}$	$+1^{\dagger 2\text{-}4 *1}$
2	Having no physical illness or disability	$+4^{\dagger ALL}$	$+3^{\dagger 1,4,5 *3}$	$+1^{\dagger 1,5 *2,4}$	$-1^{\dagger 1,2 *3}$	$-2^{\dagger 1\text{-}3}$
3	Having no physical pain	$+2^{\dagger 3,5}$	$+1$	$0^{\dagger 1}$	$+1$	$0^{\dagger 1}$
4	Having the prospect of many life years ahead	$0^{\dagger ALL}$	$-3^{\dagger 1 *3}$	$-4^{\dagger 1,5 *2}$	$-4^{\dagger 1,5}$	$-2^{\dagger 1,3,4}$
5	Living a dignified and healthy life	$+3^{\dagger 3\text{-}5}$	$+4^{\dagger 3\text{-}5}$	$+1^{\dagger 1,2}$	$+1^{\dagger 1,2}$	$0^{\dagger 1,2}$
6	Being mentally active	$+2^{\dagger 2,3}$	$-1^{\dagger ALL}$	$+4^{\dagger 1,2}$	$+2^{\dagger 2}$	$+2^{\dagger 2}$
7	Having no problems with my ability to think	$+3^{\dagger 2}$	$+1^{\dagger 1,3 *4}$	$+4^{\dagger 2 *5}$	$+3^{*2}$	$+2^{*3}$
8	Being able to communicate well with others	$-1^{\dagger 5 *2\text{-}4}$	0^{*1}	0^{*1}	0^{*1}	$+1^{\dagger 1}$
9	Not feeling anxious or depressed	0^{*4}	$0^{\dagger 3}$	$-1^{\dagger 2,4}$	$+2^{\dagger 3,5 *1}$	$-1^{\dagger 4}$
10	Not being worried about my future	$+1^{\dagger 2,3,5}$	$+2^{\dagger ALL}$	$-3^{\dagger 1,2,4}$	$0^{\dagger 2,3,5}$	$-3^{\dagger 1,2,4}$
11	Being able to look back on how my life went with satisfaction	$-3^{\dagger 2\text{-}4}$	$+2^{\dagger ALL}$	$0^{\dagger 1,2,5}$	$-1^{\dagger 1,2}$	$-3^{\dagger 2,3}$
12	Accepting my circumstances as they are	$-2^{\dagger 2,4,5}$	$-1^{\dagger 1,3,4 *5}$	$-2^{\dagger 2,4,5}$	$+2^{\dagger 1\text{-}3}$	$+2^{\dagger 1,3 *2}$
13	Being able to adjust to new circumstances	$-2^{\dagger 2,4,5}$	$-1^{\dagger 1,5 *3}$	$-2^{\dagger 4,5 *2}$	$+1^{\dagger 1,3 *5}$	$+3^{\dagger 1\text{-}3 *4}$
14	My religion or view of life	$+1^{\dagger ALL}$	$-3^{\dagger 1,4 *5}$	$-4^{\dagger 1,4}$	$+4^{\dagger ALL}$	$-4^{\dagger 1,4 *2}$
15	Having a life partner	$+4^{\dagger 2\text{-}4}$	$-4^{\dagger ALL}$	$-1^{\dagger ALL}$	$-3^{\dagger ALL}$	$+4^{\dagger 2\text{-}4}$
16	Having a good relationship with my close family	$+1^{\dagger 2,4,5 *3}$	$+4^{\dagger 1}$	$+2^{*1}$	$+3^{\dagger 1}$	$+4^{\dagger 1}$
17	Having contact with people I care about	$+2^{\dagger 3 *4}$	$+2^{\dagger 4}$	$+1^{\dagger 1,4 *5}$	$+4^{\dagger 2,3 *1}$	$+3^{*3}$

Table 4.4 | Continued

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
18 Receiving support from family, friends or neighbours when I need it	$-1^{†2,4,5}$	$+3^{†1,3}$	$0^{†2,4,5}$	$+3^{†1,3}$	$+2^{†1,3}$
19 Feeling valued by helping others	$-2^{†3,5}*4$	$-2^{†3-5}$	$+3^{†ALL}$	$0^{†2,3}*1$	$0^{†1-3}$
20 Not being dependent on others for daily activities	$0^{†3-5}$	$0^{†3-5}$	$+3^{†ALL}$	$-2^{†1-3}$	$-2^{†1-3}$
21 Being able to go wherever I want to go	$0^{†2,4}$	$+2^{†1,4,5}$	$+1^{†4,5}$	$-3^{†ALL}$	$-1^{†2-4}$
22 Having the freedom to make my own choices	$0*2$	$+1^{†3,5}*1$	$-1^{†2}$	0	$-1^{†2}$
23 Being able to spend my time as I want	$-1*2$	$-1*1$	-1	-1	0
24 Doing the things I consider important in life	$+1^{†3}*2$	$+1*1$	$+2^{†1}$	$+1$	$+1$
25 Learning new skills	$-4^{†4}$	$-4*4$	$-3^{†4}$	$-2^{†1,3,5}*2$	$-4^{†4}$
26 Meeting new people	$-4^{†2,4,5}$	$-1^{†1,3}*5$	$-3^{†2,4,5}$	$-2^{†1,3}*5$	$0^{†1,3}*2,4$
27 Having the feeling that I matter	$-3^{†3,4}*2,5$	$-2^{†3,4}*1$	$+2^{†1,2,5}$	$+1^{†1,2,5}$	$-1^{†3,4}*1$
28 Staying involved in the issues at play in my community	$-3^{†2-4}*5$	$0^{†1,3,5}$	$-2^{†1,2}*4$	$0^{†1}*3,5$	$-2^{†2}*1,4$
29 Having no financial concerns	$+3^{†3,4}*2,5$	$+1^{†3,4}*1$	$-2^{†1,2,5}$	$-1^{†1,2,5}$	$+1^{†3,4}*1$
30 Independent living	$+1^{†4,5}$	$0^{†4,5}$	$+1^{†4,5}$	$-4^{†ALL}$	$-1^{†ALL}$
31 Living in an area in which I feel comfortable	$0^{†3-5}$	$0^{†3}*5$	$+2^{†1,2}$	$+2^{†1}$	$+3^{†1}*2$
32 Living in an area with sufficient facilities	$-1^{†2}$	$-2^{†1,3,5}$	$0^{†2}$	$-2*5$	$0^{†2}*4$
33 Living in a stable environment with minor risk of social unrest	$-2^{†2}$	$-3^{†1,3,4}$	$-1^{†2}$	$-1^{†2}$	-3
34 Living in a country in which my interests are heard and represented	$-1^{†2}$	$-2^{†ALL}$	$0^{†2}$	$0^{†2}$	$+1^{†2}$
Percentage of explained variability	15.15	9.83	9.72	8.85	7.97
Correlation with Factor 2	0.29				
Correlation with Factor 3	0.36	0.37			
Correlation with Factor 4	0.10	0.22	0.17		
Correlation with Factor 5	0.35	0.15	0.46	0.27	

† $p < 0.01$; $*$ $p < 0.05$

4.3.3 | Viewpoint 1: Health, financial security and a life partner

Respondents defining this view emphasised the importance of health to their well-being (#1, +2[4,5]; #2, +4[ALL]; #3, +2[3,5]; #5, +3[3-5]; #6, +2[2,3]; #7, +3[2]). Some cited the intrinsic value of health as rationale: "[Having no physical illness or disability is] the most important thing there is!" (id 27) Also, many reported health to enable them to pursue all things that contributed to their well-being: "I am rather active in the field of sport for example, I need to be able to continue my walks, [...] to swim my laps, to participate in groups, to discuss. I want to stay involved." (id 18) In this context, financial security was stressed as well (#29, +3[ALL]): "I do not need much, I do not have to be rich, but I need to be able to meet my basic needs." (id 27)

These respondents also valued having a partner (#15, +4[2-4]) above all other social relationships (#16, +1[ALL]; #17, +2[3,4]): "You have your children, you have your family, but they are always at a distance. Your wife, or partner, is near." (id 18) They showed little interest in developing new activities, including learning new skills, staying involved in community issues and meeting new people (#25, -4[4]; #28, -3[ALL]; #26, -4[2,4,5]): "That time has passed." (id 24) "It is not a necessity in my life." (id 49) The feeling that they matter was emphasised by these respondents to a lesser extent (#27, -3[ALL]), often because their self-worth came 'from within': "I do not have to prove myself or put myself in the centre of attention. I know what I am worth." (id 27) Partly for this reason, they focused less on looking back upon life with satisfaction (#11, -3[2-4]): "That would feel self-congratulatory. Look at me, I have done this and that." (id 18) However, other reasons were mentioned as well, including not wanting to look back or finding it less relevant.

4.3.4 | Viewpoint 2: Family, support and physical functioning

Respondents defining this view attached great importance to family for their well-being (#16, +4[1]): "They are the people that are closest to me." (id 21) "I am always welcome there." (id 13) In contrast to view 1, having a partner was one of the least important aspects for their well-being (#15, -4[ALL]), mainly because their life partner had already died, they were divorced, or preferred to be single: "I can manage myself. [Being alone] also increases my autonomy." (id 30) Many experienced health limitations, which made them value having help (#18, +3[1,3]): "When I broke my arms, I asked my neighbour to put my laundry in the machine, but she said she would wash all of my laundry instead." (id 12) In that context, they valued

health (#1, +3[4,5]; #2, +3[ALL]; #3, +1; #5, +4[3-5]). Mental health was considered slightly less important (#6, -1[ALL]; #7; +1[1,3,4]): "[My ability to keep up with things] has decreased over the last couple of years. […] That is part of the ageing process." (id 12) They attached little importance to longevity (#4, -3[1,3]), partly due to (a fear of) health issues or the feeling that they had no control over that.

Compared to the other views, looking back and forward in life without any concerns was more important (#10, +2[ALL]; #11, +2[ALL]): "[Not having to worry] that you will run out of money, or that you are forced to leave your house." (id 30) "[Looking back upon life with satisfaction] makes you feel happy." (id 36) These respondents attached little importance to religion or view of life (#14, -3[1,4,5]) or issues that played at the macro level (that only affected them indirectly), including living in a stable environment and political representation (#33, -3[1,3,4]; #34, -2[ALL]): "It does not ring a bell." (id 13) "I believe that social unrest is widespread. How can I worry about that?" (id 9) As in view 1, learning new skills (#25, -4[4]) was less important for their well-being.

4.3.5 | Viewpoint 3: Autonomy, mental health and helping others

Respondents defining this view stressed the importance of autonomy to their well-being (#20, +3[ALL]). Many did not want to rely on others for help: "[I want] to be in control and decide for myself what is right." (id 5) Also, people feared not having help: "[If you are not able to do things yourself] what will you do then?" (id 6) Some discussed autonomy to be well: "You have to energise yourself and keep going." (id 25) Mental health was necessary for exercising autonomy (#1, +3[4,5]; #6, +4[1,2]; #7, +4[2,5]): "[Being mentally active means] that I can still do a lot myself." (id 23) Mental health was generally considered more important to well-being than physical health (#2, +1[ALL], #3, 0[1]): "If you have a physical disability, your mental health will help you cope with that." (id 32). "Physical pain can often be treated to greater extent. Also, there are many assistive devices." (id 19) Like longevity (#4, -4[1,2,5]), not feeling anxious or depressed (#9, -1[2,4]) was considered less important: "I think health is more important. Not feeling anxious or depressed could be a part of that, but I am not [anxious or depressed]." (id 6) "I am not very afraid of that." (id 7)

These respondents valued helping others (#19, +3[ALL]), mainly because of their taste for it, and previous experience: "I have always helped other people. […] I did not only do that for them, but also for my sense of self-worth. Because I enjoy it, but also because it is sometimes

a societal obligation." (id 6) Being able to think about the future without any concerns was believed to be less relevant (#10, $-3^{1,2,4}$), as well as religion or view of life (#14, $-4^{1,4}$): "I no longer believe in a God. […] My view of life is represented [in other statements]." (id 32) "I do believe, but I am not part of any religious community." (id 7) As in previous views, learning new skills (#25, -3^4) and meeting new people (#26, $-3^{2,4,5}$) were of little importance to their well-being.

4.3.6 | Viewpoint 4: Social contacts, support, mental health and religion

Respondents defining this view valued contact with others in general, including contact with family (#16, $+3^1$; #17, $+4^{1-3}$). As in view 2, they also valued their social contacts as a potential source for receiving help (#18, $+3^{1,3}$), if needed. Also, as in view 2, having a partner (#15, -3^{ALL}) was considered to be of little importance for their well-being. Respondents defining this view placed lower importance on longevity (#4, $-4^{1,5}$), physical health (#1, -3^{ALL}; #2, -1^{1-3}; #21, -3^{ALL}) and independence (#30, -4^{ALL}), as these aspects were generally accepted to deteriorate as part of their ageing process: "[Living independently] has passed." (id 39) In contrast, mental health was valued highly (#7, $+3^2$); in becoming physically frail, this aspect of health became much more important: "At this age, you cannot do anything because of your limited physical condition. However, mentally you can do so much." (id 40) Religion (#14, $+4^{ALL}$) was very important for them, also as a coping mechanism: "I get my strength from [religion]. […] All my life, through all periods, I got much help from that." (id 40)

4.3.7 | Viewpoint 5: A life partner, social contacts, living environment and adaptation

Respondents defining this view prioritised having a partner (#15, $+4^{2-4}$): "We have been together for 60 years, raising the kids" (id 47). In line with view 2 and 4, they valued contact with others in general, including contact with family (#16, $+4^1$; #17, $+3^3$; #18, $+2^{1,3}$). These respondents stressed the importance of feeling comfortable in the living environment (#31, $+3^{1,2}$) and, also, adaptation (#12, $+2^{1-3}$; #13, $+3^{ALL}$): "I have to accept [my health]. It is what it is." (id 46) Adaptation was also mentioned across their life course: "My whole life, I had to [take life as it comes]. When I was eleven, the war started. […] When the war ended, everything was gone. […] I also experienced the 1930s. […] That was horrible as well, with the unemployment and all things related to it." (id 47) In this context, these respondents focused less on concerns about the future (#10, $-3^{1,2,4}$): "I will adapt [to any situation]." (id 17) "I do not have any concerns about the future, since my time here is up." (id 47) Also,

looking back on their life was considered less important (#11, -3[2,3]): "I do not think it is that important, because those years have passed." (id 17) As in other views, religion (#14, -4[1,2,4]), learning new skills (#25, -4[4]) and living in a stable environment (#33, -3) were also considered to be of little importance to their well-being.

4.4 | Discussion and conclusions

4.4.1 | Main findings

This chapter revealed five shared views on what constitutes well-being among people aged 65 and older in the Netherlands. The first view focused on health, financial security and having a life partner as requirements to pursue all things that contributed to their well-being. The second view focused on physical functioning and having a social network, especially contact with family, and the possibility to receive support when needed. The third view emphasised the importance of autonomy and helping others, and mental health as a means to this end. The fourth view emphasised mental well-being, religion and a support network to help them cope with their physical frailty. Finally, the fifth view emphasised the social network as well, with an emphasis on having a life partner and a comfortable living environment, and being able to adapt oneself.

4.4.2 | Comparability with previous studies

The ranking exercise revealed that longevity did not seem to be of great importance to older people's well-being, whereas opinion statements related to good health were prioritised in four out of five views. The only exception was related to the fifth view, presumably because the respondents defining this view often had health issues which may have caused them to focus on other aspects to remain well. These findings support previous research that health is frequently mentioned when older people define well-being (Bowling, 2006; Douma et al., 2017; Wilhelmson et al., 2005). More importantly, physical health is explicitly mentioned in the first and second view and even deemed more important than mental health in the latter. All views prioritised statements related to social contacts, this variably referring to a life partner, family or friends, which is in line with previous research that reported this aspect to be at least as important as health to older people's well-being (Douma et al, 2017; Farquhar, 1995; Wilhelmson et al., 2005). As illustrated by exemplars, developing new activities, including learning new skills, staying involved in community issues and meeting new people

were placed in the middle or the left side of the sorting grid in all views. The same goes for diverse statements about environment, including having sufficient facilities nearby and living in a stable environment. Being in control of your time was not deemed important, as many felt that it was normal not to have this due to for instance, work and family obligations.

Notwithstanding the shared importance of health and social contacts, the five views did not correlate strongly with each other, partly because they focused on different statements within these domains, indicating (subtle) differences in importance of aspects within those domains. Consequently, no consensus statements were identified across the five factors. In addition, all factors had characterising statements in unique domains. To illustrate this, financial security, religion and living environment were only valued in the first, fourth and fifth view, respectively, which is in line with previous research showing that these aspects were less frequently mentioned (Jopp et al., 2015; Wilhelmson et al., 2005). Mixed results were derived regarding the importance of autonomy, as mentioned in the third view (Grewal et al., 2006; Jopp et al., 2015). Adaptation was only emphasised in the fifth view, and to a lesser extent in the fourth view, whereas it was one of the main pillars of well-being in previous research (Jopp et al., 2015).

4.4.3 | Study limitations and strengths

Some study limitations need to be considered when interpreting the results of this chapter. Due to the use of a sampling agency, we did not have access to the response rate and socio-demographic characteristics of older people who did not want to participate. We acknowledge that this may limit our insight in possible issues regarding the ranking exercise, selective non-participation, and the coverage of all possible views on well-being. We tried to determine respondents' educational level as accurately as possible. However, due to the retrospective, self-reported nature, education was sometimes less evident, especially since a few attended classes after their formative years. Respondents with severe health issues (e.g. visual or hearing impairments) received help with placing the statements on the sorting grid from the interviewer or a family member. Moreover, while we felt it was important to include respondents with different cultural and ethnic backgrounds, it proved hard to find Turkish and Moroccan respondents who mastered the Dutch language sufficiently. Consequently, these respondents were accompanied by either a close family member or a translator during the interview. Due to financial constraints, we were not able to translate the study materials a priori. The language barrier may have hampered respondents understanding of the

statement set, but we found no indication for this as respondents provided coherent explanations of their rankings. The occasional presence of family members may have influenced respondents' placement of the social contact statements. However, at least in the Dutch-spoken interviews, we again found no evidence for this.

Notwithstanding these limitations, we did succeed in including these harder-to-reach people that are often left out of this type of research. Even though some received help in the process, the lower educated and non-western older people were satisfactorily able to perform the ranking exercise. Therefore, we believe that the five distinct views on well-being presented in this chapter sufficiently capture the broad diversity in views on what is important to the well-being of people aged 65 and older in the Netherlands. The statement set was carefully constructed using diverse studies on what constitutes well-being at an older age, considering the possible heterogeneity in views. Each domain of well-being was presented at a comparable level of abstraction and detail to limit the influence of our selection on the possibility for respondents to disclose their view. In the pilot study respondents were explicitly asked if any essential aspects of well-being were left out of the statement set, and in the main study respondents were given the opportunity to reflect on the study materials as well. We have no indication that any important aspects were missing. Considering that the development of the statement set was informed by a considerable number of qualitative studies that examined what matters for the well-being of older people before (see Table 4.1), we sought no further validation for the completeness of the statement set (e.g. through in-depth interviews) after this pilot. We do emphasise that different choices could have been made in the selection of statements. However, whether a different statement set representing the same life aspects would be likely to expose substantially different views is unclear.

As a robustness check, we explored the implications for our findings of selecting a four- or six-factor solution. In the four-factor solution, aspects like not being dependent, feeling valued by helping others, and adjusting to new circumstances, which are now highlighted in factor 3 and 5, no longer emerged as important for well-being among older people. As these aspects were identified as relevant during the interviews, we considered it important to retain them (and, hence, opted for the five-factor solution). The six-factor solution did not identify a new view on well-being, and also no shared view among the negative loaders in the current factors emerged (hence, the six-factor solution was not considered an improvement).

4

4.4.4 | Research and policy recommendations

This study did not investigate the prevalence of the five views among older people in the Netherlands, since Q-methodology does not intend to do so. Both for the further development of relevant and sensitive well-being measures, as well as for the development of well-being improving care services, more knowledge on the percentage of older people holding each view may be useful. This can be done using one of many techniques available to connect Q-methodology to survey research (e.g. Baker et al., 2010). In such studies, it may also be interesting to examine the relation of views on well-being with socio-demographic characteristics of people, and their opinion with respect to specific choices in care for older people, as faced by policy makers. Moreover, while we purposively sampled respondents using a carefully designed sampling frame, it remains hard to ensure that all different viewpoints on well-being in old age have been identified here. Such a study could thus indicate that certain views have been missed in the current study, for example if sizeable proportions of the population would not associate themselves with any of the five views presented here. At the same time, such a study may probably point out how the sampling frame could be improved to uncover those views in a similar study as the present one.

As indicated by Bulamu et al. (2015) and Makai et al. (2014a), the ASCOT (Netten et al., 2012) and ICECAP-O (Coast et al., 2008; Grewal et al., 2006) may be used in economic evaluations to assess the broader well-being effects of interventions, to promote effective and efficient care services for older people. In line with our study outcomes, both measures stress the importance of social relations through one item. Only the ICECAP-O includes an item related to security, which may capture aspects of importance to specific groups of older people, including financial security, having social support and coping and adaptation mechanisms through for instance, religion. Both measures have an item about control, which is linked to autonomy in the third view we identified. Both also have an item related to role and occupation which may cover the aspect of being of value to others. The ASCOT and ICECAP-O lack direct measures of health, and may not adequately cover physical health (Davis et al., 2013; Van Leeuwen et al., 2015; and chapters 2 and 3). The current study emphasises that health is a very important determinant of well-being among older people in four out of five views reported here. We note that the ASCOT, unlike the ICECAP-O, was originally intended to be used alongside a health measure like the EQ-5D. Whether using the ASCOT or the ICECAP-O in combination with the EQ-5D in evaluative studies can be a

solution to overcome this omission, remains to be seen as it may risk double counting of elements like mental health, and would require appropriate weighting of both measures relative to each other. Therefore, the possibility to extend well-being measurement to include health next to all other aspects considered important by older people for their well-being, should be explored further.

Concluding, this study has highlighted the heterogeneity in views of older people in the Netherlands on what constitutes well-being. Ideally, instruments used to measure and value the well-being of older people, also in the context of economic evaluations, comprise the variety of relevant domains portrayed by these well-being views, as this would allow older people adhering to any well-being view to meaningfully express their well-being with the instrument. Whether existing instruments, like the ASCOT and ICECAP-O, fully capture all these domains needs to be investigated further. In that context, it is for instance important to emphasise that (physical) health is important in most views, while this may not be sufficiently covered in current instruments (Davis et al., 2013; Van Leeuwen et al., 2015; and chapters 2 and 3). Moreover, finances and living environment seemed to be important as well; however, they are often not included in current instruments (Bulamu et al., 2015; Makai et al., 2014a) or their impact is measured indirectly (Grewal et al., 2006). Hence, further validation of existing instruments in different contexts remains important. In addition, any initiatives for developing alternative measures are advised to capture the plurality in what is important for older people to their well-being, as portrayed in this study.

4

Chapter 6

Well-being of Older People (WOOP):

Quantitative validation of a new outcome measure for use in economic evaluations

Abstract

Background

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.

Methods

Between December 2017 and January 2018, an online survey was used to retrieve data from 1,113 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.

Results

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.

Conclusions

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.

6.1 | **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., 2014a). 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., 2014a). 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., 2014a; 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 (chapter 4). In recent years, 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., 2014a).

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,

6

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., 2014a). 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 (Proud et al., 2019; Van Leeuwen et al., 2015; also chapters 2 and 3). 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., 2014a), 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; Van Leeuwen et al., 2015; also chapters 2 and 3). 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 (chapter 4). Although all individuals may be assumed 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 5.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.

6.2 | **Data and methodology**

6.2.1 | **Sampling strategy**

Using an online survey, data were gathered in two stages between December 2017 and January 2018. A sampling agency recruited 1,113 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.

6

6.2.2 | 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., 2014a). For all items, response options were defined on a five-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 5.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.

6.2.3 | 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 vocation 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 €1,000 on unforeseen expenses without being in debt or taking a loan (SCP, 2014). The number of health problems was measured using the 13 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.

6.2.4 | **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., 2016b). 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 12 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 a SD of 10, with higher scores indicating better functioning.

6.2.5 | **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 seven-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).

6

6.2.6 | **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.

6.2.7 | **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's test of sphericity, the Kaiser-Meyer Olkin (KMO) 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 ≥ 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's 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 t0 and t1. 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 t0 and 41 at t1). Because the focus was on the difference in scores for single respondents on t0 and t1, 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).

6

6.3 | **Results**

6.3.1 | **Descriptive characteristics**

The descriptive statistics of the study samples at t0 and t1 are displayed in Table 6.1. At t0, 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.

Table 6.1 | Descriptive statistics of the study sample (N=1,113) and test-retest reliability sample (N=269) and the discriminatory power of the WOOP (N=1,113)

| | | 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

Table 6.1 | Continued

| | | Descriptive statistics | | | | | | | | | | Discriminatory power (t0) |
| | | Study sample (t0) | | | | | Test-retest reliability sample (t1) | | | | | |
		%	Mean	SD	Min.	Max.	%	Mean	SD	Min.	Max.	Mean WOOP score
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	

[a] Both the index score and the categorised scores are displayed.

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

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 t1 (N=269) were significantly younger, more often male, retired and less often 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 Figure 6.1). 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'.
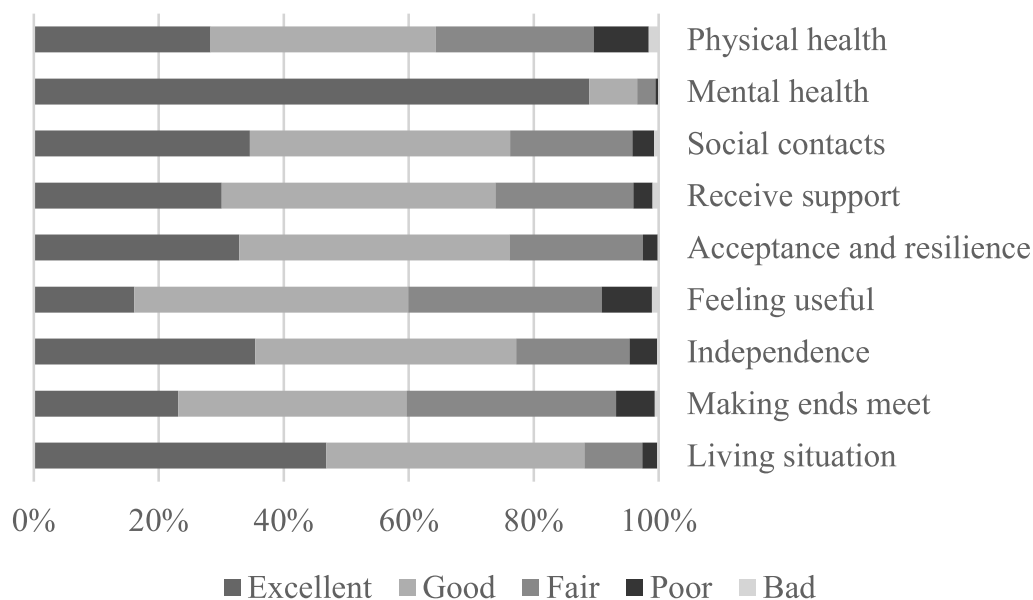

Figure 6.1 | Response distribution on the items of the WOOP (N=1,113) [a]

[a] The wording of the response options differs per item (see Appendix 5.A) but follows a scale ranging from 'excellent' to 'bad' to indicate the level of well-being.

WOOP, Well-being of Older People measure

Table 6.2 | Spearman correlations of the items of the WOOP (N=1,113)

	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

Table 6.3 | Linear regression analyses of Cantril's Ladder on the items of the WOOP (N=1113) [a]

	I	II	III	IV	V	VI	VII	VIII	IX	X
Physical health (*ref.=bad–fair*)										
Good	0.89***									0.35***
Excellent	1.16***									0.36***
Mental health (*ref.=bad–good*)										
Excellent		1.16***								0.41***
Social contacts (*ref.=bad–fair*)										
Good			0.95***							0.16
Excellent			1.46***							0.17
Receive support (*ref.=bad–fair*)										
Good				1.02***						0.26*
Excellent				1.27***						0.30*
Acceptance and resilience (*ref.=bad–fair*)										
Good					1.07***					0.16
Excellent					1.64***					0.36**
Feeling useful (*ref.=bad–fair*)										
Good						1.16***				0.44***
Excellent						1.55***				0.52***
Independence (*ref.=bad–fair*)										
Good							1.23***			0.48***
Excellent							1.59***			0.38**
Making ends meet (*ref.=bad–fair*)										
Good								1.03***		0.42***
Excellent								1.27***		0.50***
Living situation (*ref.=bad–fair*)										
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

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

WOOP, Well-being of Older People measure

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

Table 6.2 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 6.3 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 (i) 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.

6.3.2 | **Dimensionality**

Table 6.4 presents the results of the EFA to test the overlap in underlying constructs between the items of the EQ-5D-5L and the WOOP. 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.

6

Table 6.4 | Factor loadings for the items of the WOOP and the EQ-5D-5L (N=1,113)

		Uniqueness	Factor [a]		
			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	

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

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

6.3.3 | **Convergent validity**

Table 6.5 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 6.5 | 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
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***
Overall score	0.61***	0.54***	0.32***	0.51***	0.52***	0.72***	0.59***	0.62***

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

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

6

Table 6.6 | 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
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
Overall score	0.87	16.73	**0.11**	0.07	0.16	**0.89**	0.86	0.91

WOOP, Well-being of Older People measure

6.3.4 | Discriminant validity

Table 6.1 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).

6.3.5 | Test-retest reliability

Table 6.6 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 ($\kappa = 0.11$). 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 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.

6.4 | Discussion and conclusion

6.4.1 | Main findings

In a sample of 1,113 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

6

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 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.

6.4.2 | 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 general well-being (SWLS, Cantril's Ladder) are largely similar to those of the ICECAP-O and the ASCOT reported in previous studies (Van Leeuwen et al., 2015; also chapters 2 and 3). 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; also chapter 3). 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.

6.4.3 | 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 an online 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,

6

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 underexamined. 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 (chapter 4).

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, and its items and their descriptions and response levels, were based on a Q-methodology study (chapter 4), 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 chapter, 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 5.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 5.B (and in Dutch in Appendix 5.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.

6.4.4 | 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

6

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.

6

Chapter 7

Discussion

7.1 | Introduction

Policy makers in many countries face the challenge to efficiently and fairly allocate scarce health care resources. To aid them in funding those interventions that provide the most value for money, economic evaluations provide a framework to assess care services' costs and benefits. Several types of economic evaluations exist, that differ in their unit of measuring (health) benefits. The most common method applied in the health care context, cost-utility analysis (CUA), expresses benefits in terms of quality-adjusted life-years (QALYs), which comprise gains in both length and health-related quality of life (HRQoL). HRQoL is typically measured and valued using preference-based measures (Brazier et al., 2017; Drummond et al., 2015), such as the EuroQol five-dimensional questionnaire (EQ-5D) (The EuroQol Group, 1990) or the Short-Form Six-Dimension health index (SF-6D) (Brazier et al., 2002).

However, in the context of health and social care services for older people, the use of HRQoL may leave policy makers misinformed about the full benefits of such services, as their benefits may exceed health improvements. Especially in social care and long-term care, interventions may be targeted at improving older people's well-being rather than (only or primarily) their health. Broader outcome measures than the conventional HRQoL-measures are then required to allow an appropriate comparison of costs and benefits (Bulamu et al., 2015; Drummond et al., 2009; Makai et al., 2014a). Recently, such measures covering broader well-being, applicable in the context of economic evaluations, were developed, of which 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) are the most prominent examples. However, their validity has not yet been fully investigated (Bulamu et al., 2015; Makai et al., 2014a; as recently re-emphasised by Proud et al., 2019).

Therefore, the central goal of this thesis was to contribute to the measurement of well-being of older people for use in economic evaluations, by testing the validity of the ASCOT and the ICECAP-O and exploring the possibility of developing an alternative well-being outcome measure. Accordingly, the following four research questions were addressed in this thesis:

1. To what extent are the ASCOT and the ICECAP-O valid measures of the well-being of older people?

2. To what extent do views on what constitutes their well-being differ between older people?

3. How can this heterogeneity in views be integrated in a comprehensive measure of the well-being of older people?

4. How valid, feasible and reliable is this new measure of the well-being of older people?

This final chapter summarises the main findings from the studies included in this thesis and discusses the limitations of the research. The chapter ends with policy recommendations, ideas for future research and a general conclusion.

7.2 | Main findings

To answer research question 1, the validity of the ASCOT and the ICECAP-O was investigated in a sample of British social care users, aged 70 and older, as presented in chapter 2. The ICECAP-O was further examined in a broader sample of people aged 70 and older from the UK in chapter 3. The ASCOT and the ICECAP-O displayed good convergent and discriminant validity, and considerable overlap in their measurement of well-being. The ASCOT also included specific social care-related outcomes that were not captured by the ICECAP-O. The ICECAP-O largely covered health, apart from physical functioning. Also, the ASCOT and the ICECAP-O correlated only moderately with physical health.

As the limited coverage of any well-being domain will potentially lead to less attention for and impact of these domains in the evaluation and decision-making about care services, the importance of diverse well-being aspects to older people was tested in depth in chapter 4. To answer research question 2, people aged 65 and older in the Netherlands, presumably holding different views on well-being, ranked potential well-being aspects in accordance to their importance. Five shared views were identified, that mainly overlapped in the importance of social contacts and health, but each also emphasised unique elements:

I. Health, financial security and having a life partner;

II. Physical functioning and having a social network, especially contact with family, and the possibility to receive support when needed;

III. Autonomy and helping others, and mental health as a means to this end;

IV. Mental well-being and a support network to help cope with physical frailty;

7

V. Social network, with an emphasis on having a life partner, living environment, and being able to adapt oneself.

To capture and value a comprehensive set of all well-being domains relevant to (diverse groups of) older people in one instrument, an alternative measure named the Well-being of Older People measure (WOOP) was developed. To answer research question 3, the heterogeneity in older people's views on well-being, found in chapter 4, was integrated in the development of this new measure. Nine items were included, covering the following domains: (i) 'physical health', (ii) 'mental health', (iii) 'social contacts', (iv) 'receive support', (v) 'acceptance and resilience', (vi) 'feeling useful', (vii) 'independence', (viii) 'making ends meet' and (ix) 'living situation'. To ensure that the items, including their descriptions and response options, were comprehensive and clear, we used qualitative data from chapter 4, and cross-checked the design against available well-being measures in the field. A five-level response scale was developed for each item, to ensure relatively easy choices while keeping enough discriminatory power. Chapter 5 provided further details on its development.

To answer research question 4, the validity, feasibility and test-retest reliability of the WOOP were investigated in a sample of people aged 65 and older in the Netherlands, as reported in chapters 5 and 6. Overall, the qualitative validation in chapter 5 showed that people's interpretation of the WOOP items aligned with the given description for each item. Also, the corresponding response options were indicated to be clear. In addition, the WOOP seemed to capture all aspects relevant to the well-being of older people. The quantitative validation in chapter 6 showed that the WOOP and the EQ-5D-5L overlapped in their measurement of health, whereas the WOOP also included items beyond health. Convergent and discriminant validity were good; however, the correlation with physical health was mixed. The test-retest reliability in terms of the intraclass correlation coefficient (ICC) was high, whereas the kappa for the items was fair to good, except for two items. Based on the results of chapters 5 and 6, some minor changes were made to the WOOP. The official translation in English of the final version of the WOOP is included in Appendix 5.B.

7.3 | Strengths and limitations

The findings of this thesis should be interpreted in light of some limitations.

First, generalisability of the study outcomes may be hampered by using relatively small samples in only two countries. Chapters 2 and 3 were based on samples from the UK, because the development and valuation of the ASCOT and the ICECAP-O took place there. We assumed the outcomes of these chapters to also be relevant for other countries, like the Netherlands, but this needs further confirmation in future studies. We used samples from the Netherlands in chapters 4 to 6 to develop an alternative measure of well-being for older people in this context, which needs to be considered when using the results in other contexts and countries. Generalisability may also be limited by using online panels (in chapters 2, 3, 5 and 6). Although we aimed to retrieve samples as representative as possible in terms of age, sex and educational level, some groups proved to be hard to reach. Older people in higher age groups, with a lower educational level (and income) (in chapters 2 and 3), and low well-being and health levels (in chapters 5 and 6) were underrepresented. Especially for the WOOP, more validation studies are needed in these harder-to-reach groups. To have a higher chance of reaching social care users, the age range in chapters 2 and 3 was 70 years and older, instead of the 65 years and older range used in chapters 4 to 6. As the life expectancy and the retirement age keep rising, the increasing age which qualifies for senior status should remain an issue of debate.

Second, the collection of data online may have lowered respondents' level of engagement. However, in chapters 2 and 3, excluding respondents with a low and high response time did not affect our main results.

Third, the sampling agencies did not provide us the response rate and data of respondents who did not finish completing the questionnaires. Our insight into the feasibility of the different well-being measures was also limited as we did not allow any missing data to occur. However, using open-ended questions regarding the clarity of the WOOP items' descriptions and response options in chapter 5, we did not find any major issues.

Fourth, some other issues regarding our data collection procedures also deserve attention. Some older people needed help from the interviewer or a family member when ranking the well-being aspects (in chapter 4). In particular, Turkish and Moroccan respondents often needed to be assisted by a translator. The fact that the statements were not in themselves clear to these respondents may have reduced their understanding of the well-being aspects and have influenced their ranking, although we found no indication for this. Also, questions regarding the clarity of the WOOP items (in chapter 5) often prompted respondents to reflect

7

on their own well-being state (i.e. their chosen response on these items rather than the clarity of the items themselves). With hindsight, the fact that we performed a test-retest reliability study during the Christmas holiday season can also be seen as a limitation, as this period might affect the stability of responses over time, which may have contributed to the somewhat lower than expected test-retest reliability of the WOOP.

Finally, we emphasise that, besides the above-mentioned issues and the general need for more validation studies in different groups (e.g. specific disease areas, different age groups), settings (e.g. in nursing homes) and countries, the WOOP is not yet ready for use in economic evaluations of health and social care services for older people. Importantly, utility weights should be developed to be able to come to an overall score on the WOOP that expresses the relative importance of each item to overall well-being. At the moment, the alternative is to approximate overall well-being by summing the item scores.

Notwithstanding these limitations, also some strengths can be mentioned. This thesis for instance provided one of the first head-to-head comparisons of the ASCOT and the ICECAP-O. For the first time, their joint factor structure was examined (in chapter 2). The challenge in investigating their construct validity was that these two measures cover hypothetical constructs for which no gold standard exists in terms of their measurement. A further strength of this thesis is that the performance of the ASCOT and the ICECAP-O was studied in relation to a broad selection of often used and validated outcome measures. Moreover, different, but closely related, types of construct validity tests have been conducted, covering dimensionality, and convergent and discriminant validity.

To unravel which well-being domains are important to older people holding different views on well-being, we made an effort in this thesis to capture the broad diversity in views of what constitutes well-being to older people in the Netherlands. The statement set, covering well-being aspects that needed to be ranked by respondents in accordance to importance (in chapter 4), was carefully constructed by performing an extended literature review and using diverse other sources as well as interviews with older people. No essential aspects were missing according to our respondents. Moreover, we succeeded in obtaining a relatively large sample including harder-to-reach groups that are often left out of this type of research, including the eldest, those with severe health issues, and those in the four largest non-Western groups in the Netherlands.

Based on these findings, we developed a single, all-encompassing measure of well-being for use in economic evaluations of health and social care services for older people: the WOOP. With this measure, we aimed to capture a comprehensive set of well-being domains relevant to older people, accommodating the diversity in views on what they consider important to their well-being. To support its future uptake, this thesis tested the overlap between the conventional HRQoL-measure EQ-5D-5L and the WOOP (in chapter 6), as was done before for the ICECAP-O (chapter 3). Also, qualitative data from open-ended questions were analysed to provide further insight into the comprehensiveness, interpretation and clarity of the WOOP (in chapter 5). In a next step in its development, utility weights for the WOOP states should be derived.

Throughout this thesis we performed several robustness checks (e.g. for violations of underlying assumptions of our models). The similarities in findings from the diverse estimated models are encouraging. Also, the comparability with findings of other studies (also within this thesis) support their external validity. Moreover, the results with respect to the validity of the WOOP are comparable with those found in well-known studies on the ASCOT and the ICECAP-O (Bulamu et al., 2015; Cleland et al., 2019; Makai et al., 2014a). These findings provide support to the WOOP as a promising alternative measure of the well-being of older people.

7.4 | Recommendations for research and policy

Currently, the Dutch National Health Care Institute (Zorginstituut Nederland, 2016) suggests using both the EQ-5D and an ICECAP measure to capture benefits of interventions primarily aimed at gains in well-being over and above health, without much further guidance. While this is a step forward in comparison to previous guidelines, which only recommended using generic HRQoL-measures such as the EQ-5D, there are several reasons why this guideline can be viewed as an intermediate step. First, using the EQ-5D may be inappropriate in some of these contexts (Versteegh et al., 2016a), which makes the rationale for its use (even next to other outcome measures) less clear. Second, using two rather than one measure of benefit may make decision-making less straightforward, especially when the two measures would point to different policy conclusions. Third, when attempting to combine the EQ-5D with well-being measures in the context of decision-making, the possibility of double counting of some health effects needs to be factored in. As a solution, this thesis provided insight in the

7

extent to which the WOOP and the ICECAP-O may be used in isolation in future economic evaluations. With more experience in using these measures and greater understanding of their performance, future guidelines may provide clearer guidance on which measures to use for the evaluation of particular interventions, and how to use their (combined) outcomes in decision-making.

As a stepping stone, in this thesis, the validity of the ASCOT and the ICECAP-O were tested, next to the development and initial testing of our alternative well-being measure, the WOOP. All measures (the WOOP, being a new measure, in particular) need to be further investigated on their psychometric properties and how they may be included in economic evaluations. First, the feasibility of the WOOP may be further tested (as in chapter 5) using other methods in different settings to explore whether *all* older people are able to understand and complete the WOOP. For example, think-aloud studies or questionnaires that allow for missing data may lead to new insights. Replication is also encouraged to study specific findings in this thesis more in depth, such as the relationship of the WOOP with conventional HRQoL-measures and, more specifically, physical and mental health, if possible also using objective clinical measures. Moreover, the relatively low test-retest reliability of some of the WOOP items deserves further attention. Beyond the scope of this thesis, but nonetheless important, other psychometric properties of the WOOP need testing as well, such as the inter-rater reliability using proxy respondents (i.e. if WOOP scores align when filled out by older people themselves, health care professionals, caregivers or family members). Also, sensitivity to change (or responsiveness) needs to be assessed to prove that the WOOP is able to measure meaningful or clinically relevant changes caused by care interventions.

In the future, the WOOP may potentially be used in the existing framework of economic evaluations to measure and compare benefits across care sectors for older people. However, before the WOOP can be applied in such a way, two things are important: (i) developing utility weights - in line with the conventional HRQoL measures - to reflect the relative importance of each item to overall well-being in a single preference-based score; and (ii) finding an appropriate threshold value for gains in well-being. Both aspects are briefly discussed below.

Eliciting utility weights requires careful consideration as to the end goal and appropriate method to use. Different elicitation techniques can lead to different outcomes (Stamuli, 2011). On the one hand, direct elicitation methods, such as time trade-off (TTO), allow

anchoring the resulting scores on a scale from 0 to 1 (analogous to the QALY-scale), in which 0 represents the value attached to the state 'dead' and 1 represents the best possible well-being state. Alternatively, indirect elicitation methods, including discrete choice experiments (DCEs) and best-worse Scaling (BWS), could be used (Stamuli, 2011; Whitehead and Ali, 2010). However, anchoring the resulting scores on a scale with 'dead' and 'perfect well-being' is complex (Netten et al., 2012). Even though one can imagine well-being states that could be valued as being worse than dead, positioning the state 'dead' on a well-being scale can be difficult. More research is needed in this respect. In addition, there is an ongoing debate on whose preferences count in the context of resource allocation. Typically, broad samples from the general public have been used in the past for deriving QALY weights (e.g. because it would align with taking a societal and an insurance perspective – see Versteegh and Brouwer, 2016). In the context of well-being of older people, it may also be interesting to develop utility weights for subgroups of older people, for example for different age groups or subgroups based on the different views on well-being displayed in chapter 4.

Moreover, if the WOOP (or any other well-being measure) in the future is ready for use in economic evaluations, a monetary threshold value needs to be established to aid policy makers in decision-making. At the moment, the Dutch National Health Care Institute (Zorginstituut Nederland, 2018) uses a range of threshold values to determine which interventions to fund. The threshold varies with disease severity, with €80,000 per QALY as a maximum. Of course, using a well-being measure instead of a HRQoL-measure like the EQ-5D, implies the necessity to further explore whether the same threshold values can be used in this context or, plausibly, that higher values should be applied as well-being is broader in scope than health. Also, equity considerations deserve attention in this respect, to determine whether different threshold values could be used depending on who receives the well-being benefits, and the context in which they are realised (e.g. Bobinac et al., 2012). Whether the same equity principles are applicable for well-being gains as for health gains (like proportional shortfall in the Netherlands – see e.g. Reckers-Droog et al., 2018), needs further investigation as well.

Concluding, this thesis contributed to the testing and development of broader outcome measures of benefit for use in economic evaluations of health and social care services for older people. The validity of the ASCOT and the ICECAP-O were tested, alongside the development and first exploration of an alternative measure named the WOOP. In developing

7

this measure, an effort was made to include the diversity in older people's views on what constitutes well-being. Given the ageing of populations and the need to make choices about allocating scarce health care resources, also in the context of health and social care services for older people, accurately measuring and advancing the well-being of *all* older people should be high on the scientific and societal priority lists.

7

Summary

The ageing of the population leads to an increasing number of older people, with increasingly complex care needs. Partly as a result thereof, policy makers have to deal with the tension between a limited budget and a growing demand for care. Health care resources are typically insufficient to meet all care demands of the population. If policy makers allocate more funds from the available budget to care for older people, it comes at the expense of spending on care for other groups, while increasing the health care budget comes at the expense of investments in other policy areas. Alternatively, raising taxes or premiums is even more troublesome, as the economically active population is expected to decline even further. Hence, policy makers have to decide which care services for older people can be funded from the limited health care resources, and which not. An important tool supporting such decisions is the economic evaluation of interventions: a systematic and transparent framework to compare the costs and benefits of different care interventions. Various types of economic evaluations exist that especially differ in how benefits are measured. Most frequently used in the context of health care is cost-utility analysis (CUA), in which benefits are expressed in terms of quality-adjusted life-years (QALYs). These QALYs comprise both length and health-related quality of life (HRQoL). However, as the benefits of health and social care services for older people are not (confined to) health improvements, this thesis studies broader outcome measures, capturing the well-being of older people. Using such broader outcome measures in economic evaluations presumably leads to more adequate and comprehensive benefit estimates, as they facilitate comparisons of diverse types of interventions on both the health and the well-being of older people. This may ultimately contribute to funding those care services for older people that from a societal perspective provide the most value for money.

Broader outcome measures, for use in economic evaluations of health and social care services for older people, need to describe the well-being of older people in a feasible, comprehensive, reliable and valid manner. This also means that these outcome measures need to sufficiently recognise and reflect the potential heterogeneity in older people's views on well-being. Leaving out relevant well-being domains will potentially lead to less attention for these domains in the evaluation of care services for older people, and hence also to less impact on decision-making. This thesis addressed these requirements for three outcome measures: (i) the Adult Social Care Outcomes Toolkit (ASCOT), (ii) the ICEpop CAPability measure for

Older people (ICECAP-O) and (iii) a newly developed instrument called the Well-being of Older People measure (WOOP).

This thesis addressed four research questions. The first research question concerned the validity of the ASCOT and the ICECAP-O. Both measures were tested in a sample of social care users aged 70 years and older from the United Kingdom (UK) (chapter 2). The ICECAP-O was further examined in a broader sample of people aged 70 years and older from the UK (chapter 3). The ASCOT and the ICECAP-O displayed considerable overlap in their measurement of well-being, but only the ASCOT covered social care-related outcomes. In line with previous findings, the ASCOT and the ICECAP-O related only moderately to physical health. A (very) strong association was found of the ASCOT and the ICECAP-O with other validated measures of general health and mental health, and of the ICECAP-O with social functioning and happiness. Furthermore, the ASCOT and the ICECAP-O were able to distinguish subgroups of older people differing in terms of their health.

The second research question addressed the extent to which older people differ in their views on what constitutes well-being. Chapter 4 presents the results of a study conducted among a purposively sampled group of people aged 65 years and older from the Netherlands. Respondents ranked 34 potential well-being aspects, retrieved from the literature, in accordance to their importance for their own well-being, and motivated their ranking in an individual interview. Five views on well-being were found, which mainly overlap in the importance of social contacts and health, but each also emphasise unique elements, briefly summarised:

I.   Health, financial security and having a life partner;
II. Physical functioning, having a social network, especially contact with family, and the possibility to receive support when needed;
III. Autonomy and helping others, and mental health as a means to this end;
IV. Mental well-being and a support network to help cope with physical frailty;
V.   Social network, with an emphasis on having a life partner, living environment, and being able to adapt oneself.

The third research question addressed the development of a comprehensive measure of the well-being of older people, incorporating this heterogeneity in views on well-being. In chapter 5, the well-being aspects that were indicated to be most important in the five views

identified in chapter 4, were translated into nine items to be included in a new well-being measure, the WOOP. These items are: (i) 'physical health', (ii) 'mental health', (iii) 'social contacts', (iv) 'receive support', (v) 'acceptance and resilience', (vi) 'feeling useful', (vii) 'independence', (viii) 'making ends meet' and (ix) 'living situation'. For each item, five response options were defined to keep enough discriminatory power, while ensuring self-completion by older people remained feasible. Qualitative data from the individual interviews in chapter 4 and existing well-being outcome measures were used to ensure the clarity of the items, including their descriptions and response options.

The fourth research question concerned the feasibility, test-retest reliability and validity of the WOOP. Both a qualitative (chapter 5) and a quantitative validation (chapter 6) were carried out in a sample of people aged 65 years and older from the Netherlands. In chapter 5, the WOOP was shown to capture all domains relevant to the well-being of older people. Furthermore, older people's interpretations of the WOOP items sufficiently aligned with the given description for each item, and the related response options turned out to be clear. In chapter 6, the WOOP was shown to largely cover HRQoL, in addition to items beyond health. However, mixed evidence was found on the association with physical health. The associations between the WOOP and scores on other validated measures of general health and mental health were strong. The WOOP distinguished between relevant subgroups of older people based on their health. The test-retest reliability of the WOOP was studied by comparing WOOP scores collected with a two-week interval period. The intraclass correlation coefficient (ICC) was high, whereas the kappa for seven out of the nine items was fair to good. Based on the results in chapters 5 and 6, some minor changes were made to the descriptions and response options of some items of the WOOP, resulting in the final version of the instrument (see Appendix 5.B).

Concluding, the research in this thesis contributed to measuring the broader benefits of interventions in health and social care for older people in a feasible, comprehensive, reliable and valid manner. Next to testing the validity of existing instruments for measuring the well-being of older people, the ASCOT and the ICECAP-O, this thesis also presented the development and initial testing of a new measure, the WOOP. The WOOP aims to capture a comprehensive set of well-being domains relevant to older people, accommodating the diversity in views on what they consider important to their well-being. The ageing of the population is forcing governments to make choices about how to allocate the scarce resources

in care, particularly in the context of health and social care for older people. Therefore, accurately measuring and advancing the well-being of *all* older people should be high on the scientific and societal priority list.

Samenvatting

De vergrijzing van de bevolking leidt tot een toenemend aantal ouderen, met steeds vaker complexe zorgbehoeften. Mede als gevolg daarvan moeten beleidsmakers omgaan met de spanning tussen een beperkt budget en een groeiende vraag naar zorg. Doorgaans zijn er onvoldoende middelen om aan alle zorgbehoeften te voldoen. Als beleidsmakers meer geld uit het beschikbare budget toewijzen aan ouderenzorg gaat dit ten koste van uitgaven voor zorg aan andere groepen in de maatschappij, terwijl het verhogen van het budget voor zorg ten koste gaat van investeringen in andere beleidsterreinen. Het verhogen van belastingen of premies is ook niet zonder problemen, mede gezien het feit dat de beroepsbevolking steeds verder slinkt. Daarom moeten beleidsmakers beslissen welke zorg voor ouderen vanuit de beperkte middelen kan worden gefinancierd, en welke niet. Een belangrijk hulpmiddel bij dergelijke keuzes is de economische evaluatie van zorginterventies: een systematisch en transparant raamwerk om de kosten en baten van zorginterventies te vergelijken. Er bestaan verschillende soorten economische evaluaties, die met name verschillen in de manier waarop baten worden gemeten. De meest gebruikte soort binnen de gezondheidszorg is kostenutiliteitsanalyse (KUA), waarin baten worden uitgedrukt in voor kwaliteit gecorrigeerde levensjaren (quality-adjusted life-years, QALYs). Deze QALYs omvatten zowel het aantal levensjaren als de gezondheidsgerelateerde kwaliteit van leven. Echter, omdat de baten van zorg voor ouderen veelal niet (alleen) uit gezondheidsverbeteringen bestaan, bestudeert dit proefschrift bredere uitkomstmaten waarmee het effect van zorg op het welzijn van ouderen kan worden gemeten. Het gebruik van een dergelijke bredere uitkomstmaat in economische evaluaties leidt tot meer adequate en complete batenramingen, omdat ze vergelijkingen tussen diverse soorten interventies ten behoeve van zowel de gezondheid als het welzijn van ouderen mogelijk maken. Dit kan uiteindelijk bijdragen aan het financieren van zorginterventies voor ouderen die vanuit maatschappelijk perspectief het meest waar voor hun geld bieden.

Om bruikbaar te zijn in economische evaluaties van zorginterventies voor ouderen, moeten bredere uitkomstmaten het welzijn van ouderen op een praktisch haalbare, volledige, betrouwbare en valide manier meten. Dit betekent onder meer dat deze uitkomstmaten de mogelijke heterogeniteit in visies van ouderen op welzijn voldoende moeten weerspiegelen. Het niet includeren van relevante domeinen van welzijn kan immers leiden tot minder aandacht voor deze domeinen bij het evalueren van interventies voor ouderen, en daardoor

logischerwijs ook minder impact op besluitvorming. Dit proefschrift onderzocht daarom bovenstaande vereisten voor drie uitkomstmaten: (i) de Adult Social Care Outcomes Toolkit (ASCOT), (ii) de ICEpop CAPability measure for Older people (ICECAP-O) en (iii) een nieuw ontwikkeld meetinstrument genaamd de Well-being of Older People uitkomstmaat (WOOP).

Dit proefschrift heeft vier onderzoeksvragen behandeld. De eerste onderzoeksvraag betrof de validiteit van de ASCOT en de ICECAP-O. Beide meetinstrumenten werden getest in een streekproef van sociale zorggebruikers van 70 jaar en ouder uit het Verenigd Koninkrijk (VK) (hoofdstuk 2). De ICECAP-O is verder onderzocht in een grotere steekproef van ouderen van 70 jaar en ouder uit het VK (hoofdstuk 3). De ASCOT en de ICECAP-O kwamen grotendeels overeen in de manier waarop zij welzijn meten, maar alleen de ASCOT omvatte sociale zorg gerelateerde uitkomsten. Zoals al eerder werd gevonden, bleken de ASCOT en de ICECAP-O slechts een matige samenhang met lichamelijke gezondheid te vertonen. Er was een (zeer) sterke samenhang van de ASCOT en de ICECAP-O met andere gevalideerde meetinstrumenten voor algehele gezondheid en mentale gezondheid, en van de ICECAP-O met sociaal functioneren en geluk. Daarnaast waren de ASCOT en de ICECAP-O in staat om subgroepen van ouderen te onderscheiden op basis van verschillen in hun gezondheid.

De tweede onderzoeksvraag betrof de mate waarin ouderen verschillen in hun visie over wat welzijn is. Hoofdstuk 4 presenteert de resultaten van een studie onder een gerichte steekproef van mensen van 65 jaar en ouder uit Nederland. Respondenten rangschikten 34 aspecten van welzijn uit de literatuur naar hoe belangrijk zij deze vonden voor hun eigen welzijn, en lichtten dit toe tijdens een interview. Er werden vijf visies over welzijn gevonden, die met name overlappen in het belang dat wordt gehecht aan sociale contacten en gezondheid, maar elk ook unieke elementen benadrukken, kort samengevat:

I.   Gezondheid, financiële zekerheid en het hebben van een levenspartner;
II. Lichamelijk functioneren, het hebben van een sociaal netwerk, in het bijzonder contact met familie, en de mogelijkheid om steun te ontvangen wanneer dat nodig is;
III. Autonomie en het helpen van anderen, en mentale gezondheid die hiervoor nodig is;
IV. Mentaal welzijn en een ondersteunend netwerk om lichamelijke kwetsbaarheid het hoofd te bieden;
V.   Sociaal netwerk, met de nadruk op het hebben van een levenspartner, de leefomgeving en het in staat zijn om zichzelf aan te passen.

De derde onderzoeksvraag betrof de ontwikkeling van een instrument voor het meten van het welzijn van ouderen, waarin deze heterogeniteit in visies over welzijn is vertegenwoordigd. Hoofdstuk 5 beschrijft hoe de belangrijkste aspecten van welzijn uit ieder van de vijf visies beschreven in hoofdstuk 4 zijn vertaald naar negen items voor een nieuw meetinstrument, de WOOP. De negen items zijn: (i) 'lichamelijke gezondheid', (ii) 'mentale gezondheid', (iii) 'sociale contacten', (iv) 'steun ontvangen', (v) 'acceptatie en veerkracht', (vi) 'nuttig voelen', (vii) 'onafhankelijkheid', (viii) 'kunnen rondkomen' en (ix) 'leefomgeving'. Voor ieder item zijn vijf antwoordmogelijkheden gedefinieerd om voldoende onderscheid te kunnen maken in het niveau van welzijn, en tegelijkertijd het zelf invullen van de vragenlijst door ouderen praktisch haalbaar te houden. Kwalitatieve data uit de interviews gedaan in hoofdstuk 4 en bestaande welzijnsmeetinstrumenten zijn gebruikt als voorbeeld om ervoor te zorgen dat de items, inclusief hun beschrijvingen en antwoordmogelijkheden, duidelijk zijn.

De vierde onderzoeksvraag betrof de praktische haalbaarheid, test-hertest betrouwbaarheid en validiteit van de WOOP. Zowel een kwalitatieve (hoofdstuk 5) als kwantitatieve toetsing (hoofdstuk 6) werd uitgevoerd in een steekproef van mensen van 65 jaar en ouder uit Nederland. In hoofdstuk 5 werd aangetoond dat de WOOP alle domeinen omvat die relevant zijn voor het welzijn van ouderen. Daarnaast kwamen de interpretaties van ouderen van de WOOP items voldoende overeen met de beschrijving gegeven bij elk item, en bleken de bijbehorende antwoordmogelijkheden duidelijk te zijn. In hoofdstuk 6 werd aangetoond dat de WOOP grotendeels gezondheidsgerelateerde kwaliteit van leven, als niet-gezondheidsgerelateerde items dekt. Alleen wat betreft de samenhang met maten van lichamelijke gezondheid waren de resultaten gemengd. De samenhang tussen de WOOP en scores op andere gevalideerde meetinstrumenten van algehele gezondheid en mentale gezondheid was sterk. De WOOP was in staat subgroepen van ouderen te onderscheiden op basis van hun gezondheid. De test-hertest betrouwbaarheid van de WOOP werd onderzocht door twee WOOP-metingen, met een interval van twee weken, met elkaar te vergelijken. De intra-klasse-correlatie (ICC) was hoog, waarbij de kappa voor zeven van de negen items redelijk tot goed was. Op basis van de resultaten uit hoofdstuk 5 en 6 zijn kleine aanpassingen gemaakt aan de bewoording van de beschrijvingen en antwoordmogelijkheden van enkele items van de WOOP, resulterend in de uiteindelijke versie van het instrument (zie Appendix 5.C).

Samenvattend, het onderzoek in dit proefschrift heeft een bijdrage geleverd aan het op een valide en betrouwbare manier meten van de bredere baten van zorginterventies voor ouderen. Naast het testen van de validiteit van reeds bestaande instrumenten voor het meten van het welzijn van ouderen, de ASCOT en de ICECAP-O, werd een nieuw meetinstrument, de WOOP, ontwikkeld en getest. De WOOP heeft als doel de welzijnsdomeinen te meten die het meest relevant zijn voor ouderen, rekening houdend met de diversiteit in visies over wat zij belangrijk vinden voor hun welzijn. De vergrijzing van de bevolking dwingt de overheid om keuzes te maken over de aanwending van schaarse middelen binnen de (ouderen)zorg. Het nauwkeurig meten en bevorderen van het welzijn van *alle* ouderen moet dan ook hoog op de wetenschappelijke en maatschappelijke prioriteitenlijst staan.

List of publications and submissions

Chapter 2

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 in Health 20(7), 936-944.

Chapter 3

Hackert, M.Q.N., Van Exel, J., Brouwer, W.B.F., 2019. Does the ICECAP-O cover the physical, mental and social functioning of older people in the UK? Quality of Life Research 28, 761-770.

Chapter 4

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. Social Science & Medicine 240, 112535.

Chapter 5

Hackert, M.Q.N., Van Exel, J., Brouwer, W.B.F. Well-being of Older People (WOOP): Development and qualitative validation of a new outcome measure for use in economic evaluations. Submitted.

Chapter 6

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. Social Science & Medicine 259, 113109.

Other publications

Hackert, M., Linssen, R., Schmeets, H., 2012. Economische en culturele dreiging: Wie ervaart dreiging en wie niet? Sociaaleconomische Trends 4, 45-5.

PhD portfolio

PhD training

2019	MicroLab "How to create an Educational Video", Risbo, Research-Training-Consultancy, Erasmus University Rotterdam
2018	Qualitative Coding with ATLAS.ti, Erasmus Graduate School of Social Sciences and the Humanities, Erasmus University Rotterdam
2015 – 2017	University Teaching Qualification, Risbo, Research-Training-Consultancy, Erasmus University Rotterdam
2016	Academic Writing in English for iBMG staff, Language & Training Centre, Erasmus University Rotterdam
2015	PhD Course Ready in 4 Years, Erasmus University Rotterdam

Data collections

2017 – 2018	Development of an online survey including an online follow-up survey, data collection among 1,113 older people in the Netherlands
2016 – 2017	Development of a Q-methodology study, data collection among 53 older people in the Netherlands (e.g. by contacting nursing homes and day centres)
2015	Development of an online survey, data collection among 516 older people in the United Kingdom

Teaching

2018	Supervisor and co-evaluator for master theses, Master Health Economics, Policy & Law, Erasmus School of Health Policy & Management, Erasmus University Rotterdam
2015 – 2018	Co-revisor, working groups and practical lessons, Measurement of Patient Preferences using Discrete Choice Experiments, Master Health Economics, Policy & Law, Erasmus School of Health Policy & Management, Erasmus University Rotterdam

2017	Supervisor and co-evaluator for bachelor theses, Bachelor Gezondheidswetenschappen, Beleid & Management Gezondheidszorg, Erasmus School of Health Policy & Management, Erasmus University Rotterdam
2015 – 2017	Co-developer and co-evaluator, lectures and working group, Broadening Minor Quality of Life and Happiness Economics (Module 4: Quality of Life and Health), Erasmus School of Economics, Erasmus University Rotterdam
2016	Statistiek B, working groups and practical lessons, Premaster Gezondheidswetenschappen, Beleid & Management Gezondheidszorg, Erasmus School of Health Policy & Management, Erasmus University Rotterdam.
2015 – 2016	Supervisor and co-evaluator Kwantitatief Leeronderzoek, Premaster Gezondheidswetenschappen, Beleid & Management Gezondheidszorg, Erasmus School of Health Policy & Management, Erasmus University Rotterdam
2015	Co-evaluator Praktijkstage Werken in de Zorg, Bachelor Gezondheidswetenschappen, Beleid & Management Gezondheidszorg, Erasmus School of Health Policy & Management, Erasmus University Rotterdam

Conferences

Podium presentations

2018	lowlands Health Economic Study Group, Hoenderloo
2018	Dag van de Sociologie, Rotterdam
2017	iHEA World Congress of Health Economics, Boston
2017	lowlands Health Economic Study Group, Rotterdam
2017	Dag van de Sociologie, Brussel
2016	European Conference on Health Economics, Hamburg

2016	lowlands Health Economic Study Group, Gent
2016	Dag van de Sociologie, Tilburg
2016	Nederlands Congres Volksgezondheid, Rotterdam

Poster presentations

| 2018 | European Conference on Health Economics, Maastricht |

Organising committee

| 2017 | lowlands Health Economic Study Group, Rotterdam |

Presentations at other meetings

2018	PhD Platform Prize for Excellent Presentation, Erasmus School of Health Policy & Management, Erasmus University Rotterdam
2018	Well-being Group, Erasmus University Rotterdam
2017	Erasmus Choice Modelling Centre, Erasmus University Rotterdam

Other working groups, seminars

2018 – 2019	Werkplaats Health Technology Assessment & Appraisal, Zorginstituut Nederland - Utrecht University - Erasmus University Rotterdam
2018 – 2019	Well-being Group, Erasmus University Rotterdam
2017 – 2019	Erasmus Choice Modelling Centre, Erasmus University Rotterdam
2015	Seminar "Using Stated Preference methods in Health and Public Health", National Institute of Public Health and the Environment, Bilthoven

References

Al-Janabi, H., 2018. Do capability and functioning differ? A study of U.K. survey responses. Health Economics 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. Social Science & Medicine 87, 116-122. https://doi.org/10.1016/j.socscimed.2013.03.035.

Baglin, J., 2014. Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. Practical Assessment, Research, and Evaluation 19(5). Retrieved from http://pareonline.net/getvn.asp?v=19&n=5.

Baker, R., Van Exel, J., Mason, H., Stricklin, M., 2010. Connecting Q & surveys: Three methods to explore factor membership in large samples. Operant Subjectivity 34(1), 38-58. http://doi.org/10.15133/j.os.2010.003.

Basto, M., Pereira, J.M., 2012. An SPSS R-menu for ordinal factor analysis. Journal of Statistical Software 46(4), 1-29. https://doi.org/10.18637/jss.v046.i04.

Beaujean, A.A., 2013. Factor analysis using R. Practical Assessment, Research, and Evaluation 18(4). Retrieved from http://pareonline.net/getvn.asp?v=18&n=4.

Bernaards C., Jennrich, R. Package 'GPArotation'. Retrieved February 19, 2015 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.

Bobinac, A., Van Exel, N.J.A., Rutten, F.F.H., Brouwer, W.B.F., 2012. Inquiry into the relationship between equity weights and the value of the QALY. Value in Health 15(8), 1119-1126. https://doi.org/10.1016/j.jval.2012.07.002.

Bowling, A., 2006. Lay perceptions of successful ageing: Findings from a national survey of middle aged and older adults in Britain. European Journal of Ageing 3, 123-136. http://doi.org/10.1007/s10433-006-0032-2.

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). Archives of Gerontology and Geriatrics 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. Journal of Epidemiology & Community Health 65, 273-280. https://doi.org/10.1136/jech.2009.087668.

Brazier, J., Ratcliffe, J., Salomon, J.A., Tsuchiya, A., 2017. Measuring and Valuing Health Benefits for Economic Evaluation, second ed. Oxford University Press, Oxford, New York.

Brazier, J., Roberts, J., Deverill, M., 2002. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics 21(2), 271-292. https://doi.org/10.1016/S0167-6296(01)00130-8.

Brown, S.R., 2006. A match made in heaven: A marginalized methodology for studying the marginalized. Quality and Quantity 40, 361-382. https://doi.org/10.1007/s11135-005-8828-2.

Bulamu, N.B., Kaambwa, B., Ratcliffe, J., 2015. A systematic review of instruments for measuring outcomes in economic evaluation within aged care. Health and Quality of Life Outcomes 13, 179. https://doi.org/10.1186/s12955-015-0372-8.

Cantril, H., 1965. The Pattern of Human Concerns. Rutgers University Press, New Brunswick.

Chaudhry, S., Jin, L., Meltzer, D., 2005. Use of a self-report-generated Charlson Comorbidity Index for predicting mortality. Medical Care 43(6), 607-615. https://doi.org/10.1097/01.mlr.0000163658.65008.ec.

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. Applied Health Economics and Health Policy 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. Social Science & Medicine 67(5), 874-882. https://doi.org/10.1016/j.socscimed.2008.05.015.

Couzner, L., Ratcliffe, J., Crotty, M., 2012. The relationship between quality of life, health and care transition: An empirical comparison in an older post-acute population. Health and Quality of Life Outcomes 10, 69. https://doi.org/10.1186/1477-7525-10-69.

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? Quality of Life Research 22, 969-977. https://doi.org/10.1007/s11136-012-0225-4.

Devlin, N., Shah, K., Feng, Y., Mulhern, B., Van Hout, B., 2016. Valuing health-related quality of life: An EQ-5D-5L value set for England. OHE Research Paper, Office of Health Economics, London.

Douma, L., Steverink, N., Hutter, I., Meijering, L., 2017. Exploring subjective well-being in older age by using participant-generated word clouds. The Gerontologist 57(2), 229-239. http://doi.org/10.1093/geront/gnv119.

Drummond, M., Brixner, D., Gold, M., Kind, P., McGuire, A., Nord, E., 2009. Toward a consensus on the QALY. Value in 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, Oxford.

Elo, S., Kyngäs, H., 2008. The qualitative content analysis process. Journal of Advanced Nursing 62(1), 107-115. https://doi.org/10.1111/j.1365-2648.2007.04569.x.

Farquhar, M., 1995. Elderly people's definitions of quality of life. Social Science & Medicine 41(10), 1439-1446. http://doi.org/10.1016/0277-9536(95)00117-P.

Ferrans, C.E., Powers, M.J., 1985. Quality of Life Index: Development and psychometric properties. Advances in Nursing Science 8(1), 15-24. http://doi.org/10.1097/00012272-198510000-00005.

Flora, D.B., Curran, P.J., 2004. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychological Methods 9(4), 466-491. https://doi.org/10.1037/1082-989X.9.4.466.

Flynn, T.N., Chan, P., Coast, J., Peters, T.J., 2011. Assessing quality of life among British older people using the ICEPOP CAPability (ICECAP-O) measure. Applied Health Economics and Health Policy 9, 317-329. https://doi.org/10.2165/11594150-000000000-00000.

Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., et al. Package 'car'. Retrieved January 30, 2019 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.

Gabriel, Z., Bowling, A., 2004. Quality of life from the perspectives of older people. Ageing & Society 24(5), 675-691. http://doi.org/10.1017/S0144686X03001582.

Gamer, M., Lemon, J., Fellows, I., Singh, P. 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.

Gompertz, P., Pound, P., Ebrahim, S., 1994. A postal version of the Barthel Index. Clinical Rehabilitation 8, 233-239. https://doi.org/10.1177/026921559400800308.

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? Social Science & Medicine 62(8), 1891-1901. https://doi.org/10.1016/j.socscimed.2005.08.023.

Habibov, N., Afandi, E., 2016. Does life satisfaction determine subjective health? Applied Research in Quality of Life 11, 413-428. https://doi.org/10.1007/s11482-014-9371-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. Social Science & Medicine 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 in Health 20(7), 936-944. http://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? Quality of Life Research 28, 761-770. https://doi.org/10.1007/s11136-018-2042-x.

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. Social Science & Medicine 259, 113109. https://doi.org/10.1016/j.socscimed.2020.113109.

Hair, J.F., Tatham, R.L., Anderson, R.E., Black, W., 1998. Multivariate Data Analysis, fifth ed. Prentice-Hall, London.

Hartigan, I., 2007. A comparative review of the Katz ADL and the Barthel Index in assessing the activities of daily living of older people. International Journal of Older People Nursing 2(3), 204-212. https://doi.org/10.1111/j.1748-3743.2007.00074.x.

Harwood, R.H., 2008. Economic evaluations of complex services for older people. Age and Ageing 37(5), 491-493. https://doi.org/10.1093/ageing/afn154.

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). Quality of Life Research 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 & Mental 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.sportsci.org/resource/stats/effectmag.html.

Hughes, M.E., Waite, L.J., Hawkley, L.C., Cacioppo, J.T., 2004. A short scale for measuring loneliness in large surveys: Results from two population-based studies. Research on Ageing 26(6), 655-672. https://doi.org/10.1177/0164027504268574.

Hyde, M., Wiggins, R.D., Higgs, P., Blane, D.B., 2003. A measure of quality of life in early old age: The theory, development and properties of a needs satisfaction model (CASP-19). Aging & Mental Health 7(3), 186-194. https://doi.org/10.1080/1360786031000101157.

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. Quality of Life Research 22, 1717-1727. https://doi.org/10.1007/s11136-012-0322-4.

Jopp, D.S., Wozniak, D., Damarin, A.K., De Feo, M., Jung, S., Jeswani, S., 2015. How could lay perspectives on successful aging complement scientific theory? Findings from a U.S. and a German life-span sample. The Gerontologist 55(1), 91-106. http://doi.org/10.1093/geront/gnu059.

Kaambwa, B., Gill, L., McCaffrey, N., Lancsar, E., Cameron, I.D., Crotty, M., Gray, L., Ratcliffe, J., 2015. An empirical comparison of the OPQoL-Brief, EQ-5D-3L and ASCOT in a community dwelling population of older people. Health and Quality of Life Outcomes 13, 164. https://doi.org/10.1186/s12955-015-0357-7.

Keeley, T., Coast, J., Nicholls, E., Foster, N.E., Jowett, S., Al-Janabi, H., 2016. An analysis of the complementarity of ICECAP-A and EQ-5D-3L in an adult population of patients with knee pain. Health and Quality of Life Outcomes 14, 36. https://doi.org/10.1186/s12955-016-0430-x.

Kim, S. Package 'ppcor'. Retrieved December 10, 2015 from https://cran.r-project.org/web/packages/ppcor/ppcor.pdf and for a description of the package see: https://CRAN.R-project.org/package=ppcor.

Lyubomirsky, S., Lepper, H.S., 1999. A measure of subjective happiness: Preliminary reliability and construct validation. Social Indicators Research 46, 137-155. https://doi.org/10.1023/A:1006824100041.

Makai, P., Beckebans, F., Van Exel, J., Brouwer, W.B.F., 2014b. Quality of life of nursing home residents with dementia: Validation of the German version of the ICECAP-O. PLoS One 9(3), e92016. https://doi.org/10.1371/journal.pone.0092016.

Makai, P., Brouwer, W.B.F., Koopmanschap, M.A., Stolk, E.A., Nieboer, A.P., 2014a. Quality of life instruments for economic evaluations in health and social care for older people: A systematic review. Social Science & Medicine 102, 83-93. http://doi.org/10.1016/j.socscimed.2013.11.050.

Makai, P., Koopmanschap, M.A., Brouwer, W.B.F., Nieboer, A.A.P., 2013. A validation of the ICECAP-O in a population of post-hospitalized older people in the Netherlands. Health and Quality of Life Outcomes 11, 57. https://doi.org/10.1186/1477-7525-11-57.

Malley, J.N., Towers, A.-M., Netten, A.P., Brazier, J.E., Forder, J.E., Flynn, T., 2012. An assessment of the construct validity of the ASCOT measure of social care-related quality of life with older people. Health and Quality of Life Outcomes 10, 21. https://doi.org/10.1186/1477-7525-10-21.

McHugh, N., Baker, R.M., Mason, H., Williamson, L., Van Exel, J., Deogaonkar, R., Collins, M., Donaldson, C., 2015. Extending life for people with a terminal illness: A moral right and an expensive death? Exploring societal perspectives. BMC Medical Ethics 16, 14. http://doi.org/10.1186/s12910-015-0008-x.

McKeown, B., Thomas, D.B., 2013. Q Methodology, second ed. Sage Publications, Newbury Park.

Mukuria, C., Connell, J., Carlton, J., Peasgood, T., Brazier, J., Scope, A., Clowes, M., Jones, K., 2018. Developing content for a new generic QALY measure: Results from a qualitative literature review (E-QALY Project). Value in Health ABSTRACT ONLY 21(Suppl. 1), S110. https://doi.org/10.1016/j.jval.2018.04.743.

Naimi, B. Package 'usdm'. Retrieved August 1, 2015 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 Technology Assessment 16(16). http://doi.org/10.3310/hta16160.

NICE. Positioning statement on use of the EQ-5D-5L valuation set. Retrieved January 19, 2017 from https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l.

Nimrod, G., Ben-Shem, I., 2015. Successful aging as a lifelong process. Educational Gerontology 41(11), 814-824. http://doi.org/10.1080/03601277.2015.1050904.

OECD, 2017. Health at a Glance 2017: OECD Indicators. OECD Publishing, Paris. http://doi.org/10.1787/health_glance-2017-en.

OECD/EU, 2018. Health at a Glance: Europe 2018: State of Health in the EU Cycle. OECD Publishing, Paris/European Union, Brussels. https://doi.org/10.1787/health_glance_eur-2018-en.

Pavot, W., Diener, E., 1993. Review of the Satisfaction with Life Scale. Psychological Assessment 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. Nederlands Tijdschrift voor Geneeskunde 139, 1376-1380.

Power, M., Quinn, K., Schmidt, S., WHOQOL-OLD Group, 2005. Development of the WHOQOL-Old module. Quality of Life Research 14, 2197-2214. http://doi.org/10.1007/s11136-005-7380-9.

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. Quality of Life Research 28, 1429-1439. https://doi.org/10.1007/s11136-019-02114-y.

Rand, S., Malley, J., Towers, A.-M., Netten, A., Forder, J., 2017. Validity and test-retest reliability of the self-completion adult social care outcomes toolkit (ASCOT-SCT4) with adults with long-term physical, sensory and mental health conditions in England. Health and Quality of Life Outcomes 15, 163. https://doi.org/10.1186/s12955-017-0739-0.

Reckers-Droog, V.T., Van Exel, N.J.A., Brouwer, W.B.F., 2018. Looking back and moving forward: On the application of proportional shortfall in healthcare priority setting in the Netherlands. Health Policy 122(6), 621-629. https://doi.org/10.1016/j.healthpol.2018.04.001.

Revelle, W. Package 'psych'. Retrieved August 30, 2015 from https://cran.r-project.org/web/packages/psych/psych.pdf and for a description of the package see: https://CRAN.R-project.org/package=psych.

Sabatini, F., 2014. The relationship between happiness and health: Evidence from Italy. Social Science & Medicine 114, 178-187. http://doi.org/10.1016/j.socscimed.2014.05.024.

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 & Research 49(2), 156-163. https://doi.org/10.1002/art.10993.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., Bühner, M., 2010. Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. Methodology 6, 147-151. https://doi.org/10.1027/1614-2241/a000016.

SCP, 2014. Replicatie van het meetinstrument voor sociale uitsluiting. Sociaal en Cultureel Planbureau, Den Haag, The Netherlands.

Sen, A., 1982. Choice, Welfare and Measurement. Harvard University Press, Cambridge, MA.

Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin 86(2), 420-428.

Smarr, K.L., Keefer, A.L., 2011. Measures of depression and depressive symptoms. Arthritis Care & Research 63(S11), S454-S466. https://doi.org/10.1002/acr.20556.

Stamuli, E., 2011. Health outcomes in economic evaluation: Who should value health? British Medical Bulletin 97(1), 197-210. https://doi.org/10.1093/bmb/ldr001.

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.

The EuroQol Group, 1990. EuroQol – a new facility for the measurement of health-related quality of life. Health Policy 16(3), 199-208. https://doi.org/10.1016/0168-8510(90)90421-9.

The WHOQOL Group, 1998. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychological Medicine 28(3), 551-558. https://doi.org/10.1017/S0033291798006667.

Van Exel, J., Baker, R., Mason, H., Donaldson, C., Brouwer, W., EuroVaQ Team, 2015. Public views on principles for health care priority setting: Findings of a European cross-country study using Q methodology. Social Science & Medicine 126, 128-137. http://doi.org/10.1016/j.socscimed.2014.12.023.

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 in Health 18(1), 35-43. http://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.

Van Reenen, M., Janssen, B., 2015. EQ-5D-5L User Guide. EuroQol Research Foundation, Rotterdam, The Netherlands.

Versteegh, M.M., Brouwer, W.B.F., 2016. Patient and general public preferences for health states: A call to reconsider current guidelines. Social Science & Medicine 165, 66-74. https://doi.org/10.1016/j.socscimed.2016.07.043.

Versteegh, M., Knies, S., Brouwer, W., 2016a. From good to better: New Dutch guidelines for economic evaluations in healthcare. PharmacoEconomics, 34 1071-1074. http://doi.org/10.1007/s40273-016-0431-y.

Versteegh, M.M., Vermeulen, K.M., Evers, S.M.A.A., De Wit, G.A., Prenger, R., Stolk, E.A. 2016b. Dutch tariff for the five-level version of EQ-5D. Value in Health 19(4), 343-352. http://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. Medical Care 34(3), 220-233.

Watts, S., Stenner, P., 2012. Doing Q Methodological Research: Theory, Method and Interpretation. SAGE Publications, London. http://doi.org/10.4135/9781446251911.

Whitehead, S.J., Ali, S., 2010. Health outcomes in economic evaluation: The QALY and utilities. British Medical Bulletin 96(1), 5-21. https://doi.org/10.1093/bmb/ldq033.

WHO, 1948. Constitution of the World Health Organization. World Health Organization, Geneva, Switzerland.

WHO, 2015. World report on ageing and health. World Health Organization, Geneva, Switzerland.

Wilhelmson, K., Andersson, C., Waern, M., Allebeck, P., 2005. Elderly people's perspectives on quality of life. Ageing & Society 25(4), 585-600. http://doi.org/10.1017/S0144686X05003454.

Williams, B., Onsman, A., Brown, T., 2010. Exploratory factor analysis: A five-step guide for novices. Journal of Emergency Primary Health Care 8(3), 990399.

Yesavage, J.A., Sheikh, J.I., 1986. Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. Clinical Gerontologist 5(1/2), 165-173. http://doi.org/10.1300/J018v05n01_09.

Zabala, A., 2014. qmethod: A package to explore human perspectives using Q methodology. The R Journal 6(2), 163-173. http://journal.r-project.org/archive/2014-2/zabala.pdf.

Zabala, A., Held, M. Package 'qmethod'. Retrieved 2018 from: https://cran.r-project.org/web/packages/qmethod/qmethod.pdf and for a description of the package see: https://cran.r-project.org/web/packages/qmethod/index.html.

Zorginstituut Nederland, 2016. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. Zorginstituut Nederland, Diemen, The Netherlands.

Zorginstituut Nederland, 2018. Ziektelast in de praktijk: De theorie en praktijk van het berekenen van ziektelast bij pakketbeoordelingen. Zorginstituut Nederland, Diemen, The Netherlands.

Accurately measuring and advancing the well-being of *all* older people should be high on the scientific and societal priority list. This thesis contributes to measuring the broader benefits of health and social care services for older people, in a feasible, comprehensive, reliable and valid manner. Next to testing the validity of existing measures of the well-being of older people, this thesis presents the development and initial testing of a new alternative: the Well-being of Older People measure (WOOP). The WOOP aims to capture a comprehensive set of well-being domains relevant to older people, accommodating the diversity in views on what they consider important for their well-being.

Mariska Hackert (1991) graduated cum laude from the Research Master Social and Cultural Science at the Radboud University in Nijmegen. She wrote her thesis at the Erasmus School of Health Policy & Management at the Erasmus University in Rotterdam.

In loving memory of my grandparents