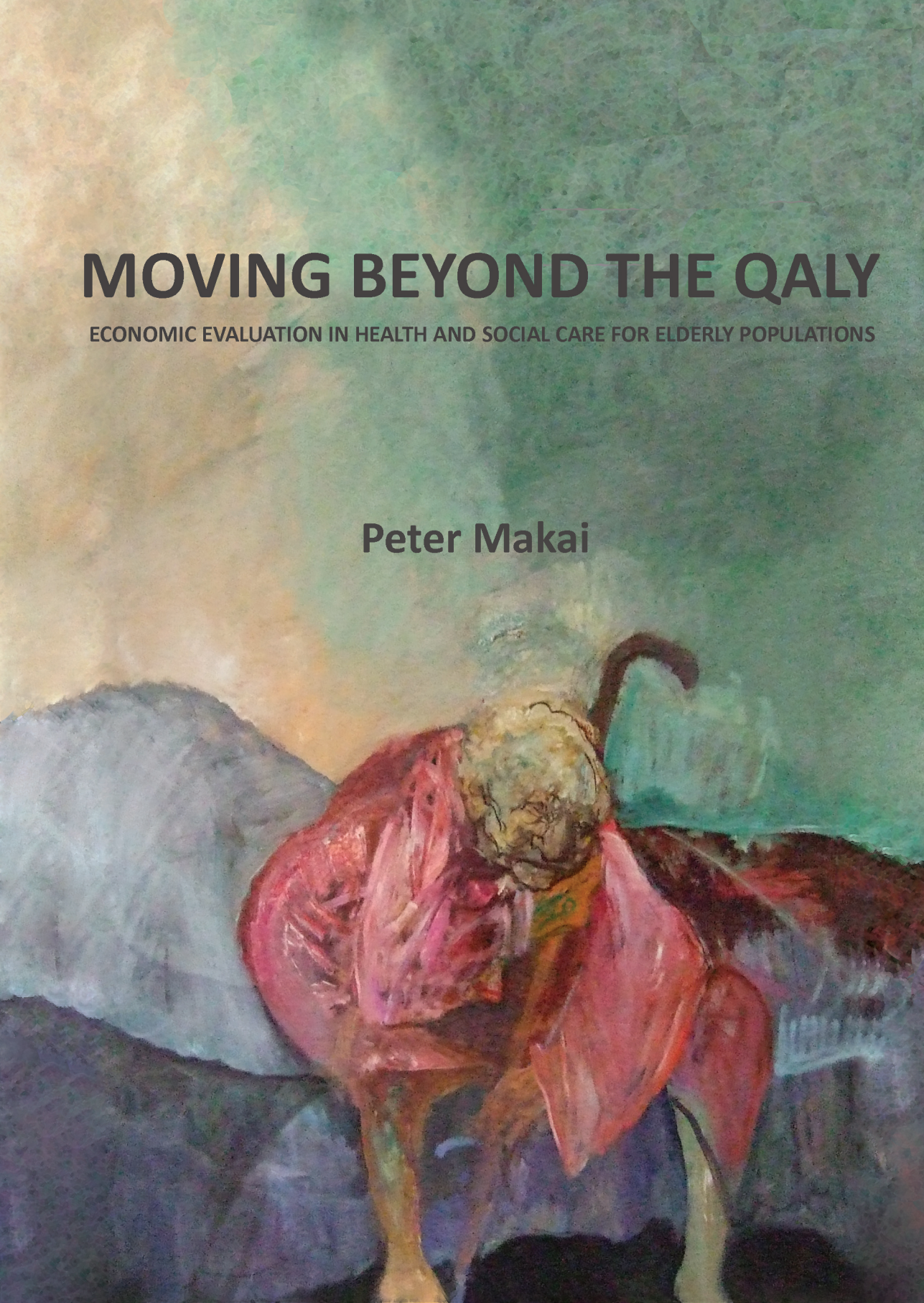
MOVING BEYOND THE QALY

ECONOMIC EVALUATION IN HEALTH AND SOCIAL CARE FOR ELDERLY POPULATIONS

Peter Makai



MOVING BEYOND THE QALY

ECONOMIC EVALUATION IN HEALTH AND SOCIAL CARE FOR ELDERLY POPULATIONS

Peter Makai

FUNDING:

The studies described in this thesis were financed by the Netherlands Organisation for Health Services Research (ZonMw).

Makai, P.

Moving Beyond the QALY: Economic Evaluation in Health and Social Care for Elderly populations

Cover Design: MeDana

Printing: Ipskamp Drukkers B.V., Enchede

Cover illustration: Hyman Bloom “Old woman in Red”, reprinted with kind permission from Hyman Bloom’s estate.

ISBN: 978-94-6259-100-4

MOVING BEYOND THE QALY:

ECONOMIC EVALUATION IN HEALTH AND SOCIAL CARE FOR ELDERLY POPULATIONS

Verder kijken dan de QALY:

Economische evaluaties in de gezondheid en langdurige zorg voor oudere populaties

Proefschrift

ter verkrijging van de graad van doctor aan de

Erasmus Universiteit Rotterdam

op gezag van de

rector magnificus
Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 10 april 2014 om 13:30 uur
door

Peter Makai

Geboren te Boedapest, Hongarije

Promotores:
Prof.dr. W.B.F Brouwer
Prof.dr. A.P. Nieboer
Copromotor:
Dr. M.A. Koopmanschap

Overige leden:
Prof.dr. M.P.M.H. Rutten-van Mölken
Prof.dr. J. Coast
Prof.dr. J.J.V. van Busschbach

Contents

Chapter 1. General Introduction

Economic evaluations

The costs of (publicly funded) healthcare have grown rapidly in the previous decades [1]. In spite of this fact, it has become increasingly clear that health care resources are limited. Policy makers therefore face the challenge of optimally allocating scarce healthcare resources over competing alternatives. The increase in health care expenditures is importantly due to the increasing medical possibilities and availability of health technologies [2] [3], such as pharmaceuticals, surgical procedures, diagnostic tests, and public health interventions. Deciding which competing interventions to fund and more importantly perhaps, which not to fund, becomes particularly pressing during budget cuts when the limits of healthcare spending become apparent. Such decisions require a transparent and systematic framework for evaluating healthcare interventions, which go beyond more traditional criteria such as safety, efficacy and effectiveness of the interventions. In doing so, economic evaluations can support policy makers to optimally allocate health and social care resources within limited budgets by comparing two or more healthcare interventions and investigating their relative value for money [4].

In the field of health care, the most common form of economic evaluation is Cost-Effectiveness Analysis (CEA) [4]. Typically, in a CEA one compares standard care and a new intervention aimed at treating a particular medical condition in terms of costs and effects. By calculating the difference in costs between standard care and new interventions and the difference in effects between standard care and the new intervention, the relationship between costs and effects can be derived. This relationship is summarized by the ICER, the ratio between the marginal costs of the new intervention divided by the marginal effects of the new intervention:

$$ICER = \frac{Ci - Cs}{Ei - Es}$$

where Ci are the costs related to the intervention, Cs the costs related to standard care, Ei represents the effects of the intervention, and Es the effects of standard care. The effects are not expressed in monetary terms, as is the case in a conventional Cost-Benefit Analysis [5], but can be expressed in a variety of ways. In CEA any clinically relevant outcome measure related to the intervention is possible, such as life years gained, hip fractures avoided or event free life years. The problem of such diverse outcome measures is that it complicates decision making, as ICERs using different outcome measures are incomparable. A dominant sub-type of CEA is Cost Utility Analysis (CUA) [4]. In a CUA Quality Adjusted Life Years (QALYs) are used as outcome measures. QALYs are a preference-based health measure comprising both length and health-related quality of life (HrQol).

To assess HrQol improvements, changes in health status are measured using preference based instruments. Preference-based instruments normally consist of two separate elements, (1) a descriptive system defined by the dimensions and answer categories of the instruments and (2) a (pre-scored) weighting system [6]. The dimensions in the descriptive system represent various health domains, such as mobility and pain, for which people can indicate their level of functioning. The weighting system allows particular states as described with the descriptive system to be transformed into a 'utility score', commonly reflecting the average strength of preferences for the various states described in the descriptive system. In case of HrQol, these scores are anchored to a standardized scale, with 1 representing the utility of the best imaginable health state, and 0 representing the value for dead or health states that are considered equivalent to dead [4] [6]. Negative values are also possible and relate to health states valued as 'worse than dead' [4] [6]. In comparison to CEAs, CUAs consistently inform health care decision-makers. While the results of CEAs using a specific clinically relevant outcome measure can only be meaningfully compared with other studies using that outcome measure, in principle the results of CUAs can be compared to any CUA within or outside of a particular medical condition. Therefore, by using CUA it is also possible to make allocation decisions in healthcare across different disease areas [4].

Ageing population

The continuing ageing of the population in many countries, particularly in Western countries has raised questions regarding the sustainability of health care systems and the optimal allocation of scarce health care resources. In numerous countries, life expectancy has risen considerably during the past decades, and can be considered a worldwide problem [7]. While this increase may partly reflect the success of the health care sector and improved medical technology, it also increases the demand for care and, hence, the pressure on health care budgets. In the Netherlands, for example, life expectancy has increased from 70.29 for men and 72.58 for women in 1950 to 78.77 for men and 82.72 for women in 2010 [8]. The majority of elderly suffers from multi-morbid and chronic conditions [9]. An important example of such a condition is dementia, with a prevalence in the Netherlands of 1% for elderly above 75 and 5% above 85 years [10], impacting both patient outcomes as well as costs. Another important group of care users is the group of frail elderly. Frailty can be defined as an accumulation of physical, mental and social deficits in functioning increasing the risk of adverse health outcomes [11], such as mortality, falls, disability, hospitalization or nursing home admission [12, 13]. The number of frail elderly above 65 in the Netherlands is expected to grow in the following years from 650.000 in 2010 to 1.160.000 [14].

The impact of multi-morbidity, frailty and chronic conditions on HrQol and wellbeing over time may have multiple forms, for example functional decline [15], loss of friends and social contacts due to decreased mobility [16], as well as depression caused by the negative changes in health or personality (e.g. in case of dementia) [17]. Decreased HrQol and wellbeing of elderly, due to chronic conditions may also translate into higher usage of health and social services. While a large number of elderly are able to remain at home, often supported by home-care services and informal care, elderly are also relatively frequently institutionalized. As a consequence of higher healthcare use, elderly consume a disproportionate amount of healthcare resources: 37,6% of total healthcare costs are directed at 15% of the population [10]. Healthcare costs associated with ageing are expected to increase in the coming decades, as the number of elderly above 65 is projected to increase to 4.2 million in 2035 from 2.4 million in 2010 [8]. Thus, the growing number of elderly raises the double challenge of providing effective services at an acceptable level of costs. This increases the necessity to base allocation decisions on a tradeoff between costs and outcomes of interventions.

Problems with outcomes in elderly populations

When searching for appropriate measures for economic evaluations of interventions for health and social care for elderly it is important that an outcome measure captures all relevant benefits that are brought about by an intervention. The starting point is the goal of the service/intervention at hand. We can illustrate this with the example of restraints in long-term care settings. In long-term care restraints are used to prevent elderly from falling [18]. While an intervention to reduce the use of physical restraints may not directly improve a patient's health [19] (in fact, it may increase risk of falling), it restores dignity, freedom of movement, and control – concepts not easily classified under health. The reason to wish to reduce physical constraints may therefore not be found in associated health gains, but in these other outcomes. Therefore, it is pivotal to have outcome measures that capture all relevant outcomes in the context of interventions aimed at (frail) elderly, including those 'beyond health', as health improvement need not be the only, or even primary, goal of such interventions.

One of the most important difficulties in performing CUAs of health and social care interventions in elderly populations thus concerns the availability of outcome measures attuned to the goals of the health and social care they receive [20] [21]. Contrary to the costs associated with social care, the benefits beyond health produced by social care remain largely unknown [22]. One way of capturing such beyond health effects in economic evaluations is through the use of wellbeing instruments

[23]. Existing wellbeing instruments have a varied theoretical basis and often focus on different well-being domains. For example, the form of wellbeing most familiar to economists is subjective wellbeing, which is mainly concerned with the measurement and determinants of happiness [24]. Besides subjective wellbeing, other forms of wellbeing concepts also exist, and focus on various other beyond-health dimensions. Psychological wellbeing focuses mainly on psychological domains such as self-acceptance, autonomy, purpose in life, control of one's surroundings [25]. Besides such traditional approaches, capability wellbeing [23] is a relatively recent conceptualization of wellbeing, focusing on what an individual can do and be in their lives. Capability theory was developed by Sen [26], distinguishing functionings from capabilities. Functionings are actual beings and doings of the individual, while capabilities are potential functioning. For example, while a man starving and a man fasting are both at the same level of functioning, one has the capability to eat and chooses not to, while the other lacks such a capability, showing an inequality of both states. Through this example, Sen emphasizes the importance of individual choice in terms of achieving a particular form of functioning, while having disparities in capabilities [26]. In this sense, capability theory is an alternative to utility theory especially, not focusing on outcomes and achievements but on freedom and possibilities. Capability theory as a conceptual framework allows both health and wellbeing outcomes to be measured, and allows the development of capability wellbeing instruments [23]. By focusing on capabilities, the evaluative space of capability wellbeing instruments is also deepened, besides broadening it by measuring a wider set of outcomes compared to conventional health-related utility measures.

The first instrument aimed explicitly to capture the benefits of health and social care interventions in elderly populations is the ICECAP-O [27]. The ICECAP-O aims to directly measure capabilities and deduce achieved capability wellbeing. The ICECAP-O was intended as an outcome measure for economic evaluation in order to integrally measure health and social care interventions used by the elderly [21]. The ability of the ICECAP-O to differentiate between people in various health states and to measure wellbeing has been established in several development papers [21, 28]. However, at the onset of this study the ICECAP-O was not widely validated, nor had it been used in economic evaluations. In this current study, several chapters use the ICECAP-O in order to address these issues.

Aim and structure of this thesis

The aim of this thesis is to address a number of issues related to outcome measurement in economic evaluations in elderly populations consuming health and social care. The focus of this

thesis concerns outcomes in economic evaluations, particularly the ICECAP-O. In order to be of use in a given population, an outcome measure has to be validated, therefore its convergent validity (whether ICECAP-O measures a similar concept as related instruments) and discriminant validity (whether the instrument is able to discriminate between groups that are expected to be different in terms of capability well-being) has to be thoroughly investigated. Furthermore, in order to investigate the policy relevance of a measure, its performance within an actual economic evaluation needs to be investigated. The latter use, depending on the intervention studied, may also demonstrate whether the use of a broader outcome measure actually leads to other conclusions about the cost-effectiveness of an intervention than when using a more restricted outcome measure.

Therefore, the research questions addressed in this thesis are as follows:

1. How is an economic evaluation performed in elderly care when using a conventional outcome measure?
2. Which instruments are potentially useful for economic evaluation in elderly care, which produce benefits beyond health?
3. Is the ICECAP-O a valid measure of capability well-being in different settings?
4. Are there differences in ICECAP-O scores between different groups of respondents?
5. How is an economic evaluation performed in elderly care when using the ICECAP-O as outcome measure?

Figure 1.1 shows the structure of the remainder of this Thesis.

The thesis consists of the following chapters. Chapter 2 presents a cost-effectiveness analysis of the "Care for better pressure ulcer quality collaborative". In this chapter, research question 1 is addressed, and some of the limitations of economic evaluations in the long-term care using conventional outcome measures are highlighted.

Chapter 3 reports on a literature review aimed at finding suitable outcome measures for economic evaluations for elderly care. The most suitable instrument identified is the ICECAP-O. Chapter 3 also gives a number of considerations for the selection of an appropriate instrument to be used in economic evaluations (Research question 2).

Chapter 4 details the validation of the ICECAP-O in a population of post-hospitalized older people, and investigates whether the ICECAP-O indeed measures various health dimensions as well as wellbeing (Research questions 3 and 4).

Chapter 5 follows up on chapter 4 and shows the validation of the ICECAP-O in a psycho-geriatric nursing-home setting in the Netherlands, using proxy respondents. This chapter gives some

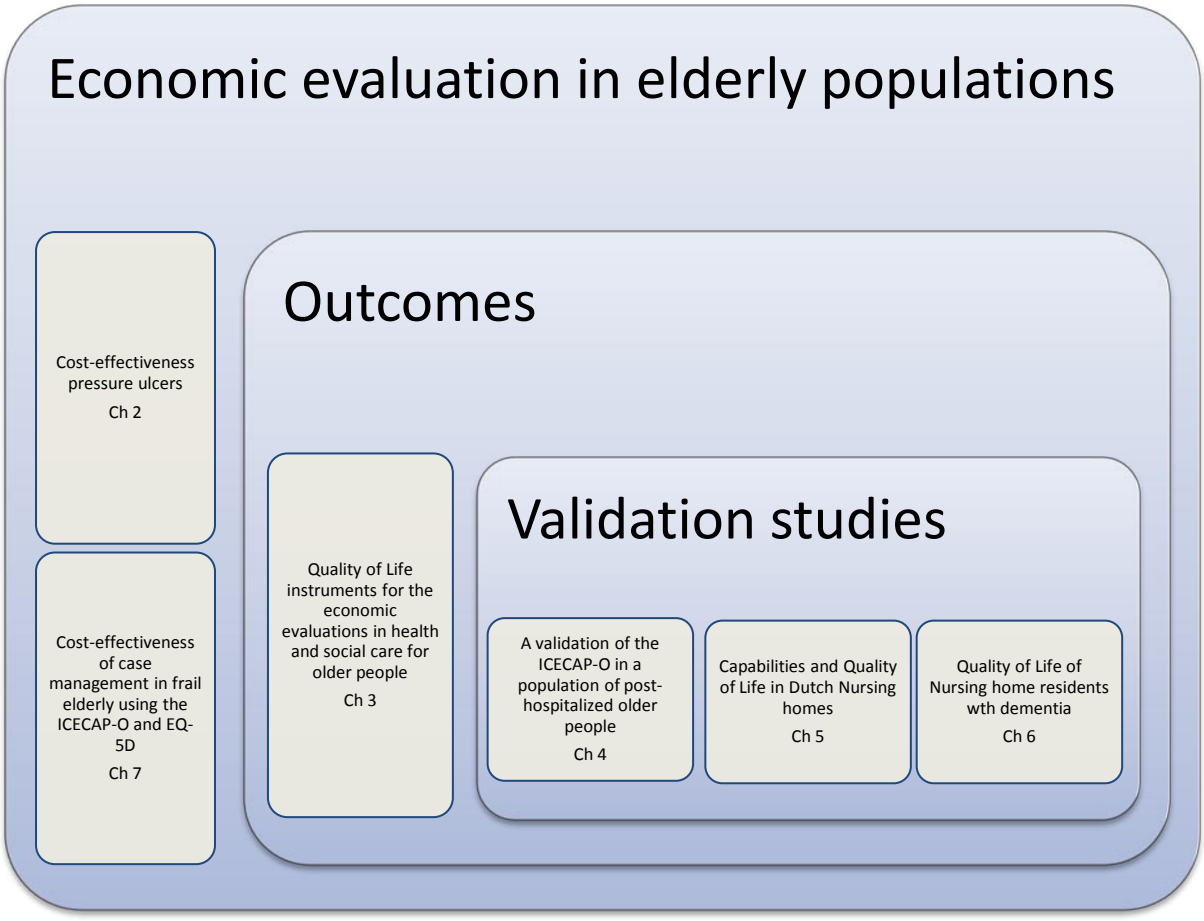
important insights into how different proxy respondents evaluate the well-being of a restrained population versus a non-restrained population (Research questions 3 and 4).

Chapter 6 reports on a validation study of the ICECAP-O in a German psycho-geriatric nursing home. This chapter focuses on how physical health is captured by the ICECAP-O and how different types of nursing proxies completed the ICECAP-O (Research questions 3 and 4).

Chapter 7 presents a cost-effectiveness study using both the ICECAP-O and the EQ-5D as outcome measures, performed in a community-dwelling population of frail elderly (Research question 5).

Chapter 8 summarizes the results, deals with methodological and theoretical implications, and gives some policy recommendations based on the research presented in this thesis.

Figure 1.1: Structure of thesis



References

1. OECD, *Health at a Glance 2011: OECD Indicators, .* 2011, OECD Publishing: Paris.
2. Chandra, A. and J.S. Skinner, *Technology growth and expenditure growth in health care*, 2011, National Bureau of Economic Research.
3. de Meijer, C., et al., *The effect of population aging on health expenditure growth: a critical review.* European Journal of Ageing, 2013: p. 1-9.
4. Drummond, M.F., et al., *Methods for the economic evaluation of health care programmes*. Vol. Third Edition. 2005, Oxford: Oxford University Press.
5. McIntosh, E., P. Clarke, and E. Frew, *Applied methods of cost-benefit analysis in health care*. Vol. 4. 2010: Oxford University Press.
6. Brazier, J., et al., *Measuring and Valuing Health Benefits for Economic Evaluation*. Vol. 1. 2007, Oxford: Oxfrod University Press.
7. Kalache, A., S.M. Barreto, and I. Keller, *Global ageing: The demographic revolution in all cultures and societies.* The Cambridge handbook of age and ageing, 2005. **606**: p. 30.
8. CBS, *Statline - Life Expectancy*, 2010, CBS (Central Bureau of Staticstics of The Netherlands): Den Haag.
9. Marengoni, A., et al., *Aging with multimorbidity: A systematic review of the literature.* Ageing Research Reviews, 2011. **10**(4): p. 430-439.
10. Slobbe L.C.J., J.M. Smit, Groen J, Poos M.J.J.C, Kommer GJ, *Kosten van Ziekten in Nederland 2007 : Trends in de Nederlandse zorguitgaven 1999-2010 [Cost of Illness in the Netherlands 2007 : Trends in healthcare expenditure 1999-2010]*, 2011, Rijksinstituut voor Volksgezondheid en Milieu RIVM: Bilthoven.
11. Gobbens, R.J., et al., *The Tilburg Frailty Indicator: psychometric properties.* Journal of the American Medical Directors Association, 2010. **11**(5): p. 344-355.
12. Lin, C.C., et al., *Reduced health-related quality of life in elders with frailty: A cross-sectional study of community-dwelling elders in Taiwan.* PLoS ONE, 2011. **6**(7): p. e21841.
13. Puts, M.T., et al., *The effect of frailty on residential/nursing home admission in the Netherlands independent of chronic diseases and functional limitations.* European Journal of Ageing, 2005. **2**(4): p. 264-274.
14. Van Campen, C., *Kwetsbare ouderen.* Den Haag: Sociaal en Cultureel Planbureau, 2011.
15. Beswick, A.D., et al., *Complex interventions to improve physical function and maintain independent living in elderly people: a systematic review and meta-analysis.* The Lancet, 2008. **371**(9614): p. 725-735.
16. De Vries, N., et al., *Outcome instruments to measure frailty: a systematic review.* Ageing Research Reviews, 2011. **10**(1): p. 104.
17. Djernes, J., *Prevalence and predictors of depression in populations of elderly: a review.* Acta Psychiatrica Scandinavica, 2006. **113**(5): p. 372-387.
18. Huizing, A.R., et al., *Preventing the use of physical restraints on residents newly admitted to psycho-geriatric nursing home wards: A cluster-randomized trial.* Int J Nurs Stud, 2009. **46**(4): p. 459-69.
19. Huizing, A.R., et al., *A Cluster-Randomized Trial of an Educational Intervention to Reduce the Use of Physical Restraints with Psychogeriatric Nursing Home Residents.* Journal of the American Geriatrics Society, 2009. **57**(7): p. 1139-1148.
20. Makai, P., et al., *Quality of life instruments for economic evaluations of elderly care: a systematic review.* Social Science & Medicine (Accepted for publication).
21. Coast, J., et al., *Valuing the ICECAP capability index for older people.* Social science & medicine (1982), 2008. **67**(5): p. 874-882.

22. Nieboer, A.P., X. Koolman, and E.A. Stolk, *Preferences for long-term care services: Willingness to pay estimates derived from a discrete choice experiment.* Social science & medicine, 2010. **70**(9): p. 1317-1325.
23. Al-Janabi, H., T. Flynn, and J. Coast, *Development of a self-report measure of capability wellbeing for adults: the ICECAP-A.* Quality of Life Research, 2012. **21**(1): p. 167-176.
24. Kahneman, D. and A.B. Krueger, *Developments in the measurement of subjective well-being.* The journal of economic perspectives, 2006. **20**(1): p. 3-24.
25. Ryff, C.D. and C.L.M. Keyes, *The structure of psychological well-being revisited.* Journal of personality and social psychology, 1995. **69**(4): p. 719.
26. Sen, A., *Choice, welfare and measurement*. 1982, Cambridge, MA: Harvard University Press.
27. Coast, J., et al., *Investigating Choice Experiments for Preferences of Older People (ICEPOP): evaluative spaces in health economics.* Journal of health services research & policy, 2008. **13 Suppl 3**(Journal Article): p. 31-37.
28. Coast, J., et al., *An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people.* Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 2008. **17**(7): p. 967-976.

Chapter 2. Cost-effectiveness of a pressure ulcer quality collaborative

This chapter is based on Peter Makai, Marc Koopmanschap, Anna Nieboer and Roland Bal:
Cost-effectiveness of a pressure ulcer quality collaborative
Cost-effectiveness and Resource Allocation 2010, 8:11

Abstract

Background

A quality improvement collaborative (QIC) in the Dutch long-term care sector (nursing homes, assisted living facilities, home care) used evidence-based prevention methods to reduce the incidence and prevalence of pressure ulcers (PUs). The collaborative consisted of a core team of experts and 25 organizational project teams. Our aim was to determine its cost-effectiveness from a healthcare perspective.

Methods

We used a non-controlled pre-post design to establish the change in incidence and prevalence of PUs in 88 patients over the course of a year. Staff indexed data and prevention methods (activities, materials). Quality of life (Qol) weights were assigned to the PU states. We assessed the costs of activities and materials in the project. A Markov model was built based on effectiveness and cost data, complemented with a probabilistic sensitivity analysis. To illustrate the results of longer term, three scenarios were created in which change in incidence and prevalence measures were (1) not sustained, (2) partially sustained, and (3) completely sustained.

Results

Incidence of PUs decreased from 15% to 4.5% for the 88 patients. Prevalence decreased from 38.6% to 22.7%. Average Quality of Life (Qol) of patients increased by 0.02 Quality Adjusted Life Years (QALY)s in two years; healthcare costs increased by €2000 per patient; the Incremental Cost-effectiveness Ratio (ICER) was between 78,500 and 131,000 depending on whether the changes in incidence and prevalence of PU were sustained.

Conclusions

During the QIC PU incidence and prevalence significantly declined. When compared to standard PU care, the QIC was probably more costly and more effective in the short run, but its long-term cost-effectiveness is questionable. The QIC can only be cost-effective if the changes in incidence and prevalence of PU are sustained.

Background

A pressure ulcer (PU) is a preventable condition that affects patients with impaired mobility, especially the elderly [1]. PUs are classified from grades 1 to 4, or least to most severe. The average prevalence of PUs in the Netherlands is 7.9% in assisted living homes and 18.3% in nursing homes [2]. Incidence varies between 2.9% and 4.5% in intensive care [3]. No incidence data are available for the Dutch long-term care sector. The probability of healing within 90 days varies with severity: 67% (grade 2), 44% (grade 3) and 32% (grade 4) [4]. PUs can interfere with recovery, cause pain and infection [1], and increase mortality (OR=1.4 after adjusting for risk factors) [5]. According to a study by Franks [6] the quality of life of PU patients is no worse than the general population of nursing home patients; a study by Fleurence, [7] however, claims that PUs decrease quality of life. The treatment of PUs costs between € 89 million and 1.9 billion, or 0.1% to 1% of total Dutch healthcare costs [8] [9]. Because they are preventable, it is safe to say that PUs should not occur in the first place.

Preventable conditions requiring a common and perhaps demanding treatment like PUs are likely candidates for Quality Improvement Collaboratives (QICs), [10] [11], in which different healthcare organizations address a certain problem by implementing specific solutions and sharing the results [12]. A QIC program team includes experts in both the health condition and methods of quality improvement. According to a recent systematic review, QICs have shown moderate effectiveness in terms of patient outcomes [10] and several studies suggest effectiveness of QICs for PUs in particular [13] [14]. Despite the popularity of QIC's, the cost-effectiveness of QICs is rarely considered [10], in fact only a study by Huang addressed this aspect [15].

This is not surprising, since the costs of quality improvement projects are not well established, and organizations generally do not or cannot assess the benefits of participation [16]. There are currently no published studies on the cost-effectiveness of a PU QIC in particular. Several studies have been published on the cost-effectiveness of the materials for PU treatment and prevention [7] [17] [18] [19], and the one study we found that focused on labor costs [20] considered only nurse staffing time and disregarded preventive activities. We did identify a cost- effectiveness study on a PU quality improvement project [21], but it did not involve a QIC. This study adds to the literature by giving a detailed account of the PU sub-program of the "Care for Better" QIC, a Dutch healthcare collaborative[22]. The aim of this article is to answer the question: Was this PU QIC cost-effective when compared to standard PU care?

Methods

Design

Our study was conducted from a healthcare perspective, considering both direct costs of PU care and costs of the QIC for a period of one year. A prospective pre-post design was used with one-month measurement periods to collect data on costs and effectiveness. We established cost effectiveness by comparing data at the end of the project year to standard care (i.e., the state of the sample before the QIC intervention). We built a Markov model to establish standard care (i.e. simulate a control group), and to determine the effect of the collaborative after a year. To extrapolate results to one additional year, we have expanded this model. Probabilistic sensitivity analysis was applied to treat uncertainty in the model parameters. QALYs and ICERs were calculated for a two year period (project year and extrapolated year).

Setting

The Care for Better QIC operates in the Dutch long term care sector (nursing homes, residential care homes, and home care). This study is limited to nursing and residential homes. Patients are not admitted with PU as a main condition, but have underlying chronic conditions affecting their daily functioning. The nursing home patients typically stay in the facilities for two to three [23] [24] years until death, and are seldom discharged.

Description of the Collaborative

The overall goal of the Care for Better PU QIC was to reduce the prevalence and incidence of PUs by 50% in 25 participating organizations over the course of a year by increasing evidence-based preventive measures and decreasing non-useful preventive measures (Table 2.1) [1], thereby reducing the need for treating PUs. The project was implemented in three consecutive rounds because not all 25 organizations could be accommodated by the Care for Better PU QIC at one time.

The Care for Better PU QIC carried out activities on three intertwining levels: program, organizational, and departmental (Figure 2.1). The program level consisted of a core team of experts who guided the organizations' project teams, defined the collaborative's goals, and organized three "learning sessions" during the year at which project teams could be taught about quality improvement methods and preventive nursing measures, and share their results with the other teams. Between the learning sessions, the core team of experts provided project teams with coaching.

Table 2.1: Patient characteristics, outcomes and changes in process

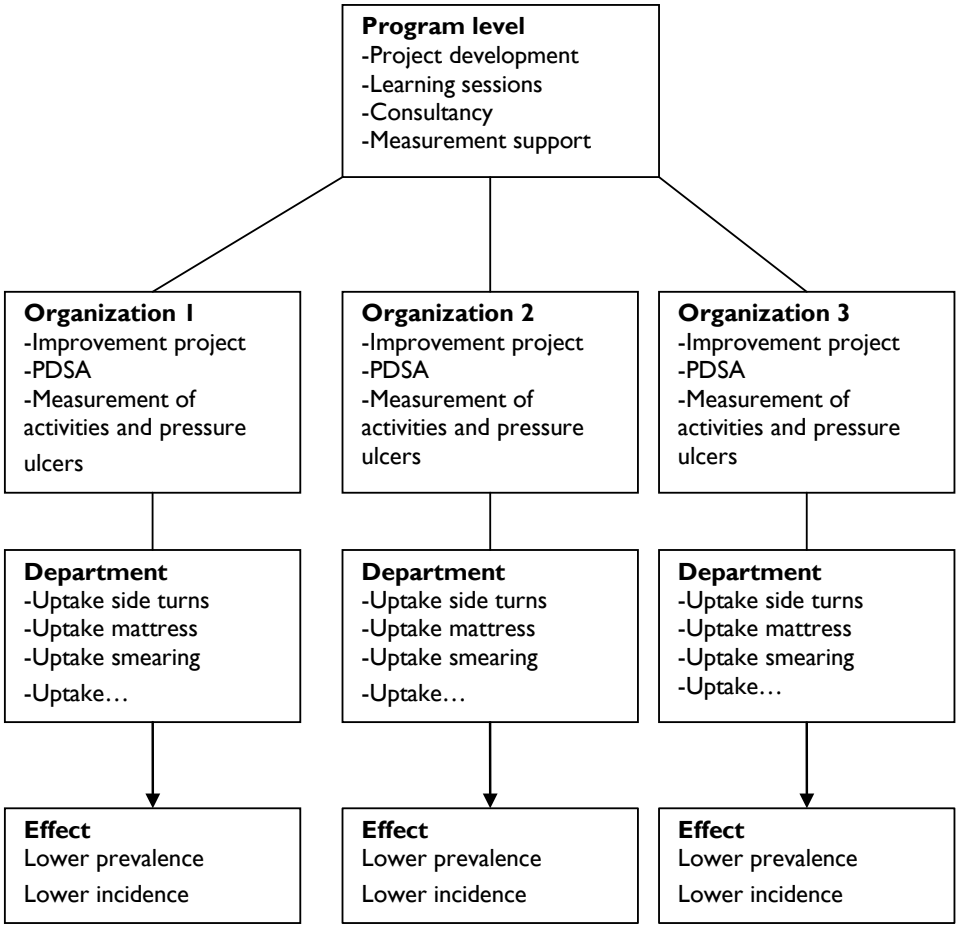
	Non-selected patients	***Selected patients***	
Number of patients	254	88	
BMI (average)	26	26 (5)	
Age (average)	80	82	
Females (average)	169 (67%) F	60 (68%) F	
Patients at risk of pressure ulcers (average)	254 (100%)	88 (100%)	
Comparison of clinical effects	*Baseline*	*Baseline*	*After*
Prevalence			
Grade 1	50 (20%)	21 (23.9%)	16 (18.2%)
Grade 2	9 (3.5%)	10 (11.4%)	2 (2.3%)
Grade 3	3 (1.2%)	1 (1.1%)	1 (1.1%)
Grade 4	5 (2%)	2 (2.3%)	1 (1.1%)
Total	59 (27%)	34 (38.6%)	20 (22.7%)*
Incidence (1 month)			
Grade 1	19 (7%)	10 (14.7%)	4 (4.5%)
Grade 2	6 (3%)	3 (3.4%)	0 (0%)
Total	25 (9%)	13 (15%)	4 (4.5%)*
Useful interventions			
Risk assessment	254 (100%)	88 (100%)	88 (100%)
Using a 30-degree side to side turn at least every 4 hours	24 (9%)	7 (8%)	9 (10%)
Preventive mattress	78 (30%)	24 (27%)	40 (45%)**
Involving patients in prevention	41 (16%)	3 (3%)	7 (8%)
Involving family/friends/caregivers in prevention	26 (10%)	3 (3%)	9 (11%)
Reactivation and mobilization by paramedics	10 (4%)	3 (3%)	11 (13%)
Smearing of the skin in case of incontinence	30 (11%)	8 (9%)	9 (11%)
Assessing nutritional state and preventing nutritional deficiency	13 (5%)	12 (14%)	4 (5%)

Inserting a catheter to prevent maceration of the skin	3 (1%)	1 (1%)	1 (1%)
Ensuring a clean, dry and square lower layer of bedclothes	52 (20%)	8 (9%)	12 (14%)
Non-useful interventions			
Smearing the skin (with topical agents) to prevent disturbance in blood supply caused by pressure	50 (20%)	23 (26%)	6 (7%)*
Massage	3 (4%)	0 (0%)	2 (2%)
Using a 90-degree side to side turn at least every 4 hours	2 (1%)	0 (0%)	3 (4%)

*P<0.05

**p<0.005

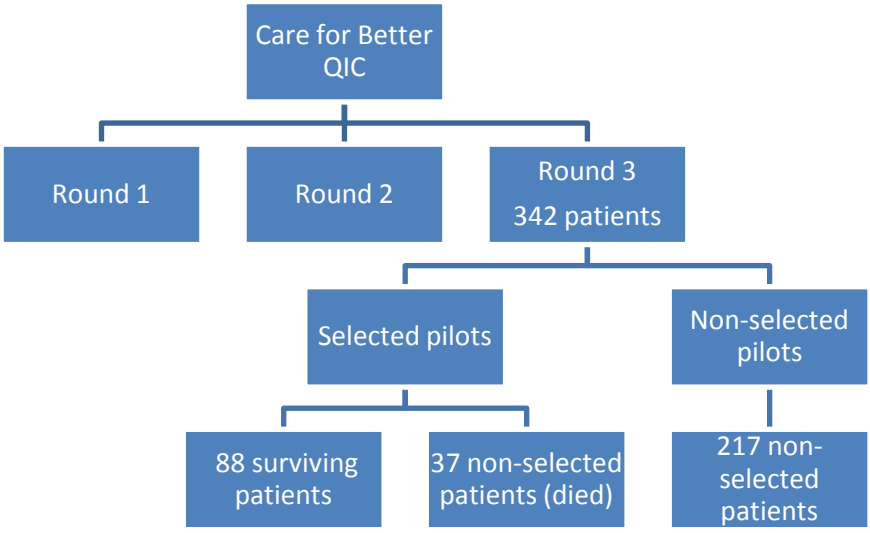
Figure 2.1: The structure of the collaborative



The participating organizations formed project teams who attended the learning sessions and were the effective drivers of the implementation in pilot departments of the organizations. Project teams had considerable freedom in the type of preventive nursing measures implemented and how they were applied, but were encouraged by the experts to formulate SMART (Specific Measureable Attainable Realistic Timely) goals and to work with PDSA (Plan Do Study Act) cycles between the learning sessions. The PDSA cycles began with "action plans" followed by introducing new interventions at the departmental level. Periodic measurement of results were documented. At the end of the cycle, the new interventions were meant to be used in the entire organization, and meant to be incorporated into the work of professionals. In this manner, successful teams standardized the new interventions and made changes permanent. In addition it was expected from the teams that they learn methods of continuous quality improvement, in other words teams were meant to continue working with the PDSA cycle after the QIC program was finished.

During the one-month measurement periods preceding the learning sessions, project teams registered 18 different preventive measures carried out by caregivers, as well as the prevalence, incidence and severity of the PUs. These registrations consisted of 12 measurement moments, measuring every patient on the pilot department every two to three days. The first measurement was conducted end October to end November 2006 or from beginning of November to the beginning of December depending on the institution. The intermittent measurement period was in June, and the last measurement period was in November 2007. The measurements were organized by the Dutch National Expertise Center for Nursing and Caring, and were carried out by the project teams themselves.

Figure 2.2: Selection process of the 88 patients



Case-selection and study population

To capture possible learning effects over the course of the year, data was used from the third round. A total of seven departments in three different organizations were investigated in detail. The following criteria were used to select cases:

1. Data was available for both first and last measurement period.
2. At least one department had a low initial PU prevalence, at least one department had an average PU prevalence, and at least one department had a high PU prevalence.

Using this criteria, 88 patients were selected – ranging from 9-19 per department – to determine cost-effectiveness (Figure 2.2). Their characteristics compared to the non-selected cases in the third round are described in Table 2.1. To determine the representativeness of the selected cases vis-à-vis the entire patient population, we compared the 88 patients' risk for PUs, age, sex, and BMI to the non-included patients in round three of the project using ANOVA at baseline.

Determination of effectiveness

We used effectiveness data on the prevalence and incidence of PUs collected by the organizational project teams. Prevalence was computed by averaging the number of patients with PU divided by 88 over the whole measurement month. Incidence was computed as the number of new PU cases during the measurement month divided by 88. To determine effectiveness, we compared the before- and after-project PU prevalence and incidence of the 88 patients using a t-test.

Assessment of costs

Cost data associated with the project and the prevention and treatment of PUs were collected for the central activities on the program level, the project activities within the organizations, and the individual treatment of patients (departmental level). Identification and valuation of costs are displayed in Table 2.2.

Program and organizational. Program costs were obtained from the central project budget. Items included expected project time, lump sums for materials, and miscellaneous costs. To ascertain organizational level costs, the organizations' project leaders supplied us with detailed plans and reports. They also furnished the individual amounts of time invested in the project by the teams and other employees for various activities (training, participation in learning sessions, writing plans, project implementation). To establish the project costs, we multiplied the number of hours spent on the project by the average hourly wages of the project team members.

Table 2.2: Activities of caregivers and treatment material used

Program level			***Amount***
Labor	Program activities (project design, expert meetings, recruitment, organizing working conferences, mid-term report, final report etc.)	Program experts	4696 hours
	Program support	Program experts	635 hours
	Knowledge management (publications, etc)	Program experts	175 hours
Materials			Lump sum
Other costs			Lump sum
Organizational level			***Average***
Labor	Project activities (coordinating the project, writing action plans, reports, etc.)	Project leader	8 hours (per week)
	Clinical level project implementation	Project member	2 hours (per week)
	Learning session participation	- Project leader - 2 Project members	76 hours (total each)
	Staff knowledge testing	- Nurses - Caregivers	30 min (total each)
	Caregiver training	- Specialized nurse - Caregivers	3.5 hours (total each)
	Specialist training	Nurses	8 hours (total each)
	Project meetings	- Project member - Nurses - Caregivers	8 hours (total each)
	Measurements	Nurses	1 hour (per month)
Departmental level			***Average/day/patient***
Useful interventions	Risk assessment	Nurses	10 sec
	30-degree side turn at least every 4 hours	Caregivers	35 min
	Involving patients in prevention	Nurses	2 sec
	Involving family/friends/caregivers in prevention	Nurses	0.4 sec
	Reactivation and mobilization by paramedics	Paramedics	4 min
	Smearing the skin with topical agents in case of incontinence	Caregivers	2 min
	Assessing nutritional state and preventing nutritional deficiency	Caregivers	4 min
	Inserting a catheter to prevent maceration of the skin	Caregivers	3 min

Non-useful interventions	Ensuring a clean, dry and square lower layer of bedclothes	Caregivers	7 min
	Smearing the skin (with topical agents) to prevent disturbance in blood supply caused by pressure	Caregivers	2 min
	Massage	Caregivers	1 min
	90-degree side turn at least every 4 hours	Caregivers	30 min
	Usual treatment grades 1-2	Caregivers	7 min
	Usual treatment grades 3-4	Caregivers	15 min
Materials		**Type**	**Number/patient**
	Basic mattress	Start	1
	Mattress (grades 1-2)	SLK 1	1
	Mattress (grades 1-2)	Dionica	1
	Mattress (grades 3-4)	SLK 2	1
	Mattress (grades 3-4)	Duo-care	1
	Mattress (grades 3-4)	Quatro-care	1
	Pillow (prevention)	Foam pillow	1
	Pillow (grades 1-2)	Normal PU pillow	1
	Pillow (grades 3-4)	ROHO	1

Departmental. We used project documentation to identify the before- and after-project differences in PU preventive measures and the number of mattresses and pillows used. The type of mattresses and pillows were taken from the organizations' treatment protocols; their rental rates were collected from the suppliers of the organizations (Table 2.3.) Since other materials used for PU care (creams, dressings, and the like) were not reliably administered, we assumed they did not change during the project. Studies have also shown these costs to be marginal compared to the total cost of care [9]. We also didn't account for changes in organizational overhead costs, because the changes all took place in the departments themselves, and had no effect on other parts of the organizations. Time spent by staff on activities related to preventive care was collected through interviews with project members, who were asked to give an average, minimum, and maximum value for each preventive measure. In the context of an average long-term care stay of 2.8 years [25], with 66% remaining until death [26], we assumed that PUs do not cause extra days of care. We computed the cost of personnel at the departmental level by multiplying the time spent on PU care by the hourly wage of caregivers in the organizations. We used the wage schedule of the 2006 collective agreement of Dutch nursing home employees [27].

Table 2.3: Wages of staff and prices of materials

Labor		
	Program experts	115.00
Project (hourly)	Project leader	36.82
	Project member	31.56
	Paramedic	31.56
	Dietician	31.56
Departmental (hourly)	Specialized nurse	34.71
	Nurse	31.56
	Other caregiver	18.94
Materials		
Project (totals)	Project materials	50,000
	Other collaborative costs	64,000
	Basic mattress	1.11
Departmental	SLK 1 (grades 1-2)	2.56
(daily rental price)	Dionica mattress (grades 1-2)	0.64
	SLK 2 (grades 3 & 4)	4.52
	Duo-care mattress (grades 3-4)	3.29
	Quatro-care mattress (grades 3-4)	13.15
	Foam pillow	0.03
	Normal PU pillow	0.04
	Special PU pillow (ROHO)	0.18

To compute an overall cost per patient value, the cost of the collaborative was evenly allocated to the participating project teams. Organizational level costs were evenly allocated to the patients. Average daily costs were computed per patient per disease state and converted into monthly values.

Decision Analytical Model

To determine the effect of the collaborative compared to standard care after a full year, we have built a decision-analytical model (Markov model) based on our data from the collaborative to simulate standard care (i.e. control group). In building the model we have used the method outlined by [28]. The model had health states consisting of no PU, single PUs grades 1-4, and multiple PUs grades 1-4. For the first year (when the collaborative ran), we used two sets of transition probabilities: one for the simulated control-group, and one for the intervention group. To establish standard care, we converted incidence and PU healing during the first measurement month into monthly transition probabilities, giving a simulation under the assumption there was no collaborative. With the intervention group we based transition probabilities on the events of the first year (based on the data from the first and last measurement month) and we transformed these

yearly transition probabilities into monthly transition probabilities. This monthly modeling was necessary to give a more precise change in effects and costs over this first year, and to make the two simulations comparable. Both arms of the model were run 12 times to simulate a one-year program.

To extrapolate the results for an additional year, we also included mortality in the model by introducing a death state into the model, and using the average mortality of nursing home patients in the Netherlands [29] as a transition probability. The simulated control-group thus consisted of no PU, single PUs grades 1-4, and multiple PUs grades 1-4 and death, with the transition probabilities adjusted accordingly. The intervention group, – in addition to a death state – three scenarios were created: total sustainability, partial sustainability and no sustainability. In the total sustainability scenario, we have assumed that the process has the same dynamic as during the first year. In the middle scenario, we have assumed that the dynamic is broken, but the new measures are sustained, as well as the achieved results. In the no sustainability scenario, we assumed that the improvement is slowly reversed, therefore we have used the inverse transition matrix of the first year.

In order to get an idea if such a collaborative is worth financing, it is important to place it in the context of a policy decision environment, to allow a tradeoff between costs and QALY-s. Quality of life (Qol) weights for PU patients and for the general geriatric population were obtained from the literature. The Qol weight was 0.703 for pressure-ulcer free nursing home patients, 0.68 for those with single PUs of grades 1 and 2; 0.5 for multiple PUs of grades 1 and 2; and 0.36 for severe PUs (grades 3 and 4) [7] [24] [30]. Cost data were the costs collected from the collaborative. To establish the effect of the uncertainty in the parameters of the base case we conducted a probabilistic sensitivity analysis, assuming a lognormal distribution for costs and effects. A Monte Carlo simulation was run with 10,000 iterations per scenario. We used standard discount rates recommended by the Dutch guideline for pharmaeconomic studies (4% for costs 1.5% for effects) [31].

Results

Patient characteristics

The 88 selected patients were not significantly different in age, sex, or BMI from the non-selected patients participating in the third round of the project. This was true for baseline and terminal measurement points.

Effectiveness

As can be seen in Table 2.1, the prevalence and incidence of PUs in the selected patient group is lower after the collaborative, primarily due to reduction of less serious ulcers (grades 1 and 2). The participating patient group also had a lower prevalence and incidence of PUs compared to the non-participating patients. The uptake of useful interventions generally increased or did not change significantly over time. We also observed the uptake of non-useful interventions.

Costs

Table 2.2 shows a breakdown of materials used and time spent on activities by all participants. The most time-consuming activity was intermittently turning the patient to the side. Materials and time are translated into costs in Table 2.3. The program experts have the highest hourly wage, the caregivers the lowest. The daily rental price of mattresses varies substantially. Table 2.4 shows that the project created a savings in variable nursing costs while increasing costs of preventing and treating PUs. Most of the cost goes to personnel, followed by mattress rental. Costs fluctuated primarily by the reduction of grades 1 and 2 PUs, since the number of severe ulcers did not change. In addition, the one-year project costs for the organizations were larger than the possible savings of a reduction of PUs. Therefore, the initial investment can only be recovered over a longer time period.

Figure 2.3: Number of patients with pressure ulcers for two years after the start of the QIC

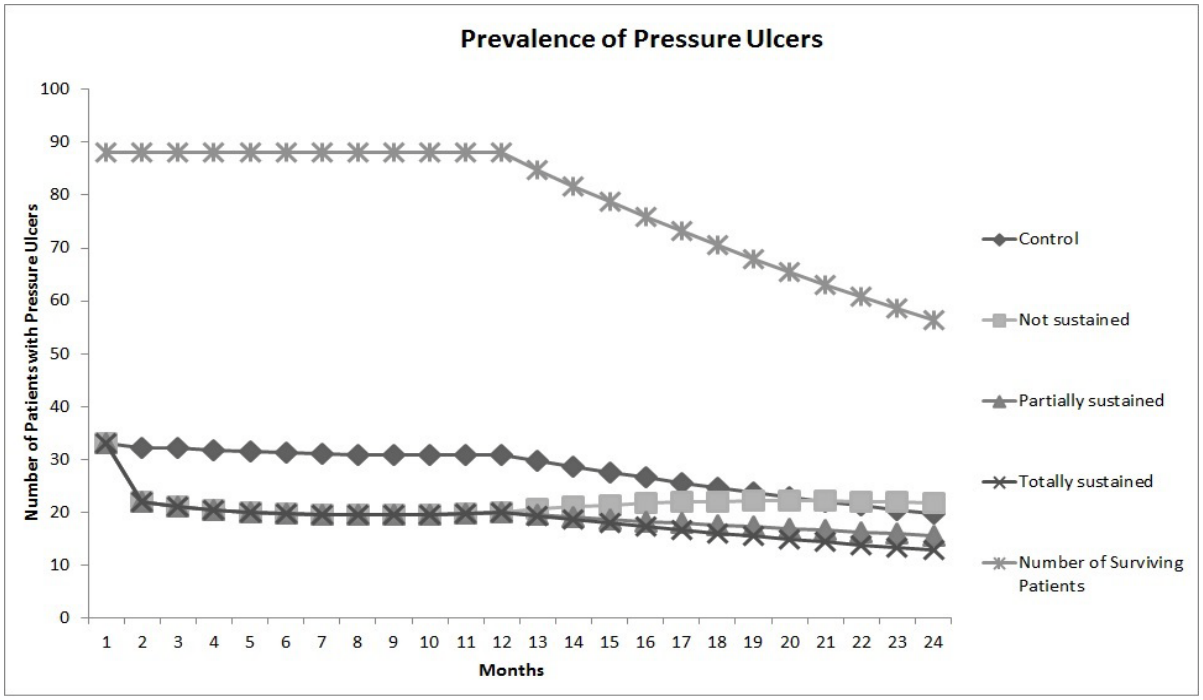


Table 2.4: Costs per person treated for selected patients

						PU Grade						***Average costs***	
STANDARD CARE			0	1		2		3		4			
			n/a	single	multiple	single	multiple	single	multiple	single	multiple	monthly	yearly
	Prevention	Labor	11.02	13.00	23.29	148.43	47.62	n/a	132.55	n/a	279.82		
		Mattress	2.22	19.18	28.53	200.01	41.57		394.52		232.21		
		Pillows	0.00	0.18	0.32	0.43	0.63		0.00		3.05		
	Treatment		0.00	68.38	64.82	59.17	68.64	n/a	142.01		118.34		
	Total standard care costs		*13.15*	*100.73*	*114.57*	*408.04*	*115.89*	*n/a*	*669.09*	*n/a*	*657.10*	*84*	*1026*
QIC	Prevention	Labor	30.80	45.86	110.25	123.30	n/a	n/a	192.60	278.10	n/a		
		Mattress	7.38	42.37	49.57	47.95			82.19	98.63			
		Pillows	0.34	0.45	0.51	0.41			0.00	1.32			
	Treatment		0.00	59.17	66.27	56.81	n/a	n/a	142.01	142.01	n/a		
	Total QIC clinical costs		*38.52*	*147.86*	*226.86*	*228.46*	*n/a*	*n/a*	*416.71*	*520.10*	*n/a*	*79*	*969*
	Program costs												323
	Organizational costs												1550
	Total QIC costs												*2842*

Modeling and sensitivity analysis

The prevalence of PUs over the course of the extrapolated year depends on whether or not the change in incidence and prevalence are sustained (Figure 2.3). If changes are not sustained at all, any success realized during the year in terms of prevalence is lost. If changes are partially sustained, prevalence slightly increases in the second year; in the scenario where changes are fully sustained, prevalence remains low.

From a healthcare perspective, the costs of PU care increased as a result of the project. At the same time, the project raised the average QoI of patients. Although the exact value of the QALY is debatable, there is a Dutch policy advice [32] stating that the values should be maximally €80,000 for patients with high disease severity. The QIC's incremental cost-effectiveness ratio after two years is above this limit of 80,000 €/QALY except for the most optimistic scenario where changes are completely sustained (Table 2.5).

Figure 2.4: Incremental costs and effects from Monte-Carlo simulation for three sustainability scenarios

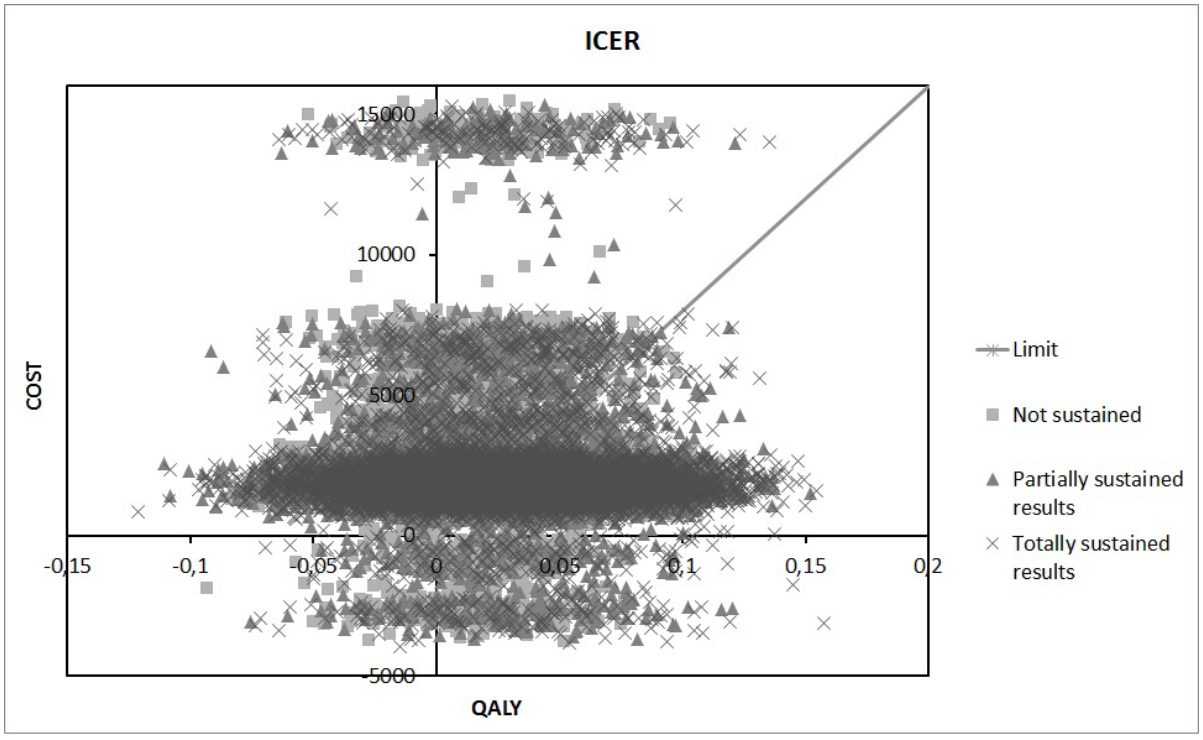


Table 2.5: Incremental costs, quality of life and cost-effectiveness ratio

	Not sustained	***Partially sustained***	***Totally sustained***
Difference in cost per person	€ 2.208 Probability=0.97	€ 2.072 Probability=0.97	€ 2.037 Probability=0.97
Difference in QoI per person	0.016820965 Probability=0.74	0.023361 Probability=0.74	0.02594592 Probability=0.75
ICER	131 253	88 692	78 517

The sensitivity analysis (Figure 2.4) allows us to investigate the robustness of our results. The joint probability of the ICER being below 80,000 along with a positive effect on QoI is 37% for the not sustained scenario, 47% for the partially sustained scenario, and 50% for the totally sustained scenario. Therefore there is no clear indication of the collaborative being effective after two years, and there is a high probability that it is more costly in every scenario.

Discussion

Summary of main results

The QIC significantly reduced the PU prevalence when the measurements before and after the collaborative are compared. This decrease was mainly due to the decrease of non-severe PUs (grades 1 and 2). The QoI of patients probably did not increase significantly.

Even though the variable costs of the organizations decreased, the large project costs of the QIC increased healthcare costs overall. Therefore, a QIC can only be cost-effective if the efforts to reduce PUs are sustained. In other words, short-term effectiveness is a necessary, but not a sufficient condition for long-term cost-effectiveness.

Sensitivity of the results

The sensitivity analysis showed considerable uncertainty in the results of the model and thus it is not possible to indicate clearly that the intervention was cost-effective. The uncertainty lies in the effects of the collaborative; it is only moderately probable that the patient's quality of life will increase. This may be caused by the fact that the difference in quality of life of a regular nursing home patient and a PU patient (independent of severity) is very small [6], which makes detection of change difficult. In this study, the difference in QoI between a patient without a PU and a patient with a low-grade PU was minimal.

It is likely that the intervention is more costly than standard PU care; this study, however, works with a different assumption than previous studies, therefore the savings reached by preventing PUs are lower than that which can be found in the literature [9]. This study assumed that PUs in the long term care sector do not cause extra patient days because 66% of nursing home patients receive long-term care [26] or die as in-patients. Therefore, we considered only the costs associated with PUs and their prevention. This is contrary to a previous Dutch study [9] that assumed PUs caused additional patient days in the long term care sector.

Limitations and Strength

The main limitation of this study is that it was based on an observational study. This limitation has far-reaching consequences. Because of the lack of case-mix measures for the population, we were only able to include the small number of cases that survived the duration of the study, while ignoring cases that died during the study. In addition, overrepresentation may be a problem because we worked with self-reported data. Therefore we cannot say with certainty that the selected cases were representative of the whole population. Furthermore the results are prone to the biases of any observational study, namely, secular trends; therefore it is not certain that this decline actually happened because of the collaborative. It should be noted that secular trends were far slower then the improvement in the selected patients: according to the LPZ panel data from 2006 and 2007[33-34], the prevalence of pressure ulcers decreased from 24% to 18.3% in Dutch nursing homes and from 11% to 7.9% in assisted living facilities. Therefore it is not plausible that the decline in PU-s in the collaborative was caused exclusively by secular trends. Besides secular trends, selection of the cases may have had an effect on the precise cost per patient ratio. First including the costs of the remaining teams (9 successful and 6 unsuccessful teams) would have slightly increased the central cost per collaborative per patient. Second the project costs made by unsuccessful teams would slightly raise the average project cost, but since these teams did not complete the project these costs are small in comparison to the costs made by the successful departments. Therefore large biases are unlikely in the average cost/patient ratio.

Caution is called for when interpreting the long term effects of a collaborative. On one hand the small number of cases made the decision-analytic modeling difficult because the probabilities of incidence and healing in the model may not be representative for the whole group. On the other hand there is the question of which sustainability scenario is most realistic. There is scarce evidence in the literature about sustaining the changes of a QIC when the project is over [10], raising the question of whether a collaborative would ever be cost-effective. Even in organizations where the results are sustained for an additional year, the question of how far in the future the changes can

be sustained remains. This is especially important because sustaining the changes is a prerequisite for the organizations participating in the QIC to regain the initial investment. The PU QIC involved staff training, and the significant rate of labor fluctuation characteristic of Dutch caregivers (10% annually) [35] may endanger sustainability in the long run.

The major strength of this study is that it is one of the first attempts to address the cost-effectiveness of a PU QIC. This study gives detailed information on the costs of the program level, the project costs within the organizations, as well as the differences in the costs of nursing activities. In addition we have put serious effort into decreasing the effect of design limitations. By simulating a control-group based on the real data of first measurement month we could visualize a situation where no attention would have been paid to PU-s, a situation in which all conditions are held the same. In other words we have been able to control for every variable except for changes caused by secular trends. Since control-groups are usually not feasible for QICs, simulating control-groups may be a feasible and promising approach to evaluate their cost-effectiveness, naturally with this limitation in mind.

Additional research using an appropriate-case-mix adjustment is needed to determine the effects of a PU QIC and to establish incidence and healing rates in a larger sample that includes the home care sector. Furthermore, additional research is needed on the effects of PU collaboratives using cluster-randomization and Qol measurements sensitive enough to detect changes in nursing home patients. Finally, the long term effects are also worthy of investigation, focusing especially on effective methods for sustaining beneficial changes.

Conclusions

During the PU QIC the incidence and prevalence of PUs significantly declined thus reducing variable costs of organizations and probably realized small gains in quality of life. From a healthcare perspective, the collaborative was probably more costly and more effective in the short run than standard PU care. Long term effects are highly sensitive to the sustainability of the changes in nursing method. Running a collaborative costs money and profitability depends on the extent to which teams manage and sustain new working methods. Further research is needed to know how the improvement cycle plays out over a longer time period.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PM has acquired and analyzed the data, drafted the manuscript and approved the final version. MK has made substantial contribution to the interpretation of the data, critically revised the manuscript for important intellectual content and approved the final version. RB revised the manuscript and approved the final version. AN has contributed to the study design, critically revised the manuscript for important intellectual content, and approved the final version.

Acknowledgements

We thank Mathilde Strating and Pepijn Vemer for their valuable comments.

References

1. Reddy, M., S.S. Gill, and P.A. Rochon, *Preventing Pressure Ulcers: A Systematic Review.* JAMA, 2006. **296**(8): p. 974-984.
2. Halfens R.J.G., J.M.M. Meijers, J.C.L. Neyens, M.P.W. Offermans, *Rapportage Resultaten: Landelijke Prevalentiemeting Zorgproblemen 2008*. Landelijke Prevalentiemeting Zorgproblemen. 2008, Maastricht: Universiteit Maastricht.
3. Shahin, E.S., T. Dassen, and R.J.G. Halfens, *Pressure ulcer prevalence and incidence in intensive care patients: a literature review.* Nursing in Critical Care, 2008. **13**(2): p. 71-79.
4. Kapoor, A., et al., *Using the Case Mix of Pressure Ulcer Healing to Evaluate Nursing Home Performance.* American Journal of Medical Quality, 2008. **23**(5): p. 342-349.
5. Berlowitz, D.R., et al., *Effect of pressure ulcers on the survival of long-term care residents.* J Gerontol A Biol Sci Med Sci, 1997. **52**(2): p. M106-110.
6. Franks, P.J., H. Winterberg, and C.J. Moffatt, *Health-related quality of life and pressure ulceration assessment in patients treated in the community.* Wound Repair and Regeneration, 2002. **10**(3): p. 133-140.
7. Fleurence, R.L., *Cost-effectiveness of pressure-relieving devices for the prevention and treatment of pressure ulcers.* International Journal of Technology Assessment in Health Care, 2005. **21**(03): p. 334-341.
8. Poos, M.J.J.C., et al., *Kosten van Ziekten in Nederland 2005*. 2008, Bilthoven: RIVM.
9. Severens, J., et al., *The cost of illness of pressure ulcers in The Netherlands.* Adv Skin Wound Care, 2002. **15**(2): p. 72-77.
10. Schouten, L.M.T., et al., *Evidence for the impact of quality improvement collaboratives: systematic review.* BMJ, 2008. **336**(7659): p. 1491-1494.
11. Ovretveit, J. and D. Gustafson, *Evaluation of quality improvement programmes.* Qual Saf Health Care, 2002. **11**(3): p. 270-275.
12. Mittman, B.S., *Creating the Evidence Base for Quality Improvement Collaboratives.* Ann Intern Med, 2004. **140**(11): p. 897-901.
13. Lynn, J., et al., *Collaborative Clinical Quality Improvement for Pressure Ulcers in Nursing Homes.* Journal of the American Geriatrics Society, 2007. **55**(10): p. 1663-1669.
14. Baier, R.R., et al., *Quality Improvement for Pressure Ulcer Care in the Nursing Home Setting: The Northeast Pressure Ulcer Project.* Journal of the American Medical Directors Association, 2003. **4**(6): p. 291-301.
15. Huang, E.S., et al., *The Cost-Effectiveness of Improving Diabetes Care in U.S. Federally Qualified Community Health Centers.* Health Services Research, 2007. **42**(6p1): p. 2174-2193.
16. Kilpatrick, K.E., et al., *The insufficiency of evidence to establish the business case for quality.* Int J Qual Health Care, 2005. **17**(4): p. 347-355.
17. Iglesias, C., et al., *Pressure relieving support surfaces (PRESSURE) trial: cost effectiveness analysis.* BMJ, 2006. **332**(7555): p. 1416-.
18. Thomson, J.S. and R.G. Brooks, *The economics of preventing and treating pressure ulcers: a pilot study.* J Wound Care, 1999. **8**(6): p. 312-316.
19. McInnes, E., et al., *Support surfaces for pressure ulcer prevention.*, in *Cochrane Database of Systematic Reviews* 2008.

20. Dorr, D.A., S.D. Horn, and R.J. Smout, *Cost Analysis of Nursing Home Registered Nurse Staffing Times.* Journal of the American Geriatrics Society, 2005. **53**(5): p. 840-845.
21. Vu, T., et al., *Cost-effectiveness of multidisciplinary wound care in nursing homes: a pseudo-randomized pragmatic cluster trial.* Fam. Pract., 2007: p. cmm024.
22. Strating, M., et al., *Evaluating the Care for better Collaborative Results of the First year of evaluation*. 2008: Institute of Health Policy and Management, Rotterdam. .
23. de Veer, A.J.E. and A. Kerkstra, *Feeling at home in nursing homes.* Journal of Advanced Nursing, 2001. **35**(3): p. 427-434.
24. Nijs, K.A.N.D., et al., *Effect of family style mealtimes on quality of life, physical performance, and body weight of nursing home residents: cluster randomised controlled trial.* BMJ, 2006. **332**(7551): p. 1180-1184.
25. *RIVM Zorgkompas*. 2005 [cited 2010 2010.03.03]; Available from: http://www.rivm.nl/vtv/object_document/o5021n20327.html.
26. Schols J.M.G.A., C.H., van Weel C, *Nursing Home and Nursing Home Physician: The Dutch Experience.* Journal of the American Medical Directors Association, 2004. **5**(3): p. 207-212.
27. SOVV, *Collectieve Arbeidsovereenkomst Verpleeg en Verzorgingshuizen*, 2006.
28. A Briggs, M.S., K Claxton, *Decision Modelling for Health Economic Evaluation*. 2006, Oxford: Oxford University Press. 256.
29. CBS, *Statline*, 2003.
30. Drageset, J., et al., *The impact of social support and sense of coherence on health-related quality of life among nursing home residents--A questionnaire survey in Bergen, Norway.* International Journal of Nursing Studies, 2009. **46**(1): p. 66-76.
31. CVZ, *Richtlijnen voor farmaco-economisch onderzoek, geactualiseerde versie*, CVZ, Editor 2006.
32. RvZ, *Rechtvaardige en Duurzame Zorg*, R.v.V.e. Zorg, Editor 2007.
33. Halfens R.J.G., J.M.M. Meijers, S.W. Wansink, *Landelijke Prevalentiemeting Zorgproblemen: Resultaten negende jaarlijkse meting 2006*. Landelijke Prevalentiemeting Zorgproblemen. 2006, Maastricht: University of Maastricht.
34. Halfens R.J.G., J.M.M. Meijers, J.C.L Neyens, M.P.W. Offermans, *Rapportage resultaten: Landelijke Prevalentiemeting Zorgproblemen 2007*. Landelijke Prevalentiemeting Zorgproblemen. 2007, Maastricht: Universiteit Maastricht.
35. LEVV, *Feiten & Cijfers 2008.* 2008.

Chapter 3. Quality of Life instruments for economic evaluations in health and social care for older people: a systematic review

This chapter is based on Peter Makai, Werner Brouwer, Marc Koopmanschap, Elly Stolk and Anna Nieboer

Quality of life instruments for economic evaluations in health and social care for older people: a systematic review

Social Science and Medicine, February 2014; 102: 83-93.

Abstract

Gaining health may not be the main goal of healthcare services aimed at older people, which may (also) seek to improve wellbeing. This emphasizes the need of finding appropriate outcome measures for economic evaluation of such services, particularly in long-term care, capturing more than only health-related quality of life (HrQol). This review assesses the usefulness of HrQol and wellbeing instruments for economic evaluations specifically aimed at older people, focusing on generic and preference-based questionnaires measuring wellbeing in particular.

We systematically searched six databases and extracted instruments used to assess HrQol and wellbeing outcomes. Instruments were compared based on their usefulness for economic evaluation of services aimed at older people (dimensions measured, availability of utility scores, extent of validation).

We identified 487 articles using 34 generic instruments: 22 wellbeing (two of which were preference-based) and 11 HrQol instruments. While standard HrQol instruments measure physical, social and psychological dimensions, wellbeing instruments contain additional dimensions such as purpose in life and achievement, security, and freedom.

We found four promising wellbeing instruments for inclusion in economic evaluation: Ferrans and Powers QLI and the WHO-Qol OLD, ICECAP-O and the ASCOT. Ferrans and Powers QLI and the WHO-Qol OLD are widely validated but lack preference-weights while for ICECAP-O and the ASCOT preference-weights are available, but are less widely validated. Until preference-weights are available for the first two instruments, the ICECAP-O and the ASCOT currently appear to be the most useful instruments for economic evaluations in services aimed at older people. Their limitations are that (1) health dimensions may be captured only partially and (2) the instruments require further validation. Therefore, we currently recommend using the ICECAP-O or the ASCOT alongside the EQ-5D or SF-6D when evaluating interventions aimed at older people.

Key words: quality of life, cost-utility analysis, older people, long-term care, review

Introduction

The growing number of older people worldwide and the associated higher demand for healthcare increasingly puts pressure on public funds. Hence, there is a growing need to make funding decisions about various health and social services aimed at older people. Cost-utility analysis (CUA) can support policy makers to optimally allocate health and social care resources within limited budgets by comparing two or more healthcare interventions to investigate their relative value for money [1]. CUA is increasingly used in the curative sector for such comparisons. In CUA, the benefits of these interventions are commonly expressed in Quality-Adjusted Life-Years (QALYs), a utility-based health measure comprising both length and health-related quality of life (HrQol). To assess HrQol improvements, typically patients' health states are measured (using standardized instruments) using health dimensions such as mobility, pain and anxiety. Subsequently, these health states are valued (on a scale from 0 – dead – to 1 – perfect health). Such outcome measures are appropriate for curative services, where the goal is to improve health. However, in other fields of healthcare, such as mental health, social care, public health, and care for older people, a focus on health dimensions of quality of life (Qol) may be less appropriate if health improvement is not the only or even the main goal of the services provided [2]. A relevant question is how to broaden the scope of outcome measurement within a CUA to include Qol domains that are intentionally affected by interventions in other fields of healthcare, in particular care for older people.

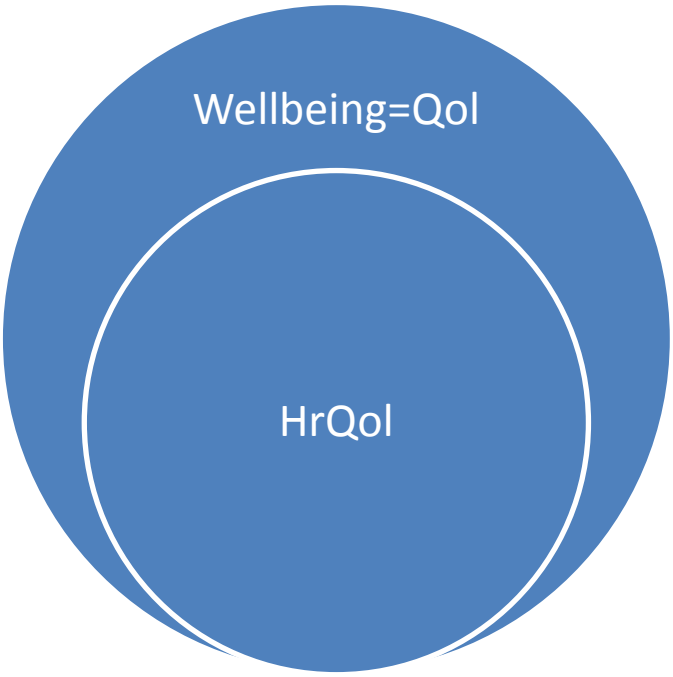
Current QALY measures using a quality adjustment factor that is based on domains of HrQol only, may not be appropriate to evaluate interventions for older people such as long-term care. This holds since the latter interventions may be aimed at improving non-health aspects of Qol, such as maintaining independence, dignity, comfort or social interaction. Evaluating such interventions using HrQol-instruments would likely undervalue the benefits. One of the most important challenges for performing CUA in the context of interventions aimed at older people thus concerns the availability of outcome measures attuned to the goals of services consumed by older people [3]. The aim of this paper was to review the literature in order to investigate the existence of such appropriate outcome measures, which would facilitate CUA in the context of health and social care for older people.

Older people consume a variety of health and social services. These may be curative services such as hospital care, as well as long-term care services provided by nursing homes, residential homes, and home care. Often, elderly consume a combination of such services within an illness episode.

The benefits of such a varied list of services should be evaluated using outcome measures that adequately capture the value of all services provided [3]. This may be particularly difficult in long-term care. To illustrate this, consider an intervention aimed at reducing the frequency of restraining older people in a nursing home setting to prevent them from falling [4]. While reducing the use of physical restraints may not directly improve a patient's health [4], such an intervention aims to restore dignity, freedom of movement, and control, outcomes that transcend health. If such an intervention were to be evaluated in a CUA, it is pivotal that outcome measures allow for capturing benefits 'beyond health' in order to provide adequate information on the costs and benefits of the intervention. Below we discuss some of the desirable characteristics of such instruments.

A first desirable characteristic of instruments attuned for evaluation of care for older people, is that such instruments should capture QoI dimensions transcending health. HrQoI instruments commonly used in CUAs measure health as a multi-dimensional construct minimally measuring psychological, physical and social dimensions [5], while for economic evaluation of services aimed at older people, particularly in long-term care other dimensions may also be relevant, such as affection or control. Instruments covering such dimensions 'beyond health' can be labeled as wellbeing instruments. There are two main conceptualizations relevant for the scope of wellbeing instruments. The first one focuses on wellbeing as an inherently subjective concept and thus holds that wellbeing does not contain health dimensions [6]. By distinguishing between functional HrQoI dimensions and subjective wellbeing dimensions, both HrQoI and wellbeing are components of the overarching concept of QoI. The second conceptualization treats wellbeing as representing individuals' welfare [7], which is dependent on individuals' functioning, thus encompassing HrQoI dimensions (see Figure 3.1). In this view, wellbeing can be seen as synonymous with overall QoI. In this paper, wellbeing will be referred to in the latter meaning.

Figure 3.1: Conceptualization of wellbeing and Quality of life



This second conceptualization may offer the opportunity to jointly explore treatment effects on health with other impacts on wellbeing. By broadening the evaluative space of a CUA [3], wellbeing instrument are, in principle, better equipped than HrQol measures to capture the full benefit of interventions aimed at older people, also when these aim at outcomes beyond health. However, wellbeing instruments based on the subjective notion of wellbeing may not explicitly or completely capture health. This deserves attention, since the aim must be to adequately capture all relevant outcomes of interventions in order to come to a complete comparison of costs and benefits in an economic evaluation. While some wellbeing instruments may include health as an underlying concept [8], it remains unclear whether existing outcome measure capture all wellbeing domains adequately and in such a way that allows inclusion in CUAs. To overcome this problem, it has been suggested that combinations of HrQol and wellbeing instruments could be used in economic evaluations in older people [9]. Moreover, the lines between HrQol and wellbeing measures may not always be easy to draw nor have been consistently drawn (when definitions of HrQol or wellbeing differ between measures). Therefore, in reviewing measures that may be useful in economic evaluation of services aimed at older people, particularly in long-term care, we will include both measures labeled as HrQol as well as measures of wellbeing. This allows an open and consistent categorization of instruments.

A second desirable characteristic of outcome measures for application in CUA in older people is that the classification system of health or well-being states is combined with a preference-based scoring system, as is the case for popular HrQol instruments like the EQ-5D and SF-36. Preference-based instruments normally consist of (1) a descriptive system defined by the dimensions and answer categories of the instruments (states), and (2) a (pre-scored) weighting system reflecting the valuation of the states described with element (1) by a relevant population (e.g. general public) [10]. The weighting system thus allows particular states as described with the descriptive system to be transformed into a 'utility score', commonly reflecting the average strength of preference for the various states described. In case of HrQol, these scores are typically anchored to a standardized scale, with 1 representing the utility of the best imaginable health state, and 0 representing the value for the state 'dead'. Negative values relate to health states valued as 'worse than dead'. For wellbeing instruments, anchoring on a 0 to 1 scale is also possible, 1 representing the best imaginable wellbeing instead of best imaginable health, while 0 can represent 'dead' or, more logically perhaps, the value for the worst level of all included domains in the descriptive system. Additionally, negative values for wellbeing instruments may also be allowed depending on the theory behind the instrument. Here, we will not limit our search to preference based instruments of wellbeing, as it is imaginable that utilities are attached in a later stage to promising measures of wellbeing that are currently not preference based (similar to development of SF-6D from SF-36 [11]).

A third desirable characteristic of instruments to be used in evaluations in older people is that their feasibility of use and psychometric properties are well-established. [10]. Instruments which measure what they intend to measure and those which do so with a smaller error seem to be more preferable to instruments lacking such properties. This aspect will also be considered in the review.

The number of instruments developed specifically to address and evaluate outcomes of health care services targeted at older people is growing. However, guidance is lacking on which instruments can or should be used for CUA of (long-term care) services aimed at older people. Such guidance depends on knowledge regarding the existing instruments, their ability to capture relevant outcomes and their feasibility and validity. The results of this review allow us to formulate (preliminary) advice on the choice of outcome measure for conducting economic evaluation of services aimed at older people.

Hence, we set out to perform a systematic literature search to identify generic outcome measures used in older people, which are applicable to all people irrespective of the type and nature of

diseases they have, thus, in principle, facilitating comparisons between people, treatments and services. Thus, we excluded disease specific instruments, which specifically aim to measure HrQol in well-defined populations. Instruments were included irrespective of whether they were labeled as HrQol or wellbeing instruments, and whether they were preference-based or not. When possible, the usefulness of these instruments for CUA was assessed. For the review we used a structured, three-step approach. First, we extracted the relevant generic HrQol and wellbeing instruments used in the studies. Second, we assessed their current and potential degree of suitability for economic evaluations in older people, particularly in long-term care. Finally, we examined the most promising instruments in more detail, with specific attention paid to their psychometric properties.

Methods

Search Methods

Database sources

We searched Pubmed, EMbase and CINALH for the English-language literature using the same keywords. In addition to standard medical and healthcare databases, we searched major psychological, sociological, and economic databases, namely Psycinfo, and Econlit, and Social Science Citation Abstracts to account for the multidisciplinary nature of Qol research [9].

Search terms

We used the following search terms, their synonyms and their combinations to find validated, generic, preference-based HrQol and wellbeing measures: “elderly”, “older”, “geriatric”, “quality of life”, “HrQol”, “wellbeing”, “validated questionnaire”, “validated measure”, “utilities”, and "preference-based valuation". For a complete description of the search terms see Appendix. The search strategy was customized for all databases.

Selection criteria

Selected articles met the following inclusion criteria. First, HrQol or wellbeing was an explicit outcome measure of an empirical article or validation study and its target population was characterized by ‘older persons’ or ‘elderly’, above 65 years. Second, the articles were written in English, measured HrQol or wellbeing through a questionnaire, and were published after 2000. Third, we excluded studies (1) using an instrument measuring only symptoms or instruments measuring only one dimension, (2) reported decision-analytical modelling, or (3) used disease-specific measures. Finally, we investigated the selected articles’ reference lists to identify the original development articles of the instrument used in the identified article.

Evaluation of the instruments

The following aspects of the instruments were evaluated in June 2012 for all identified instruments: scope of the instrument (HrQol or wellbeing), dimensions measured, availability of utility tariffs, and frequency of use. PM evaluated the titles and abstracts of all the studies, assessed articles for inclusion and exclusion; AN has independently checked the accuracy of this assessment. PM and AN then classified instruments independently. Differences in opinion were resolved by consensus.

We examined all selected instruments to identify those potentially useful for economic evaluations in older people, particularly in long-term care. Since there is no consensus definition of wellbeing or HrQol across disciplines (health science, psychology, sociology, economics) in order to classify instruments we examined their operationalizations. For our purpose, we classified instruments as HrQol or wellbeing according to the following definitions : HrQol instruments measure health as minimally measuring psychological, physical and social dimensions [5], while wellbeing instruments measure broader Qol domains (as well). We also classified the instruments by the availability of utility tariffs and frequency of use.

Finally, we investigated the most promising instruments (preference-based or commonly used wellbeing instruments) for economic evaluation in more detail, looking at feasibility of use in older people, paying attention to age-related cognitive decline and psychometrics in order to determine which instrument had the most potential for actual usage in economic evaluations in older people. Psychometrics were evaluated according to the criteria outlined for the critical appraisal of psychometric properties[12]. McDowell distinguishes between the thoroughness and the results of instrument validity and reliability. Validity can be defined as the extent to which a test measures what it intends to measure. The major forms of validity are content validity, which is assessed by the dimensions present; construct validity, which is tested in terms of convergent and discriminant validity; and sensitivity to change. Reliability is the capacity of a measure to provide consistent and stable estimates. Reliability has two major forms relevant here: responsiveness over time (test retest reliability), and internal consistency in case of multiple factor variables.

Results

Search outcome

Figure 3.2: Search Strategy

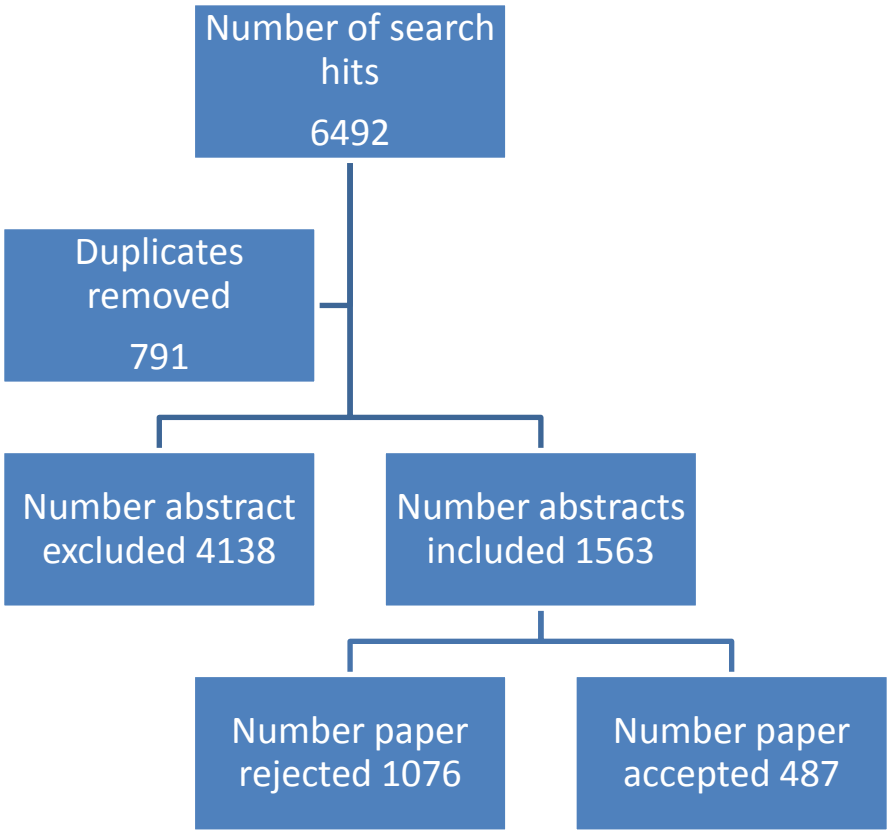


Figure 3.2 shows the search strategy. From the initial 6492 hits, we removed 791 duplicates. From the remaining articles, 1563 met the inclusion criteria. Of these, 1076 were subsequently assessed as meeting the exclusion criteria. The systematic search therefore resulted in 487 included articles. We identified 34 generic (non-disease specific) HrQol and wellbeing instruments in these articles, which were subsequently evaluated.

Dimensions measured

The systematic search uncovered 34 generic (i.e. non-disease-specific) instruments of which 23 were classified as wellbeing instruments and 11 as HrQol instruments. The HrQol instruments operationalized health with an average of eight dimensions, which pertained to only physical, psychological, and/or social functioning. Therefore, the HrQol instruments in Table 3.1 did not meet our definition of wellbeing.

Table 3.1: HrQol and wellbeing instruments with their classifications

Instrument	*Scope*	*Dimensions*	*Utilities*	*N of studies**
Health utility index 2[13]	HrQol	sensation, mobility, emotion, cognitive, self-care, pain, fertility	Yes	Many 32
Health Utility Index 3[14]	HrQol	vision, hearing, speech, ambulation, dexterity, emotion, cognition, pain	Yes	Many 11
Quality of well-being[15]	HrQol	mobility, physical activity, social activity, symptoms,	Yes	Several 5
Herdecke Quality of life questionnaire[16]	HrQol	Initiative power and interest, social interaction, mental balance, mobility, physical complaints, digestive well-being,	No	Single 1
Duke health profile[17]	HrQol	physical, mental, social, general, perceived health, self-esteem, anxiety, depression, pain, disability	No	Single 1
Nottingham health profile[18]	HrQol	energy level, pain, emotional reaction, sleep, social isolation, physical abilities	No	Several 5
Sickness Impact Profile [19]	HrQol	sleep and rest, mobility, emotional behavior, body care and movement, work, home management, eating social interaction, ambulation, alertness behavior, communication, recreation and pastimes	No	Few 3

Instrument	*Scope*	*Dimensions*	*Utilities*	*N of studies**
Assesment of Quality of life [20]	HrQol	psychological wellbeing, physical senses, social relationship, independent living, illness,	Yes	Several 5
SF-6D[11] (SF-36, Sf-12 SF-8)	HrQol	physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, mental health	Yes	Many 189
EQ-5D[21]	HrQol	anxiety/depression, mobility, self-care, usual activities, pain, discomfort	Yes	Many 117
15D[22]	HrQol	mobility, vision, hearing, breathing, sleeping, eating, speech, elimination, usual activities, mental function, discomfort and symptoms, depression, distress, vitality, sexual activity	Yes	Many 10
Ferrans and Powers QLI [23]	Wellbeing	health and functioning, social and economic, psychological/spiritual, family	No (but weighted)	Several 6
ICECAP-O [3]	Wellbeing	attachment, security, role, enjoyment, control	Yes	Few 4
OPUS[24]	Wellbeing	food and nutrition, personal care, safety, social participation and involvement, control over daily living	Yes	Single 1
CASP-19 [8]	Wellbeing	control, autonomy, self-realization and pleasure	No	Single 1

Instrument	*Scope*	*Dimensions*	*Utilities*	*N of studies**
ASCOT [25]	Wellbeing	control over daily life, personal cleanliness and comfort, food and drink, accommodation cleanliness and comfort, safety, social participation, occupation, dignity, living situation	Yes	Single 1
Older people quality of life profile [26]	Wellbeing	life overall health and functioning social relationships leisure and social activities independence, control over life, freedom, home and neighborhood psychological and emotional well-being financial circumstances religion and culture	No	Single 1
WHO-Qol Old [27]	Wellbeing	sensory abilities, autonomy, past, present, future activities, death and dying	No	Several 7
WHO-Qol Bref [28]	Wellbeing	overall Qol, general health, physical, psychological, social relationships, environment	No	Many 13
WHO-Qol 100[29]	Wellbeing	physical health, psychological health, social relationships, independence, environment, spirituality	No	Many 5

Instrument	*Scope*	*Dimensions*	*Utilities*	*N of studies**
Comprehensive quality of life Scale[30]	Wellbeing	material well-being health, productivity, intimacy, safety, place in community, emotional well-being	No	Single 1
Personal Well-being index [31]	Wellbeing	satisfaction with health, personal relationships, community, overall satisfaction, standard of living, achievement, safety, spirituality future security,	No	Single 1
Interactive Computerized Quality of life Scale (ICQOL)(SF)[32]	Wellbeing	overall life satisfaction, day-to day functioning, self-esteem, health status, work, energy level, home life, social life, mood, interacting with others, body image, illnesses, effect of stress/fear, sense of achievement, life expectancy, aches/pains, sleep/rest comfort, activity level, sex life, stamina, pleasures	No	Few 2
MANSA (Manchester Short Assessment of Quality of Life) [33]	Wellbeing	satisfaction with life as a whole, job, financial situation, friendships, leisure activities, accommodation, personal safety, people that the person lives with, family, health, mental health	No	Single 1
SPF-IL [34]	Wellbeing	behavioral confirmation, affection, status, comfort, stimulation	No	Single 1

Instrument	*Scope*	*Dimensions*	*Utilities*	*N of studies**
McGill quality of life scale [35]	Wellbeing	physical well-being physical symptoms, psychological, existential, support	No	Few 2
Quality of Life Questionnaire[36]	Wellbeing	social support, general satisfaction, physical well-being, free time	No	Single 1
Quality of life inventory[37]	Wellbeing	self-esteem, goals/values, health, learning, work, creativity, play, helping, friends, neighborhood, community, home, children, love, money, relatives	No	Few 2
Social Wellbeing of Nursing home residents-scale [38]	Wellbeing	affection, behavioural confirmation, status social wellbeing	No	Single 1
Quality of Life Scale[39]	Wellbeing	material and physical wellbeing, relationships with other people, social, community and civic activities, personal development and fulfillment, recreation	No	Single 1
National wellbeing index [31]	Wellbeing	country's economic situation, state of the environment, social conditions, government, business, and national security	No	Single 1

Instrument	*Scope*	*Dimensions*	*Utilities*	*N of studies**
Quality of life in elders with multiple morbidities[40]	Wellbeing	Family, own health, friendship, cognitive abilities mobility and physical functioning, hobbies, social contacts, cultural and aesthetic matters, developing new abilities, own abode, spouse, (social) participation, finances (social) commitment, well-being and sensual experience, travel autonomy and self-determination weltanschauung/philosophy incontinence/continence	No	Single 1
Life in General scale [41]	Wellbeing	general satisfaction, fatigue, fear, anxiety, unhappy, depressed, shaking/trambling loneliness, friends, social life, world too complicated	No	Single 1
Personal wellbeing index[42]	Wellbeing	standard of living, health, achievements in life, relationships, safety, community connectedness, and future security, spirituality/religion	No	Few 2
Satisfaction with Life scale[43]	Wellbeing	general satisfaction, life conditions, life close to ideal, goal achievement, lack of regret	No	Single 1

* Number of studies validated in older people few: 1-4 studies; Several: 5-8 studies; Many: more than 9 studies;

The wellbeing instruments usually measured some health dimensions [5] also included in HrQol instruments, next to broader domains of Qol. The exception was CASP, an instrument exclusively measuring non-health dimensions of Qol [8]. These additional, non-health dimensions included in the identified wellbeing instruments could be classified into four main concepts: (1) purpose in life and achievement (wishes, goals, values, spirituality, self-realization, activity level, achievements, work); (2) worries about security and safety (present and future); (3) financial well-being (money, financial situation, standard of living); and (4) personal freedom (control, autonomy, independence). Less frequently mentioned dimensions were related to pleasure; creativity and play, or related to the environment such as physical environment, community, and neighborhood.

Preference-based instruments

We found five HrQol measures (EQ-5D, AQol, Quality of Well Being, SF-36 and HUI) and two wellbeing measures (ASCOT, formerly the OPUS, and ICECAP-O) for which utility scores existed [1] [20] [24] [3, 25] [44] [45]. The ASCOT was developed based on earlier experience with the OPUS instrument. The development of the latter instrument was not without problems [24], e.g. in a first valuation study no differences in utility scores for the different levels of OPUS's safety dimension were detected. For the ASCOT, however, the development, validation and valuation of the instrument seemed successful with the following eight dimensions included: control over daily life, personal cleanliness and comfort, food and drink, accommodation cleanliness and comfort, safety, social participation, occupation, dignity and an additional question on living situation. The ICECAP-O was developed in several steps, jointly leading to the currently available instrument [3] with five dimensions: attachment, security, role, enjoyment, control. The ASCOT (with the exception of the dichotomous 'living at home' dimension) and the ICECAP-O both use four answering levels per included domain [3] [45]. As for the preference elicitation techniques and anchoring used in the two wellbeing instruments, both used discrete choice experiments (DCE), with the ICECAP-O using Best-Worst Scaling (BWS), while the ASCOT used both BWS as well as the more traditional DCE to elicit preferences for health and wellbeing states [46]. The ICECAP-O utility scores were normalized with 0 indicating no capabilities, while 1 denotes full capabilities. [3]. Dead and states worse than dead are not defined on the scale. The ASCOT is similarly anchored at 1 and 0, although here 0 is anchored to 'dead' and negative values (states worse than dead) are possible. A detailed discussion on preference elicitation approaches for HrQol instruments can be found elsewhere [1].

Table 3.2: Potentially most relevant instruments for CUA in elderly care

Dimensions	***ICECAP-O***	***ASCOT***	***Ferrans and Powers QLI***	***WHO-Qol old+WHO-Qol Bref***
Physical	-	+	+	+
Psychological	+	-	+	+
Social	+	+	+	+
Purpose in life and achievement	+	+	+	+
Financial	-	-	+	-
Security	+	+	-	-
Personal Freedom	+	+	-	+
Psychometrics and other criteria	*ICECAP-O*	*ASCOT*	*Ferrans and Powers QLI*	*WHO-Qol old+WHO-Qol Bref*
Validation	+	-	+	+
*Reliability - Thoroughness	0	0	+++	++
**Reliability – Results	0	0	+++	+++
*Validity – Thoroughness	++	+	+++	++
**Validity - Results	++	++	+++	+++
Item number	5	9	64	24+25
Useful in cognitively declined populations	+	+	-	-
Utilities	+	+	-	-

&*+ = included in the instruments - = not included in the instrument

* In case of thoroughness, 4 categories are distinguished: 0 = no reported evidence of reliability or validity, + = very basic information only

++= several types of tests, or several studies have reported reliability or validity +++= all major forms of reliability/validity tested

** In case of results of the validation, the categories are: 0= no numerical results reported ?=results uninterpretable, +=weak reliability/validity,

++ = adequate reliability/validity, +++= excellent reliability/validity

Validation

Based on the studies included in this review, the most widely used HrQol instruments were the EQ-5D and the SF-36, which have been extensively validated across a wide range of conditions and countries. The most widely used wellbeing instruments were Ferrans and Powers Qol index (QLI) and the WHO-Qol instruments. To date, the ICECAP-O preference-based instrument has been more widely validated than the ASCOT. A more thorough comparison of the two preference-based and the two most frequently used questionnaires is given in Table 3.2.

Comparison of the most promising instruments

The most extensively used instruments, Ferrans and Powers QLI and the WHO-Qol OLD, both have as important disadvantages for economic evaluations that preference based weights for outcomes are lacking. These preferences could, however, be obtained in future studies, which would improve their usefulness in economic evaluations. Both instruments have been tested for reliability and validity with good or excellent results. In older people, Ferrans and Powers QLI had excellent reliability, with an internal consistency of 0.86 to 0.96 [47]. Test-retest reliability was tested in the general population with a test-retest correlation of 0.87 using a two-week interval [47]. The instrument showed good validity; it was moderately to highly and positively correlated with life satisfaction and general health perception. Moreover, it was moderately negatively correlated with disease burden, and showed sensitivity to change in 27 intervention studies [47]. Ferrans and Powers QLI, however, misses some of the frequently measured dimensions of wellbeing. A more comprehensive instrument is the WHO-Qol OLD, which has shown good reliability in older people (internal consistency of 0.88 to 0.89 and test-retest reliability of 0.91 after two weeks) [27]. Good validity was suggested with medium to strong negative correlations with different depression measures and moderately positive correlations with general health perception. Additionally, there is increasing evidence for its sensitivity to change for a number of conditions [12] [48]. A major disadvantage of both instruments was that they are relatively long. The WHO-Qol OLD, in fact, is an extension of the WHO-Qol BREF, having no less than 24 additional questions [28]. Furthermore, proxy versions were not available for either instrument.

The ICECAP-O instrument has only five items, while the ASCOT has nine items, making their use in older people quite feasible. The ICECAP-O measures five of the seven most frequently identified dimensions of wellbeing, but it has the (potential) disadvantage of not measuring a physical health dimension directly. It is possible that physical health is captured indirectly by the other dimensions, which is suggested by several empirical findings [49] [50] [51] [52]. The ASCOT also measures five of

the seven most frequently identified dimensions of wellbeing. The ASCOT has as a potential disadvantage that it does not explicitly measure a psychological dimension.

Another advantage of the ICECAP-O is its more widespread validation as compared to the ASCOT. It has been applied in different settings and cultures such as the UK, Australia the Netherlands and Canada [49] [51, 52] [50] [53] [54] [9, 55-58] . Although clearly related to its early stage of development and use, this implies that the validity of the ASCOT is more uncertain, especially in different settings and cultures than used in so far [25]. Its psychometric properties also require further testing. A disadvantage of both instruments is that they lack explicit assessments of their reliability.

Discussion

Key findings

This study reviewed the literature to search for outcome measures which can be used in economic evaluations of interventions in older people, particularly in long-term care. To avoid leaving out potentially useful and relevant outcome measures, we included instruments labelled as HrQol instruments in the first stage of the review. This was deemed important since the classification of instruments as HrQol or wellbeing need not have been done consistently. We retrieved 34 generic instruments, of which 23 were classified as wellbeing instruments and 11 as HrQol instruments. Additional dimensions of wellbeing instruments that emerged from the review included purpose in life and achievement, security, financial well-being and personal freedom. Of the wellbeing instruments, two had utility scores available, allowing use in economic evaluations: the ICECAP-O and the ASCOT. The two most widely validated wellbeing instruments, the WHO-Qol OLD and Ferrans and Powers QLI, do not have utility scores.

Wellbeing instruments enable researchers to evaluate a wider range of benefits of services for older people, thus more closely conforming to the goals of some interventions especially in long-term care (e.g. less restraints or a better living environment in nursing homes). They typically go beyond measuring HrQol alone and can measure the benefits of interventions that aim to produce value beyond HrQol domains.

Methodological issues

There are some limitations of this review worth considering. Any classification of instruments based on the dimensions measured is inherently subjective. To overcome this problem as much as

possible, the instruments were classified by two reviewers in a structured manner using standardized criteria, while the additional authors were consulted for additional expert guidance. Nonetheless, other categorizations than the one presented here are possible, especially in light of the inclusive nature of this review, which allowed an exceptionally broad range of instruments to be included. Another limitation is that the review excluded grey literature. Therefore, instruments still in development may have been missed.

Choice of instruments

A first noteworthy point is that even though wellbeing or Qol is a difficult theoretical concept [6, 59], its actual measurement converges to a limited number of dimensions. Such a convergence could form the basis of an operational definition of wellbeing, although there is no consensus at this point.

Most wellbeing instruments measure a combination of health and non-health consequences, making them potentially suitable for evaluating interventions that result in a combination of health and non-health consequences. The exception is the CASP, which exclusively measures non-health outcomes [8]. Nonetheless, even though it seems to be rooted in a more subjective notion of wellbeing which is distinct from health, the CASP may still capture health consequences indirectly, if the measured domains are influenced by health status.

A thorough exploration of how the individual dimensions of HrQol and wellbeing relate to each other is an important yet difficult conceptual and empirical puzzle beyond the scope of this paper. For such work, additional conceptual and integrative reviews based on qualitative studies may be necessary as well. A few important features of wellbeing instruments are nonetheless worth noting for potential users and developers.

First, different instruments measure the dimensions of well-being on different levels. We can attempt to classify the dimensions according to Wilson's taxonomy [59], where outcome measures are placed on a continuum from medical variables to overall Qol. Outcome measures have five levels: physiological, symptomatic, functional, perceptive, and overall Qol. For example, ICECAP-O and ASCOT both measure a dimension of control, but seem to do this on different levels. The ICECAP-O asks respondents if they are able to be independent, which can be viewed as measuring on the perceptive level. The ASCOT asks if they have control over daily lives, which can be viewed as measuring on the functional level. Such distinctions have an influence on how the measure aims to capture benefits in a comprehensive manner. While wellbeing measures on the perception level

may be more abstract attempting to capture benefits with broad dimensions, wellbeing measures on the functioning level often are more specific and may be aimed at explicitly capturing the dimensions relevant for (i.e. influenced by) health and social care. By measuring on the functional level, a wellbeing instrument may be more comparable to current HrQol measures which also typically measure the dimensions on this level. At the same time, comprehensiveness may then require a large number of dimensions. In contrast, dimensions measured on the perception level may reach comprehensiveness through a smaller list of dimensions. The latter may improve the feasibility of use in elderly populations.

With respect to the relationship between health and non-health dimensions, certain HrQol dimensions underlie wellbeing dimensions completely (as may be the case for the CASP) or partially, as may be the case with ICECAP-O, where the physical health dimension can be thought of as underlying, for instance, the dimensions of control and role. Even though validation work shows that the ICECAP-O reflects and captures all three health dimensions [49] [50] [51] [52] there is also some indirect evidence that the ICECAP-O may not measure physical health as fully as a HrQol measure like the EQ-5D does, but more research on this issue remains necessary [9]. Similarly, further research is especially encouraged on the ASCOT, to investigate whether the lacking psychological dimension is indirectly captured by some other dimension(s) [45].

The majority of the reviewed wellbeing instruments did not appear to be directly useful for economic evaluations in older people consuming health and social care. Many are not preference-based, and would thus require a utility-elicitation procedure to be more readily useful to CUA, following the example of the SF-36. This might be relevant to Ferrans and Powers QLI and the WHO-Qol OLD, since both have been widely used and extensively validated. If utility scores would be derived for the states described by these instruments, the findings of previous studies using these instruments could also be revisited. As development of utility scores for existing lengthy questionnaires typically involves reducing the number of included items in the descriptive system [11], the feasibility of including these instruments (in shortened form) in economic evaluation would improve as well. However, these shortened instruments would require additional validation in order to ascertain that they retain their psychometric properties.

The preference-based instruments ICECAP-O and ASCOT have been developed more recently, and thus have not been extensively validated. A major drawback of the current preference-based measures is that they do not integrally measure health and non-health consequences to the extent that for instance the Ferrans and Powers QLI and the WHO-Qol OLD do. This can be a problem for

interventions that, while not exclusively aimed at non-health dimensions, have an effect on health as well.

Currently, the ICECAP-O and the ASCOT may be useful in the context of economic evaluations. While both are promising, the validation of the ASCOT at this point lags behind that of the ICECAP-O. If further studies provide support for their sensitivity to change, and clarify the relationship between the ICECAP-O, ASCOT and various health dimensions both instruments could be a suitable wellbeing instrument for economic evaluations in older people, particularly in long-term care. In fact, currently the Social Care Institute for Excellence[60] and the NICE [61] guidelines recommend using these two instruments for measuring and valuing effects in the United Kingdom. Sensitivity to change of wellbeing instruments seems particularly relevant, as interventions are one factor amongst many which influence wellbeing. Therefore, further research on sensitivity to change of wellbeing instruments to a number of interventions is particularly encouraged. In the context of sensitivity to change, the general design of instruments may also matter. The ASCOT for example measures the effect of particular services more specifically than the ICECAP-O does. Hence, one might expect the ASCOT to potentially be more sensitive to changes in the provision of these services. On the other hand, the ICECAP-O may be more sensitive to changes in the provision of other social care services or the general care context, which are not included in the ASCOT dimensions. Such hypotheses have to be rigorously tested, using pre-specified hypotheses [62, 63]. This may also shed more light on the question when to use which instrument.

Additionally, given the fact that, at present, the ability of the ICECAP-O and ASCOT to (completely or adequately) capture all relevant health dimensions remains unclear, it seems advisable to use a health measure such as the EQ-5D or SF-6D along with the ICECAP-O or ASCOT instrument in economic evaluations of interventions aimed at older people in order to explicitly capture health benefits alongside broader benefits. We note that health-related and wellbeing-based utilities should not be condensed into a single utility index. First of all because they relate to two different scales and concepts that cannot simply be added. Secondly, because it is currently unclear which dimensions would be double-counted by the different instruments. More research is required to investigate the potential degree of double-counting when using these measures simultaneously and the degree of missed health effects when using only wellbeing measures. If and when it becomes clear which if any health dimensions may be missed by the wellbeing measures, then a common valuation of different measures may be attempted [64], resulting in an instrument capturing all relevant dimensions. However, this requires much research, potentially involving the development of methods for combining capabilities and functionings. So far, CUAs using such broad measures

assessing the full benefits of elderly care have yet to be published. We should emphasize that such studies would be using a far broader concept of utility than one that is solely health-related, better suiting the aims of many interventions aimed at older people, particularly in long-term care.

Choice of maximand

When attempting to support optimal allocations of scarce resources for older people, it is pivotal to include all costs and all benefits of interventions, and this can be achieved with wellbeing instruments. One theoretically well-developed approach towards quantifying wellbeing in that context stems from the capability approach. The capability approach is claimed to underlie the ICECAP-O as well as the ASCOT [45]. While most health measures seek to measure functionings, the capability approach focuses on capabilities, which are two different concepts [60]. According to capability theory, functionings can be defined as beings or doings of the individual, while capabilities are potential beings and doing, or potential functionings. For example, a classic distinction is made between a person who is starving or fasting. While both are equal in terms of functionings, the former lacks capabilities while the latter simply chooses not to engage in a functioning [60]. It is important to recognize such differences in choosing outcome measures, as they (implicitly) define the maximand of interventions.

In the context of receiving health and social services, services can be seen to expand peoples capabilities through either directly allowing people to function (for example washing them) or indirectly through mitigating an impairment [45]. Following Forder's reasoning, outcome measures in older people should be able to measure improvement in wellbeing even if personal functioning is not improved, as long-term care services allow individuals to achieve outcomes that they would not be able to achieve themselves.

Although in this review we have limited ourselves to a review of wellbeing instruments in older people, particularly in long-term care, the problem of the evaluative space goes beyond services for older people, and is applicable to the whole healthcare sector. In order to maintain the possibility to evaluate interventions across the whole healthcare sector, ideally comparable wellbeing measures (or better still, one overall wellbeing measure) should be available for the entire healthcare sector. It appears that there is a great need for appropriate wellbeing instruments, since different fields of healthcare, such as mental health, social care, public health may not be directly (solely or mainly) aimed at improving health. In fact, there are preference-based wellbeing instruments being developed for the general adult population, for example the ICECAP-A [2]. It is an interesting area of future research to investigate whether such measures adequately capture all dimensions relevant

for older people as well or whether specific measures for them remain necessary. Such issues should be considered and explored further in development and validation of preference-based wellbeing instruments.

Conclusion

The development and use of wellbeing instruments for CUA in older people aiming to capture the benefits of both health and social care is a new, developing and important area of research. In the short-run, two preference-based instruments may be useful in the context of economic evaluations: the ICECAP-O and the ASCOT. The validation of the ASCOT at this point lags behind that of the ICECAP-O, although both require substantial validation. During this validation work, attention should be paid to the exact relationships between the ICECAP-O, ASCOT and various health dimensions contained in widely validated preference-based HrQol measures, such as the SF-6D or the EQ-5D. An alternative direction forward would be to develop scoring algorithms for extensively validated non-preference-based measures that encompass more dimensions, following the example of the SF-36 [11]. Thus, utility weights could be attached to the results of earlier studies with these instruments. Irrespective of future direction, the conceptual puzzle of which dimensions need to be covered by wellbeing instruments for CUA remains unsolved. This also holds for how the dimensions should be measured and at what level. While further instrument validation and development remain crucial to capture the benefits of all services aimed at older people within CUA, with the availability of preference-based wellbeing instruments, reaching such a goal has become more feasible. This, in turn, has the potential of allowing a more optimal and fair allocation of services aimed at older people.

References

1. Drummond, M.F., et al., *Methods for the economic evaluation of health care programmes*. Vol. Third Edition. 2005, Oxford: Oxford University Press.
2. Al-Janabi, H., T. Flynn, and J. Coast, *Development of a self-report measure of capability wellbeing for adults: the ICECAP-A.* Quality of Life Research, 2012. **21**(1): p. 167-176.
3. Coast, J., et al., *Valuing the ICECAP capability index for older people.* Social science & medicine (1982), 2008. **67**(5): p. 874-882.
4. Huizing, A.R., et al., *A Cluster-Randomized Trial of an Educational Intervention to Reduce the Use of Physical Restraints with Psychogeriatric Nursing Home Residents.* Journal of the American Geriatrics Society, 2009. **57**(7): p. 1139-1148.
5. Who, *Official Records of the World Health Organisation*, 1948.
6. Morgan, S., et al., *The cost of drug development: a systematic review.* Health Policy, 2011. **100**(1): p. 4-17.
7. Nussbaum, M., *The quality of life*. 1993, Oxford: Clarendon Press.
8. Hyde, M., et al., *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, 2003. **7**(3): p. 186-194.
9. Davis, J., et al., *A comparison of the ICECAP-O with EQ-5D in a falls prevention clinical setting: are they complements or substitutes?* Quality of Life Research: 2013. **22**(5):p. 969-977.
10. Brazier, J., et al., *Measuring and Valuing Health Benefits for Economic Evaluation*. Vol. 1. 2007, Oxford: Oxfrod University Press.
11. Brazier, J., J. Roberts, and M. Deverill, *The estimation of a preference-based measure of health from the SF-36.* Journal of health economics, 2002. **21**(2): p. 271-292.
12. McDowell, I. and C. Newell, *Measuring health: a guide to rating scales and questionnaires*. Vol. 268. 2006: Oxford University Press New York.
13. Lutomski, J.E., et al., *The validation of a frailty index from the*
    The Older Persons and Informal Caregivers Survey Minimal DataSet (TOPICS-MDS). Journal of the American Geriatrics Society, 2013. **accepted for publication**.
14. Fried, L.P., et al., *Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care.* The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 2004. **59**(3): p. M255-M263.
15. Rawls, J., *A theory of justice*. 1999: Belknap Press.
16. Ostermann, T., et al., *The Herdecke Questionnaire on Quality of Life (HLQ): Validation of factorial structure and development of a short form within a naturopathy treated in-patient collective.* Health and Quality of Life Outcomes, 2005. **3**(Journal Article).
17. Novella, J.L., et al., *Evaluation of the quality of life in dementia with a generic quality of life questionnaire: The Duke Health Profile.* Dementia and geriatric cognitive disorders, 2001. **12**(2): p. 158-166.
18. Baro, E., et al., *Using the Nottingham Health Profile (NHP) among older adult inpatients with varying cognitive function.* Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 2006. **15**(4): p. 575-585.
19. Rothman, M.L., S. Hedrick, and T. Inui, *The Sickness Impact Profile as a measure of the health status of noncognitively impaired nursing home residents.* Medical care, 1989. **27**(3 Suppl): p. S157-67.

20. Osborne, R.H., et al., *Quality of life assessment in the community-dwelling elderly: validation of the Assessment of Quality of Life (AQoL) Instrument and comparison with the SF-36.* Journal of clinical epidemiology, 2003. **56**(2): p. 138-147.
21. The EuroQol, G., *EuroQol--a new facility for the measurement of health-related quality of life.* Health policy (Amsterdam, Netherlands), 1990. **16**(3): p. 199-208.
22. Sintonen, H., *Comparing properties of the 15D and the EQ-5D in measuring health-related quality of life.* Archives of Hellenic Medicine, 2001. **18**(2): p. 156-160.
23. Ferrans, C.E. and M.J. Powers, *Psychometric assessment of the quality of life index.* Research in nursing & health, 1992. **15**(1): p. 29-38.
24. Ryan, M., et al., *Using discrete choice experiments to estimate a preference-based measure of outcome--an application to social care for older people.* Journal of health economics, 2006. **25**(5): p. 927-944.
25. Malley, J.N., et al., *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, 2012. **10**(Journal Article).
26. Bowling, A., *The Psychometric Properties of the Older People's Quality of Life Questionnaire, Compared with the CASP-19 and the WHOQOL-OLD.* Current gerontology and geriatrics research, 2009. **2009**(Journal Article): p. 298950.
27. Peel, N.M., H.P. Bartlett, and A.L. Marshall, *Measuring quality of life in older people: reliability and validity of WHOQOL-OLD.* Australasian Journal on Ageing, 2007. **26**(4): p. 162-167.
28. Harper, A.P.N., *Development of the World Health Organization WHOQOL-BREF Quality of Life Assessment.* Psychological medicine, 1998. **28**(03): p. 551-558.
29. Power, N.K.W.T.W.G., *The World Health Organization quality of life assessment (WHOQOL): Development and general psychometric properties.* Social science & medicine, 1998. **46**(12): p. 1569-1585.
30. Gullone, E.C.R.A., *The Comprehensive Quality of Life Scale: A Psychometric Evaluation with an Adolescent Sample.* BEHAVIOUR CHANGE, 1999. **16**(2): p. 127-139.
31. Cummins, R.A., et al., *Developing a National Index of Subjective Wellbeing: The Australian Unity Wellbeing Index.* Social Indicators Research, 2003. **64**(2): p. 159-190.
32. Jamison, R.N., et al., *Validation of the short-form interactive computerized quality of life scale (ICQOL-SF).* Pain medicine (Malden, Mass.), 2007. **8**(3): p. 243-250.
33. Bjorkman, T. and B. Svensson, *Quality of life in people with severe mental illness. Reliability and validity of the Manchester Short Assessment of Quality of Life (MANSA).* Nordic journal of psychiatry, 2005. **59**(4): p. 302-306.
34. Nieboer, A., et al., *Dimensions Of Well-Being And Their Measurement: The Spf-Il Scale.* Social Indicators Research, 2005. **73**(3): p. 313-353.
35. Cohen, S.R., et al., *The McGill Quality of Life Questionnaire: a measure of quality of life appropriate for people with advanced disease. A preliminary study of validity and acceptability.* Palliative medicine. **9**(3): p. 207-219.
36. Ruiz, M.A. and E. Baca, *Design and validation of the "Quality of Life Questionnaire" ("Cuestionario de Calidad de Vida", CCV): A generic health-related perceived quality of life instrument.* European Journal of Psychological Assessment, 1993. **9**(1): p. 19-32.
37. Frisch, M.B., et al., *Clinical Validation of the Quality of Life Inventory: A Measure of Life Satisfaction for Use in Treatment Planning and Outcome Assessment.* Psychological assessment, 1992. **4**(1): p. 92-101.
38. Gerritsen, D.L., et al., *Social well-being and its measurement in the nursing home, the SWON-scale.* Journal of Clinical Nursing, 2010. **19**(9-10): p. 1243-1251.

39. Flanagan, J.C., *Measurement of quality of life: current state of the art.* Archives of Physical Medicine and Rehabilitation, 1982. **63**(2): p. 56.
40. Holzhausen, M., A. Kuhlmey, and P. Martus, *Individualized measurement of quality of life in older adults: Development and pilot testing of a new tool.* European Journal of Ageing, 2010. **7**(3): p. 201-211.
41. Campbell, A., P.E. Converse, and W.L. Rodgers, *The quality of American life: Perceptions, evaluations, and satisfactions*. 1976: New York: Russell Sage Foundation.
42. (IWbG), I.W.G., *Personal wellbeing index – adult, manual 4th edition. Manual*, 2006, Australian Centre on Quality of Life, Deakin University: Melbourne.
43. Diener, E., et al., *The satisfaction with life scale.* Journal of personality assessment, 1985. **49**(1): p. 71-75.
44. Potoglou, D., et al., *Best–worst scaling vs. discrete choice experiments: An empirical comparison using social care data.* Social science & medicine, 2011. **72**(10): p. 1717-1727.
45. Forder, J.E. and J. Caiels, *Measuring the outcomes of long-term care.* Social science & medicine, 2011. **73**(12): p. 1766-1774.
46. Netten, A., et al., *Outcomes of social care for adults: developing a preference-weighted measure.* Health technology assessment, 2012. **16**(16): p. 1-166.
47. Ferrans, C.E. and M. Powers. *Reliability and validity of the Ferrans and Powers Quality of Life Index*. 2013 [cited 2013 14.08]; Available from: http://www.uic.edu/orgs/qli/reliability/reliabilityhome.htm.
48. Figueira, H.A., et al., *Effects of a physical activity governmental health programme on the quality of life of elderly people.* Scandinavian Journal of Public Health, 2012. **40**(5): p. 418-422.
49. Coast, J., et al., *An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people.* Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 2008. **17**(7): p. 967-976.
50. Flynn, T.N., et al., *Assessing quality of life among British older people using the ICEPOP CAPability (ICECAP-O) measure.* Applied Health Economics and Health Policy, 2011. **9**(5): p. 317-329.
51. Makai, P., et al., *Capabilities and quality of life in Dutch psycho-geriatric nursing homes: an exploratory study using a proxy version of the ICECAP-O.* Quality of Life Research, 2012. **21**(5): p. 801-812.
52. Makai, P., et al., *A validation of the ICECAP-O in a population of post-hospitalized older people in the Netherlands.* Health and Quality of Life Outcomes, 2013. **11**(1): p. 1-11.
53. Comans, T.A., et al., *Quality of life of older frail persons receiving a post-discharge program.* Health and Quality of Life Outcomes, 2013. **11**(1): p. 58.
54. Couzner, L., J. Ratcliffe, and M. Crotty, *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, 2012. **10**(1): p. 69.
55. Davis, J.C., et al., *Exploration of the association between quality of life, assessed by the EQ-5D and ICECAP-O, and falls risk, cognitive function and daily function, in older adults with mobility impairments.* BMC Geriatrics, 2012. **12**(1): p. 65.
56. Couzner, L., et al., *Measuring and valuing quality of life for public health research: application of the ICECAP-O capability index in the Australian general population.* International Journal of Public Health, 2012: p. 1-10.
57. Mitchell, P.M., et al., *Predicting the ICECAP-O Capability Index from the WOMAC Osteoarthritis Index Is Mapping onto Capability from Condition-Specific Health Status Questionnaires Feasible?* Medical Decision Making, 2013.
58. Henderson, C., et al., *Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial.* BMJ, 2013. **346**.

59. Wilson, I.B. and P.D. Cleary, *Linking Clinical Variables With Health-Related Quality of Life: A Conceptual Model of Patient Outcomes.* JAMA: The Journal of the American Medical Association, 1995. **273**(1): p. 59-65.

60. Francis, J., *SCIE's Approach to Economic Evaluation in Social Care.* Social Care Institute for Excellence, London, 2009.

61. NICE. *The social care guidance manual* 2013 [cited 2013 11.11]; Available from: http://publications.nice.org.uk/the-social-care-guidance-manual-pmg10/incorporating-economic-evaluation.

62. Terwee, C.B., et al., *Quality criteria were proposed for measurement properties of health status questionnaires.* Journal of clinical epidemiology, 2007. **60**(1): p. 34-42.

63. Mokkink, L.B., et al., *The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study.* Quality of Life Research, 2010. **19**(4): p. 539-549.

64. Alava, M.H., et al., *Common Scale Valuations across Different Preference-Based Measures Estimation Using Rank Data.* Medical Decision Making, 2013.

65. Sen, A., *Choice, welfare and measurement*. 1982, Cambridge, MA: Harvard University Press.

Appendix

Search terms for all databases

Population	*Outcome*	*Questionnaire*	*Exclusion*
Elderly	Quality of life	Validation	Gene
Older	Overall Quality of Life	Validation questionnaire	genetic
Older people	Health-related Quality of Life	Validated measure	clinical utility
Geriatric	Health Related Quality of Life	Validated measures	utilization
Gerontology	Generic Quality of Life	Validated questions	Risk factor
Nursing	Health Perceptions	Validity and reliability	Risk factors
Old age	Functional Status	Reliability and validity	Semi-structured interview
Old-age specific	Functional State	Utilities	
Old age specific	Environment	Utility	
Older adults	Health Status	Util	
Older people	Well-being	Preference based valuation	
Frail elders	Welfare	Valuation	
Frail elderly	Wellbeing	Preference-based health measure	
Older patient	Quality weight	Capabilities	
	States	Choice experiment	
	State	Choice experiments	
	Health outcome measurement	Discrete choice experiments	
	GEN-QOL	Preferences	
	GEN-QOLQ	Economic evaluation	
	OQOL	Health outcome measure	

Chapter 4. A validation of the ICECAP-O in a population of post-hospitalized older people in the Netherlands

This chapter is based on Peter Makai, Marc Koopmanschap, Werner Bouwer, and Anna Nieboer
A validation of the ICECAP-O in a population of post-hospitalized older people in the Netherlands
Health And Quality of Life Outcomes 2013, 11:57

Abstract

Background

Various healthcare and social services may impact not only health, but wellbeing as well. Such effects may be more fully captured by capability-wellbeing instruments than with Health-related Quality of Life (HrQol) instruments. The aim of this study is to validate the ICEpop (Investigating Choice Experiments for the Preferences of Older People) CAPability measure for Older people (ICECAP-O) capability wellbeing instrument in a population of post-hospitalized older people admitted to a hospital 3 months earlier.

Methods

296 post-hospitalized older people in the Netherlands were interviewed 3 months after admission between September 2010 and January 2011. We investigated the convergent validity of the ICECAP-O and overall wellbeing measures (Cantril's ladder and Social Production Function: Instrument for Level of Well-being (SPF-IL)), as well as with various health measures (EQ-5D, Katz-15 Instrumental Activities of Daily Living (IADL) scale, Geriatric Depression Scale (GDS) and the Medical Outcomes Study Short form (SF-20) social functioning dimension). Additionally, we assessed discriminant validity by comparing several relevant subgroups in our sample (based on age, depression, IADL dependency, living situation, etc.). We also investigated the relationship between overall wellbeing and the ICECAP-O, controlling for HrQol and background characteristics.

Results

This study suggests that the ICECAP-O has good convergent validity with wellbeing measures as well as health measures and discriminates between various groups of post-hospitalized older people. Wellbeing measured by both Cantril's ladder and SPF-IL is associated with the ICECAP-O in a multivariate analysis controlling for HrQol as well.

Conclusion

The ICECAP-O seems to be a valid instrument of capability-wellbeing in older, post-hospitalized people, showing good convergent validity with health and wellbeing instruments, and is able to discriminate between elderly with various health profiles. The ICECAP-O measure seems to capture both health and wellbeing. Therefore it is a promising instrument for assessing the outcomes of health and social services aimed at older people.

Background

Economic evaluation of healthcare services aims to inform policy makers by comparing the costs and benefits of alternative health care interventions. In such an evaluation, it is crucial that besides all costs, all benefits of healthcare services are captured. Capturing such benefits can be challenging, since healthcare services such as elderly care, long-term mental health, and public health may impact individuals health and health related quality of life, as well as their wellbeing more generally [1] [2] [3] [4].

Health can be defined as a multidimensional construct of physical, psychological and social dimensions [5]. These health dimensions can be inter-related, for example decreased mobility may lead to a decrease in social contacts and depression [6] [7], subsequently impacting social and psychological dimensions of health [7]. Health related quality of life (HrQol) tries to capture how health impacts individuals' Quality of Life (Qol) [8]. In economic evaluations, benefits are frequently assessed by changes in health-related quality of life combined with the duration an individual spends in various health states. Duration and HrQol are then subsequently combined in Quality-Adjusted Life-Years (QALYs), and thus arguably capture the effect of healthcare services on physical, psychological and social dimensions of health. Aspects of broader wellbeing, such as maintaining independence, dignity, and comfort [1], however, arguably are not captured by the concept of HrQol in its entirety. This can cause problems in capturing the full benefits of interventions, in particular in the evaluation of social care interventions, as well as integrated health and social care services [9]. For example, specific social care interventions like day care and meals on wheels may improve wellbeing, but not health, or at least not only health [9]. As a consequence, such services cannot be evaluated in the same manner as other healthcare services such as medicines [9] where using HrQol seems more appropriate in many cases. Otherwise, the benefits of these provisions may be undervalued [10].

Therefore, broadening the evaluative space of economic evaluations by a wider measurement of benefits has been suggested in evaluation of elderly care [1] [11], using dimensions of wellbeing such as independence, attachment, or the ability to pursue valued activities [10] in addition to health dimensions. In that context, a proposed alternative to measuring HrQol is to measure capabilities. Capabilities may be seen as a conceptualization of wellbeing [1], defined as the capacity to perform certain actions and achieve certain states (irrespective of actually doing so). Capability wellbeing assesses what individuals can do instead of focusing on functioning, i.e. what individuals

actually do [1]. Capability-wellbeing captures a variety of health and non-health dimensions, which may be difficult to separate [12].

In order to measure capability wellbeing, two instruments have been developed to date, the ICECAP-O [10,13] (ICEpop (Investigating Choice Experiments for the Preferences of Older People)) CAPability measure for Older people above 65 and the ICECAP-A [1] for the general population. Both instruments are intended as outcome measures for economic evaluations of both health and social services, where beyond health, wellbeing aspects have to be considered as well [1] [9,10]. In order to be useful for economic evaluations, instruments should be sufficiently validated in terms of their convergent and discriminant validity. While the ICECAP-A has been validated in the UK only [14], the ICECAP-O has been validated in a number of settings: in the British general elderly population [10], in an Australian population of post-hospitalized elderly receiving residential care [15], in a Canadian population of elderly visiting a fall-prevention clinic [16] and a proxy version has been validated in Dutch nursing home settings [17].

However, to date, the ICECAP-O has not been validated in a population of post-hospitalized older-people in the Netherlands. Post-hospitalized elderly are increasingly recognized as a population in which health improvements can be achieved [18] through geriatric interventions. In the Netherlands, in the context of the National Care for the Elderly Program significant efforts are made to improve health and quality of life outcomes in frail elderly, for instance through the Prevention and Reactivation Care Programme among older patients who are admitted to a hospital [19]. For elderly populations, hospitalization increases the risk of functional decline, defined customarily as a decrease in (instrumental) activities of daily living ((I)ADL) [20]. Although elderly may be hospitalized due to function decline resulting from illness, such functional decline is also frequent after admission: 35% of 70 year olds and 65% of 90 year olds experience such a decline. Functional decline is therefore influenced by hospital care as well [20], through increased complications [21] or through less aggressive treatment regimens than customary in younger populations [18]. In a group of post-hospitalized older people, a wide range of differences in health, capabilities and well-being problems may be expected due to (differences in) age, physical function, and other characteristics of the elderly such as multi-morbidity and support from their direct environment. As a result, this population is likely to receive various forms of publicly funded healthcare, as well as being the recipients of other social services. Furthermore, there is little research on how the ICECAP-O is related to other conceptualizations of wellbeing and the relationships between the ICECAP-O and measures of health (physical, psychological and social) remain underexplored. Exploring such issues is preferably done in a group in which a variety of health and well-being problems may be expected

such as post-hospitalized elderly. Therefore, the aim of this study is to validate the convergent and discriminant validity of the ICECAP-O in a Dutch community-dwelling population discharged from a hospital in the prior three months. We further study the discriminant validity of the ICECAP-O by performing sub-group analyses, highlighting the differences in ICECAP-O scores between groups of elderly.

Method

Design, participants and setting

This validation study was based on a pilot study of the Transition-experiment Geriatric Network Rotterdam Prevention and Reactivation of Care program. The aim of the pilot was to select outcome measures and triage instruments for the actual trial [19]. In order to be able to select appropriate instruments, several instruments measuring similar constructs were included in the pilot. As some instruments such as the ICECAP-O were not widely validated, their validity was further examined on the basis of the pilot. This helped to reduce the number of instruments measuring the same concepts in the actual trial. This pilot study was conducted among all older people admitted to the Vlietland hospital between June and October 2010. The sample included 500 older people (>65 years of age) who were interviewed using face to face questionnaires. Three months after hospital admission, a total of 296 discharged patients (59% response rate) completed questionnaires using face to face administration and were included in the analysis. Reasons for dropout were: death (n=49), lost interest to participate (n=52), too ill (n=35), terminally ill (n=5), objection by partner/family (n=14), mentally not able (n=8), private reasons (e.g. death of spouse; n=4), questions not applicable (n=8), no contact/unable to reach respondent (n=12), and reason unknown (n=22). The study protocol was approved by the medical ethics committee of the Erasmus Medical Centre, Rotterdam, the Netherlands, under protocol number MEC2011-041. Informed consent was obtained from all participants. The study protocol is extensively described in Asmus-Szepesi [19].

Measures

To investigate the convergent and discriminant validity of the ICECAP-O, we used a wide variety of outcome measures. To measure different conceptualizations and operationalizations of wellbeing we used three wellbeing measures. First, capability wellbeing was measured using the ICECAP-O capability measure for older people. The ICECAP instruments can be seen as measuring capability wellbeing [1] achieved by the capacity to perform certain actions and achieve certain states [9]. The ICECAP-O measures five capability dimensions – attachment, security, role, enjoyment, and control

– with one question per dimension. Each dimension can be scored on four levels, thus distinguishing 1024 possible 'capability states'. The ICECAP-O was developed using rigorous qualitative and quantitative approaches [22] [9] [10] [13]. In order to obtain tariffs for the ICECAP-O, the attributes were valued using best-worst scaling, a special type of discrete choice analysis [9]. The ICECAP-O tariffs have values between 0 (no capability) and 1 (full capability). Second, wellbeing was measured using the Cantril's ladder life satisfaction scale, a one-dimensional index ranging from zero (completely dissatisfied) to 10 (completely satisfied) [23]. Third, we also used a multi-dimensional measure of wellbeing, the Social Production Function: Instrument for Level of Well-being (SPF-IL), to assess wellbeing. The SPF-IL measures affection, behavioral confirmation, status, comfort and stimulation on a 4 point scale, ranging from 1 (never) to 4 (always) [24], providing an overall index of wellbeing, with higher scores indicating higher levels of wellbeing.

To measure HrQol we used the EQ-5D [25]. The EQ-5D measures HrQol in terms of five dimensions (mobility, self-care, daily activities, pain and discomfort, anxiety and depression) with three levels each (1=no problems, 2=moderate problems, and 3=extreme problems) describing 243 health states. The EQ-5D health states can be converted into a utility score by applying the scoring values (tariff) for the Dutch population [25]. The EQ-5D utility scores range from 1 (perfect health) through 0 (death) and has negative values accounting for health states worse than dead [25]. The EQ-5D is one of the most widely used measures of HrQol, and is extensively used in economic evaluations [25]. To assess physical functioning, we used the combined ADL (Activities of Daily Living)-IADL (Instrumental Activities of Daily Living) scale (Katz-15) consisting of yes or no responses on IADL items such as bathing, dressing and abilities such as using the telephone and managing money [26]. The IADL scores range from 0–15 with higher scores indicating higher dependency. Three cutoff-scores are commonly used, 7 (severely IADL dependent), 4 (moderately IADL dependent) and 1 (mildly dependent) [27]. In this current study we used the cutoff score for mildly dependent.

To assess depressive symptoms, we used the Geriatric Depression Scale-15 (GDS-15). The GDS-15 consists of 15 items, measuring psychological function and mood swings. The instrument has been widely validated in older people [28]. The cutoff score of 10 is a reliable cut-off score for major depression, while a score below five is considered to indicate the absence of clinically significant depressive symptoms. Scores between 5 and 10 indicate mild depression [29] [30]. In this current study we used the cutoff score of five.

To assess social functioning, we used the social activity limitation item from the SF-20 [31]. This item measures the frequency with which respondents experienced social activity limitations due to

health. The item runs from 1 (none of the time) to 6 (all of the time), and converts to a 0–100 scale. In this current study we have used a cutoff score at the middle of the scale, i.e. 50, to distinguish elderly who have frequent limitations (limitations a good bit of the time or more frequently) from those with less frequent limitations.

Finally, we investigated the presence of multi-morbidity. Multi-morbidity was defined as having two or more chronic disease conditions, as is common in the literature [32] [33]. We included the following chronic illnesses in our multi-morbidity count: diabetes, stroke (cerebral haemorrhage , cerebral infarction or TIA), heart failure, cancer (malignant condition), asthma or chronic bronchitis or lung emphysema or COPD, incontinence, degenerative arthrosis of hip or knee, osteoporosis, prostate symptoms caused by benign prostate enlargement, dementia, hearing problems, problems with vision.

Hypotheses

For convergent validity, we expect the ICECAP-O capability wellbeing measure to correlate more strongly with Cantril's ladder and the SPF-IL wellbeing measures, than with the EQ-5D HrQol measure and with the IADL, GDS and the SF-20's social activity limitation health measures, because the ICECAP-O is intended as measure of well-being that transcends measuring HrQol [13]. For discriminant validity, we expect to find higher ICECAP-O scores in older people living with others as compared to living alone due to higher affection [10] [34]. We also expect to find higher scores in IADL independent as compared to IADL dependent older people, and for non-depressed as compared to depressed older people as well as in older people with no social activity limitations vs. those with such limitations. This was based on earlier work showing strong relationships between the ICECAP-O role, enjoyment and control dimensions and physical problems, and between the ICECAP-O dimensions attachment and enjoyment and mental health measures, and between a number of social measures and the ICECAP-O dimensions role and enjoyment [10]. Furthermore, we will explore differences on the ICECAP-O in older people living at home compared to those in a nursing home, in the young-old (<75 years old) compared to the old-old (≥75 years old) and in multi-morbid older people versus those without multi-morbidity (the latter expected to score higher on the ICECAP-O). In order to gain further insight into how the ICECAP-O and health are related to older and more accepted wellbeing measures, we will explore if the ICECAP-O is related to other measures of wellbeing in a multivariate model controlling for health.

Analysis

All analyses were performed in STATA 11. Item level analysis of non-response was carried out. For all analyses, available cases were used.

We calculated descriptive statistics. In establishing convergent validity we used correlation analyses. Correlations above 0.5 are referred to as strong, between 0.3 and 0.5 as moderate, and below 0.3 as weak. Differences in strength of correlation between ICECAP-O and EQ-5D, and between ICECAP-O and the wellbeing measures were assessed with Steiger's Z [35]. For discriminant validity we used t-tests for two group comparisons and one-way ANOVA for comparisons between multiple groups. To further explore discriminant validity of the ICECAP-O, we also performed stepwise regression analyses with a p-value of 0.2. To analyze to what degree the ICECAP-O is related to the Cantril's ladder and SPF-IL wellbeing measures, we have performed stepwise multivariate regressions including all variables with a p-value below 0.2. Regression assumptions were checked. In the subgroup analysis, categorical groups were compared using chi-squared tests.

Results

Response

296 clients completed face to face questionnaires three months after admission, and were included in the analysis. For these included clients, demographic characteristics had no missing values, while for other variables missing values ranged from 2 (0,7%) in case of Cantril's ladder to 12 (4%) in case of the ICECAP-O tariffs. Response on the ICECAP-O dimensions was quite good, ranging from 97% on the role dimension to 99% on the control dimension, demonstrating good feasibility. All analyses below were conducted on a net sample using complete case analysis (n=275).

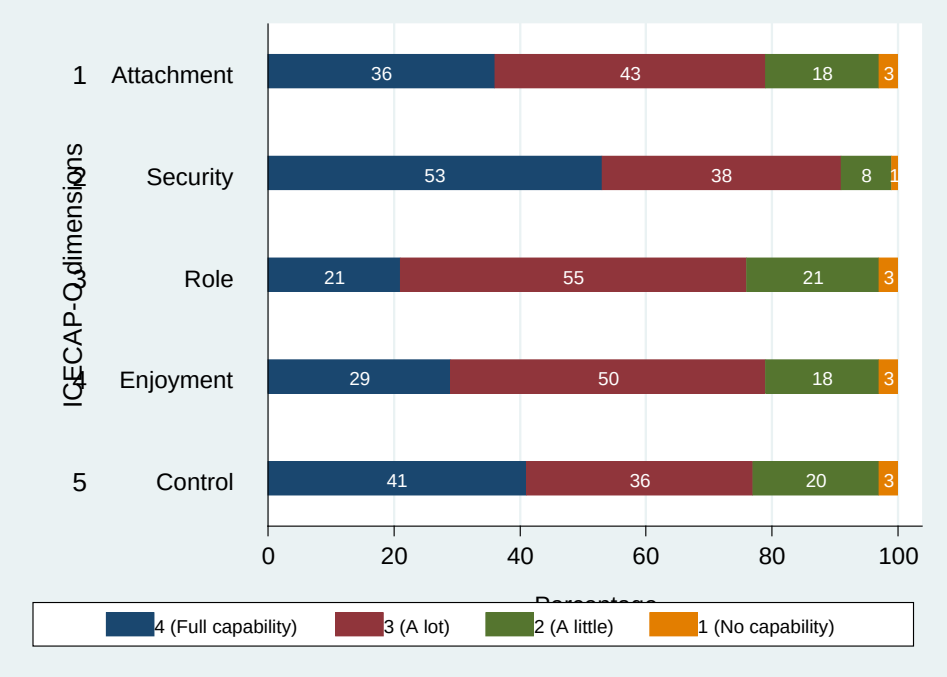
Table 4.1 below shows the demographic characteristics of post-hospitalized elderly, as well as their health status, HrQol, and wellbeing. Figure 4.1 below details the response to the individual ICECAP-O dimensions.

Table 4.1: Sample characteristics

Variable	*Complete-case analysis (n=275)*	*Mean (SD)*	*Percentage*
Age	Young old (65–75)	76.21 (6.79)	46.55
	Old-old (75+)		53.45
Sex	Female		53.82
	Male		46.28
Education	None		6.55
	Primary school		26.90
	Lower vocational		18.18
	General secondary education		34.18
	Grammar school		9.09
	Polytechnic/Higher vocational education/University		5.09
Maritial Status	Married/Other living together		57.46
	Divorced		5.82
	Widow(er)		30.90
	Never married		5.82
Living arrangement	Home alone		37.09
	Home with partner or children		56.73
	Nursing home/Elderly home		6.18
Diagnoses at admission	Diabetes		20.96%
	Stroke, cerebral haemorrhage (bleed in the brain), cerebral infarction (blocked blood vessel in the brain) or TIA		9.97%
	Heart failure		38.49%
	A type of cancer (malignant condition)		16.49%
	Asthma, chronic bronchitis, lung emphysema or COPD		22.68%
	Incontinence		20.27%
	Degenerative arthritis of hip or knee		49.48%
	Osteoporosis		27.49%

	Hip fracture		5.50%
	Other fractures		9.97%
	Dizziness with falling		16.15%
	Prostate symptoms caused by benign prostate enlargement		8.59%
	Depression		7.56%
	Anxiety/panic disorder		4.12%
	Dementia		0.69%
	Hearing problems		23.37%
	Problems with vision		15.81%
Multimorbidity	Maximum 1 chronic condition		34.55
	Multimorbid (2 or more chronic conditions)		65.45
ICECAP-O tarrifs		0.84 (0.14)	
Cantril's ladder		7.43 (1.32)	
SPF-IL		2.85 (0.43)	
EQ-5D	Mobility –some problems		49.45
	Self-care –some problems		11.64
	Self-care – severe problems		2.18
	Daily activities		25.45
	Daily activities – severe problems		5.09
	Pain and discomfort – some problems		40.73
	Pain and discomfort- severe problems		8.00
	Anxiety and depression – some problems		12.00
	Anxiety and depression – severe problems		0.73
	EQ-5D utilities	0.80 (0.17)	
Health measures	SF-20 social activity limitations	74.18 (26.18)	
	GDS	2.55 (2.61)	
	IADL (average dependency)	2.47 (2.59)	

Figure 4.1: Response on the ICECAP-O



Convergent and discriminant validity

Correlation analysis shows, that the ICECAP-O overall tariffs were significantly and strongly correlated with Cantril's ladder, while the ICECAP-O dimensions were generally moderately correlated with Cantril's ladder. The SPF-IL total scores were generally moderately correlated with the ICECAP-O dimensions and strongly correlated with the ICECAP-O tariffs. The overall EQ-5D utility score was also moderately correlated with the ICECAP-O tariffs. The EQ-5D dimensions were mostly weakly correlated with the ICECAP-O tariffs, with the exception of Self-Care and Control, Usual activities and Role, and Usual activities and Control for which moderate correlations were found. Correlations between other health measures and the ICECAP-O tariffs were generally moderate, with the correlation between GDS and Attachment being weak. GDS and IADL were both strongly correlated with the ICECAP-O tariffs. The social activity limitations dimension was moderately correlated with Role, Enjoyment, Control and the ICECAP-O tariffs. Using Steiger's Z, we found that the difference in strength of the correlation between the ICECAP-O and the wellbeing measures on the one hand and between the ICECAP-O and the EQ-5D on the other hand was not statistically significant (Table 4.2).

Table 4.2: Correlations between capability, wellbeing and health dimensions

	ICECAP-O capability dimensions					*Weighted capabilities*
	Attachment	Security	Role	Enjoyment	Control	ICECAP-tariffs
Wellbeing						
Cantril's ladder	0.31**	0.22**	0.46**	0.46**	0.28**	0.51**
SPF_IL	0.47**	0.27**	0.43**	0.48**	0.34**	0.60**
Health						
EQ-5D Mobility	−0.17**	−0.08	−0.35**	−0.20**	−0.32**	−0.30**
EQ-5D Self-care	−0.16**	−0.12	−0.35**	−0.25**	−0.42**	−0.39**
EQ-5D Usual Activities	−0.17**	−0.19**	−0.47**	−0.31**	−0.43**	−0.47**
EQ-5D Pain/Discomfort	−0.13	−0.13*	−0.28**	−0.25**	−0.25**	−0.25**
EQ-5D Anxiety/Depression	−0.07	−0.25**	−0.18**	−0.30**	−0.16**	−0.25**
EQ-5D utilities	0.12*	0.20**	0.40**	0.30**	0.40**	0.40**
SF-20 social activity limitations	0.19**	0.22**	0.46**	0.34**	0.42**	0.47**
GDS	−0.29**	−0.35**	−0.42**	−0.46**	−0.36**	−0.57**
IADL	−0.24**	−0.16*	−0.47**	−0.31**	−0.60**	−0.51**

* p value<0.05 ** p-value <0.01

Results regarding discriminant validity of the ICECAP-O are shown in Table 4.3. In the bivariate analysis the ICECAP-O significantly discriminated between young-old and old-old, between multi-morbid and single-morbid respondents, depressed and non-depressed respondents, between IADL dependent and non dependent respondents as well as between respondents with frequent social activity limitations and those without. Furthermore, the ICECAP-O discriminated between people with higher and lower EQ-5D scores. This is similar to the other wellbeing instruments as shown in Table 4.4, although only the ICECAP-O discriminated the young-old and the old-old. In the multivariate stepwise regression, the ICECAP-O discriminated groups based on IADL dependency, depressive symptoms, social activity limitations and EQ-5D scores (operationalized as dummies).

Table 4.3: Discriminant validity of the ICECAP-O in select groups

Variable	*Level*	*ICECAP-O*			
		Bivariate group		Multivariate group comparisons	
Demographic		Mean	p-value	Standardized coefficients	p-value
Age	Older people below 75	0.86*	0.01		
	Elderly above 75	0.83			
Sex	Female	0.85	0.35		
	Male	0.84			
Education	Pre-secondary	0.86	0.10		
	Post-secondary	0.83			
Married	Married or other living	0.85	0.13		
	Divorced	0.76			
	Widow	0.85			
	Never married	0.83			
Living situation	Alone	0.84	0.13		
	With partner/children	0.84			
	Nursing home	0.78			
Health					
Multimorbid	Maximum 1 chronic	0.89**	0.00		
	More than 2 conditions	0.82			
IADL	Independent	0.92**	0.00	−0.21**	0.00
	Dependent	0.81			
SF-20 social activity	No limitations	0.90**	0.00	−0.27**	0.00
limitations	Limited	0.77			
GDS	Not depressed	0.88**	0.00	−0.29**	0.00
	Depressed	0.73			
EQ-5D	Top 50%	0.90**	0.00	0.13	0.01
	Bottom 50%	0.80			
R squared					0.38

* p-value <0.05 ** p-value<0.01.

Table 4.4: Comparison of the discriminant validity of the wellbeing instruments

Variable	*Level*	*Cantril's ladder*	*SPF-IL*	*ICECAP-O*
Demographics		Mean Cantril's ladder score	Mean SPF-IL score	Mean ICECAP-O score
Age	Older people below 75	7.51	2.86	0.86*
	Elderly above 75	7.36	2.81	0.83
Sex	Female	7.44	2.82	0.85
	Male	7.41	2.84	0.84
Education	Pre-secondary	7.49	2.89	0.86
	Post-secondary	7.36	2.77	0.83
Married	Married or other living together	7.58	2.86	0.85
	Divorced	6.13	2.60	0.76
	Widow	7.38	2.86	0.85
	Never married	7.56	1.84	0.83
Living situation	Alone	7.17	2.80	0.84
	With partner/children	7.67	2.87	0.84
	Nursing home	6.76	2.66	0.78
Health				
Multimorbid	Maximum 1 chronic condition	7.76**	2.99**	0.89**
	More than 2 conditions	7.26	2.75	0.82
IADL	Independent	7.99**	3.01**	0.92**
	Dependent	7.17	2.74	0.81
SF-20 social activity limitations	No limitations	7.78**	2.97**	0.90**
	Limited	6.92	2.63	0.77
GDS	Not depressed	7.73**	2.95**	0.88**
	Depressed	6.48	2.47	0.73
EQ-5D	Top 50%	7.87**	2.97**	0.90**
	Bottom 50%	7.10	2.72	0.80

Indicating bivariate significance: * p-value <0.05 ** p-value<0.01.

Subgroups

Differences in demographic characteristics between the population with the highest ICECAP-O scores (highest third, n=111) and the lowest ICECAP-O scores (lowest third, n=94) were also investigated (analysis not shown here). Significant differences were found for age (older people having lower ICECAP-O scores), place of residence (living in a nursing home being associated with lower scores) and multi-morbidity (which is associated with lower scores). As for the other measures, a low ICECAP-O score is significantly associated with lower Cantril's ladder scores, SPF-IL scores and EQ-5D scores. As for GDS and IADL, depressed respondents and those with functional limitations were more likely to be in the group with low ICECAP-O scores.

Relationship between the ICECAP-O and measures of overall wellbeing

In a multivariate analysis of other measures of wellbeing, capability wellbeing as measured by the ICECAP-O tariffs was significantly and positively associated with wellbeing as measured by Cantril's ladder and the SPF-IL. HrQol as measured by the EQ-5D utility scores was not independently associated with SPF-IL or Cantril's ladder after ICECAP-O tariffs were included in the regression analyses. Being depressed was independently associated with lower Cantril's ladder as well as SPF-IL scores. Marital status and living arrangement were significantly related to Cantril's ladder but not to SPF-IL. Multimorbidity was associated with lower SPF-IL scores, but not significantly associated with Cantril's ladder scores (Table 4.5).

Table 4.5: Stepwise regression between Cantril's ladder, ICECAP-O and health*

	Dep: Cantril's ladder		Dep: SPF-IL	
	Standardized regression	p-value	Standardized regression	p-value
ICECAP-O tariffs	0.26	0.00	0.35	0.00
EQ-5D utilities			0.07	0.17
GDS	−0.38	0.00	−0.36	0.00
Divorced	−0.06	0.48		
Widow	0.15	0.03		
Never married	0.07	0.11		
Living alone at home	0.25	0.00		
Living in a nursing home	−0.05	0.48		
Multimorbidity			−0.10	0.04
Adj. R-square		0.41		0.46

*Demographic variables have been converted to dummies, and inserted in batch.

Discussion

Summary of main results

As hypothesized, the capability wellbeing instrument ICECAP-O tariffs were significantly correlated with other measures of wellbeing (Cantril's ladder, the SPF-IL) as well as with all health measures (EQ-5D dimensions and utilities, IADL, GDS, SF-20 Social Activity limitation). Contrary to expectations based on the type of instrument, the strength of the correlation between the ICECAP-O and the wellbeing measures was fairly similar as that with health measures. The individual ICECAP-O dimensions were also correlated with the overall scores of the different health and wellbeing measures. Overall, we found significant correlations between the ICECAP-O dimensions and the individual EQ-5D dimensions, with the exception of Attachment, which was not significantly correlated with the Pain/Discomfort and Anxiety/Depression dimensions of the EQ-5D and Security, which was not significantly correlated with the EQ-5D dimensions Mobility and Self-care. As hypothesized, the ICECAP-O discriminated between the following measures in the bivariate and multivariate analyses: depressed and non-depressed elderly, IADL dependent and non IADL dependent elderly and between those with social activity limitations and without social activity limitations. In the exploratory analysis the ICECAP-O discriminated between multi-morbid and other elderly and between elderly with high and low EQ-5D scores. Regarding measures of wellbeing, the ICECAP-O is significantly related to both Cantril's ladder and the SPF-IL, even when correcting for health variables.

Limitations

This study has a number of limitations worth mentioning. First, our sample of elderly was not representative, but consisted of post-hospitalized elderly, who were previously admitted to a single hospital, living in one region of the Netherlands. Elderly in our sample are frailer than the general community-dwelling elderly population, reporting lower levels of mobility on the EQ-5D [36] [37] [38] than customary for the age group. Such reduced mobility suggests that our population is characterized by functional decline, consistent with frailty. In addition, patients in our sample were characterized by a broad range of diseases and multiple chronic conditions, with heart failure and osteoporosis being the most common diagnoses. Such a relative high number of elderly with multi-morbidity is also consistent with frailty. Associations between capabilities, health and well-being may be weaker in a general sample of frail elderly due to less variation in measurements. However, we have no indication that the selection of respondents drives the results regarding validity. Future research in other community-dwelling elderly populations also in other countries than the UK is necessary to further test this and validate the instrument. Second, we used a stepwise regression to

identify explanatory variables of the ICECAP-O scores, which has limitations. In order to avoid rejecting possible significant variables, we used a relatively high p-value (0.2) for excluding variables. Additionally, we performed a regression analysis with all possible independent variables, which confirmed the results from the stepwise regression. It is worth noting moreover that, given the modest sample size, some subgroups were relatively small. This may lead to lack of power in establishing significant relationships.

Comparability with other findings

Compared to previous studies [10] [15] [16] [17] [34], the values for the individual dimensions and overall scores of the ICECAP-O in this current study are similar to those obtained in the general elderly population and substantially higher than those obtained in a Dutch nursing home [17]. The current scores are comparable to the British and Australian reference values [10] [15] [34], with the exception of the attachment dimension, where the British and Australian studies [10] [15] [34] report a higher percentage of older people at full capability (57% British and Australian studies vs. 36% current study) and the security dimension, where this current study has a far higher percentage of older people at full capability (53% current study vs. 18% British study vs. 37% Australian study). The differences in the attachment dimension cannot be explained by differences in the fraction of married elderly, which is quite similar across the studies. However, the elderly in the current study are a worse-off group (i.e. in terms of mobility) than the general elderly population in the UK, which may partly explain the lower scores on the attachment dimension. Differences on the security dimension may be explained by cultural differences in answering this question. Indeed, this is the second study in the Netherlands in which relatively high scores were found for the security dimension [17]. Hence, Dutch elderly either have fewer concerns about the future than UK elderly or are less likely to share their concerns about the future. It also seems important to further investigate whether the translation of the description of the security dimension may lead to the observed differences. The average overall scores found here i.e. 0.84 were comparable to those obtained in the British and Canadian population (0.82), the Australian population (0.81) and substantially higher than for older people in a nursing home (0.63). Comparison of the overall scores suggest that on average the ICECAP-O scores of the Dutch community-living elderly are comparable to the general population in Australia and the UK, and are substantially better than elderly living in nursing homes in the Netherlands.

Furthermore, the correlations between the ICECAP-O, Cantril's Ladder and EQ-5D show broadly similar results as reported in previous studies, with a number of exceptions. Unlike the British validation study [10] but in line with the Australian study [15], we found a statistically significant

though moderate correlation between ICECAP-O attachment dimension and the EQ-5D dimensions mobility, self-care and usual activities. In addition, unlike the British study we found a significant correlation between the ICECAP-O's security dimension and the EQ-5D dimensions usual activities and pain. It must be noted that these are quite weak correlations, and significance may or may not be reached due to minor differences in sampling variation. Such minor differences in sample variation may be related to differences in the respective samples; here we approached previously hospitalized elderly, while the British study was performed in a sample from the general elderly population. Our correlation results were also comparable to a Dutch study using proxy respondents in nursing homes [17]. There, however, the correlation between the ICECAP-tariffs and the EQ-5D was somewhat stronger then found here, which may be due to differences between self-report and proxy responses. In this study the ICECAP-O is unrelated to Sex and Education level, which is consistent with previous findings.

Relationship between health and wellbeing and the ICECAP-O

Comparing the performance of the ICECAP-O to that of other health and wellbeing instruments, some aspects deserve mentioning. Given the strong correlations between the ICECAP-O measure of capability wellbeing and the other two wellbeing measures, as well as between the ICECAP-O measure and the EQ-5D HrQol measure, ICECAP-O scores are related to both health and other wellbeing scores. The ICECAP-O scores are moreover related to individual health dimensions in terms of physical functioning, psychological functioning and social functioning. The tests of discriminant validity confirm this relationship between health measures and the ICECAP-O scores. Even though the ICECAP-O does not have an explicit physical dimension [39], it seems that it is capable of capturing the effect of decreased physical function on capability wellbeing to a large degree, primarily through the control and role dimensions. With respect to the wellbeing instruments, the strong correlation between the ICECAP-O and Cantril's ladder as well as the SPF-IL suggests that the ICECAP-O is related to these wellbeing measures as well, which is also confirmed in multivariate analyses. Table 4.4 does suggest that GDS has an influence on SPF-IL and Cantril's ladder beyond what is captured by the ICECAP-O. This may be related to the concept of capability wellbeing or to the ICECAP-O instrument's insensitivity for depression.

Implications for policy and future research

The ICECAP-O is a measure of wellbeing, and therefore has the potential to broaden the evaluative space of economic evaluations in health care by focusing on more than health alone. As such, it can potentially compare the benefits across a large number of sectors which (primarily) aim to improve wellbeing, such as (parts of) social care [2], institutionalized elderly care [40], public health [3], and

mental health [4]. This is a particularly useful property in case of populations such as frail elderly characterized by decreasing independence and multi-morbidity, potentially across different health dimensions. The ICECAP-O measures (one conceptualization of) wellbeing. In doing so, its outcomes are, expectedly, related to health outcomes. The ICECAP-O moreover discriminates between various better off and worse-off groups. In this current study, in a post-hospitalized group significant insights were gained in terms of the relationship between capability wellbeing, life satisfaction, SPF_IL and various health measures. On the basis of our findings, we advocate the further use of the ICECAP-O measure in the context of economic evaluations, especially in those circumstances where broader well-being effects are expected and in combination with other measures. It can also be used in large scale surveys aimed at identifying deprived populations in order to identify groups which may benefit from interventions, as has been done previously [34]. Nonetheless, a number of issues need to be explored further.

Further research is required to confirm the current favorable findings and to further explore the feasibility, validity and usefulness of the ICECAP-O instrument, also in the context of economic evaluations. In that context, larger studies would be helpful, allowing more subgroup analyses, as well as studies in different contexts (e.g. specific disease areas, living environments or cultural settings). Further research is especially encouraged in more homogeneous population characterized by a single disease. Furthermore, since the performance of the ICECAP-O has not been widely explored in longitudinal studies, the sensitivity to changes of the ICECAP-O is currently unclear. Whether the ICECAP-O comprehensively captures health and wellbeing changes, including depression, also deserves further attention. Additionally, further research is necessary to establish a causal relationship between health and wellbeing as measured by the ICECAP-O, and to explore ways in which the capabilities of older people can be improved.

Conclusion

The ICECAP-O is an outcome measure which may be particularly useful in the context of (economic) evaluations of health care services such as long-term elderly care, where broader effects are expected than those captured with conventional HrQol measures. In the current study, the ICECAP-O showed good convergent validity with validated measures of health and well-being as well as good discriminant validity in a heterogeneous population of post-hospitalized elderly. As such, the ICECAP-O seems to be a promising instrument. Additional research is required to not only confirm these findings in other settings and samples, but also to study the sensitivity to change of the instrument as well as its comprehensiveness in all relevant wellbeing effects.

Abbreviations

HrQol, Health-related Quality of life; Qol, Quality of Life; QALY, Quality-Adjusted Life Years; ICECAP-O, ICEpop (Investigating Choice Experiments for the Preferences of Older People); SPF-IL, Social Production Function: Instrument for Level of Well-being; IADL, Instrumental Activities of Daily Living; GDS, Geriatric Depression Scale

References

1. Al-Janabi, H., T. Flynn, and J. Coast, *Development of a self-report measure of capability wellbeing for adults: the ICECAP-A.* Quality of Life Research, 2012. **21**(1): p. 167-176.
2. Ryan, M., et al., *Using discrete choice experiments to estimate a preference-based measure of outcome--an application to social care for older people.* Journal of health economics, 2006. **25**(5): p. 927-944.
3. Chalkidou, K.K., *Cost-effective public health guidance: asking questions from the decision-maker's viewpoint.* Health Economics, 2008. **17**(3): p. 441-448.
4. Chisholm, D.D., *QALYs and mental health care.* Social psychiatry and psychiatric epidemiology, 1997. **32**(2): p. 68-75.
5. Who, *Official Records of the World Health Organisation*, 1948.
6. Bowling, A.P., et al., *Loneliness, mobility, well-being and social support in a sample of over 85 year olds.* Personality and Individual Differences, 1989. **10**(11): p. 1189-1192.
7. Jones, D.A., C.R. Victor, and N.J. Vetter, *The problem of loneliness in the elderly in the community: characteristics of those who are lonely and the factors related to loneliness.* The Journal of the Royal College of General Practitioners, 1985. **35**(272): p. 136.
8. Bowling, A. and S. Ebrahim, *Handbook of health research methods*. 2005, Berkshire, England: Open University Press.
9. Coast, J., et al., *Valuing the ICECAP capability index for older people.* Social science & medicine (1982), 2008. **67**(5): p. 874-882.
10. Coast, J., et al., *An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people.* Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 2008. **17**(7): p. 967-976.
11. Coast, J., R. Smith, and P. Lorgelly, *Should the capability approach be applied in health economics?* Health economics, 2008. **17**(6): p. 667-670.
12. Cookson, R., *QALYs and the capability approach.* Health economics, 2005. **14**(8): p. 817-829.
13. Coast, J., et al., *Investigating Choice Experiments for Preferences of Older People (ICEPOP): evaluative spaces in health economics.* Journal of health services research & policy, 2008. **13 Suppl 3**(Journal Article): p. 31-37.
14. Al-Janabi, H., et al., *An investigation of the construct validity of the ICECAP-A capability measure.* Quality of Life Research, 2012: p. 1-10.
15. Couzner, L., J. Ratcliffe, and M. Crotty, *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, 2012. **10**(1): p. 69.
16. Davis, J., et al., *A comparison of the ICECAP-O with EQ-5D in a falls prevention clinical setting: are they complements or substitutes?* Quality of Life Research, 2013. **22**(5): p. 969-977.
17. Makai, P., *Capabilities and quality of life in Dutch psycho-geriatric nursing homes: an exploratory study using a proxy version of the ICECAP-O.* Quality of life research, 2012. **21**(5): p. 801-812.
18. Covinsky, K.E., E. Pierluissi, and C.B. Johnston, *Hospitalization-Associated Disability.* JAMA: The Journal of the American Medical Association, 2011. **306**(16): p. 1782-1793.
19. Asmus-Szepesi, K.K., *Evaluation design of a reactivation care program to prevent functional loss in hospitalised elderly: A cohort study including a randomised controlled trial.* BMC geriatrics, 2011. **11**(1): p. 36.
20. Covinsky, K.E., et al., *Loss of Independence in Activities of Daily Living in Older Adults Hospitalized with Medical Illnesses: Increased Vulnerability with Age.* Journal of the American Geriatrics Society, 2003. **51**(4): p. 451-458.

21. Creditor, M.C., *Hazards of Hospitalization of the Elderly.* Annals of Internal Medicine, 1993. **118**(3): p. 219-223.
22. Grewal, I., et al., *Developing attributes for a generic quality of life measure for older people: Preferences or capabilities?* Social Science and Medicine, 2006. **62**(8): p. 1891-1901.
23. Cantril, H., *The pattern of human concerns*. 1966, New Brunswick: Rutgers University Press.
24. Nieboer, A., et al., *Dimensions Of Well-Being And Their Measurement: The Spf-Il Scale.* Social Indicators Research, 2005. **73**(3): p. 313-353.
25. Szende, A. and M.D.N. oppe, *EQ-5D value sets - inventory, comparative review and user guide*. Vol. 1. 2007, Dordrecht: Springer.
26. Katz, S., *Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living.* Journal of the American Geriatrics Society, 1983. **31**(12): p. 721-727.
27. De Rooij, S.E.J.A., *Cognitive, Functional, and Quality-of-Life Outcomes of Patients Aged 80 and Older Who Survived at Least 1 Year After Planned or Unplanned Surgery or Medical Intensive Care Treatment: Cognitive, Functional, and Quality of Life Outcomes.* Journal of the American Geriatrics Society (JAGS), 2008. **56**(5): p. 816-822.
28. Mehta, M.M., *Depressive symptoms in late life: associations with apathy, resilience and disability vary between young-old and old-old.* International journal of geriatric psychiatry, 2008. **23**(3): p. 238-243.
29. Almeida, O.O.P., *Short versions of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV.* International journal of geriatric psychiatry, 1999. **14**(10): p. 858-865.
30. Van Marwijk, H., et al., *Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the geriatric depression scale.* The British Journal of General Practice, 1995. **45**(393): p. 195.
31. Stewart, A.L., D.H. Ron, and J.E. Ware, Jr., *The MOS Short-Form General Health Survey: Reliability and Validity in a Patient Population.* Medical care, 1988. **26**(7): p. pp. 724-735.
32. Marengoni, A., et al., *Aging with multimorbidity: A systematic review of the literature.* Ageing Research Reviews, 2011. **10**(4): p. 430-439.
33. Mercer, S.S.W., *Multimorbidity in primary care: developing the research agenda.* Family practice, 2009. **26**(2): p. 79-80.
34. Flynn, T.N., *Assessing Quality of Life among British Older People Using the ICEPOP CAPability (ICECAP-O) Measure.* Applied health economics and health policy, 2011. **9**(5): p. 317.
35. Steiger, J.J.H., *Tests for comparing elements of a correlation matrix.* Psychological bulletin, 1980. **87**(2): p. 245-251.
36. Szende A, Williams A: *Measuring self-reported population health: an international perspective based on EQ-5D.* SpringMed Publishing 2004 115p.
37. König, H.-H., et al., *Health status of the advanced elderly in six European countries: results from a representative survey using EQ-5D and SF-12.* Health and Quality of Life Outcomes, 2010. **8**(1): p. 1-11.
38. Hoeymans, N., et al., *Measuring mental health of the Dutch population: A comparison of the GHQ-12 and the MHI-5.* Health and Quality of Life Outcomes, 2004. **2**(Journal Article).
39. Makai, P., et al., *Quality of life instruments for economic evaluations in health and social care for older people: a systematic review* Social Science & Medicine, 2014. **102**: p. 10.
40. Donaldson, C., et al., *Should QALYs be programme-specific?* Journal of health economics, 1988. **7**(3): p. 239-257.

Chapter 5. Capabilities and Quality of Life in Dutch psycho-geriatric nursing homes: An exploratory study using a proxy version of the ICECAP-O

This chapter is based on Peter Makai, Werner Brouwer, Marc Koopmanschap and Anna Nieboer Capabilities and Quality of life in Dutch psycho-geriatric nursing homes: An exploratory study using a proxy version of the ICECAP-O
Quality Of Life Research 2012 Jun;21(5):801-12

Abstract

Purpose

To validate the ICECAP-O capability measure in psycho-geriatric elderly in nursing homes, we compared the capability scores of restrained and unrestrained clients. Both nursing staff and family were used as proxies for assessing clients' capabilities.

Method

For 122 psycho-geriatric elderly a total of 96 nursing professionals and 68 family members completed a proxy questionnaire. We investigated the convergent and discriminant validity of the ICECAP-O and measures of care dependency, health-related quality of life, and overall quality of life. We also directly compared ICECAP-O scores of the 56 clients for whom both nursing staff and family members had completed the questionnaire.

Results

Convergent validity between ICECAP-O and care dependency, health-related and overall quality of life measures could be established, as well as discriminant validity for the restrained and unrestrained groups. Nursing and family proxy ICECAP-O tariffs were not significantly correlated.

Discussion

ICECAP-O measures a more general concept than health-related quality of life, and can differentiate between restrained and non-restrained psycho-geriatric clients. Since nurses seem to be able to assess the current quality of life of clients using the ICECAP-O more precisely than the family proxies, for now the use of nursing proxies is recommended in a nursing home setting.

Background

Services aimed at elderly living in psycho-geriatric (PG) wards who mainly suffer from advanced dementia, are not often evaluated using cost-effectiveness analysis. Still, in general, trading off costs and benefits is as important in long-term care as it is in curative care. Especially in times of budget cuts or when care innovations find their way into the long term care sector, considering the costs and benefits of interventions is important. In curative care, this is commonly done using cost-utility analysis, where the primary outcome is health-related quality of life (HrQol). Preference-based health-related quality of life measures attach utility weights to specific health states in order to be able to compute utility gains from health changes. Such gains are then compared to the (incremental) costs of an intervention [1]. Using this approach to evaluate services for the PG elderly, however, is problematic.

A major problem is that health related quality of life measures aim to detect and value changes in health and functioning, while services for the elderly may (be aimed to) affect quality of life more broadly [2] [3] [4]. For example, it is not uncommon to physically restrain the PG elderly to prevent them from falling [5], but doing so restricts freedom of movement, autonomy, and enjoyment of life. Removing such restraints would restore some control over their lives and allow more enjoyment through an increased capacity to fill their day with more varied activities. Whether the health of unrestrained patients would also improve, however, is questionable since freedom of movement may not directly affect existing health problems. Therefore, in determining the value for money of interventions aimed to reduce restraints, HrQol is likely to be a too restrictive evaluative space, since it does not (directly) value self-control or enjoyment of life. HrQol measures (such as the EQ-5D and SF-6D) may therefore not fully account for all benefits of such interventions, and using them in these contexts could misinform decision makers.

A promising approach to measure Qol more comprehensively in the PG elderly is to use the newly developed ICEpop (Investigating Choice Experiments for the Preferences of Older People) capability measure for older people (ICECAP-O). The ICECAP instruments can be seen as measuring capability Qol [6] achieved by the capacity to perform certain actions and achieve certain states [7]. The ICECAP-O measures five capability dimensions – attachment, security, role, enjoyment, and control – with one question per dimension. Each dimension can be scored on four levels. The ICECAP-O was developed using rigorous qualitative and quantitative approaches [8] [9] [10] [11]. In order to obtain tariffs for the ICECAP-O, the attributes were valued using best-worst scaling, a special type of discrete choice analysis. The ICECAP-O has been used in the British general elderly population,

demonstrating that it is related to, but not exclusively dependent on HrQol [9]. The overwhelming majority of the included elderly lived at home and did not receive long-term health or social care. To date, ICECAP-O has not been used in populations receiving long-term care. This lack of validation is especially problematic for the vulnerable PG elderly populations, who consume substantial amounts of health and social services [12].

It needs noting that substantial effort has been put in recent years into developing dementia-specific Qol instruments for use in patients with mild to moderate dementia [13] [14] [15][16][17] [18] [19][20]) and in severe dementia [21]. However, not only do these instruments normally not have related utility weights, limiting their usefulness in cost-utility analysis, they also are, by definition, disease specific rather than generic, which limits their usefulness in decision making across diseases and sectors. Moreover, disease specific measures can still focus on health related, rather than general quality of life. Hence, here we focus on the generic ICECAP-O, with its preference based tariffs.

The use of Qol instruments in a PG patient population is difficult, since due to their cognitive limitations, patients may not be able to assess their Qol accurately. It has been shown to be possible to develop user-friendly (disease specific) instruments for self-completion in this context, especially for mild to moderate dementia patients. However, with diminishing cognitive ability, this becomes increasingly difficult. Currently, to our knowledge, there are no generic Qol instruments with accompanying utility weights that are recommended for use in people with dementia. The lack of validation in this particular population, i.e., the PG elderly, is likely to be related to limited cognitive ability due to severe dementia [22], hampering self-completion of questionnaires. We therefore decided to use proxies, who complete the questionnaire on the patient's behalf. An important issue with proxies is that they may not complete the questionnaire as the client would have. A prerequisite in using proxies is that they can at least provide reasonable approximations of the patient's Qol [23]. Proxy measurement has been associated with a consistent negative bias in Qol measurement [13], although this may be more typical in case of informal carers of dementia patients [14]. It has been suggested that such proxy effects can be minimized using substituted judgement [14], asking the proxies to fill out the questionnaire *as if* they were the person with dementia.

The aim of this study is to explore the validity of the ICECAP-O for the PG elderly. To that end, we first investigated the convergent validity of the ICECAP-O by comparing it to other care-related HrQol and overall Qol instruments. We used a sample of elderly in Dutch psycho-geriatric nursing

homes to establish the discriminant validity of the ICECAP-O by (1) comparing a restrained group to a non-restrained group, and (2) investigating whether the ICECAP-O was indeed measuring a concept broader than health. To complete the validation exercise, we compared questionnaires filled out by two appropriate proxies, namely nursing staff and family members. This is to our knowledge the first study of its kind.

Method

Design

The ICECAP-O questionnaire was forward-backward translated into Dutch by two independent translators. For our study we used the baseline measurement from an economic evaluation study of a quality improvement intervention that aimed to reduce restraints in the Care for Better quality collaborative in Dutch long-term care [24,25]. Four nursing homes and a total of 122 clients from different geographic regions in the Netherlands participated in the study. All 72 clients in restraints participated and 50 randomly selected non-restrained clients in the same departments served as a control group. We distributed two copies of the questionnaire for each client, one to be filled out by nursing staff that personally cared for the client (nursing version) and one for family members (family version) asking proxies to use substituted judgement. Since data collection of the nursing version was carried out in the context of a national quality improvement program, no ethical committee approval was necessary under Dutch law [26] [27]. Informed consent was obtained for the family version. The researchers received no personal information about the clients during the study.

Measures

Besides the ICECAP-O (as shown in Appendix), the questionnaire contained the following QoL measures: the EQ-5D, EQ-VAS instrument, Cantril's ladder, and overall life satisfaction. It also contained the Hospital Anxiety and Depression Scale (HADS). The nursing version contained the care dependency scale (CDS), which needs to be completed by care professionals. The EQ-5D [28] measures HrQoL along five dimensions (mobility; self-care; daily activities; pain and discomfort; and anxiety and depression) with three levels each (1 = no problems, 2 = moderate problems, and 3 = extreme problems). It has been used with proxies in a large number of studies, including clients with Alzheimer and severe dementia [29]. The EQ-VAS is a one-dimensional HrQoL measure frequently used alongside the EQ-5D in validation studies and has also been used with proxies [23]. The EQ-VAS comprises a single scale ranging from zero (worst imaginable health) to 100 (best imaginable health). Cantril's ladder is a classic one-dimensional overall quality of life scale [30], with

the bottom rung representing no quality of life and the top representing full quality of life. It has been used with proxies [31]. We also used an overall life satisfaction scale, a one-dimensional index ranging from zero (completely dissatisfied) to 10 (completely satisfied) [32]. The HADS scale was originally developed for use in hospitals, but has since been used in various populations [33] and with proxies [34] to assess anxiety and depression symptoms. HADS consists of two 7-item scales, one for depression and one for anxiety, which can be also used in a composite index (Cronbach's alpha = 0.82 nursing version, 0.87 family version) in this study comparable to self-reported values in dutch elderly [35] with values ranging from 0 (no problems) to 42 (severe depression and anxiety). The care dependency scale (CDS), developed by Dijkstra [36] contains 15 dimensions measuring the amount of independence the patient has retained with regard to dimensions such as eating and drinking, body posture, incontinence, learning ability, ability to structure the day, communication or autonomy. The CDS has scores that range from 15 (completely care dependent) to 75 (completely care independent). The CDS has been used and validated extensively [37,38] and is a useful instrument for assessing need for care. CDS scores have been shown to be associated with a number of problems in elderly care, such as fall-risk, pressure ulcers, and so on, and are designed to be completed by nurses and professional caregivers [37,38].

Hypotheses

For convergent validity, we expect the ICECAP-O to correlate with overall measures (Cantril's ladder, and overall life satisfaction) and HrQol measures, as well as with CDS scores, since all measurement instruments differentiate between better and worse states. With respect to discriminant validity, we expect to find differences between the non-restrained and restrained groups in terms of ICECAP-O scores and other overall Qol measures, but not in HrQol measures, since we expect the two groups to be in a similar health state, while their non-health circumstances differ. To test whether capabilities are indeed measuring a concept broader than health, we expect to observe a difference in ICECAP-O scores between the restrained and the non-restrained clients even when controlling for HrQol, demographic variables, and care dependency. For proxy agreement, we expect the nursing and family proxies for each client to be correlated and the scores to be not significantly different from each other.

Analysis

We performed an item-level analysis to determine non-response for all scales in the questionnaire. We used multiple imputations to treat item non-response for nursing and family questionnaires separately with the Markov chain Monte-Carlo method (MCMC) [39]. We also tested the assumption of multivariate normality underlying the MCMC method. Following multiple

imputations, utility and sum-scores were computed where relevant (see Appendix). For the CDS, which was only included in the nursing version, the nursing scores for the patients for which a family version was also present, were also used in the analysis pertaining to the family version. Remaining missing observations were imputed.

We used descriptive statistics to analyze demographic characteristics. Means and standard deviations were computed for continuous variables, medians for ordinal variables. All comparisons between demographic variables were performed using the Mann-Whitney-U test, except in the case of education, where a Chi-square was performed. Data were analyzed using STATA 11.

Concurrent validity was assessed using correlations in the nursing and family versions separately. To test discriminant validity we employed chi-square tests and Mann-Whitney-U tests to compute mean differences between the restrained and non-restrained groups. We performed this comparison on the nursing and the family proxy separately. To further investigate whether the ICECAP-O could both discriminate between the groups and measure a concept broader than HrQol, we performed multivariate regressions. For this purpose we controlled for demographic variables and care dependency. Two multivariate ordinary least squares regression (OLS) models were fitted on the ICECAP-O index, one using nursing proxy variables and one using family proxy variables. Regression assumptions were checked. To test agreement between the two proxy groups, we used the Mann-Whitney-U tests and correlations for the questionnaires for which both proxies were available.

Results

Response

For a total of 122 clients, 96 nurses and 68 family members completed the questionnaires, implying response rates of 78% and 56% respectively. For the 96 nursing questionnaires, 62 clients (64%) were in restraints; for the 68 family questionnaires, 47 clients (69%) were in restraints. For 56 clients we received both types of proxy questionnaires. Item non-response was not systematic, and averaged around 2% across all items in the nursing questionnaires, and 4% in the family questionnaires. In the nursing version, multiple imputations allowed for using 96 cases instead of 88-91 in bivariate analysis and 87 in multivariate analyses. In the family version, multiple imputations allowed using 68 cases instead of 58-61 in bivariate analysis and 47 in multivariate analysis.

Table 5.1: Demographic and care-related characteristics

Demographics	**All cases**	
	Nursing version n = 96	Family version n = 68
Age[&]	82 (9.1)	82 (7.3)
Sex[&]	68% female	67% female
Education		
Primary	45.8%	51.5%
Secondary (general)	27.1%	14.7%
Secondary (vocational)	14.6%	16.2%
Secondary (scientific)	5.2%	7.4%
Tertiary (university, college)	7.3%	10.4%
CDS[&§]	31.14 (15.09)	33.53 (15.16)
HADS[&]	23.68 (3.91)	24.30 (3.17)
ICECAP-O		
Attachment[&$]	2.34 (0.78) 2	2.79 (0.89) 3
Security[&$]	3.38 (0.79) 4	3.15 (1.10) 4
Role[&$]	1.78 (0.81) 2	1.43 (0.68) 1
Enjoyment[&$]	2.21 (0.81) 2	1.88 (0.82) 2
Control[&$]	1.40 (0.71) 1	1.15 (0.50) 1
ICECAP-O Tariffs[&&]	0.50 (0.20)	0.43 (0.17)
EQ-5D		
Mobility[&$]	2.09 (0.62) 2	2.25 (0.53) 2
Self-Care[& $]	2.71 (0.50) 3	2.82 (0.46) 3
Usual Activities[&$]	2.75 (0.52) 3	2.88 (0.37) 3
Pain, Discomfort[&$]	1.81 (0.57) 2	1.69 (0.58) 3
Anxiety,Depression[&$]	1.70 (0.63) 2	1.76 (0.65) 2
EQ-5D Tariffs[&]	0.49 (0.21)	0.46 (0.20)
EQ-VAS[&]	55.33 (17.24)	45.87 (16.56)
Cantrill's Ladder[&]	4.62 (2.01)	4.11 (4,02)
Overall life satisfaction[&]	4.72 (2.34)	4.76 (3.22)

[&]mean, (Standard deviation) [$] median [§] Only included in nursing version

Descriptive characteristics and relationship between different proxies

Client's demographic and care-related characteristics can be seen in Table 5.1, split according to the two versions of the questionnaire.

Table 5.2: Convergent and discriminant validity nursing version

Nursing version/ nursing version n = 96	***ICECAP-O tariffs nursing version***	***EQ-5D nursing version***	***EQ-VAS nursing version***	***Cantril's ladder nursing version***	***Overall life satisfaction nursing version***	***Care Dependency scale nursing version***	***HADS nursing version***
ICECAP-O tariffs nursing version	1.00						
EQ-5D nursing version	0.48(**)	1.00					
EQ-VAS nursing version	0.55(**)	0.49(**)	1.00				
Cantril's ladder nursing version	0.60(**)	0.51(**)	0.70(**)	1.00			
Overall life satisfaction nursing version	0.52(**)	0.34(**)	0.65(**)	0.70(**)	1.00		
Care Dependency scale nursing version	0.56(**)	0.50(**)	0.34(**)	0.47(**)	0.23(*)	1.00	
HADS nursing version	-0.18	0.08	-0.10	-0.19	-0.08	-0.16	1.00

* significance on the 5% level ** significance on the 1% level

Convergent validity

As can be seen in Tables 5.2 and 5.3, there was a significant correlation between capabilities and HrQol, as shown by the significant correlation between the ICECAP-O tariffs and the EQ-5D and the EQ-VAS health measures. The correlation, however, was not particularly strong. The ICECAP-O tariffs were also correlated with Cantril's ladder and the overall life satisfaction measures. There was also a significant relationship between ICECAP-O tariffs in both versions of the questionnaire and CDS, though the correlation was stronger in the nursing version. The HADS was not correlated with the ICECAP-O tariffs in either the nursing or family questionnaires.

Table 5.3: Convergent and discriminant validity family version

Family version/ family version n = 68	***ICECAP-O tariffs family version***	***EQ-5D family version***	***EQ-VAS family version***	***Cantril's ladder family version***	***Overall life satisfaction family version***	***Care Dependency scale nursing version***	***HADS family version***
ICECAP-O tariffs family version	1.00						
EQ-5D family version	0.57(**)	1.00					
EQ-VAS family version	0.43(**)	0.36(**)	1.00				
Cantril's ladder family version	0.33(**)	0.20	0.32(**)	1.00			
Overall life satisfaction family version	0.48(**)	0.37(**)	0.28(*)	0.80(**)	1.00		
Care Dependency scale nursing version	0.32(**)	0.10	0.45(**)	0.20	0.16	1.00	
HADS family version	-0.01	0.32(**)	0.01	0.07	0.11	-0.21	1.00

* Significance on the 5% level ** Significance on the 1% level

Discriminative validity

The demographic and care-related characteristics for the respondents of the restrained and unrestrained client groups can be seen in Table 5.4. Age and gender were not significantly different for the two groups. There was no significant association between education and being in restraints. The mean CDS score was significantly lower in the group in restraints, indicating higher dependency. HADS scores differed significantly for clients in restraints in the nursing version; they were more depressed and anxious. In the nursing version, there was a significant difference between the groups in all ICECAP-O dimensions except for security. In the family version, two dimensions – role and enjoyment – were significantly different. A difference was also observed in the ICECAP-O tariffs. Clients without restraints score somewhat higher on HrQol as measured by EQ-5D and EQ-VAS, but the difference was not significant at the 5 percent confidence level. A Mann-Whitney-U test indicated that there was a significant difference in terms of capabilities. This was also true for the overall Qol as measured by Cantril's ladder and overall life satisfaction.

Table 5.4: Comparison of restrained and non-restrained clients

	Nursing version N=96			***Family version N=68***		
	Restrained	Not-restrained	P value	Restrained	Not restrained	P value
Age[&]	83 (7.7)	80 (10)	0.232	82 (8.1)	83 (5.3)	0.456
Sex[&]	67% female	68% female	0.407	64% female	72% female	0.975
Education			0.154			0.495
Primary	37.1%	61.8%		55.3%	42.9%	
Secondary (general)	30.6%	20.6%		10.6%	23.8%	
Secondary (vocational)	16.1%	11.8%		12.8%	23.8%	
Secondary (scientific)	8.1%	6.7%		8.5%	4.8%	
Tertiary (university, college)	8.0%	5.8%		12.8%	4.8%	
CDS[&§]	27.58 (11.62)	40.47 (15.09)	0.000(**)	28.70 (12.17)	44.33 (15.89)	0.000(**)
HADS[&]	24.16 (4.18)	22.82 (3.25)	0.048(*)	24.38 (3.34)	24.14 (2.83)	0.826
ICECAP-O						
Attachment[&$]	2.17 (0.78) 2	2.64 (0.69) 3	0.001(**)	2.74 (0.92) 3	2.90 (0.83) 3	0.533
Security[&$]	3.43 (0.80) 4	3.29 (0.76)3	0.266	3.13 (1.10) 4	3.19 (1.17) 4	0.744
Role[&$]	1.53 (0.78) 1	2.20 (0.69) 2	0.000(**)	1.32 (0.63) 1	1.67 (0.73) 2	0.028(*)
Enjoyment[&$]	2.05 (0.76) 2	2.53 (0.75) 2	0.005(**)	1.74 (0.70) 2	2.19 (0.98) 2	0.034(*)
Control[&$]	1.26 (0.65) 1	1.65 (0.73) 2	0.000(**)	1.11 (0.48) 1	1.24 (0.54) 1	0.121
ICECAP-O Tariffs[&]	0.43 (0.19)	0.63 (0.16)	0.000(**)	0.40 (0.16)	0.51 (0.18)	0.033(*)
EQ-5D						
Mobility[&$]	2.21 (0.55) 2	1.88 (0.69) 2	0.016(*)	2.30 (0.59) 2	2.14 (0.36) 2	0.194
Self-Care[&$]	2.84 (0.41) 3	2.47 (0.56) 2	0.000(**)	2.87 (0.40) 3	2.71 (0.56) 3	0.160

Usual activities[&$]	2.82 (0.46) 3	2.61 (0.60) 3	0.043(*)	2.89 (0.37) 3	2.86 (0.36) 3	0.496
Pain, discomfort[&$]	1.76 (0.62) 2	1.91 (0.45) 2	0.156	1.72 (0.62) 2	1.62 (0.50) 2	0.589
Anxiety, Depression[&$]	1.77 (0.62) 2	1.61 (0.65) 2	0.260	1.83 (0.64) 2	1.62 (0.67) 2	0.193
EQ-5D Tariffs[&]	0.46 (0.22)	0.53 (0.18)	0.200	0.43 (0.21)	0.53 (0.18)	0.073
EQ-VAS[&]	53.18 (15.52)	59.26 (19.66)	0.072	43.96 (16.50)	50.14 (16.28)	0.230
Cantrill's Ladder[&]	4.15 (2.00)	5.49 (1.96)	0.000(**)	3.96 (4.05)	4.45 (4.04)	0.931
Overall life satisfaction[&]	4.48 (2.16)	4.16 (2.63)	0.118	4.46 (2.23)	5.42 (3.14)	0.319

* Significance on the 5% level ** Significance on the 1% level [$] median [&]Mean, (SD) [#] Chi-square test
[§] Only included in nursing version

Table 5.5 shows how the ICECAP-O tariffs discriminated between clients with and without restraints using a multivariate analysis. Being in restraints independently discriminated between capability Qol in the nursing version, but not in the family version, when controlling for HrQol measures, demographic measures, and care dependency. The individual influences of the EQ-5D and CDS on the ICECAP-O tariffs are pronounced in both versions.

Table 5.5: Regression results

Independent variables	***Dependent variable: ICECAP-O***	
	Nursing version (n = 96)	Family version (n = 68)
Constant	3.61 (0.326)	-5.68 (0.242)
Restrained	-0.141 (0.001)(**)	-0.043 (0.403)
Age	0.002 (0.349)	0.003 (0.225)
Sex	-0.008 (0.804)	0.029 (0.526)
Education level	0.012 (0.289)	0.002 (0.871)
Organization dummy 1	0.055 (0.357)	-0.067 (0.293)
Organization dummy 2	0.062 (0.125)	0.030 (0.642)
Organization dummy 3	0.059 (0.274)	-0.037 (0.518)
Pilot or control department	0.041 (0.437)	-0.073 (0.172)
CDS[§]	0.003 (0.007)(**)	0.003 (0.033)(*)
EQ-5D	0.325 (0.000)(**)	0.411 (0.000)(**)
R square (adj.)	0.500 (0.441)	0.263 (0.373)

* Significance on the 5% level ** Significance on the 1% level
[§] Only included in nursing version

Table 5.6: Analysis of selection of respondents for whom both versions were available

	Both responded N=56			
	Nursing version	Family version	Mann-Whitney U	Correlation
CDS[&§]	32.00 (15.47)	32.00 (15.47)	n.a.	n.a.
HADS[&]	24.39 (3.22)	24.33 (3.23)	0.96	0.33(*)
ICECAP-O				
Attachment[&$]	2.34 (0.78) 2	2.79 (0.89) 3	0.06	0.14
Security[&$]	3.38 (0.79) 4	3.15 (1.10) 4	0.75	0.04
Role[&$]	1.78 (0.81) 2	1.43 (0.68) 1	0.01(*)	0.23
Enjoyment[&$]	2.21 (0.81) 2	1.88 (0.82) 2	0.03(*)	0.19
Control[&$]	1.40 (0.71) 1	1.15 (0.50) 1	0.04(*)	0.30(*)
ICECAP-O Tariffs[&]	0.49 (0.19)	0.44 (0.17)	0.24	0.15
EQ-5D				
Mobility[&$]	2.09 (0.62) 2	2.25 (0.53) 2	0.02(*)	0.38(**)
Self-Care[&$]	2.71 (0.50) 3	2.82 (0.46) 3	0.18	0.23
Usual activities[&$]	2.75 (0.52) 3	2.88 (0.37) 3	0.05	0.18
Pain, discomfort[&$]	1.81 (0.57) 2	1.69 (0.58) 2	0.29	0.18
Anxiety, Depression[&$]	1.70 (0.63) 2	1.76 (0.65) 2	0.85	-0.34
EQ-5D Tariffs[&]	0.46 (0.22)	0.47 (0.20)	0.63	0.01
EQ-VAS[&]	54.48 (14.64)	46.56 (1.57)	0.00(**)	0.33(*)
Cantrill's Ladder[&]	4.46 (2.24)	4.11 (4.14)	0.41	0.01
Overall life satisfaction[&]	4.44 (2.45)	4.76 (3.29)	0.62	0.03

* Significance on the 5% level ** Significance on the 1% level [$] median [&]Mean, (SD) [#] Chi-square test
[§] only included in nursing version

Relationship between the two proxies

Table 5.6 shows the agreement between the nursing and family assessment of the variables using Mann-Whitney-U and correlations for the 56 clients for whom both proxy versions were available. The results of the Mann-Whitney-U show that three out of five ICECAP-O dimensions had a significantly different distribution between the proxy groups, while the average tariffs were not significantly different. The distributions of the EQ-5D were not significantly different except for the mobility dimension. The EQ-VAS was significantly different. Overall Qol measures were the same in both proxy groups.

Agreement between the nursing and family proxies was low for the ICECAP-O dimensions. A significant correlation existed only between the two versions for the control dimension. Neither the ICECAP-O tariffs for both proxy groups nor the EQ-5D scores were significantly correlated. Measures of overall quality of life were also uncorrelated between the two proxy groups; this is true for both Cantril's ladder and overall life satisfaction. On the other hand, there was a slightly significant correlation between the EQ-VAS scores in both proxy groups. The HADS score was significantly correlated in the two proxy groups.

Discussion

Summary of main results

Our study is the first attempt to measure Qol and capabilities of physically-restrained psycho-geriatric nursing home clients. It was performed in the context of a validation exercise of the ICECAP-O. The ICECAP-O seems a promising, generic, preference-based instrument in the context of evaluating interventions in the psycho-geriatric context. Our study showed reasonable convergent and discriminant validity. Although related to HrQol, the relationship did not turn out to be very strong in our study. Given that and the multivariate regression results, the ICECAP-O appears to encompass a broader evaluative space than health alone. As expected, when the two groups were compared, clients in restraints had a lower Qol than clients without restraints. Being in restraints discriminated in capability Qol in the nursing version, even when correcting for the influence of other variables. This was not the case in the family version. In general, little agreement between the family and nursing versions was found for the different variables, raising important questions about which proxy version to consider superior or most reliable.

Methodological limitations

There are some noteworthy methodological limitations to our study. Ours has been a relatively small-scale study in a particular setting, limiting the generalization of results. This is especially true since collection of additional data on diagnosis and disease severity was not feasible, given limited space in the questionnaire. Only the functional consequences of the disease were measured through the CDS. Unfortunately, using dementia-specific QoI measures as well was not possible. This study was performed alongside a real-world economic evaluation, and adding additional instruments to the questionnaire would have decreased the number of participating organizations even further due to the increased burden caused by the study. Therefore, we necessarily restricted the scope of this study, focusing on generic quality of life instruments that are particularly useful in economic evaluation. Another limitation concerns the use and interpretation of the HADS. This instrument to our knowledge has not been validated in people with severe dementia. We nonetheless opted for inclusion of the HADS because symptoms of depression and anxiety could be particularly important in the context of restraints. Also, HADS is widely used making comparisons to other populations straightforward. Moreover, it is a relatively short instrument compared to instruments like the Cornell Scale for Depression and Dementia[40], while showing similar reliability. [40]

Moreover, given the limited number of respondents, we used multiple imputations to retain the full sample in the analyses. Multiple imputations allow for a more valid statistical inference [37] than full-case analysis, as long as only a small percentage of the data are imputed even if the assumption of multivariate normality is not met, as in this case. In the current study imputed results are comparable with a full-case analysis (not shown here). OLS estimates in the nursing version had non-normally distributed error terms; in our analysis of it, we thus used robust estimation techniques. Clearly, therefore, repeating studies like this, using larger samples is encouraged.

Security dimension

The nursing version of the ICECAP-O discriminated between restrained and non-restrained clients on all dimension levels except for security. This may be related to the fact that the scores on this item were relatively high. This was somewhat surprising since the average score on the security dimension was low in the study among general British elderly [8]. This difference may have to do with the study setting or item phrasing. Regarding the former, it is quite possible that nursing home clients suffering from dementia really did not seem worried. This may imply that nursing homes provide a safe environment. It is also possible, however, that these patients may not have been (seen as being) able to worry about or have a grasp of their future. In future (proxy) studies, it may

be worthwhile to further investigate this by, for instance, using alternative wording. Then, the underlying reasons for indicating being able to think about the future without worry (i.e. because there is nothing to worry about or one has lost the ability to worry) can be distinguished. Additionally, the proximity to death of some clients may influence the security dimension for the proxies completing the questionnaire. We have received anecdotal information from some respondents that the proximity to death makes questions regarding the future difficult to answer. It is also noteworthy in this context that in another version of the ICECAP, the ICECAP-A, the wording of the security dimension reads "feeling settled and secure" rather than “thinking about the future”[41].

Clients in restraints

According to the nursing version of the ICECAP-O, clients in restraints are indeed worse off than non-restrained clients in terms of capabilities, indicating that physical freedom seems to be an empirically important element of Qol. This finding is in line with earlier studies that indicated that being restrained is not beneficial to the elderly [5] [42]. We do note that, since the current study did not measure cognition directly, only indirectly through the CDS, it is possible that unobserved differences in cognition between two groups may have influenced our results. Cognition, however, is not consistently identified as a predictor of using physical restraints [5]. Our study, in that sense, gives further rationale for efforts toward reducing the use of physical restraints in psycho-geriatric nursing homes [42].

Differences between proxies

The differences between the nursing and family versions of the questionnaire raise important, yet difficult to answer questions regarding suitable (and valid) proxies. The observed differences may well relate to a difference in reference points. Nursing staff might answer the questions with similar clients in mind, while family members may assess the client’s current capability Qol in relation to former capability Qol, i.e., before psycho-geriatric services were necessary. While both viewpoints can be relevant in their own right, for evaluating interventions aimed at improving the situation of clients in the care context, the nursing proxy seems the most logical choice.

An important limitation of the ICECAP-O to date is that its sensitivity to change has not been explored. Indirectly, our study provides some indication of it in that nursing proxies distinguish between restrained and non-restrained clients. The fact that family proxies apparently did not may strengthen the choice for using nursing staff as proxies. Still, it is necessary to test and explain the discrepancy between proxies further, also in relation to sensitivity to change.

Decision-making and transferability of tariffs

Concerning the use of ICECAP-O in cost-utility studies, it should be noted that on a theoretical level, the ICECAP instruments are rooted in capability theory rather than utility theory. In capability theory, developed by Sen [7], people's wellbeing is measured in terms of their capacity to perform certain actions and achieve certain states [7]. Its prescription for societal redistribution may be seen as maximizing capabilities or as guaranteeing basic capabilities for everyone [43]. Until recently, the approach did not have an empirically tested, well-defined list of capabilities [44], which can be a weakness if societal redistribution is an issue. It would also be possible to use capability-based QoI instruments in cost-effectiveness studies [45]. In such an evaluation the final outcome would be based on capability attributes instead of HrQoI attributes, allowing the computation of 'capability QALYs' [6]. Such an approach could be considered to be consistent with the extra-welfarist framework [46] underlying cost-effectiveness analysis, which allows the broadening of the evaluative space to include (also) non-utility information [46]. On the other hand, there is considerable theoretical and empirical uncertainty about how such an approach might work [8] [45] with respect to the valuation of health and capabilities.

Our study used the British tariffs to compute capability valuation since Dutch tariffs are not (yet) available. Using Dutch tariffs would probably not have led to vastly different results, since, in the nursing proxy questionnaire, already four of the five dimensions of the ICECAP-O had significantly different scores for the restrained versus non-restrained group. Still, the weights attached to different capabilities may vary between countries.

Besides the problem of tariffs, the transferability of the capability dimensions themselves can also be a point of discussion. According to Sen [7], who does not list specific capabilities, relevant capabilities should be tailored to the local population and hence generating a list should be performed on a more local level. On the contrary, Nussbaum [43] proposed that basic capabilities exist and can be used globally. Since the capability measure is a possible outcome used in optimization and redistributive policies, using a standardized descriptive system across health systems (and countries) to evaluate similar interventions aimed at basic capabilities is clearly advantageous. On the other hand, specific (non-basic) capabilities may be valuable for the relevant target group of a particular intervention. The issue here is whether the ICECAP-O measures basic capabilities, or at least transferable capabilities, or more specifically capabilities important to British elderly. The fact that the dimensions of the ICECAP-O resemble frequently-reported universal subjective well-being measures [47] is indicative of the former, although the physical dimension is not measured directly. It seems, therefore, that the ICECAP-O is suitable as a more generic outcome

measure in elderly care. As such, it may assist decision makers to make choices based on ensuring and enhancing basic capabilities for this group.

Conclusion

The ICECAP-O instrument appears to be a promising tool for use in evaluations of interventions in psycho-geriatric care that do not necessarily or primarily improve health. The nursing proxy version of the questionnaire particularly demonstrated convergent and discriminant validity. Future research will have to confirm these findings in other settings, with particular attention paid to dementia severity, diagnosis, and validation alongside dementia-specific Qol measures. Additional research is also required on (1) the ICECAP-O's sensitivity to change, especially in evaluating interventions, (2) the relationship between overall quality of life, utilities, and capabilities for different settings, and (3) eliciting valid proxy information. With respect to the clients involved in this study, the ICECAP-O makes it clear that interventions aimed at removing restraints may well be worthwhile if capabilities are deemed important.

Acknowledgements

We are grateful to Hareth Al-Janabi and Mathilde Strating for their useful suggestions on earlier versions of this paper, as well as the anonymous reviewers for their suggestions on the current one. This study was financed by ZonMw in the context of the Care for Better evaluation project.

References

1. Drummond, M.F., et al., *Methods for the economic evaluation of health care programmes*. Vol. Third Edition. 2005, Oxford: Oxford University Press.
2. Ryan, M., et al., *Using discrete choice experiments to estimate a preference-based measure of outcome--an application to social care for older people.* Journal of health economics, 2006. **25**(5): p. 927-944.
3. Makai, P., et al., *Quality of life instruments for economic evaluations in health and social care for older people: a systematic review* Social Science & Medicine, 2014. **102**: p. 10.
4. Nieboer, A.P., X. Koolman, and E.A. Stolk, *Preferences for long-term care services: Willingness to pay estimates derived from a discrete choice experiment.* Social science & medicine, 2010. **70**(9): p. 1317-1325.
5. Hamers, J.P. and A.R. Huizing, *Why do we use physical restraints in the elderly?* Z Gerontol Geriatr, 2005. **38**(1): p. 19-25.
6. Cookson, R., *QALYs and the capability approach.* Health economics, 2005. **14**(8): p. 817-829.
7. Sen, A., *Choice, welfare and measurement*. 1982, Cambridge, MA: Harvard University Press.
8. Coast, J., et al., *Valuing the ICECAP capability index for older people.* Social science & medicine (1982), 2008. **67**(5): p. 874-882.
9. Coast, J., et al., *An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people.* Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 2008. **17**(7): p. 967-976.
10. Flynn, T.N., et al., *Rescaling quality of life values from discrete choice experiments for use as QALYs: A cautionary tale.* Population Health Metrics, 2008. **6**(Journal Article).
11. Grewal, I., et al., *Developing attributes for a generic quality of life measure for older people: Preferences or capabilities?* Social Science and Medicine, 2006. **62**(8): p. 1891-1901.
12. Macdonald, A. and B. Cooper, *Long-term care and dementia services: an impending crisis.* Age and Ageing, 2007. **36**(1): p. 16-22.
13. Kwasky, A.N., B.E. Harrison, and A.L. Whall, *Quality of Life and Dementia: An Integrated Review of Literature.* Alzheimer's Care Today, 2010. **11**(3): p. p 186-195.
14. Banerjee, S. and R. Wittenberg, *Clinical and cost effectiveness of services for early diagnosis and intervention in dementia.* International journal of geriatric psychiatry, 2009. **24**(7): p. 748-754.
15. Brod, M., et al., *Conceptualization and Measurement of Quality of Life in Dementia: The Dementia Quality of Life Instrument (DQoL).* The Gerontologist, 1999. **39**(1): p. 25-36.
16. Ready, R.E., *Quality of life measures for dementia.* Health and quality of life outcomes, 2003. **1**(11).
17. Logsdon, R.G., et al., *Quality of life in Alzheimer's disease: patient and caregiver reports.* Journal of Mental Health & Aging, 1999. **5**(1): p. 21-32.
18. Selai, C.C.E., *Assessing quality of life in dementia: Preliminary psychometric testing of the Quality of Life Assessment Schedule (QOLAS).* Neuropsychological rehabilitation, 2001. **11**(3): p. 219-243.
19. Smith, S.C., et al., *Development of a new measure of health-related quality of life for people with dementia: DEMQOL.* Psychological medicine, 2007. **37**(05): p. 737.
20. Ettema, T.P., et al., *QUALIDEM: development and evaluation of a dementia specific quality of life instrument--validation.* International journal of geriatric psychiatry, 2007. **22**(5): p. 424-430.
21. Hoe, J., et al., *Use of the QOL-AD for measuring quality of life in people with severe dementia--the LASER-AD study.* Age and Ageing, 2005. **34**(2): p. 130-135.

22. Neyens, J.C.L., et al., *A multifactorial intervention for the prevention of falls in psychogeriatric nursing home patients, a randomised controlled trial (RCT).* Age and Ageing, 2009. **38**(2): p. 194-199.
23. Pickard, A.S. and S.J. Knight, *Proxy evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives.* Medical care, 2005. **43**(5): p. 493-499.
24. Strating, M.M.H., et al., *Evaluating the Care for Better collaborative. Results of the first year of evaluation*, 2008, Institute of Health Policy and Management: Rotterdam. p. 1-132.
25. Strating, M.M.H., et al., *Creating effective quality-improvement collaboratives: a multiple case study.* BMJ Quality & Safety, 2011(Journal Article).
26. *Burgerlijke Wetboek [Civil code of the Netherlands]*, 1992, The Netherlands: The Hague. p. Article 446-Article 468.
27. Miller, F.G. and E.J. Emanuel, *Quality-improvement research and informed consent.* The New England journal of medicine, 2008. **358**(8): p. 765-767.
28. The EuroQol, G., *EuroQol--a new facility for the measurement of health-related quality of life.* Health policy (Amsterdam, Netherlands), 1990. **16**(3): p. 199-208.
29. Jonsson, L., et al., *Patient- and proxy-reported utility in Alzheimer disease using the EuroQoL.* Alzheimer Disease and Associated Disorders, 2006. **20**(1): p. 49-55.
30. Cantril, H., *The pattern of human concerns*. 1966, New Brunswick: Rutgers University Press.
31. Bengtsson-Tops, A., et al., *Subjective versus interviewer assessment of global quality of life among persons with schizophrenia living in the community: a Nordic multicentre study.* Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 2005. **14**(1): p. 221-229.
32. Veenhoven, R., *World database of Happiness*.
33. Bjelland, I., et al., *The validity of the Hospital Anxiety and Depression Scale: An updated literature review.* Journal of psychosomatic research, 2002. **52**(2): p. 69-77.
34. Carod-Artal, F.J., et al., *Self- and Proxy-Report Agreement on the Stroke Impact Scale.* Stroke, 2009. **40**(10): p. 3308-3314.
35. Spinhoven, P., et al., *A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects.* Psychological medicine, 1997. **27**(2): p. 363-370.
36. Dijkstra, A., G. Buist, and T. Dassen, *Nursing-care dependency. Development of an assessment scale for demented and mentally handicapped patients.* Scandinavian journal of caring sciences, 1996. **10**(3): p. 137-143.
37. Valimaki, M., et al., *Elderly patient's care dependency.* Hallym International Journal of Aging, 2004. **6**(1): p. 79-93.
38. Dijkstra, A., et al., *Cross-cultural psychometric testing of the Care Dependency Scale with data.* Journal of advanced nursing, 2003. **43**(2): p. 181-187.
39. Lee, K.J. and J.B. Carlin, *Multiple Imputation for Missing Data: Fully Conditional Specification Versus Multivariate Normal Imputation.* American Journal of Epidemiology, 2010. **171**(5): p. 624-632.
40. Alexopoulos, G.S., et al., *Cornell scale for depression in dementia.* Biological psychiatry, 1988. **23**(3): p. 271-284.
41. Al-Janabi, H., T. Flynn, and J. Coast, *Development of a self-report measure of capability wellbeing for adults: the ICECAP-A.* Quality of Life Research, 2012. **21**(1): p. 167-176.
42. Huizing, A.R., et al., *A Cluster-Randomized Trial of an Educational Intervention to Reduce the Use of Physical Restraints with Psychogeriatric Nursing Home Residents.* Journal of the American Geriatrics Society, 2009. **57**(7): p. 1139-1148.
43. Nussbaum, M., *The quality of life*. 1993, Oxford: Clarendon Press.

44. Robeyns, I., *The capability approach: a theoretical survey*. Journal of Human Development, 2005(6): p. 93-114.
45. Coast, J., R. Smith, and P. Lorgelly, *Should the capability approach be applied in health economics?* Health economics, 2008. **17**(6): p. 667-670.
46. Brouwer, W.B.F., et al., *Welfarism vs. extra-welfarism.* Journal of health economics, 2008. **27**(2): p. 325-338.
47. Gullone, E. and R.A. Cummins, *Universality of subjective wellbeing indicators*. Social indicators reasearch series. 2002, Dordrecht: Kluwer. 196.

Appendix 1 The ICECAP-O instrument proxy version

This appendix highlights the ICECAP-O instrument proxy version used in this study. It was based on the original ICECAP-O version, as developed by Joanna Coast and Terry Flynn, and introduced and validated in [9]. The original version can be found at http://www.icecap.bham.ac.uk/documents/icecapquest.doc

General instructions for the proxy questionnaire:

We would like to ask you to fill out the questions below for the client/family member. Please try to answer the questions in manner as the client/family member would if he/she would be able to answer the questions. With every question, please tick the answer that the client/family member would give.

ABOUT THE CLIENT'S QUALITY OF LIFE

By placing a tick (✓) in ONE box in EACH group below, please indicate which statement best describes the clients quality of life at the moment.

1. Love and Friendship		
The client can have all of the love and friendship that he/she wants		4
The client can have a lot of the love and friendship that he/she want		3
The client can have a little of the love and friendship that he/she wants		2
The client cannot have any of the love and friendship that he/she wants		1

2. Thinking about the future		
The client can think about the future without any concern		4
The client can think about the future with only a little concern		3
The client can only think about the future with some concern		2
The client can only think about the future with a lot of concern		1

3. Doing things that make you feel valued		
The client is able to do all of the things that make him/her feel valued		4
The client is able to do many of the things that make him/her feel valued		3
The client is able to do a few of the things that make him/her feel valued		2
The client is unable to do any of the things that make him/her feel valued		1

4. Enjoyment and pleasure		
The client can have all of the enjoyment and pleasure that he/she wants		4
The client can have a lot of the enjoyment and pleasure that he/she wants		3
The client can have a little of the enjoyment and pleasure that he/she wants		2
The client cannot have any of the enjoyment and pleasure that he/she wants		1

5. Independence		
The client is able to be completely independent		4
The client is able to be independent in many things		3
The client is able to be independent in a few things		2
The client is unable to be at all independent		1

Tick one box only in each section

Tariffs and scoring algorithm in STATA

The ICECAP-O has 5 attribute dimensions each having 4 levels, 44444 representing full capability and 11111 representing no capability. The tariffs for the ICECAP-O, based on the preferences of the 65+ population in the United Kingdom were presented by Coast et al. [8]. The tariffs basically provide preference weights for the different ICECAP-O states, which are normalised in such a way that 0 represents the worst situation described on the ICECAP-O (11111) and 1 represents the best situation described on the ICECAP-O. Lower scores thus represent fewer, preference based, capabilities.

The complete explanation as to how to calculate them is fully described on the ICECAP website: http://www.icecap.bham.ac.uk/tariffs.shtml.

Chapter 6. Quality of Life of nursing home residents with dementia: Validation of the German version of the ICECAP-O

This chapter is based on Peter Makai,* Franziska Beckebans*, Job van Exel and Werner Brouwer Quality of life of nursing home residents with dementia: validation of the German version of the ICECAP-O *PloS One(Accepted for publication)*

* Joint first authorship

Abstract

Objectives

To validate the ICECAP-O's German translation in people with dementia living in a nursing home, and investigate the influence of proxy characteristics on response.

Method

For 95 residents living in a German nursing home, questionnaires were completed by nursing professionals serving as proxy respondents. We investigated the convergent validity of the ICECAP-O with other generic Qol measures (EQ-5D+C and ADRQL) as well as with a measure of Activities of Daily Living (ADL) (Barthel-index). Discriminant validity was investigated by comparing the ICECAP-O scores in various subgroups of dementia severity using a measure based on the Mini Mental Score Examination (MMSE), a measure of care dependency, ADL status and demographic characteristics using bivariate tests and multivariate stepwise regression analysis.

Results

Convergent validity between the ICECAP-O, EQ-5D+C, ADRQL and Barthel-Index scores was moderate to good, but differed considerably between dimensions of the instruments. Discriminant validity was confirmed by investigating subgroups based on ADL scores and other characteristics. The ICECAP-O scores based on available tariffs were related to proxy characteristics (gender and work experience).

Discussion

The results of this pilot study in Germany suggest that the ICECAP-O is a promising generic measure for Qol of people with dementia living in a nursing home. Validity tests generally yielded favorable results. Work experience and gender appeared to influence proxy response, which raises questions regarding appropriate proxies especially since the ICECAP-O may be completed by proxies relatively often. Research is necessary to further validate the German version of the ICECAP-O, with specific attention for proxy completion in people with dementia.

Introduction

Growing life expectancy leads to higher numbers of people with dementia due to increasing risk of incidence of dementia with age [1]. Most people with dementia initially receive informal care at home, but with the progression of the disease, the amount of professional care typically increases. Frequently, a sufficient amount and quality of professional care can only be provided in an institutional long-term care setting in advanced stages of the disease, making admissions inevitable for a growing number of people with dementia [2]. Faced with increasing demand, the long-term care sector in many countries may experience strong economic pressure, raising questions of optimal resource allocation and affordability of care.

Economic evaluation has traditionally assisted allocation decisions by integrally measuring health status and mortality using the QALY (Quality Adjusted Life Year) concept. In QALY calculations values (often referred to as utility scores) are assigned to different health states, which allow quantifying health gains in terms of length and quality of life from interventions [3]. These health states are commonly measured using Health-Related Quality of Life (HrQol) instruments, which are used for computing the quality adjusted component of QALYs, making HrQol instruments an essential outcome measure for economic evaluation. Measurement of HrQol is important for chronic diseases such as dementia, which impair the quality of life of affected patients in addition to their length of life [4]. HrQol is most commonly measured with the EQ-5D [5]. Economic evaluation is increasingly used in the curative sector as a decision support tool for resource allocation, but may aid the allocation of resources in the long-term care as well [5] [6] [7].

However, quality of life of individuals does not only depend on generic assessed HrQol, as for instance measured by the EQ-5D, but also depends on other dimensions [3]. This is important in the context of economic evaluations when interventions do not (only) affect HrQol but also these other factors of overall quality of life. For example, people with dementia living in nursing homes may have less contact with their family members, which may reduce their feelings of attachment. Additionally, people with dementia forget where they are, lose their sense of time or do not recognize their own family members [8] [9], which may lead to a decreased sense of control, and may inhibit their feeling of being valued. Therefore, to ensure a sense of accomplishment and independence for people with dementia, other activities matching their abilities and remaining resources are offered in nursing homes, for example through providing engaging activities [10] [11]. Such activities do not necessarily lead to an improvement in health but will improve nursing home residents Qol more broadly by increasing their enjoyment of life, feeling of control and may

contribute to a feeling of being valued. HrQol instruments like the EQ-5D are therefore not sufficient for a complete economic evaluation in the long-term care.

In order to be able to perform a complete economic evaluation the full benefit of the evaluated intervention or service should be measured. For this purpose, broader Qol measures, often termed wellbeing measures should be used which capture more facets of people's lives rather than health status alone. A recently developed wellbeing instrument, the ICECAP-O (ICEpop CAPability measure for Older people), aims to incorporate such aspects beyond health [12] [13]. These broader wellbeing aspects are captured through the notion of capabilities. The capacity to live life the way one desires is obviously important, also to older people, and reductions of this capability limit their wellbeing [14] [15]. Derived from Sen's capabilities approach [15], the instrument was originally developed to provide a set of general capability values for use in economic evaluations for people above 65 in the UK. Previous validation studies confirmed that the ICECAP-O evaluates a spectrum of outcomes beyond HrQol [16] [12]. So far, the ICECAP-O has been used in the general UK population, and in the Netherlands a proxy version has been used in psycho-geriatric elderly nursing homes [16] [12].

Measuring HrQol and wellbeing in elderly suffering from dementia raises special challenges. At the stage of intermediate and advanced dementia the disease affects cognitive abilities and people lack the capacity of self-completing questionnaires (even in an interview setting) due to loss of memory, attention and language [17]. For all instruments in this study, we therefore used the proxy-report as suggested in the literature among people with moderate to severe levels of cognitive disorders [17] [18] [19] [20] [21] [22]. The choice of proxy may influence response, as professional and family proxies respond differently to HrQol and wellbeing questionnaires in general [22] [23] and specifically for the ICECAP-O [12]. In case of psycho-geriatric residents, nursing professionals may be recommended as proxy to complete the ICECAP-O [12]. However, the exact influence of specific respondent characteristics beyond being a family member or a professional caregiver on the ICECAP-O is unknown.

Measuring wellbeing is important in the German long-term care as well. Around 1.3 million Germans suffer from dementia and this figure is expected to reach almost 2 million by 2040 [24]. In addition, institutionalization of people with dementia is quite common in the German context [25] [9]. About 60% of nursing home residents in Germany suffer from dementia and require appropriate care [25].

The aim of this study was to investigate the convergent and discriminant validity of the ICECAP-O in a population of elderly with dementia living in a nursing home. Furthermore, we wish to explore whether proxy characteristics influence response.

Methods

Setting, study population and data collection

The study was conducted in two separate sites of a specialized nursing facility for dementia patients between May and August 2011 in North Rhine-Westphalia, Germany. The sample size consisted of 95 residents diagnosed with dementia, who were older than 55 and had been living in the nursing home for longer than two months. Nurses were selected as proxy respondents if they were primary caregivers. This was defined as who had the most experience with taking care of particular residents and were involved in their care at least four times a week. In total, 11 nurses completed between 4 and 20 written questionnaires. Informed consent was obtained from legal guardians for all 95 residents. To ensure privacy, the researchers did not see the name list of the residents at any time in the study.

Measures

Dementia status

Dementia status was measured using the general practitioner's diagnosis: type according to the ICD-10 (F00.-, F01.- or F02.-) [26] and severity according to the German guideline for dementia [27]. This classification is based on the Mini Mental Score Examination (MMSE), with mild dementia corresponding to MMSE scores between 20 and 26, moderate dementia corresponding to MMSE scores from 10 to 19 and Severe dementia corresponds to MMSE scores below 10 [27]. Furthermore, care dependency (1 lowest /2 medium/3 high dependency) was measured using the care-level classification of the German National Association of Statutory Health Insurance Funds [28]. According to this classification, people in care level 1 need help once a day in some ADL activities, people in care level 2 need help three times a day, while people in care level 3 need continuous nursing care [28].

Wellbeing

The ICECAP-O has five attributes (attachment, security, role, enjoyment and control) each with four levels, thus distinguishing 1,024 wellbeing states in total [3] [16]. In order to obtain tariffs for the ICECAP-O, the attributes were valued using best-worst scaling, a special type of discrete choice analysis [3]. The ICECAP-O tariffs have values between 0 (no capability) and 1 (full capability). In this

study British tariffs were applied as German tariffs are lacking. For this first use of the ICECAP-O in Germany, the questionnaire was forward-backward translated from English into German by two independent translators.

Health-related Quality of life

We used the revised 40-Item version of the Alzheimer Disease Related Quality of Life (ADQRL) instrument, which allows for the assessment of Qol for people at intermediate or late-stage dementia using proxy response [17] [23] [29] [30] [31]. The dementia-specific, multi-dimensional ADRQL instrument can be completed by family or professional caregivers [29] [4] [32] [33]. The ADRQL measures the dimensions Social Interaction, Awareness of Self, Enjoyment of Activities, Feelings and Mood and Response to Surrounding [32]. The various dimensions range from 4 to 12 items on a dichotomous scale and each item is weighted in a range between 9.15 and 13.75, based on a judgment of importance by caregivers [34]. For each dimension a separate subscale can be calculated and summed up in one total score ranging from 0 (lowest quality of life) to 100 (highest quality of life) [35]. The instrument exhibits good psychometric properties having adequate validity, good internal-consistency reliability, very low missing data and good sensitivity to change [36] [37]. The authorized German edition of the ADRQL was used [33].

The EQ-5D as developed by the EuroQol group is a common instrument to measure generic HrQol [38]. The EQ-5D measures five dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) on three levels (no problems, some problems, extreme problems) [38] [5], describing 243 health states. The EQ-5D health states can be converted to a utility score by applying the German EQ-5D index, based on TTO values [39] [40]. The EQ-5D utility scores range from 1 (perfect health) through 0 (dead) and has negative values accounting for health states worse than dead. For use in people with dementia, the EQ-5D was extended with a cognitive dimension, for which utility scores are unavailable [41] [42]. In this study the official German proxy version 2 of the EQ-5D was used [43] and a German translation of the question pertaining to the cognitive dimension was added.

Activities of daily living

The Barthel-Index is a well-established instrument that measures residents' ability to perform activities of daily living (ADL) by proxy- or self-report. Decrease in ADL is one of the visible manifestations of dementia, and the subsequent loss of independence [44]. The ADL-score is mainly used in geriatric fields and is a strong predictor of Qol scores across several outcome measurements, including the ADRQL [23] [45]. The Barthel-Index includes items such as personal

care and moving from wheelchair to bed and back, measured on two to four levels depending on the item. The available scores per question are 0 and 5 for two-level items, 0, 5, and 10 for three-level items and 0, 5, 10 and 15 for four level items, ranging from inability to independence. The total achievable score ranges between 0 (completely dependent) and 100 (completely independent) [46] [47] with a cutoff score of 65 indicating need for ADL assistance [48]. In this study the validated German version was used [49].

Patient and proxy characteristics

The questionnaire contained questions on patient's age, sex, marital status, length of stay in the nursing home, and frequency of visits by family members. Finally, the questionnaire contained questions on age, role, work experience and length of time the nurse selected as proxy respondent knew the resident, since previous studies have shown that proxy characteristics may influence responses [22] [35].

Hypotheses

To establish convergent validity we expected moderate to strong and positive correlations between the ICECAP-O, the EQ-5D and ADRQL because all of these instruments measure operationalizations of Qol (H1). Furthermore, we expected a moderate and positive correlation between the ICECAP-O dimensions, tariffs and the Barthel-index (H2).

For discriminant validity we expected to find differences in terms of ICECAP-O scores between residents suffering from severe and mild/moderate dementia (based on the MMSE), between ADL dependent (Barthel-score <65) and ADL independent (Barthel score ≥ 65) residents, between different care dependency groups and between older (75+) and younger residents (H3). A higher score on the ICECAP-O was expected for the better-off groups.

We expected that the proxy characteristics function (leading/non-leading), work experience (more or less than 2 years) and time knowing the resident (more or less than a year) would influence response on the ICECAP-O instrument (H4).

Data analysis

Descriptive statistics of resident and proxy characteristics were calculated. Correlations between the outcomes of the ICECAP-O and dimensions of the ADQRL, EQ-5D and the ADL were used to estimate the convergent validity. Correlations above 0.5 are referred to as strong, between 0.3 and 0.5 as moderate, and correlations below 0.3 are considered weak [50]. Discriminant validity was

analyzed using T-test and one-way-ANOVA to explore differences in means of the ICECAP-O between different demographic and dementia-related groups. Discriminant validity was also examined using two stepwise multivariate regressions, in the first model controlling for demographic variables, and in the second model for proxy characteristics as well. For the stepwise analyses we used a cutoff of 0.1 for entering variables. There was no missing data, so there was no need to correct for this in the study. For all analyses the level of significance was $p < 0.05$. Data was analyzed using STATA 11.

Results

Descriptive characteristics

Descriptive statistics of the 95 residents and the proxies are presented in Table 6.1. Average age of the residents was 77 years, with 54% being female and 55% of residents living in the nursing home for more than 2 years. 60% had Alzheimer's dementia, and dementia severity could be categorized as severe in 60% of the cases. The majority of the residents (56%) had visitors less than once a week. As for the characteristics of the proxy respondents, the majority of the proxy respondents were female, and they, on average had worked at the nursing home for a 3.5 years. Figure 6.1 illustrates the response of the ICECAP-O. On most dimensions, the majority of the residents had at least some deficits in terms of capabilities.

Figure 6.1: Response on the ICECAP-O

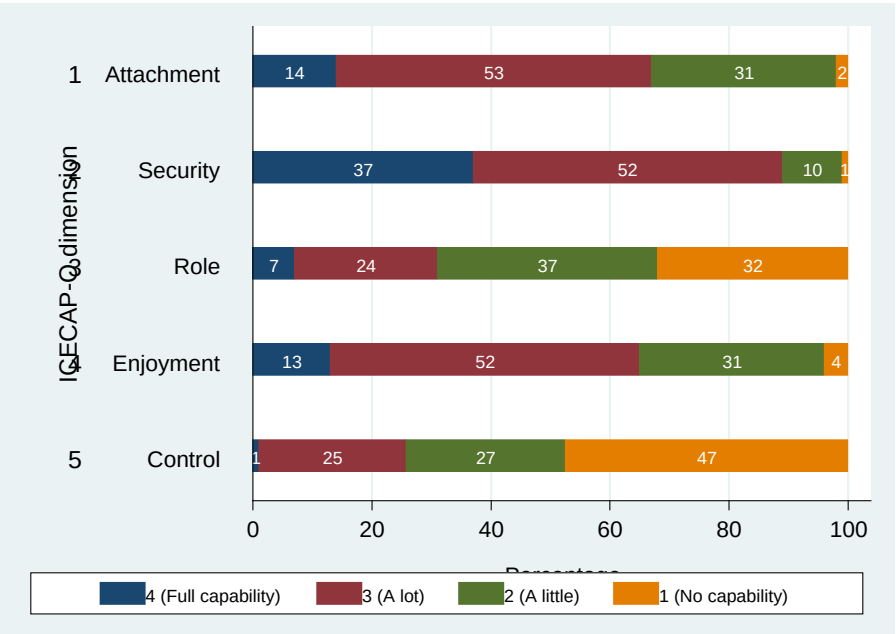


Table 6.1: Demographic characteristics of residents and proxy's (n=95)

Variable		*statistic*
Resident characteristics		
Age		76.7 (8.5)
Sex (female)		56.8%
Type of dementia (Alzheimer's)		60.0%
Dementia Severity	Mild	5.3%
	Moderate	34.7%
	Severe	60.0%
Length of stay in nursing home	0 ≤ 6 months	8.4%
	6 ≤ 12 months	13.7%
	12 ≤ 24 months	23.2%
	> 24 months	54.7%
Marital Status	Unmarried	21.1%
	Married	23.2%
	Divorced	18,9.%
	Widowed	36.8%
Frequency of visits by family members	once a week or more	39.9%
	less than once a week	55.2%
	never	4.9%
Care Level	Level 1 (Low)	15.8%
	Level 2 (Medium)	33.7%
	Level 3 (High)	50.5%
Proxy characteristics		
Age		44.8 (11.5)
Sex (female)		87.0%
Working time (months)		43.4 (32.2)
Leading function		47.4%
Time knowing the resident (months)		19.2 (19.5)

Table 6.2: Description of measurement instruments (n=95)

Instrument			*Mean (SD)*	*Median*
ICECAP-O	- tariffs		0.63 (0.20)	
	- dimension scores	Attachment	2.79 (0.70)	3
		Security	3.24 (0.68)	3
		Role	2.07 (0.92)	2
		Enjoyment	2.73 (0.74)	3
		Control	1.78 (0.83)	2
Barthel-Index (ADL)	- score		41.18 (30.65)	
	- need for ADL assistance (73.7% with score <65)		27.21 (22.13)	
EQ-5D (+C)	- utilities		0.52 (0.34)	
	- dimension scores	Mobility	1.78 (0.87)	1
		Self-Care	2.52 (0.62)	3
		Usual activities	2.51 (0.56)	3
		Pain/Discomfort	1.35 (0.54)	1
		Anxiety/Depression	1.17 (0.43)	1
		Cognition (C)	2.69 (0.46)	3
ADRQL	- tariffs		70.36 (15.69)	
	- dimension scores	Social Interaction (SI)	73.64 (26.63)	
		Awareness of Self (AS)	47.29 (28.19)	
		Feelings and Mood (FM)	83.83 (17.69)	
		Enjoyment of Activities (EA)	50.17 (28.69)	
		Response to Surroundings (RS)	90.56 (17.12)	

Table 6.2 describes the dimensions and tariffs of the measurement instruments used. The overall average scores for the instruments were as follows: average ICECAP-O score (based on the tariffs) was 0.63, EQ-5D score was 0.53, and the ADRQL score (based on tariffs) was 70.36.

Convergent validity

Table 6.3 shows that the ICECAP-O scores were strongly correlated with EQ-5D scores, ADRQL scores and Barthel scores. Correlations between the ICECAP-O tariffs and the different dimensions of the EQ-5D+C were generally strong and significant, except for the EQ-5D+C dimensions "pain" and "anxiety". Correlations between the ICECAP-O and the ADRQL proved to be similarly strong and significant, with the exception of the ADRQOL dimensions "Feeling and Mood" (FM) and "Response to the Surroundings" (RS). The individual ICECAP-O dimensions Role and Control were strongly and significantly correlated with the EQ-5D+C dimensions mobility, self-care, usual activities and cognition. Role was also significantly and strongly correlated with AS (ADRQL). The Barthel index was significantly correlated with all ICECAP-O dimensions except for security, with correlations between the Barthel index and the role and control dimensions being particularly strong.

Discriminant validity

Table 6.4 shows the means of the ICECAP-O tariffs in various subgroups defined by resident and proxy characteristics. The results of the t-tests for the ICECAP showed significant differences in ICECAP scores between patients with different dementia severity (mild/moderate, severe), ADL scores (<65, ≥ 65) and ages (i.e., above or below 75). ANOVA results showed that the ICECAP-O tariffs differentiated between residents classified into different care dependency levels. As expected, lower scores were observed for the more severe groups, and higher for the less severe groups. Additionally, the ICECAP-O tariffs varied with two proxy characteristics: gender and work experience.

Table 6.4 also shows the discriminant validity of the ICECAP-O scores in a multivariate analysis. A relatively weak, but significant association was observed between the ICECAP-O scores and ADL scores in both the model with only patient characteristics (analysis not shown) and in the model also including proxy characteristics. ADL coefficients, standard deviations and p-values were identical in both models. From the proxy characteristics, nurses' gender and work experience were associated with the ICECAP-O scores.

Table 6.3: Convergent validity (n=95)

			ICECAP tariff	*ICECAP dimension scores*				
				Attachment	**Security**	**Role**	**Enjoyment**	**Control**
Barthel-Index (ADL)	- score		0.72**	0.25*	0.04	0.72**	0.40**	0.69**
EQ-5D	- utilities		0.69**	0.21*	-0.03	0.69**	0.35**	0.67**
	- dimension scores (+C)	Mobility	0.64**	0.17	0.00	0.59**	0.33**	0.60**
		Self-Care	0.61**	0.24*	-0.05	0.63**	0.27**	0.61**
		Usual activities	0.56**	0.24*	-0.09	0.54**	0.30**	0.58**
		Pain/Discomfort	0.23	-0.05	0.09	0.26**	0.13	0.25*
		Anxiety/Depression	0.16	0.27**	0.21*	0.09	0.26**	0.01
		Cognition (C)	0.48**	0.14	-0.07	0.52**	0.25*	0.54**
ADRQL	- overall		0.53**	0.48**	0.04	0.49**	0.56**	0.30**
	- dimension scores	Social Interaction (SI)	0.39**	0.43**	0.09	0.32**	0.46**	0.18
		Awareness of Self (AS)	0.56**	0.38**	-0.20	0.59**	0.43**	0.47**
		Feelings and Mood (FM)	0.13	0.14	0.15	0.10	0.27**	0.00
		Enjoyment of Activities (EA)	0.37**	0.28**	-0.09	0.34**	0.27**	0.23*
		Response to Surroundings (RS)	0.04	0.16	0.30**	0.02	0.15	-0.08

Note: * significance on the 5% level; ** significance on the 1% level

Table 6.4: Discriminant validity ICECAP-O tariff

Variable	*Level*	*Bivariate analysis*			*Stepwise multivariate analysis*		
		Mean	**SD**	**P-value**	**Beta**	**SD**	**P-value**
Resident characteristics							
Age	75+	0.59	0.19	0.007			
	55-75	0.69	0.20				
Gender	Female	0.64	0.17	0.446			
	Male	0.63	0.23				
Dementia type	Alzheimer	0.61	0.19	0.103			
	Other	0.67	0.21				
Time in nursing home	<12 months	0.67	0.19	0.209			
	>12 months	0.62	0.23				
Marital Status	Married	0.65	0.22	0.400			
	Not married [a]	0.63	0.19				
Visits	Once a week or more	0.63	0.19	0.361			
	Less than once a week	0.64	0.21				
Care level	Low	0.80	0.17	0.000			
	Medium	0.70	0.17				
	High	0.54	0.17				
Dementia severity	Mild/moderate	0.78	0.19	0.000			
	Severe	0.54	0.12				
ADL	Below 65	0.58	0.19	0.000	0.00	0.00	0.000
	Above 65	0.80	0.12				
Proxy characteristic							
Gender	Male	0.52	0.20	0.010	0.09	0.04	0.041
	Female	0.65	0.15				
Work experience	Less than two years	0.61	0.21	0.049	0.00	0.01	0.012

	More than two years	0.68	0.18		
Months knowing resident	Less than 12 months	0.63	0.21	0.474	
	Longer than 12 months	0.64	0.20		
Function	Leading	0.63	0.21	0.432	
	Non-leading	0.64	0.18		
R-square					0.55

Note: [a] Unmarried, Divorced, Widowed.

Discussion

Main results

In this study the ICECAP-O was used and validated for the first time in Germany, in a specialized nursing home for dementia patients. Our results indicate that the ICECAP-O has good convergent validity. Hypotheses were supported by the significant and strong correlation between the ICECAP-O scores and HrQol scores (both EQ-5D and ADRQL scores) (H1)as well as between ICECAP-O scores and ADL scores (H2). Moreover, as hypothesized (H3), the ICECAP-O significantly discriminated between dementia severity (mild/moderate and severe), ADL-status (<65; ≥ 65), care level (low/middle/high) and between residents younger and older than 75 years, supporting discriminant validity. In the stepwise multivariate model, the ICECAP-O discriminated between nursing home residents with different ADL status. The exploration of the relationship between the proxy responses on the ICECAP-O showed a significant influence of proxy characteristics on the ICECAP-O scores (confirming H4).

Methodological limitations

Some limitations of this study deserve mentioning. First, residents all lived in two sites of the same nursing home facility and were not randomly selected; therefore they might have characteristics that differ from the typical population with dementia in German nursing homes. Hence, the results presented here are not necessarily representative nor generalizable. However, the focus of the study was the validation of the properties of a wellbeing instrument in relation to various HrQol instruments. For that purpose, the current sample seems adequate. Obviously, confirmation of these findings in other samples and settings remains important.

Second, the sample size was relatively limited. Hence, also in light of the promising results reported here, further research in larger samples is encouraged. Specific attention should also be paid in future research to the influence of proxy characteristics.

Third, nursing proxies completed varying numbers of questionnaires, which may have influenced our results. However, due to sample size considerations this could not be investigated in detail. A fourth limitation is that only nursing proxies were used, while family proxies were not approached. Family members or spouses may assess resident's Qol differently on the ICECAP-O than nurses do [12]. Nurses care for the residents on a day to day basis and thus have more contact with the residents than family members (frequently observing physical and mental conditions of patients, not only during visiting hours). Therefore, as suggested previously [12], in this care setting the nurse as proxy respondent seems to be the logical choice.

Finally, since German tariffs for the ICECAP-O were not available, British tariffs were used in this study. Although preference weights for capability dimensions may vary between countries, it is questionable whether using German tariffs (if available) would have led to different results regarding the validity of the ICECAP-O instrument. At the time of the study no ADRQL tariffs were available for Germany either, therefore we used the official American tariffs [34]. In order to investigate possible cultural effects on the valuation of the ADRQL we performed a sensitivity analysis (results not shown) using weights from the German-speaking region of Switzerland, obtained in a pilot study [33]. Using these ADRQL weights in the sensitivity analysis did not yield different results.

Convergent and discriminant validity

The strong correlation between the ICECAP-O and EQ-5D scores shows that generic HrQol is captured to a wide extent by the ICECAP-O, which is consistent with other findings [16] [12] [51] [52]. The results also confirmed the expected significant correlation between the ICECAP-O scores and the ADRQL scores, which shows that the ICECAP-O captures both generic HrQol and dementia specific Qol. Additionally, the correlation between ADL-scores and the ICECAP-O scores reflected that a loss of independence in ADL was associated with a decline in wellbeing. Decreased ADL was also associated with lower scores on HrQol instruments, confirming previous results that reduced ADL leads to a decrease in Qol [44]. Overall, these significant findings point in the direction of favorable convergent validity.

The ICECAP-O discriminated between patients based on the variables age, dementia severity, care dependency and ADL. This suggests that the ICECAP-O is sensitive to age differences indicators of health. In a multivariate setting, ICECAP-O scores were only significantly influenced by ADL, while dementia severity, care dependency and age were not significant. A possible explanation for this may be that ADL, dementia severity [53] and care dependency are related, while Qol is not necessarily determined by biological age. Dementia severity is one explanatory variable for the ADL-status [53], which in turn determines care-dependency [28]. Another explanation for this finding may be a lack of power to detect all existing relationships between relevant variables. Indeed, in the univariate analysis ICECAP-O scores varied with different dementia severity and ADL status, supporting discriminant validity of the German version of the ICECAP-O.

Influences of proxy characteristics

That the choice of professional or family proxy matters in the measurement of Qol has already been observed in other studies [22] [12] [18] [54]. Specific proxy characteristics such as gender or work experience were not examined previously, especially not in relation to the ICECAP-O. Our results suggest that nurses' gender and work experience influence their response on the ICECAP-O scores.

Controlling for residents' characteristics, proxy gender and work experience were related to the ICECAP-O scores. In absence of a golden standard, it is difficult to judge which proxies provided the most accurate description of residents' Qol. It may be hypothesized that in assessing Qol, nurses benefit from more experience with caring for dementia patients. Male nurses assessed residents Qol significantly higher than female nurses did, controlling for ADL status of residents. This difference may either be due to the small number of questionnaires answered by male nurses, or by a genuine gender difference in assessing residents' Qol. The relationships between other proxy characteristics and proxy responses should be explored further in larger samples in future research. Although a golden standard for the resident population included in this study is difficult to obtain, by comparing scores of proxies to those of patients obtained in early stages of dementia, one may perhaps shed more light on accuracy of QoL assessment of different groups of proxies.

Conclusion

The German version of the ICECAP-O was used for the first time in this study and appeared to be a reliable wellbeing measurement instrument showing good convergent and discriminant validity for people with dementia. The influence of proxy characteristics like gender and work experience suggests potentially fruitful avenues of further research in determining the influence of proxy

characteristics on response. In order to further investigate the findings of this study, additional validation studies using larger samples and in different settings are required and encouraged.

Validating the ICECAP-O as a generic wellbeing instrument which has the capacity to capture broader outcomes might contribute to enabling economic evaluation of long-term care services and interventions, also in Germany. This seems to be especially relevant for informed decisions in the long-term care sector where an increase in healthcare spending is expected due to the growing number of elderly with dementia. In such a setting, appropriately measuring the potential benefits of care and comparing them to the costs is pivotal for optimal healthcare provision. By capturing the relevant outcomes in long-term care, the ICECAP-O seems to be a suitable wellbeing instrument for residents with dementia, though further validation work is encouraged.

References

1. McCullagh, C.D., et al., *Risk factors for dementia.* Advances in Psychiatric Treatment, 2001. **7**(1): p. 24-31.
2. Verbeek, H., et al., *Small-scale, homelike facilities versus regular psychogeriatric nursing home wards: a cross-sectional study into residents' characteristics.* BMC Health Services Research, 2010. **10**(1): p. 30.
3. Coast, J., et al., *Valuing the ICECAP capability index for older people.* Social science & medicine (1982), 2008. **67**(5): p. 874-882.
4. Ettema, T.P., et al., *A review of quality of life instruments used in dementia.* Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 2005. **14**(3): p. 675-686.
5. Drummond, M.F., et al., *Methods for the economic evaluation of health care programmes*. Vol. Third Edition. 2005, Oxford: Oxford University Press.
6. Hächer J, R.B., *Die Pflegeversicherung in der Krise - Renditen, Leistungsniveau und Versorgungslücken*. 2008, Köln: Deutsches Institut für Altersversorge.
7. Jonker, C., *A model for quality of life measures in patients with dementia: Lawton's next step.* Dementia and geriatric cognitive disorders, 2004. **18**(2): p. 159.
8. Robert Koch Institut in Zusammenarbeit mit dem statistischen Bundesamt, e., *Gesundheitsberichtserstattung des Bundes - Alterdemenz*, 2005.
9. Bowling, A., *Lay theories of quality of life in older age.* Ageing and society, 2007. **27**(6): p. 827.
10. Cohen-Mansfield, J., et al., *Can persons with dementia be engaged with stimuli?* The American journal of geriatric psychiatry: official journal of the American Association for Geriatric Psychiatry, 2010. **18**(4): p. 351.
11. Cohen-Mansfield, J., M. Dakheel-Ali, and M.S. Marx, *Engagement in persons with dementia: the concept and its measurement.* The American journal of geriatric psychiatry: official journal of the American Association for Geriatric Psychiatry, 2009. **17**(4): p. 299.
12. Makai, P., *Capabilities and quality of life in Dutch psycho-geriatric nursing homes: an exploratory study using a proxy version of the ICECAP-O.* Quality of life research, 2012. **21**(5): p. 801-812.
13. Al-Janabi, H., T. N Flynn, and J. Coast, *Development of a self-report measure of capability wellbeing for adults: the ICECAP-A.* Quality of Life Research, 2011(Journal Article).
14. Grewal, I., et al., *Developing attributes for a generic quality of life measure for older people: Preferences or capabilities?* Social Science and Medicine, 2006. **62**(8): p. 1891-1901.
15. Sen, A., *Choice, welfare and measurement*. 1982, Cambridge, MA: Harvard University Press.
16. Coast, J., et al., *An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people.* Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 2008. **17**(7): p. 967-976.
17. Rabins, P.V., *Measuring quality of life in dementia: purposes, goals, challenges and progress.* International psychogeriatrics, 2007. **19**(03): p. 401.
18. Pickard, A.S. and S.J. Knight, *Proxy evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives.* Medical care, 2005. **43**(5): p. 493-499.
19. Coucill, W., et al., *EQ-5D in patients with dementia: An investigation of inter-rater agreement.* Medical care, 2001. **39**(8): p. 760-771.
20. Ankri, J.J., *Use of the EQ-5D among patients suffering from dementia.* Journal of clinical epidemiology, 2003. **56**(11): p. 1055-1063.

21. Novella, J.L., et al., *Agreement between patients' and proxies' reports of quality of life in Alzheimer's disease.* Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 2001. **10**(5): p. 443-452.
22. Bryan, S., et al., *Proxy completion of EQ-5D in patients with dementia.* Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 2005. **14**(1): p. 107-118.
23. Kasper, J.D., et al., *Evaluation of the validity and reliability of the Alzheimer Disease-related Quality of Life Assessment Instrument.* Alzheimer Disease and Associated Disorders, 2009. **23**(3): p. 275-284.
24. Bundesministerium für, G., *Demenz: Eine Herausforderung für die Gesellschaft*.
25. Weyerer, S.S., *Die Versorgung dementer Patienten in Deutschland ausepidemiologischer Sicht.* Zeitschrift für Gerontopsychologie & -psychiatrie, 2004. **17**(1): p. 41-50.
26. Deutsches Institut für Medizinische Dokumentation und, I., *ICD-10-GM Vorabversion 2012*.
27. Deutsche Gesellschaft für Neurologie, D.G.f.P.P.u.N., *S3- Leitlinie - "Demenzen" (Kurzversion)*, 2009.
28. GKV-Spitzenverband, *Richtlinien des GKV-Spitzenverbandes zur Begutachtung von Pflegebedürftigkeit nach dem XI. Buch des Sozialgesetzbuches*, 2009: Berlin.
29. Brod, M., et al., *Conceptualization and Measurement of Quality of Life in Dementia: The Dementia Quality of Life Instrument (DQoL).* The Gerontologist, 1999. **39**(1): p. 25-36.
30. Ready, R.E. and B.R. Ott, *Integrating patient and informant reports on the Cornell-Brown Quality-of-Life Scale.* American Journal of Alzheimer's Disease and Other Dementias, 2008. **22**(6): p. 528-534.
31. Missotten, P., *Relationship between quality of life and cognitive decline in dementia.* Dementia and geriatric cognitive disorders, 2008. **25**(6): p. 564.
32. Rabins, P.V., et al., *Conxepts and Methods in the Development of the ADRQLD: an Instrument for Asessing Health-Related Quality of Life in Persons with Altzheimer Disease.* Journal of Mental Health and Ageing, 1999. **5**(1): p. 33.
33. Menzi-Kuhn, C., *Lebensqualität von Menschen mit Demenz in stationären Langzeitpflegeeinrichtungen*, 2006, University of Maastricht: Maastricht.
34. Black, B.S., P.V. Rabins, and J.D. Kasper, *Alzheimer Disease Related Quality of Life User's Manual*. Vol. 2. 2009, Baltimore, Maryland.
35. Sloane, P.P.D., *Evaluating the Quality of Life of Long-Term Care Residents With Dementia.* The Gerontologist, 2005. **45**(supplement 1): p. 37-49.
36. González-Salvador, T., et al., *Quality of life in dementia patients in long-term care.* International journal of geriatric psychiatry, 2000. **15**(2): p. 181-189.
37. Lyketsos, C.C.G., *A follow-up study of change in quality of life among persons with dementia residing in a long-term care facility.* International journal of geriatric psychiatry, 2003. **18**(4): p. 275-281.
38. The EuroQol, G., *EuroQol--a new facility for the measurement of health-related quality of life.* Health policy (Amsterdam, Netherlands), 1990. **16**(3): p. 199-208.
39. Greiner, W.W., *Validating the EQ-5D with time trade off for the German population.* The European journal of health economics, 2005. **6**(2): p. 124-130.
40. Kunz, S.S., *Psychometric properties of the EQ-5D in a study of people with mild to moderate dementia.* Quality of life research, 2010. **19**(3): p. 425-434.
41. Wolfs, C.A., et al., *Performance of the EQ-5D and the EQ-5D+C in elderly patients with cognitive impairments.* Health and quality of life outcomes, 2007. **5**(Journal Article): p. 33.

42. Krabbe, P.P.F.M., *The Effect of Adding a Cognitive Dimension to the EuroQol Multiattribute Health-Status Classification System.* Journal of clinical epidemiology, 1999. **52**(4): p. 293-301.
43. Greiner, W. and C. Claes, *Der EQ-5D der EuroQol-Gruppe*, in *Gesundheitsökonomische Evaluationen*, O. Schöffski and J.M.G. Schulenburg, Editors. 2007, Springer Berlin Heidelberg. p. 403-414.
44. Andersen, C.C.K., *Ability to perform activities of daily living is the main factor affecting quality of life in patients with dementia.* Health and quality of life outcomes, 2004. **2**(1): p. 52.
45. Kwasky, A.N., B.E. Harrison, and A.L. Whall, *Quality of Life and Dementia: An Integrated Review of Literature.* Alzheimer's Care Today, 2010. **11**(3): p. p 186-195.
46. Mahoney, F.I. and D.W. Barthel, *Functional Evaluation: the Barthel Index.* Md State Med J, 1965. **14**(Journal Article): p. 61-65.
47. Collin, C.C., *The Barthel ADL Index: A reliability study.* Disability and rehabilitation, 1988. **10**(2): p. 61-63.
48. Berliner Schlaganfall, A., *Medizinische Rehabilitation nach Schlaganfall*.
49. Heuschmann, P.P.U., *Untersuchung der Reliabilität der deutschen Version des Barthel-Index sowie Entwicklung einer postalischen und telefonischen Fassung für den Einsatz bei Schlaganfall-Patienten.* Fortschritte der Neurologie, Psychiatrie, 2005. **73**(2): p. 74-82.
50. Cohen, J., *Set Correlation and Contingency Tables.* Applied Psychological Measurement, 1988. **12**(4): p. 425-434.
51. Couzner, L., J. Ratcliffe, and M. Crotty, *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, 2012. **10**(1): p. 69.
52. Davis, J., et al., *A comparison of the ICECAP-O with EQ-5D in a falls prevention clinical setting: are they complements or substitutes?* Quality of Life Research: 2013. **22**(5):p. 969-977
53. Carpenter, C.R., et al., *Four sensitive screening tools to detect cognitive dysfunction in geriatric emergency department patients: Brief Alzheimer's screen, short blessed test, Ottawa 3DY, and the caregiver-completed AD8.* Academic Emergency Medicine, 2011. **18**(4): p. 374-384.
54. Albert, S.M., *Proxy-reported quality of life in Alzheimer's patients: Comparison of clinical and population based samples.* J Ment Health Aging, 1999. **5**(Journal Article): p. 49-68.

Appendix 1: The ICECAP-O instrument proxy version

This appendix highlights the ICECAP-O instrument proxy version used in this study. It was based on the original ICECAP-O version, as developed by Joanna Coast and Terry Flynn, and introduced and validated in [9]. The original version can be found at http://www.icecap.bham.ac.uk/documents/icecapquest.doc

General instructions for the proxy questionnaire: We would like to ask you to fill out the questions below for the client/family member. Please try to answer the questions in manner as the client/family member would if he/she would be able to answer the questions. With every question, please tick the answer that the client/family member would give.

ABOUT THE CLIENT'S QUALITY OF LIFE

By placing a tick (✓) in ONE box in EACH group below, please indicate which statement best describes the clients quality of life at the moment.

1. Love and Friendship		
The client can have all of the love and friendship that he/she wants		4
The client can have a lot of the love and friendship that he/she want		3
The client can have a little of the love and friendship that he/she wants		2
The client cannot have any of the love and friendship that he/she wants		1
2. Thinking about the future		
The client can think about the future without any concern		4
The client can think about the future with only a little concern		3
The client can only think about the future with some concern		2
The client can only think about the future with a lot of concern		1
3. Doing things that make you feel valued		
The client is able to do all of the things that make him/her feel valued		4
The client is able to do many of the things that make him/her feel valued		3
The client is able to do a few of the things that make him/her feel valued		2
The client is unable to do any of the things that make him/her feel valued		1
4. Enjoyment and pleasure		
The client can have all of the enjoyment and pleasure that he/she wants		4
The client can have a lot of the enjoyment and pleasure that he/she wants		3
The client can have a little of the enjoyment and pleasure that he/she wants		2
The client cannot have any of the enjoyment and pleasure that he/she wants		1
5. Independence		
The client is able to be completely independent		4
The client is able to be independent in many things		3
The client is able to be independent in a few things		2
The client is unable to be at all independent		1

Tick one box only in each section

Tariffs and scoring algorithm in STATA

The ICECAP-O has 5 attribute dimensions each having 4 levels, 44444 representing full capability and 11111 representing no capability. The tariffs for the ICECAP-O, based on the preferences of the 65+ population in the United Kingdom were presented by Coast et al. [8]. The tariffs basically provide preference weights for the different ICECAP-O states, which are normalised in such a way that 0 represents the worst situation described on the ICECAP-O (11111) and 1 represents the best situation described on the ICECAP-O. Lower scores thus represent fewer, preference based, capabilities. The complete explanation as to how to calculate them is fully described on the ICECAP website: http://www.icecap.bham.ac.uk/tariffs.shtml.

Appendix 2: German version of the ICECAP-O

1. Liebe und Freundschaft

Der Bewohner kann all die Liebe und Freundschaft haben, die er will	4
Der Bewohner kann viel von der Liebe und Freundschaft haben, die er will	3
Der Bewohner kann ein wenig von der Liebe und Freundschaft haben, die er will	2
Der Bewohner kann keinerlei von der Liebe und Freundschaft haben, die er will	1

2. Gedanken über die Zukunft

Der Bewohner kann über die Zukunft ohne Sorgen nachdenken	4
Der Bewohner kann mit wenig Sorgen über die Zukunft nachdenken	3
Der Bewohner kann über die Zukunft nur mit einigen Sorgen nachdenken	2
Der Bewohner kann über die Zukunft nur mit großen Sorgen nachdenken	1

3. Dinge tun, durch die ich man sich geschätzt fühlt

Der Bewohner ist in der Lage alle Dinge zu tun, durch die er sich geschätzt fühlt	4
Der Bewohner ist in der Lage viele Dinge zu tun, durch die er sich geschätzt fühlt	3
Der Bewohner ist in der Lage einige Dinge zu tun, durch die er sich geschätzt fühlt	2
Der Bewohner ist nicht in der Lage irgendwelche Dinge zu tun, durch die er sich geschätzt fühlt	1

4. Freude und Vergnügen

Der Bewohner kann all die Freude und das Vergnügen haben, die er will	4
Der Bewohner kann viele der Freuden und Vergnügen haben, die er will	3
Der Bewohner kann nur wenig der Freuden und Vergnügen haben, die er will	2
Der Bewohner kann keinerlei Freude und Vergnügen haben, die er will	1

5. Unabhängigkeit

Der Bewohner ist in der Lage, völlig unabhängig zu sein	4
Der Bewohner ist in der Lage, in vielen Dingen unabhängig zu sein	3
Der Bewohner ist in der Lage, in einigen Dingen unabhängig zu sein	2
Der Bewohner ist nich in der Lage, unabhängig zu sein	1

Chapter 7 Cost-effectiveness of integrated care in frail elderly using the ICECAP-O and EQ-5D: does choice of instrument matter?

This chapter is based on Peter Makai, Willemijn Looman, René Melis, Eddy Adang, Elly Stolk, and Isabelle Fabbricotti
Cost-effectiveness of integrated care in frail elderly using the ICECAP-O and the EQ-5D; does choice of instrument matter? (Revised, European Journal of Health Economics)

Abstract

Economic evaluations likely undervalue the benefits of interventions in populations receiving both health and social services, such as frail elderly, by measuring only Health-Related Quality of Life (HrQol). For this reason, alternative preference-based instruments have been developed for economic evaluations in the elderly, such as the ICECAP-O. The aim of this paper is to investigate the cost-effectiveness of the Walcheren Integrated Care Model and if using the ICECAP-O in an economic evaluation leads to a different outcome in terms of cost-effectiveness than the EQ-5D, using data from the Walcheren Integrated Care Model. We performed univariate and multivariate analyses on costs and outcomes separately. We also performed incremental net monetary benefit (INMB) regressions using QALYs based on the ICECAP-O and EQ-5D. In terms of QALYs as measured with the EQ-5D and the ICECAP-O there were small and insignificant differences between the instruments, due to negligible effect size. Therefore, widespread implementation of the WICM would be premature based on these results. All results suggest that using the ICECAP-O the intervention has a higher probability of cost-effectiveness than with the EQ-5D at the same level of WTP. Further research is necessary in order to compare the two instruments in effective interventions.

Introduction

Especially during budget cuts, it is important to compare competing health and social services in terms of value for money. Economic evaluation increasingly aids decision makers in such comparisons by comparing the costs and effects of various healthcare interventions. In this comparison, Quality Adjusted Life Years (QALY) are used to integrally measure health status and mortality. For the measurement of the quality adjusted component of the QALY, Health-related Quality of Life (HrQol) measures are used, such as the EQ-5D [1]. However, various populations may be the recipients of social care as well, which does not aim to improve health status directly, but may contribute to broader wellbeing. For example, a taxi service taking elderly to church or bingo is unlikely to improve health, but will likely contribute to broader wellbeing. In addition, elderly may receive various forms of long-term care such as home care and residential care, both combining health and social care [2]. Therefore, using only HrQol instruments in this population may undervalue the range of services provided, biasing allocation decisions. Therefore, the impact of these services should be measured and valued accordingly, using measures capturing benefits of both health and social services. One solution to capture such benefits can be found in capability theory, through the notion of capability wellbeing rooted in Sen's capability theory [3].

Capability theory as developed by Sen distinguishes between functioning and capability, defined as the ability to function in a certain manner. The classic example for this distinction is a comparison between an individual who chooses to fast and an individual who is starving [4] [5]. Although the two individuals are identical in terms of functioning, the former clearly has the capability to move to a higher level of functioning by obtaining food, while the latter does not. Although capability theory is a popular and influential theory of redistribution and can be seen as influencing the extra-welfarist framework [6], explicit attempts to incorporate capability theory into economic evaluations and to directly measure capabilities are relatively recent [7] [8] [9]. The first capability-based instrument, the ICECAP-O was developed as an outcome measure for economic evaluations in health and social care [9] [10], measuring capability wellbeing [3] with a combination of health and non-health dimensions [11]. By directly measuring capabilities, the instrument allows computing capability QALYs in contrast with health QALYs based on the EQ-5D [7]. The ICECAP-O has been widely validated, and its properties are increasingly well established in comparison to the EQ-5D in cross-sectional and longitudinal studies [12-20]. However, the ICECAP-O is rarely used in economic evaluations, and its properties are not investigated in detail [21]. Furthermore, it has not been used in cost-effectiveness analyses in a population of frail elderly, which is a population well

suited to assess its usefulness, as this population consumes a wide variety of health and social care resources.

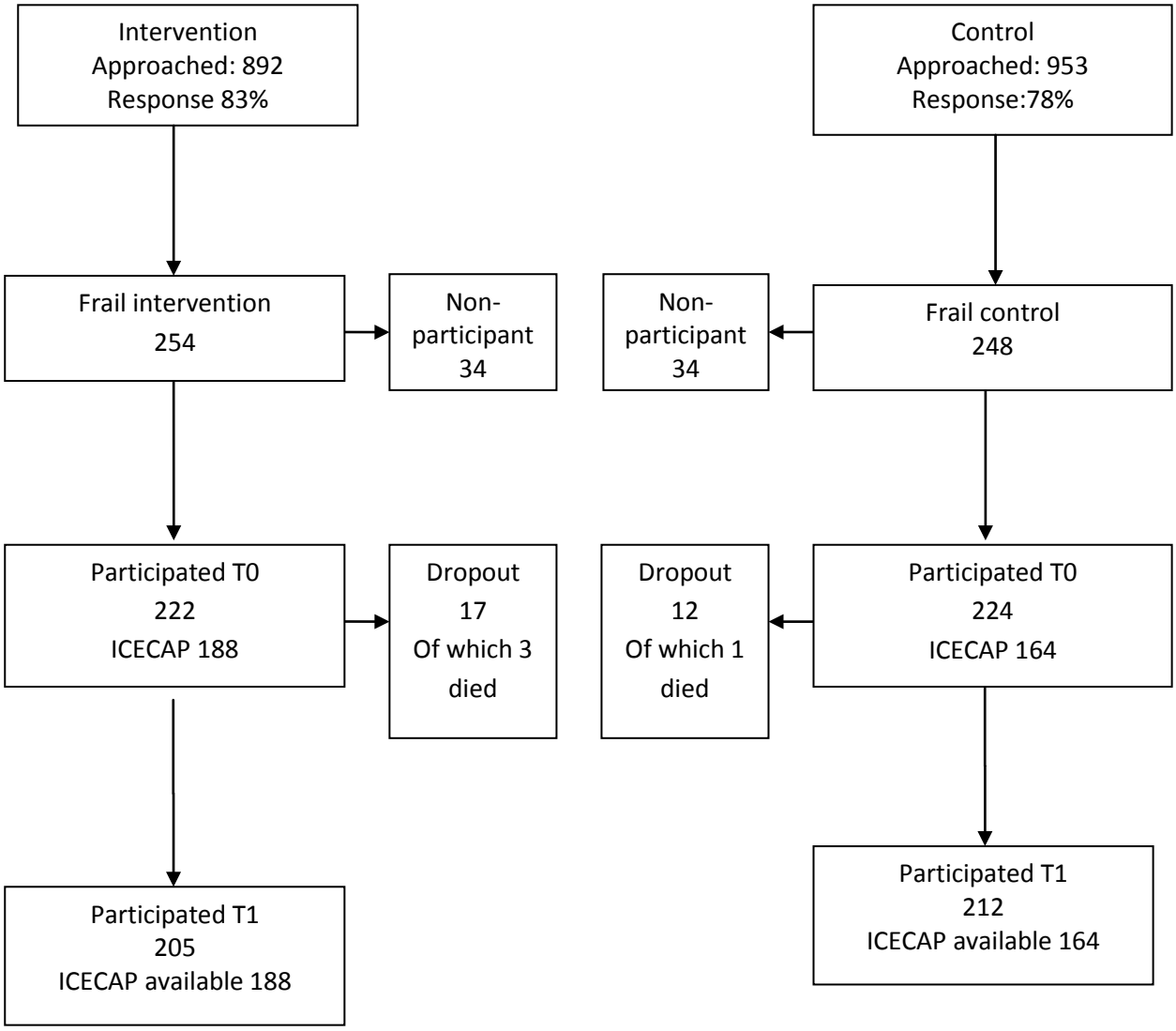
Frail elderly have an increased risk of health decline due to the accumulation of deficits such as limitations of functional decline, decreased social contacts or depression and are at risk of declining health and wellbeing [22] [23] [24]. To prevent such a decline, frail elderly often receive integrated care interventions. There is evidence that integrated care for frail elderly decreases institutionalization and may decrease associated costs [25], positive health effects may be expected after 3 months [26], while wellbeing effects are unknown. Using the Walcheren Integrated Care Model (WICM) [27] tailors the mix of health and social care services around the elderly in order to meet his or her needs. Therefore WICM is likely to influence health and wellbeing outcomes resulting from alterations in such care, alongside efficiency gains [27]. The WICM has previously shown a significant difference on the ICECAP-O dimension attachment between the intervention and control group, while scoring no difference on any EQ-5D dimension after a study period of three months [28], suggesting a small improvement in wellbeing, but not on health. The aim of this paper is twofold: (1) to evaluate the cost-effectiveness using a short run time frame for an integrated care model for frail elderly and (2) to investigate if using a broader measure of (capability) wellbeing in an economic evaluation leads to a different outcome in terms of cost-effectiveness. For this purpose, we compare the capability QALY based on the ICECAP-O and the health QALY based on the EQ-5D using the case of the WICM. For this current cost-effectiveness study, we use data from the baseline and three month follow up of the WICM [27].

Method

Design and participants

A controlled before-after design was used, comparing WICM to a control group receiving standard care. Treatment assignment was performed on the level of the GP practice. 3 GP practices were assigned to the intervention group and 5 to the control group. In both the intervention and control practices elderly above 75 years were screened using the Groningen frailty indicator (GFI), and if elderly were frail they were eligible for inclusion. Frailty was defined as having a score of 4 on a scale of 1-15 [29]. Exclusion criteria were being on a waiting list for a nursing home, a life expectancy of below 6 months, or terminally ill. The design of the study was extensively described elsewhere [27]. Participants were interviewed in their homes at the beginning of the study and after 3 months. Figure 7.1 shows the flow of participants during the first 3 months of the study, taking place in 2011.

Figure 7.1: Flowchart



For the current analyses, the group for whom the ICECAP-O was available was used, which did not contain dropouts.

Intervention and comparator

WICM is a complex, multicomponent intervention, with elements such as a single entry point system, case-management, geriatric assessment, multidisciplinary teams, and use of incentives for substitution [30]. The GP functions as a coordinator, a partner in prevention, and a single entry point of care. GPs detect frailty using the GFI [29]. Frail elderly are visited by their GP's nurse practitioner, who assesses their functional, cognitive, mental, and psychological functioning using EASYcare, an evidence-based instrument used for geriatric assessment [26]. The assessment is discussed in a multidisciplinary meeting, and a multidisciplinary treatment plan is then formulated in consultation with the elderly person and his or her informal caregiver(s). Case management is

provided by a specialized nurse practitioner or a secondary-line geriatric nursing practitioner, depending on case complexity. Multidisciplinary meetings are organized to monitor and adjust the treatment plan, attended by all professionals involved in the care of a particular elderly. The entire process is supported by an electronic patient record and multidisciplinary protocols. The WICM requires task reassignment and delegation between nurses and doctors and between GPs, nursing home doctors, and geriatricians. Consultations take place between primary, secondary, and tertiary care providers [27].

In contrast to the preventive intervention, which includes screening and assessments of requirements, usual care is reactive. Frail elderly patients only consult with their GP on their own initiative. The GP has a gate keeper role in the Dutch health care system [31]. This means that the GP is an important guide for access to the care system and assigns frail elderly patients to both care and cure in the secondary and tertiary echelons [31]. Care as usual does not include case management or multidisciplinary cooperation by protocols and meetings.

Outcomes and covariates

Capability wellbeing [3] was measured using the ICECAP-O capability measure for older people. The ICECAP instruments can be seen as measuring capability wellbeing [3] achieved by the capacity to perform certain actions and achieve certain states [9]. The ICECAP-O measures five capability dimensions – attachment, security, role, enjoyment, and control – with one question per dimension. Each dimension can be scored on four levels, thus distinguishing 1024 possible 'capability states'. The ICECAP-O was developed using rigorous qualitative and quantitative approaches [9] [12] [10] [32]. In order to obtain tariffs for the ICECAP-O, the attributes were valued using best-worst scaling, a special type of discrete choice analysis [9]. The ICECAP-O tariffs have values between 0 (no capability) and 1 (full capability). ICECAP scores were computed using the British tariffs of the ICECAP-O [9].

To measure HrQol we used the EQ-5D [33]. The EQ-5D measures HrQol in terms of five dimensions (mobility, self-care, daily activities, pain and discomfort, anxiety and depression) with three levels each (1=no problems, 2=moderate problems, and 3=extreme problems) describing 243 health states. The EQ-5D health states can be converted into a utility score by applying the scoring values (tariff) for the Dutch population [33]. The EQ-5D utility scores range from 1 (perfect health) through 0 (death) and has negative values accounting for health states worse than dead [33]. The EQ-5D is one of the most widely used measures of HrQol, and is extensively used in economic evaluations [33]. Health state utilities were computed using the Dutch tariffs of the EQ-5D [34].Health QALYs

based on the EQ-5D utility scores and capability QALYs based on the ICECAP-O scores were computed for a 3 month period based on the number of days between baseline and the 3 month measurement.

Costs were computed based on the Dutch guideline of costing studies [35]. For the collection of cost data, we adopted a societal perspective. The majority of costs categories were identified using care use questionnaires for the elderly and their informal caregivers at both baseline and 3 month follow-up, with a number of categories investigated in detail. Volume of care for the first month was based on care use of baseline, while for subsequent months on the follow-up measures, as in the first month case managers only assessed the condition of patients, and only in the second months did they initiate changes in care use. Hospital costs were calculated based on the GP's patient files, while intervention costs were based on the time registration of the case-managers. Accounting for inflation, standard cost-prices for 2011 were used for the following cost-categories: GP, inpatient and outpatient hospital costs, emergency care use, ambulance, revalidation and permanent residence costs in nursing and residential homes, home care, costs of allied health professionals (such as physiotherapy), social care, and informal caregivers. This was supplemented by an own micro costing exercise for the intervention costs and for the activities of the nurse practitioner.

Furthermore, we have computed a frailty index to account for case-mix variation based on 46 health deficits such as morbidity or ADL deficits[36] [37]. The frailty index is the number of deficits presents divided by a total possible number of deficits [37]. As such, it can account for all kinds of health related imbalance between the intervention and control group, and provides a detailed assessment of individuals frailty. For this reason the frailty index (FI) was used as a covariate. Other covariates were sex [38] (due to HrQol differences between men and women), living independently [39] (institutionalized elderly have lower Qol), days in the study [40] (due to unequal observation time between patients)) and baseline Qol, for the ICECAP and the EQ-5D [41] (in order to account for regression to the mean). In observational studies, the results of cost-effectiveness studies may be misleading [42, 43] due to selection bias if imbalanced covariates are correlated with either costs or outcomes [41, 44-46]. In order to avoid selection bias, we performed covariate adjustment [47, 48]. Selection bias is caused by systematic differences between the intervention and control groups, which are also correlated with the outcomes [42, 43]. However, controlling for imbalanced covariates is necessary, but not sufficient to obtain correct p-values of treatment effect. Baseline variables can also have a disproportionate effect on the treatment outcomes [49] even when not imbalanced, which also need to be adjusted for. Therefore we also included covariates in all

regression analyses if they were correlated with either the capability or health QALY outcome, with a correlation coefficient above 0.5 as recommended by Pocock [49] (baseline QoI on the EQ-5D and ICECAP-O, FI, days in the study). In short, we include case-mix variables (sex, living independently, days in the study and FI), and we adjust for baseline (EQ-5D and ICECAP-O).

Analyses

We performed the analyses on all cases presented with the ICECAP-O. Descriptive statistics were computed for all baseline variables and treatment usage in order to explore differences between the intervention and the comparator, and the missingness structure. Subsequently, we have performed multiple imputations using chained equations [50] [51] in order to treat item-nonresponse [52], creating 25 datasets in order to have stable estimates and 99% efficiency. For the imputation model, we used the following variables: age, sex, education, living in a nursing home, IADL status, mental health status, social functioning, self-reported health and quality of life, ICECAP-O dimensions, EQ-5D dimensions, as well as major costs categories.

Means and standard deviations for major costs categories were calculated. Treatment effects for QALYs, capability QALYs and costs were analyzed using multi-level regression models, in order to account for the clustered nature of the data due to treatment assignment [53, 54] [55, 56]. We used PROC MIXED specifying a random intercept model with an unstructured covariance matrix for all multilevel analyses below. In order to investigate the influence of the treatment on costs, capability QALYs and QALYs as outcomes, we performed unadjusted and covariate adjusted analyses on the outcomes separately, taking into account the following case-mix variables: sex, living at home, frailty index, and baseline EQ-5D and ICECAP-O [43, 45, 54]. Additionally, we adjusted for baseline tariffs for the EQ-5D and the ICECAP-O [41].

To determine actual cost-effectiveness, we performed unadjusted and adjusted Incremental Net Monetary Benefit (INMB) regressions using both QALYs as the effect measure [57]. In the INMB framework the covariates were entered directly into a regression with the INMB as an outcome [45]. We performed the INMB regression on a range of willingness to pay (WTP) values between 0 and 100000€. All analyses were performed individually on each multiply imputed dataset, and were combined using Rubin's rules to obtain means and standard deviations [52]. To indicate the probability of cost-effectiveness, Cost Effectiveness Acceptability Curves (CEAC) curves were constructed. A CEAC curve shows the probability of cost-effectiveness given the data plotted against the willingness to pay [58]. We constructed the curves based on the mean and the standard errors of the individual INMB regressions, and presented both unadjusted and covariate adjusted

analyses [59]. Results were presented for a range of WTP values between 0 and 100000€. Computations were carried out in SAS 9.2, and multiple imputations were carried out in IVEware 0.2, a SAS extension for imputations.

Results

Descriptives and response

Table 7.1 shows that there was a significant imbalance at baseline between the intervention and control group on a number of important covariates. In the intervention group, elderly were more likely to be women, to be less educated, and to live in an assisted living facility. Twenty-two percent of the participants had missing data on costs or outcomes, in both the intervention and control group. Elderly with fullcase data were better off in terms of frailty and multimorbidity, and were less likely to use care, while imbalance was the same as in the study population, suggesting there was no differences in dropout (selective dropout) between the intervention and control group (analysis not shown). The population where the ICECAP-O was available and therefore included in this study and the population not included were highly comparable.

Care use, costs and effects

Table 7.2 shows unadjusted care use in the intervention and control group at three months. Elderly in the intervention group were more likely to live permanently in an assisted living facility, to be visited by the GP after hours, and had more contact with the practice assistant nurse than the control group. The intervention group received more nursing care than the control group, while the control group is more likely to receive outpatient psychological care. Table 7.3 shows the costs in the intervention and QALY values at three months. There were significant differences in the costs of the practice assistant between the two groups, costs of practice assistant, costs of nursing care at home, costs of social work are higher in the intervention group. The control group had higher costs of psychological help. There were no significant differences between QALYs and capability QALYs, or total costs between the two groups.

Table 7.1: Baseline characteristics of the sample

Demographic Characteristic	*Category*	*Total*	*Intervention N=188 n (%) Mean, (SD)*	*Control N=164 n (%) Mean, (SD)*	*p-value (2 sided)*
Gender	Male	130	59 (31.4)	71 (43.3)	0.0209*
	Female	222	129 (68.6)	93 (56.7)	
In which country were you born?	Netherlands	335	175 (93.1)	160 (97.6)	0.0507
	Other country	17	13 (6.9)	4 (2.4)	
What is the highest education you completed?	Less than 6 classes	15	8 (4.3)	7 (4.3)	0.021*
	6 primary school classes	148	87 (46.3)	61 (37.7)	
	More than primary school	32	21 (11.2)	11 (6.8)	
	Practical training	36	18 (9.6)	18 (11.1)	
	Secondary vocational	89	47 (25)	42 (25.9)	
	Pre-university education	19	5 (2.7)	14 (8.6)	
	University/ higher practical	11	2 (1.1)	9 (5.6)	
What is you marital status?	Married	144	69 (37.1)	75 (46)	0.4307
	Divorced	8	4 (2.2)	4 (2.5)	
	Widow/Widower	184	106 (57)	78 (47.9)	
	Single	7	3 (1.6)	4 (2.5)	
	Sustainable living together	6	4 (2.2)	2 (1.2)	

What is your living arrangement/ situation?	Independent, alone	153	80 (42.6)	73 (44.5)	0.0011*
	Independent, with others	138	63 (33.5)	75 (45.7)	
	Home for the aged	61	45 (23.9)	16 (9.8)	
Osteoporosis (osteoporosis)	Not ticked	271	136 (72.3)	135 (82.3)	0.0265*
	Ticked	81	52 (27.7)	29 (17.7)	
Fractures other then hip	Not ticked	324	167 (88.8)	157 (95.7)	0.017*
	Ticked	28	21 (11.2)	7 (4.3)	
Living situation	Independent	291	143 (76.1)	148 (90.2)	0.0005*
	Nursing home	61	45 (23.9)	16 (9.8)	
Age	Mean, (SD)	352	81.76 (4.63)	81.66 (5.10)	0.8613
Frailty index	Mean, (SD)	352	0.22 (0.08)	0.21 (0.08)	0.6947
Mental health	Mean, (SD)	351	22.60 (4.50)	22.80 (4.24)	0.6701
Multimorbidity	Mean, (SD)	352	3.73 (1.89)	3.82 (1.78)	0.6538
Social function	Mean, (SD)	351	67.55(34.27)	67.48 (38.50)	0.9859
Days in study	Mean, (SD)	352	98.78(14.83)	95.88 (13.30)	0.0554
Katz ADL	Mean, (SD)	352	0.85 (1.17)	0.76 (1.27)	0.4912
Katz IADL	Mean, (SD)	352	4.00 (3.19)	3.62 (3.36)	0.2722
EQ-5D baseline	Mean, (SD)	348	0.64 (0.26)	0.67 (0.27)	0.2055
ICECAP-O baseline	Mean, (SD)	334	0.78 (0.16)	0.79 (0.15)	0.4623

Table 7.2: Care use

Variable	*Care use Total*	*N=188 Intervention Mean (Std. Dev)*	*N=164 Control Mean (Std. Dev)*	*P-value (2 sided)*
Hospital				
Admission general hospital	352	0.27 (1.84)	0.31 (1.95)	0.8444
Outpatient general hospital	352	0.61 (1.04)	0.42 (0.89)	0.0631
Outpatient academic hospital	352	0.02 (0.16)	0.01 (0.08)	0.4606
Day surgery general hospital	352	0.02 (0.13)	0.04 (0.22)	0.2878
Emergency ward	352	0.04 (0.20)	0.03 (0.17)	0.5463
Ambulance	352	0.00 (0.00)	0.01 (0.08)	0.3188
Residential care/nursing home				
Temporary stay assisted living facility	339	0.02 (0.22)	0.02 (0.24)	0.9101
Temporary stay nursing home	351	0.04 (0.58)	0.01 (0.08)	0.398
Permanent stay assisted living facility	352	2.60 (15.93)	1.18 (10.69)	0.3223
Permanent stay nursing home	352	0.09 (1.24)	0.66 (8.43)	0.3939
GP	352	2.06 (2.34)	1.64 (2.52)	0.1027
Consultation on the telephone	352	0.26 (0.65)	0.36 (0.93)	0.2519
GP consultation	352	0.84 (1.21)	0.70 (1.11)	0.2655
GP visit at home	352	0.96 (1.58)	0.58 (1.94)	0.0449*
Practice assistant	352	0.86 (1.42)	0.79 (1.59)	0.6401
Consultation on the telephone	352	0.10 (0.37)	0.10 (0.52)	0.8708
Consultation practice assistant	352	0.26 (0.56)	0.43 (0.84)	0.0265*
Practice assistant visit	352	0.51 (1.07)	0.26 (0.86)	0.0139*
HAP	352	0.07 (0.29)	0.02 (0.19)	0.0868
Emergency GP telephone consultation	352	0.01 (0.10)	0.01 (0.11)	0.891
Emergency GP consultation	352	0.01 (0.10)	0.00 (0.00)	0.1579
Emergency GP visit	352	0.05 (0.26)	0.01 (0.11)	0.0871
Daily activity	350	0.05 (0.28)	0.05 (0.29)	0.9531
Day care	349	0.04 (0.37)	0.00 (0.05)	0.2522
Home care household activities component	338	2.19 (3.05)	1.95 (2.28)	0.4118

Home care personal care component	341	0.94 (2.14)	1.10 (3.88)	0.6367
Home care nursing care component	344	0.23 (0.91)	0.14 (0.55)	0.2379
Physiotherapy	343	0.12 (0.30)	0.12 (0.26)	0.7912
Occupational therapy	344	0.00 (0.00)	0.00 (0.03)	0.1946
Psychological care	345	0.00 (0.01)	0.04 (0.30)	0.1043
Social care	342	0.00 (0.01)	0.00 (0.00)	0.3186
Informal care				
Household activities	341	2.57 (7.21)	2.09 (5.71)	0.4949
Personal care	349	0.62 (2.26)	0.70 (3.14)	0.7856
Nursing care	342	0.66 (1.24)	0.79 (2.75)	0.557
Intervention costs				
Casemanager	188	265.63 (89.87)	. (.)	
Easycare Assesment	188	99.36 (34.80)	. (.)	
Care plan	188	81.45 (46.24)	. (.)	
Other meetings	188	6.78 (13.47)	. (.)	
Multi-disciplinary meeting casemanager	188	49.63 (37.10)	. (.)	
Case management	188	28.40 (43.76)	. (.)	
Hours multidisciplinary meeting	188	60.72 (48.97)	. (.)	
Hours GP multidisciplinary meeting	188	33.75 (23.03)	. (.)	
Hours nursing home physician multidisciplinary meeting	188	17.24 (15.56)	. (.)	
Hours geriatric specialist multidisciplinary meeting	0	. (.)	. (.)	
Hours geriatric physiotherapist	188	6.11 (15.99)	. (.)	
Hours practice assistant multidisciplinary meeting	0	. (.)	. (.)	
Hours district nurse multidisciplinary meeting	188	3.62 (9.53)	. (.)	

Table 7.3: Costs, utilities at follow-up and QALYs

Variable	*Total*	*N=188 Intervention Mean (Std. Err)*	*N=164 Control Mean (Std. Err)*	*P-value (2 sided)*
Costs				
Admission general hospital	352	122 (60)	140 (69)	0.8444
Outpatient general hospital	352	41 (5)	28 (5)	0.0660
Outpatient academic hospital	352	2 (2)	1 (1)	0.4794
Day surgery general hospital	352	4 (2)	10 (4)	0.2710
Emergency ward	352	8 (3)	5 (2)	0.3784
Ambulance	352	0 (0)	2 (2)	0.2850
Temporary stay assisted living facility	352	62 (25)	107 (35)	0.2887
Temporary stay nursing home	352	73 (73)	21 (15)	0.5122
Permanent stay assisted living facility	352	203 (101)	110 (78)	0.4769
Permanent stay nursing home	352	22 (22)	162 (162)	0.3621
GP costs	352	81 (7)	62 (9)	0.1217
Practice assistent costs	352	16 (2)	10 (2)	0.0371
Emergency GP costs	352	1 (0)	0 (0)	0.3152
Daily activity	352	18 (7)	19 (8)	0.9028
Day care	352	84 (67)	38 (19)	0.5299
Home care household activities component	352	907 (166)	738 (78)	0.3607
Home care personal care component	352	683 (129)	904 (286)	0.4765
Home care nursing care component	352	310 (159)	137 (42)	0.3036
Physiotherapy	352	68 (12)	64.82 (14)	0.8545
Occupational therapy	352	0.00 (0.00)	1 (1)	0.1645
Psychological care	352	2 (1)	86 (37)	0.0137*
Social care	352	2 (1)	0.00 (0.00)	0.0210*
Informal care household activities	352	994 (204)	822 (235)	0.5862
Informal care personal care	352	434 (127)	521 (187)	0.6977
Informal care nursing care	352	145 (46)	221 (72)	0.3758
Costs case management	352	145 (3)	0.00 (0.00)	<.0001
Costs multi-disciplinary meetings	352	52 (3)	0.00 (0.00)	<.0001
COSTS care use HC perspective	352	2574 (357)	2518 (454)	0.9218
COSTS with intervention HC perspective	352	2771 (357)	2518 (454)	0.6581
COSTS societal perspective	352	4344 (501)	4082 (605)	0.7294
Outcomes				
ICECAP score follow-up	352	0.76 (0.01)	0.76 (0.01)	0.9828
EQ-5D score follow-up	352	0.63 (0.02)	0.66 (0.02)	0.3062
Capability QALY	352	0.21 (0.00)	0.20 (0.00)	0.1953
Health QALY	352	0.17 (0.01)	0.17 (0.01)	0.6732

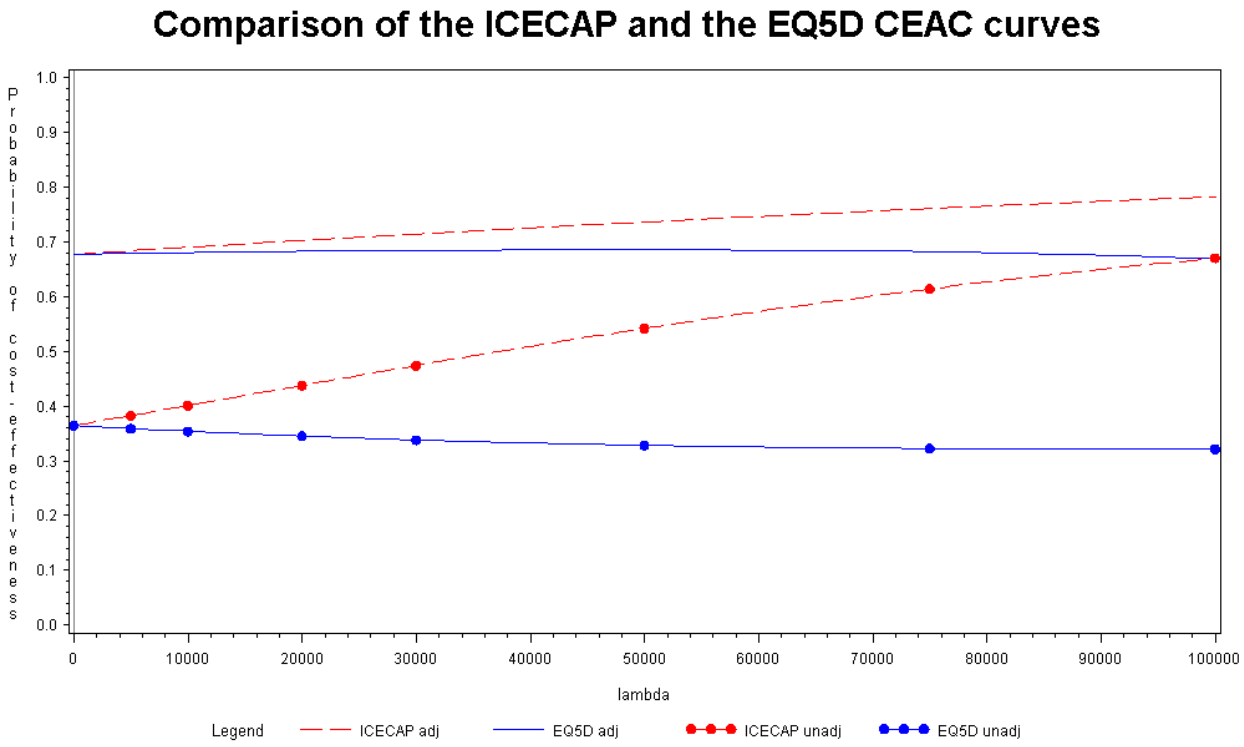
Table 7.4 shows the effect of the adjustment on the differences between intervention and control groups in costs and QALYs. The differences in the unadjusted analysis between the intervention and control group are shown by the coefficients of the unadjusted difference variable in Table 7.4. The differences in costs in the unadjusted analysis favored the comparator by 261€, as well as the differences in health QALY (-0,003), while capability QALYs favored the intervention (0,007), although these effects were not significant. The differences in the adjusted analysis between the intervention and control group are shown by the coefficients of the unadjusted difference variable in Table 7.4. In the adjusted analysis, which made the intervention and control groups more comparable due to the adjusting for a number of differences at baseline, the intervention group was favored (-247€). The difference between the adjusted and unadjusted cost analysis is due to the higher unadjusted costs in the intervention group caused by differences between the intervention and control group. The differences are caused by the higher frailty index, spending more days in the study, the higher percentage of people living in a nursing home, and by having a higher percentage of women in the intervention group. The influence of the individual covariates on costs can be seen in the appendix. The effect of the intervention expressed in health QALY was close to 0.0009, while the capability QALYs still favored the intervention with a difference of 0.003 QALY. In both cases, the difference after adjustment in both QALYs is mainly due to correcting for differences in baseline capability-Qol and HrQol status respectively, and correcting for the number of days in the study, which also varied between individuals. In addition, the standard error surrounding the difference in capability QALYs is smaller than the health QALYs in all analyses, leading to more precise estimates. Correlations between baseline EQ-5D and costs (-0.33, p-value <0.001) were somewhat higher than between costs and baseline ICECAP-O (-0.26, p-value <0.001).

Table 7.4: Base case and adjusted results for costs and effects

Multiply imputed N=352	*Outcome*	*Capability QALY Mean (Std. Err) p-value*	*QALY Mean (Std. Err) p-value*	*Costs Mean (Std. Err) p-value*
	Unadjusted intercept	0.1723 (0.0063) 0.0000	0.2010 (0.0040) 0.0000	4082 (562) 0.0000
	Unadjusted Difference	0.0070 (0.0054) 0.1953	-0.0026 (0.0090) 0.7673	262 (757) 0.7294
	Adjusted intercept	-0.1198 (0.0165) 0.0000	-0.1820 (0.0105) 0.0000	-7803.83 (2418) 0.0013
	Adjusted difference	0.0028 (0.0019) 0.1382	-0.0009 (0.0052) 0.8603	-247 (742) 0.7390
Adjusted for:	Days in study	0.0022 (0.0001) 0.0000	0.0017 (0.0001) 0.0000	78 (24) 0.0012
	Living situation	-0.0043 (0.0025) 0.0909	-0.0084 (0.0041) 0.0393	2284 (900) 0.0112
	Sex	0.0010 (0.0020) 0.6271	-0.0001 (0.0031) 0.9784	-1012 (709) 0.1534
	Frailty index	-0.0200 (0.0158) 0.2052	-0.0674 (0.0246) 0.0062	22160 (4167) 0.0000
	ICECAP baseline	0.2210 (0.0073) 0.0000	0.0185 (0.0106) 0.0806	n.a.
	EQ-5D baseline	0.0115 (0.0049) 0.0186	0.2021 (0.0078) 0.0000	n.a.

*Significant below p<0.05

Figure 7.2: CEAC curves based on the Incremental Net Monetary Benefit analysis using QALYs as outcomes



Cost-effectiveness

Figure 7.2 shows the CEAC curves where incremental QALYs are used as outcomes in the INMB regression. Unadjusted INMB regressions showed that INMB measured with capability QALYs had a higher probability of cost-effectiveness than INMB with health QALYs, especially at higher WTP values. Additionally, health QALY estimates were negative in the unadjusted analysis, and the majority of the confidence interval is below 0, the CEAC curves were bounded at a p-value of around 0,27. The CEAC curves of the unadjusted QALYs did not increase sharply, as there were small, and insignificant differences in both costs and effects. In the adjusted analysis, there was a smaller difference in probability of cost-effectiveness between the two instruments. The adjustment reverses the conclusion because it changed the cost differences from favoring the control group to favoring the intervention. In all analyses, using the capability QALYs result in consistently higher probability of cost-effectiveness for the WICM than the health QALYs.

Discussion

Summary of main results

This is the first study to investigate if using a broader outcome measure, the ICECAP-O in a frail elderly population within a cost-effectiveness analysis leads to different outcomes as the EQ-5D.

There were small and insignificant differences in QALY between the intervention and the control group both in terms of capability and health QALYs. Additionally, the costs differences between the intervention and control group were also small and not significant. The probability of cost-effectiveness of the WICM in the adjusted analysis was similar when based on health QALYs and capability QALYs as outcome, with capability QALYs showing a higher probability of cost-effectiveness in all analyses. As for the traditional cost-effectiveness of the WICM after 3 months, as measured by the EQ-5D: adjusted point estimates the intervention is less costly and no more effective than standard care. At the same time, both effects are marginal and not statistically significant, and covariate adjustments have a large influence on the cost-effectiveness results. Therefore, based on these results widespread implementation would be premature.

Strengths and limitations

This current study has a number of strengths. With a sample size of 352 participants, this study is relatively large in investigation of cost-effectiveness of frail elderly, from a societal perspective. Furthermore, this study accounted for the clustered nature of the data using the recently developed appropriate methods [53] [54], which is largely ignored in economic evaluations [60]. It also has a number of limitations. First, this was an observational study with small effect sizes within an imbalanced frail elderly population, which is likely to bias unadjusted comparisons. The adopted procedure of covariate adjustment reduced the imbalance and improved efficiency, but also had a large impact on the results, especially in terms of incremental costs, where covariate adjustment lead to savings instead of additional costs for the intervention group. Therefore obtaining precise estimates is crucial for comparing INMB regressions based on health QALYs and capability QALYs. For this reason, covariates were included into the model with great care using a systematic procedure, removing bias and improving efficiency of the estimates. However, it may be possible that some unobserved imbalance remained which may influence the results, and it is also possible to introduce some bias with covariate adjustment, although the bias is typically smaller than the bias in an unadjusted analysis [61]. As the effect sizes are small, both sources of bias - even if small - raise important concerns. Therefore further research is necessary in more balanced groups to confirm these results on the effects of case-management and for the comparison of the instruments in economic evaluations. Second, care use for the cost categories day care, home care, physiotherapy, occupational therapy, psychological care, social care and informal care were identified at baseline and follow-up at the given time point, and not for the previous period as was the case for other cost categories. As change in care in such categories are initiated by the case-manager after the second month of the intervention, we used the baseline values for the first month and follow-up values for the remaining 2 months. Because costs may dominate INMB

results, this may potentially bias the comparison of the INMB regressions based on the health and capability QALY. However, there is no significant difference in care use between baseline and FU in either the intervention or control group, therefore this is unlikely to bias cost results (analysis not shown).

Finally, our study period was likely too short to account for all potential costs and benefits of a preventive intervention such as the WICM. While often effective on the short-term, because frailty is a gradual process[62, 63], it is likely that more profound effects of geriatric assessment on prevention of nursing home admissions and maintained/increased independence are only apparent over a longer time period [64]. This also holds for physical and mental health, and for quality of life. While complex interventions, such as the WICM cost a lot of time and adjustment of those involved, healthcare professionals need to master a new method of working[65, 66]. It is likely, that time spent on multidisciplinary consultation or EasyCare assessment will decrease as people become used to this method of working, thus reducing associated costs. However, the focus of this study was on the comparison of the health QALY and the capability QALY, and not on giving definitive results on cost-effectiveness of the intervention, therefore this is unlikely to bias our results.

Comparison of the instruments

Using QALYs as outcomes, there is little difference in probability of cost-effectiveness between the capability QALY and health QALYs in the context in this study due to negligible effect sizes, as the adjusted analysis showed. This is somewhat contrary to the expectations, as there is widespread evidence that the ICECAP-O measures a broader set of outcomes [12-20]. Therefore, selection of instruments are unlikely to influence policy, and the acceptance of WICM in the adjusted analysis. This conclusion holds irrespective of the value of a QALY. This raises the question if we are looking at a lack of difference between the two instruments, or at a lack of differences between the interventions. The lack of difference between the two instruments was found in a largely ineffective intervention, the WICM [28], and it would be troubling if the ICECAP-O would show differences which are not supported by other measures. In case of an effective intervention and a longer study period larger differences between the two QALY estimates can be expected based on these results, thus differences in effect size may be more pronounced between the instruments in a covariate-adjusted analysis of QALYs as well, as shown by previous studies [21] [20].

It is important to note that it is currently unclear what the WTP for a capability QALY is. As health QALYs can be seen as a part of the capability QALY [7], it may be possible that the WTP for the capability QALY is higher than for the health QALY, although by which magnitude is currently

unknown. Therefore it seems unlikely that valid comparisons can be made between the two instruments at a given level of WTP.

Due to the limitations of the study and specific context of the WICM, further research is necessary to compare the ICECAP-O and the EQ-5D in order to disentangle which effects of health and social services are missed by using health alone. This current study suggest, that this amount may be quite subtle, at least with complex, multicomponent interventions such as the WICM [67]. Globally, such interventions are characterized by small effect sizes [25], also in terms of QALYs [68][69]. Additionally, cost-effectiveness studies are increasingly performed in this setting [70, 71] in order to identify interventions providing good value for money. Therefore, research on the properties of the different outcome measures in such a setting is relevant, as interventions leading to changes in both health and social care outcomes are likely multicomponent interventions targeting different forms of healthcare provision. However, it is not clear which parts of the complex interventions impact the different outcomes, and which parts of such interventions improve health and which ones capability wellbeing. In order to shed light on how much benefits may be missed by focusing on health alone, health QALYs and capability QALYs should ideally be compared in an RCT of a simple intervention which likely has enormous wellbeing effects and negligible health effects.

Conclusion

From the perspective of economic evaluations, we compared an instrument which can integrally measure the benefits of both health and social care, the ICECAP-O to an instrument that measures health care benefits, the EQ-5D. Using QALYs as outcomes, there were little differences between capability QALY and health QALY after adjusting for baseline covariates. Capability QALYs found little additional benefit for the WICM intervention as compared to the health QALYs due to negligible effect sizes of the intervention. As for the cost-effectiveness of the intervention, widespread implementation of the WICM would be premature based on these results. In case an interventions' health and wellbeing effects are not significant, as in this study, using the ICECAP-O will not lead to a false claim of cost-effectiveness of the intervention. On the other hand if differences on capability QALYs are meaningful and significant, the ICECAP-O may have the potential to measure broader outcomes and be more sensitive to differences between intervention and comparators.

Acknowledgements

We would like to thank Erwin Birnie and Marc Koopmanschap for their useful suggestions on a previous version of this paper.

References

1. Drummond, M.F., et al., *Methods for the economic evaluation of health care programmes*. Vol. Third Edition. 2005, Oxford: Oxford University Press.
2. Medicare, *Long-term care definition*.
3. Al-Janabi, H., T. Flynn, and J. Coast, *Development of a self-report measure of capability wellbeing for adults: the ICECAP-A.* Quality of Life Research, 2012. **21**(1): p. 167-176.
4. Sen, A., *Choice, welfare and measurement*. 1982, Cambridge, MA: Harvard University Press.
5. Nussbaum, M., *The quality of life*. 1993, Oxford: Clarendon Press.
6. Brouwer, W.B.F., et al., *Welfarism vs. extra-welfarism.* Journal of health economics, 2008. **27**(2): p. 325-338.
7. Cookson, R., *QALYs and the capability approach.* Health economics, 2005. **14**(8): p. 817-829.
8. Anand, P.P., *Capabilities and health.* Journal of medical ethics, 2005. **31**(5): p. 299-303.
9. Coast, J., et al., *Valuing the ICECAP capability index for older people.* Social science & medicine, 2008. **67**(5): p. 874-882.
10. Coast, J., et al., *Investigating Choice Experiments for Preferences of Older People (ICEPOP): evaluative spaces in health economics.* Journal of health services research & policy, 2008. **13 Suppl 3**(Journal Article): p. 31-37.
11. Makai, P., et al., *Quality of life instruments for economic evaluations of elderly care: a systematic review.* Social science & medicine (Accepted for publication).
12. Coast, J., et al., *An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people.* Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 2008. **17**(7): p. 967-976.
13. Makai, P., et al., *A validation of the ICECAP-O in a population of community-dwelling older people.* Health and Quality of Life Outcomes, 2013. **Accepted for publication**.
14. Makai, P., et al., *A validation of the ICECAP-O in a population of post-hospitalized older people in the Netherlands.* Health and Quality of Life Outcomes, 2013. **11**(1): p. 1-11.
15. Couzner, L., J. Ratcliffe, and M. Crotty, *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, 2012. **10**(1): p. 69.
16. Couzner, L., et al., *Measuring and valuing quality of life for public health research: application of the ICECAP-O capability index in the Australian general population.* International Journal of Public Health, 2012: p. 1-10.
17. Davis, J.C., et al., *A comparison of the ICECAP-O with EQ-5D in a falls prevention clinical setting: are they complements or substitutes?* Quality of Life Research, 2011: 2013. **22**(5):p. 969-977.
18. Davis, J.C., et al., *Exploration of the association between quality of life, assessed by the EQ-5D and ICECAP-O, and falls risk, cognitive function and daily function, in older adults with mobility impairments.* BMC Geriatrics, 2012. **12**(1): p. 65.
19. Mitchell, P.M., et al., *Predicting the ICECAP-O Capability Index from the WOMAC Osteoarthritis Index Is Mapping onto Capability from Condition-Specific Health Status Questionnaires Feasible?* Medical Decision Making, 2013.
20. Comans, T.A., et al., *Quality of life of older frail persons receiving a post-discharge program.* Health and Quality of Life Outcomes, 2013. **11**(1): p. 58.
21. Henderson, C., et al., *Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial.* BMJ, 2013. **346**.
22. Rockwood, K., et al., *A global clinical measure of fitness and frailty in elderly people.* CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne, 2005. **173**(5): p. 489-495.

23. Ferrucci, L., et al., *Designing Randomized, Controlled Trials Aimed at Preventing or Delaying Functional Decline and Disability in Frail, Older Persons: A Consensus Report.* Journal of the American Geriatrics Society, 2004. **52**(4): p. 625-634.
24. Djernes, J., *Prevalence and predictors of depression in populations of elderly: a review.* Acta Psychiatrica Scandinavica, 2006. **113**(5): p. 372-387.
25. Eklund, K. and K. Wilhelmson, *Outcomes of coordinated and integrated interventions targeting frail elderly people: a systematic review of randomised controlled trials.* Health & social care in the community, 2009. **17**(5): p. 447-458.
26. Melis, R., et al., *A randomized study of a multidisciplinary program to intervene on geriatric syndromes in vulnerable older people who live at home (Dutch EASYcare Study).* The journals of gerontology. Series A, Biological sciences and medical sciences, 2008. **63**(3): p. 283.
27. Fabbricotti, I.N., et al., *Integrated care for frail elderly compared to usual care: a study protocol of a quasi-experiment on the effects on the frail elderly, their caregivers, health professionals and health care costs.* BMC Geriatrics, 2013. **13**(1): p. 31.
28. Looman, W. and I. Fabbricotti, *The effects of an integrated care model on frail elderly.* International Journal of Integrated Care, 2012. **12**(Suppl3).
29. Schuurmans, H., et al., *Old or frail: what tells us more?* The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 2004. **59**(9): p. M962-M965.
30. Johri, M., F. Beland, and H. Bergman, *International experiments in integrated care for the elderly: a synthesis of the evidence.* International journal of geriatric psychiatry, 2003. **18**(3): p. 222-235.
31. Schäfer, W., et al., *The Netherlands: health system review.* Health systems in transition, 2010. **12**(1): p. xxvii, 1-228.
32. Grewal, I., et al., *Developing attributes for a generic quality of life measure for older people: Preferences or capabilities?* Social Science and Medicine, 2006. **62**(8): p. 1891-1901.
33. Szende, A. and M.D.N. Oppe, *EQ-5D value sets - inventory, comparative review and user guide*. Vol. 1. 2007, Dordrecht: Springer.
34. Lamers, L.M., et al., *The Dutch tariff: Results and arguments for an effective design for national EQ-5D valuation studies.* Health economics, 2006. **15**(10): p. 1121-1132.
35. Tan, S.S., et al., *Update of the Dutch manual for costing in economic evaluations.* International Journal of Technology Assessment in Health Care, 2012. **28**(02): p. 152-158.
36. Lutomski, J.E., et al., *Validation of a Frailty Index from The Older Persons and Informal Caregivers Survey Minimum Data Set.* Journal of the American Geriatrics Society, 2013. **61**(9): p. 1625-1627.
37. Mitnitski, A.B., A.J. Mogilner, and K. Rockwood, *Accumulation of deficits as a proxy measure of aging.* The Scientific World Journal, 2001. **1**: p. 323-336.
38. Hanmer, J., et al., *Report of nationally representative values for the noninstitutionalized US adult population for 7 health-related quality-of-life scores.* Medical Decision Making, 2006. **26**(4): p. 391-400.
39. Rodriguez-Blazquez, C., et al., *Health status and well-being of older adults living in the community and in residential care settings: Are differences influenced by age?* Aging & Mental Health, 2012. **16**(7): p. 884-891.
40. Ganiats, T.G., C.J. Miller, and R.M. Kaplan, *Comparing the quality-adjusted life-year output of two treatment arms in a randomized trial.* Medical care, 1995: p. AS245-AS254.
41. Manca, A.A., N. Hawkins, and M.J. Schulpher, *Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility.* Health Economics, 2005. **14**(5): p. 487-496.

42. Kakkar, A. and B.C. Jacobson, *Failure of an Internet-Based Health Care Intervention for Colonoscopy Preparation: A Caveat for Investigators.* JAMA Internal Medicine: p. 1-2.
43. Nixon, R.M. and S.G. Thompson, *Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations.* Health Economics, 2005. **14**(12): p. 1217-1229.
44. Westbrook, J.I. and J. Braithwaite, *Will information and communication technology disrupt the health system and deliver on its promise?* Medical Journal of Australia, 2010. **193**(7): p. 399.
45. Hoch, J.S., A.H. Briggs, and A.R. Willan, *Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis.* Health Economics, 2002. **11**(5): p. 415-430.
46. Samoocha, D., et al., *Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis.* Journal of Medical Internet Research, 2010. **12**(2).
47. Craig, P., et al., *Developing and evaluating complex interventions: the new Medical Research Council guidance.* BMJ: British Medical Journal, 2008. **337**.
48. Vedel, I., et al., *Health information technologies in geriatrics and gerontology: a mixed systematic review.* Journal of the American Medical Informatics Association, 2013.
49. Davis, J., et al., *Guidelines for conducting and reporting economic evaluation of fall prevention strategies.* Osteoporosis International, 2011. **22**(9): p. 2449-2459.
50. Raghunathan, T.E., et al., *A multivariate technique for multiply imputing missing values using a sequence of regression models.* Survey methodology, 2001. **27**(1): p. 85-96.
51. White, I.I.R.R.P. and A.M. Wood, *Multiple imputation using chained equations: Issues and guidance for practice.* Statistics in medicine, 2011. **30**(4): p. 377-399.
52. Little, R.J.A. and D.B. Rubin, *Statistical analysis with missingdata*. 1987, New York: John Wiley and Sons.
53. Gomes, M., et al., *Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials.* Medical Decision Making, 2012. **32**(2): p. 350-361.
54. Gomes, M., et al., *Methods For Covariate Adjustment In Cost-Effectiveness Analysis That Use Cluster Randomised Trials.* Health Economics, 2012. **21**(9): p. 1101-1118.
55. Campbell, M.K., et al., *Consort 2010 statement: extension to cluster randomised trials.* BMJ: British Medical Journal, 2012. **345**.
56. Manca, A., et al., *Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models.* Health Economics, 2005. **14**(5): p. 471-485.
57. Bachmann, M.O., et al., *Methods for analyzing cost effectiveness data from cluster randomized trials.* Cost Effectiveness and Resource Allocation, 2007. **5**(1): p. 12.
58. Fenwick, E., B.J. O'Brien, and A. Briggs, *Cost-effectiveness acceptability curves–facts, fallacies and frequently asked questions.* Health Economics, 2004. **13**(5): p. 405-415.
59. Hoch, J.S., M.A. Rockx, and A.D. Krahn, *Using the net benefit regression framework to construct cost-effectiveness acceptability curves: an example using data from a trial of external loop recorders versus Holter monitoring for ambulatory monitoring of.* BMC Health Services Research, 2006. **6**(1): p. 68.
60. Sumukadas, D., et al., *Effect of perindopril on physical function in elderly people with functional impairment: a randomized controlled trial.* Canadian Medical Association Journal, 2007. **177**(8): p. 867-874.
61. Walter, S.D., et al., *When should one adjust for measurement error in baseline variables in observational studies?* Biometrical Journal, 2011. **53**(1): p. 28-39.
62. Slaets, J.P., *Vulnerability in the elderly: frailty.* Medical Clinics of North America, 2006. **90**(4): p. 593-601.

63. Gobbens, R., et al., *Towards an integral conceptual model of frailty.* The journal of nutrition, health & aging, 2010. **14**(3): p. 175-181.

64. Norman, G.R., J.A. Sloan, and K.W. Wyrwich, *Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation.* Medical care, 2003. **41**(5): p. 582-592.

65. Kodner, D.L., *Consumer-directed services: lessons and implications for integrated systems of care.* International Journal of Integrated Care, 2003. **3**.

66. Mur-Veeman, I., I. Eijkelberg, and C. Spreeuwenberg, *How to manage the implementation of shared care-A discussion of the role of power, culture and structure in the development of shared care arrangements.* Journal of management in medicine, 2001. **15**(2): p. 142-155.

67. Jones, D.A., C.R. Victor, and N.J. Vetter, *The problem of loneliness in the elderly in the community: characteristics of those who are lonely and the factors related to loneliness.* The Journal of the Royal College of General Practitioners, 1985. **35**(272): p. 136.

68. Ratcliffe, J., et al., *Health economics and geriatrics: challenges and opportunities.* Geriatrics: InTech, available from: http://www. intechopen. com/articles/show/title/health-economic-evaluation-and-geriatricschallenges-and-opportunities, 2012.

69. Forder, J.E. and J. Caiels, *Measuring the outcomes of long-term care.* Social science & medicine, 2011. **73**(12): p. 1766-1774.

70. Meeuwsen, E.J., et al., *Cost-effectiveness of post-diagnosis treatment in dementia coordinated by multidisciplinary memory clinics in comparison to treatment coordinated by general practitioners: an example of a pragmatic trial.* JNHA-The Journal of Nutrition, Health and Aging, 2009. **13**(3): p. 242-248.

71. Graff, M.J., et al., *Community based occupational therapy for patients with dementia and their care givers: randomised controlled trial.* BMJ: British Medical Journal, 2006. **333**(7580): p. 1196.

Appendix 1 Regression equations for Costs regressions for the adjusted and unadjusted analysis

Unadjusted:

Intervention-group: Costs=4082 (intercept)+1 (treatment dummy)*262=4344

Costs control group: Costs= 4082 (intercept)+0*262=4082

Adjusted:

As several covariates were unevenly distributed between the intervention and control group (male / female etc.), it is desirable to make the groups comparable, in order to have improved estimates between the cost differences. With the regression equations, costs of such groups become comparable by multiplying the coefficients with the population values from the intervention group. If you would have in the intervention arm and the control arm the same distribution of the covariates, the only remaining difference would be caused by the intervention itself.

Intervention group= -7803 (intercept)-+1(treatment dummy)*(-247)+98,79 (days in study)*78+2284 *0,23(living in an assisted living facility)-1011*0,69(female)+0.22(FI)*22159=4079

Individual contribution of predictors: -7803 (intercept)-247(treatment dummy)+7704,84 (days in study)+546,876 (living in an assisted living facility)-694,232 (female) + 4875,2 (FI)= 4381,684

Costs control group with the population weights of the intervention group:

Costs control group=-7803+0(treatment dummy)*(-247)+ 98,79 (number of days in the study)*78+2284*0,23(living in an assisted living facility)-1011*0,69(female)+0.22(average FI)*22159=4628,684

Individual contribution of predictors: Costs control group=-7803 (intercept)-0 (treatment dummy)+ 7704,84 (days in study)+ 546,876 (living in a nursing home)- 694,232 (female) +4875,2 (FI)= 4628,684

Predicted costs control group without correction:

Costs control group=-7803+0(treatment dummy)*(-247)+ 95,88 (number of days in the study)*78+2284*0,098(living in an assisted living facility)-1011*0,567(female)+0.21(average FI)*22159=3979,268

Individual contribution of predictors: Costs control group=-7803 (intercept)-0 (treatment dummy)+ 7478,64 (days in study)+ 223,876 (living in a nursing home)- 573,804 (female) +4653,6 (FI)= 3979,268

Chapter 8. General Discussion

The aim of this thesis was to address a number of issues related to outcome measurement in economic evaluations in elderly populations consuming health and social care. This thesis focused on instruments which are able to capture broader benefits than solely health benefits of such services, with a particular emphasis on the ICECAP-O. In order to assess the usefulness in a given population, the ICECAP-O was validated in various settings and its convergent validity and discriminant validity were thoroughly investigated. Furthermore, the potential policy relevance of the outcome measure was investigated by using both the ICECAP-O and the EQ-5D in an economic evaluation. This allowed exploring the feasibility and added value of using the ICECAP-O in an economic evaluation. This chapter first presents the main results of the thesis, after which a discussion of the methodological limitations follows. Next, a theoretical discussion of the here highlighted outcome measure is provided. Subsequently, further issues in economic evaluations of interventions aimed at frail elderly are discussed, and some policy implications are put forward.

Main results

1. How is an economic evaluation performed in elderly care using conventional outcome measures?

Chapter 2 presented an example of an economic evaluation of an intervention in social care using health QALYs as outcome measure. Even though the intervention turned out to be highly effective and the prevalence of pressure ulcers decreased, this did not translate into significant differences in health utilities between patients in the Quality Improvement Collaborative (QIC) and those receiving standard care. As a result, the collaborative was more costly and slightly more effective than standard care. It did not show a high probability of cost-effectiveness, even when assuming a high societal willingness to pay for QALY gains. In addition, the expected long-term effects of the intervention were highly sensitive to the sustained effectiveness of the QIC care. Hence, in order to give more definite estimates of cost-effectiveness, a longer time period should be considered.

2. Which instruments are potentially useful for economic evaluation in elderly care, which produce benefits beyond health?

Chapter 3 showed the results of a review of instruments potentially useful for economic evaluations in elderly populations using health and social care services. Our systematic search uncovered the ICECAP-O and the ASCOT as potentially useful preference-based instruments to measure the broad benefits of health and social care in economic evaluations of interventions in elderly care. However,

both instruments lack thorough validation and their ability to also capture health benefits remains unclear. Therefore, it was recommended to use the ICECAP-O and the ASCOT alongside more conventional (health-related) outcome measures such as EQ-5D or the SF-6D, in economic evaluations of interventions aimed at (also) producing broader well-being benefits, at least until the properties of these instruments are well established.

3. Is the ICECAP-O a valid measure of capability well-being in different settings?

Chapters 4 to 6 presented the results of validation studies of the ICECAP-O in different settings. The ICECAP-O appeared to be a valid measure of capability-wellbeing, in post-hospitalized- frail elderly, as well as in patients in psycho-geriatric nursing homes in the Netherlands and Germany. In all three studies, the ICECAP-O was closely related to HrQol and important measures of physical functioning, indicating convergent validity. The ICECAP-O thus appeared to be able to also measure health outcomes (at least partially). Additionally, in a Dutch nursing home, ICECAP-O scores were shown to be related to the Care Dependency Scale. The ICECAP-O was moreover associated with wellbeing measures, as expected. Therefore, the ICECAP-O appears to also measure wellbeing. Furthermore, the ICECAP-O was found to discriminate between groups based on the presence of multi-morbidity, depression and limitations in social activity. In addition, it discriminated between elderly with and without constraints in psycho-geriatric nursing homes. Moreover, in a German nursing home setting, the ICECAP-O was shown to be related to the ADRQL, and discriminated between groups with different care levels and dementia severities. Hence, the ICECAP-O appeared able to discriminate between elderly with different health states as well as between elderly with different levels of wellbeing.

4. Are there differences in ICECAP-O scores between different groups of respondents?

Chapters 4 to 6 also shed some light on differences in response patterns between various groups of responders to the ICECAP-O. On average, the ICECAP-O scores (based on the existing tariffs) of elderly with dementia (obtained in proxies) [1] were 0.20 points lower than the scores provided by frail, community dwelling elderly without dementia (obtained through self-completion) [2]. We observed differences between different kinds of proxy respondents. In the Dutch study in a psycho-geriatric nursing home, family respondents to the ICECAP-O did not provide different scores for restrained elderly than for non-restrained elderly, while nursing professionals did provide different scores for the two groups. Additionally, in Germany, nursing proxy gender and work experience was shown to influence responses. Therefore, the choice of proxy respondent can have important

consequences for an (economic) evaluation, and remains a highly relevant topic in the context of care provided to elderly.

5. How is an economic evaluation performed in elderly care when using the ICECAP-O as an outcome measure?

In chapter 7, we compared an instrument which was designed to integrally measure the benefits of both health and social care (the ICECAP-O) to an instrument that measures health-related benefits only (the EQ-5D) within an economic evaluation. Using QALYs as outcomes, we found little difference between capability QALYs based on the ICECAP-O scores and health QALYs based on the EQ-5D scores. After adjusting for baseline covariates, we found a small additional benefit of the investigated intervention using capability QALYs. This was not observed when using health QALYs. Although it is difficult to draw firm conclusions on the basis of one study, the study demonstrated the feasibility of performing an economic evaluation using the ICECAP-O. Moreover, the instrument appears to have the potential to measure broader outcomes than health alone, and thus may be more sensitive to differences between the intervention and usual care groups in case of effective interventions, especially when the gains relate to non-health benefits.

Limitations

This thesis has explored the important issue of outcome measurement in economic evaluations of interventions aimed at producing more than health benefits alone. A number of important limitations need noting. First, the studies in this thesis did not investigate whether the EQ-5D and the ICECAP-O measure the same or different concepts, i.e. whether they are complements or substitutes [3]. An interesting topic for future research is to explore whether the ICECAP-O captures all relevant health dimensions. Second, further research is necessary on the ICECAP-O's sensitivity to change within the context of an effective intervention, since this issue could not be explored sufficiently in the context of this thesis. Third, the comparisons of outcome measures in this thesis mainly related to the ICECAP-O and the EQ-5D. Obviously, there are other, similarly interesting and important outcome measures for economic evaluations for elderly populations receiving health and social care, such as the ASCOT, which could not be investigated in this thesis. Also, in order to have a more complete understanding of the nature of the capability QALY, it would be advisable to investigate the performance of the ICECAP-O alongside other capability measures (including the ICECAP-A), and compare them with a wide variety of other preferences based instruments, including the SF-6D and the HUI. The latter instruments potentially may be more suited to capture a

broader range of functionings and may empirically be more similar to the capability measures in a range of settings. This is especially important as the ICECAP-O has been developed for elderly, while the ICECAP-A has been developed for younger populations, and thus it should be investigated which version is suitable for which age group.

Fourth, the longitudinal studies in this thesis (chapters 2 and 7) suffer from relatively high percentages of missing data, due to dropout and item nonresponse, as is common in studies conducted in elderly populations. Such missing data limits the validity and the range of conclusions which can be drawn from such studies. Dropout is due mainly to mortality [4], which is especially strong in nursing homes reaching 30-40% [5] (chapter 2) [6, 7], resulting in censored data [8]. Censored and missing data decrease statistical power, and may bias results, as dropouts are likely to be sicker than the full-case population. Therefore, the data are likely not missing completely at random, which is required in order for full-case analyses to be valid [9]. Furthermore, in observational designs missing baseline data may also occur, causing left censoring, for example if patients enter the nursing home during the evaluation of a running quality collaborative, further decreasing the available data. Additionally, such patients probably are healthier than the full-case population, thus limiting the generalizability of the results to the less healthy drop-outs. Although there is ample research on missing data in economic evaluations [8] [10, 11] [12], currently there is no guidance on how to deal with missing data on such scale in health and social care. Further research should address this in order to make cost-effectiveness studies more feasible in long-term care, and in order to make more stringent comparisons of outcome measures such as the ICECAP-O and the EQ-5D.

Fifth, this thesis has not included studies using the ICECAP-O in economic evaluations of interventions within an RCT, which would have allowed more accurate comparisons of the ICECAP-O and the EQ-5D instruments. In fact, all the studies in this thesis were based on complex interventions [13]. All components of these interventions commonly interact with each other, making it difficult to separate the individual effects of components, and study them separately. Due to this non-separability of individual effects, and practical problems of performing RCTs of complex interventions in frail elderly populations [14], the evaluation of such interventions often follows other designs than an RCT. One way of dealing with complex interventions in economic evaluations is through accounting for cluster-effects at the level of the GPs or other institutions [15] [16, 17] as done in chapter 7. However, accounting for clustering only adequately addresses a relatively low level of complexity. In case the goal of the intervention is widespread organizational change, as was the case in the pressure ulcer collaborative in chapter 2, accounting for clustering on the ward level

may not be sufficient. In projects like quality collaboratives, 'contamination' between various wards is the explicit goal of the intervention, therefore clustering on the ward level would be inappropriate. In such cases observational designs or stepped wedged designs have to be used, which again make it more difficult to draw valid inferences on cost-effectiveness because such designs are more prone to bias. Comparing different outcome measures in the context of such interventions thus has its limitations. Moreover, while using appropriate outcomes is a necessary condition for economic evaluation of health and social services aimed at elderly, it is not a sufficient condition. A careful consideration of design and feasibility issues is also required in order for economic evaluations in health and social care for older people to become routine practice.

Theoretical considerations

The main instrument for measuring broad outcomes in economic evaluations aimed at elderly populations used in this thesis is the ICECAP-O, which is based on the capability theory. While capabilities certainly are not the only conceptualization of wellbeing, capability theory does have the useful characteristic that by measuring capability wellbeing all relevant health and non-health benefits can be captured. Nonetheless, capability theory was developed as an alternative to utility theory, and thus may be (considered to be) at odds with certain requirements for the use of instruments in the context of economic evaluations. For example, instruments then need to define a fixed number of capabilities and weigh them, which is often done on the basis of preferences. During the development of the ICECAP, such concerns have been addressed to some degree.

Conventional HrQol measures used in CUAs typically measure and value (health) functionings [18]. The ICECAP-O in contrast aims to measure and value capabilities. In order to do so, it was necessary to translate capability theory into practice, which requires a number of methodological choices, some of which may be controversial. In Sen's view [19], capability theory consists of the following main tenets: people desire goods because of their characteristics, which in turn allow people to reach certain functionings. Having such goods may also allow them to expand their capabilities, which in Sen's view is more important than the utility gained by functionings. In terms of resource allocation, according to Sen it is therefore more important to ascertain that people have and can expand their capabilities, than to try to maximize the satisfaction that they derive from functionings.

Capability theory has strong philosophical underpinnings from a liberal tradition [20] [21], and there have been extensive debates on how capability theory can actually guide resource allocation in

general. From the perspective of its usage in economic evaluations in health and social care in particular, four issues are worth mentioning. First, whether the list of capabilities to be included in an instrument is universal or context-specific. Second, whether capabilities can be traded off against each other as implicitly done within an instrument like the ICECAP-O. Third, the relationship between health and capability wellbeing deserves attention, also to ensure adequate capturing of health benefits through a capability instrument. Fourth, the impact of health and social services on capability wellbeing is important to establish. The first issue concerns the debate if lists of capabilities are appropriate to use, and if so, which elements such a list of capabilities should contain. [21]. Sen himself never provided a definitive or exhaustive list of capabilities, leaving the operational definition of capabilities to others. Nussbaum [20] did provide a list of essential capabilities which she considered as universal human rights. However, there are also other approaches of coming to relevant lists of capabilities. Some claim that capabilities have to be defined locally through participatory approaches, in order to account for the preferences of local populations [21]. These two approaches may perhaps be reconciled by looking outside capability theory, distinguishing between basic universal human needs and instrumental goals [22]. The former are relatively stable worldwide and concern physical and social dimensions [23], while the latter may vary between cultures and individuals [22]. In order to guide resource allocation within a health system, economic evaluations at a minimum require a list of capabilities which are consistent and comparable within a given health system, for example across different nursing homes. In a special case a single Visual Analogue Scale measuring capabilities worded in a consistent manner across applications may suffice. Furthermore, a universal list is preferable in order to be able to perform multi-country evaluations and to compare results between different applications. Therefore using a standardized instrument such as the ICECAP-O seems valuable.

As for the second issue, tradability of capabilities, according to Sen and Nusbaum [20] capabilities are untradeable human rights. Following this view, trading off different capabilities or ‘capability states’ is particularly problematic. This is at variance with the QALY framework, which requires weighted preference-based measurement of the quality adjusted component, which necessarily involves that capabilities are at least implicitly traded off against each other. Cookson [24] integrated the capability approach in the QALY framework, and solved this dilemma by treating weights elicited from the population as value judgments elicited from the population instead of utilities, and thus allowing capabilities to be weighted [24] [25]. This still involves an implicit trade-off between capabilities, though. Additionally, value judgments play a role in anchoring. Value judgments can also be used to normalize capability wellbeing scores to 0 (defined as the worst possible wellbeing state or capability wellbeing equal to that in the state ‘dead’) and 1 for the best

possible wellbeing state. This is similar to the normalization of health-related utility measures. How capability-based instruments can be best anchored, and how the valuation of the duration aspect of the different capability states can be incorporated into valuations, is currently unresolved theoretically and methodologically, and this needs further attention. At the same time, the methodological problem of correct anchoring is not specific for capability-based instruments, but for all ordinal preference-elicitation techniques [26].

The third issue, the relationship between health and non-health elements is also important for economic evaluations in health and social care. Cookson introduced the notion of the non-separability of health and non-health components, which were termed the health QALY and the capability QALY in chapter 7. Following Cookson, health and wellbeing – achieved through capabilities – are not separable. For example, mobility of a physically impaired person may be influenced by wealth, through the ability to hire taxis. Hence, the intrinsic value of health for the capability-QALYs depends largely on the values of other factors. In other words, the health QALY is one variable within a multivariate wellbeing function [24]. From this perspective of inseparability of health and non-health components, it is important to investigate to which degree capability-wellbeing instruments such as the ICECAP-O actually capture all relevant health dimensions.

Fourth, the impact of health and social services on capabilities and functionings is important to establish within the capability framework, in order to understand the consequences of evaluating interventions based on either capabilities or functionings. The impact of health and social services on functionings, capabilities and utilities was theoretically specified by Forder [27]. Forder considers the influence of health and social services on individuals' functionings, capabilities and utilities explicitly, as well as the relationship between these concepts. In Forder's model, utility is defined as the individual valuation of particular functionings. In the valuation of particular functionings, individual preferences play a major role. Functioning, in turn is determined by a person's capability set, for example by their economic and health circumstances. A person's preferences are determined by their capabilities and expectations regarding achieving a given functioning. Changes in capabilities have two effects on utilities: improvement in capabilities directly allows a higher level of functioning, and indirectly, through adaptation: expectations are changed, leading to a change in preferences. As peoples circumstances, or capabilities improve, - for example being able to walk after being confined to a bed - their valuation of additional improvement will be more modest. Services, in turn add a second loop: services may impact functioning directly, such as dressing or washing someone – thus making a range of capabilities in other dimensions accessible - or they can restore their capability directly, as rehabilitation services or hip operations may do. In case

individuals do not exercise a particular functioning, the indirect effect of improved capabilities (potential functionings) may be missed by only measuring functionings. Therefore, in order to capture the full effect of services, the direct measurement of capabilities seems useful.

Policy implications and further research

Economic evaluations are performed in order to advise policy-makers on the allocation of scarce healthcare resources. Such guidance is urgently necessary for social care resources as well, and even more challenging, for allocating combinations of health and social care resources. Deciding on the type of benefits that should guide such allocation decisions has profound implications for the efficient and fair allocation of resources. Concerning benefits, two issues need to be considered: (i) should health and social care benefits be considered separately, or should all health and wellbeing benefits be considered integrally, and (ii) should distribution be based on functionings or capabilities?

As to the first point of considering health and social care benefits separately or integrally, it is useful to look firstly at the goal of interventions under evaluation. If the goal is to produce wellbeing through non-health dimensions solely, non-health wellbeing measures could suffice. Likewise, if the sole goal is to produce health, measuring health-related quality of life suffices. However, when evaluating interventions in elderly populations, who typically make use of both healthcare services and social services, often both goals are present. Then, integral measurement of health and wellbeing (or wellbeing encompassing health domains), is appropriate. At this point it remains unclear, whether this can be achieved through using one wellbeing instrument, or that a combination of two or more instruments is required. This is an interesting and important avenue for further research. In order to determine if a single instrument can be the basis of resource allocation in healthcare, it is important to investigate if the currently developed wellbeing instruments are able to capture all the benefits including health benefits.

The second issue concerns whether resource allocation should be based on functionings or capabilities. Currently, economic evaluations inform policy makers on resource allocation based on functionings. The availability of capability-based instruments and the growing influence of capability theory in economic evaluations [28] [29] raises the question whether capabilities are a more appropriate basis for allocation decisions. Theoretical arguments have been made that using capabilities may be better in line with societal values, and capabilities may capture aspects of wellbeing which functioning may not [29]. There are theoretical arguments therefore to base

redistribution on capabilities instead of functionings. However, further theoretical and empirical work is necessary in order to explore if such a distinction is desirable and relevant. If there is strong evidence that capability-based measures, such as the ICECAP-O, lead to other conclusions than functioning-based instruments do, it is important to explicitly re-evaluate the basis of resource allocation.

Concluding remarks

Development, validation and ultimately using wellbeing instruments in economic evaluations remain relatively unexplored topics. On the one hand, this thesis showed that wellbeing measures are often able to capture additional benefits beyond health. On the other hand, further research is required to demonstrate the added value of such instruments in economic evaluation of health care interventions. As seen in the previous sections, the availability and usage of wellbeing instruments in economic evaluations raise a number of dilemmas for research and policy in terms of the technical properties of wellbeing instruments, and may have implications for the basis of resource allocation in healthcare, especially in social care. If appropriate outcome measures measuring wellbeing would be used, and such dilemmas are adequately resolved, then economic evaluations can fulfill the promise of giving guidance for a transparent method of redistribution in all forms of health care and social care. Such a process would significantly strengthen the basis of the decisions underlying introduction, cutting and retention of health, as well as social care services.

References

1. Makai, P., *Capabilities and quality of life in Dutch psycho-geriatric nursing homes: an exploratory study using a proxy version of the ICECAP-O.* Quality of life research, 2012. **21**(5): p. 801-812.
2. Makai, P., et al., *A validation of the ICECAP-O in a population of post-hospitalized older people in the Netherlands.* Health and Quality of Life Outcomes, 2013. **11**(1): p. 1-11.
3. Davis, J.C., et al., *A comparison of the ICECAP-O with EQ-5D in a falls prevention clinical setting: are they complements or substitutes?* Quality of Life Research, 2013: 2013. **22**(5):p. 969-977.
4. de Klerk, M., *Ouderen in instellingen: landelijk overzicht van de leefsituatie van oudere tehuisbewoners*. 2005: Sociaal en Cultureel Planbu.
5. Makai, P. *Cost-effectiveness of a pressure ulcer quality collaborative.* Cost effectiveness and resource allocation, 2010. **8**(1): p. 11.
6. Neyens, J.C.L., et al., *A multifactorial intervention for the prevention of falls in psychogeriatric nursing home patients, a randomised controlled trial (RCT).* Age and Ageing, 2009. **38**(2): p. 194-199.
7. Huizing, A.R., et al., *A cluster-randomized trial of an educational intervention to reduce the use of physical restraints with psychogeriatric nursing home residents.* Journal of the American Geriatrics Society, 2009. **57**(7): p. 1139-1148.
8. Glick, H.A., *Economic evaluation in clinical trials*. 2007: Oxford University Press.
9. Rubin, D.B., *Multiple imputation for nonresponse in surveys*. Vol. 307. 2009: Wiley. com.
10. McGhan, W.F., et al., *The ISPOR good practices for quality improvement of cost-effectiveness research task force report.* Value in Health, 2009. **12**(8): p. 1086-1099.
11. Noble, S.M., W. Hollingworth, and K. Tilling, *Missing data in trial-based cost-effectiveness analysis: the current state of play.* Health Economics, 2012. **21**(2): p. 187-200.
12. Wijeysundera, H.C., et al., *Techniques for estimating health care costs with censored data: an overview for the health services researcher.* ClinicoEconomics and outcomes research: CEOR, 2012. **4**: p. 145.
13. Craig, P., et al., *Developing and evaluating complex interventions: the new Medical Research Council guidance.* BMJ: British Medical Journal, 2008. **337**.
14. Reelick, M.F., et al., *How to perform a preplanned process evaluation for complex interventions in geriatric medicine: exemplified with the process evaluation of a complex falls-prevention program for community-dwelling frail older fallers.* Journal of the American Medical Directors Association, 2011. **12**(5): p. 331-336.
15. Bachmann, M.O., et al., *Methods for analyzing cost effectiveness data from cluster randomized trials.* Cost Effectiveness and Resource Allocation, 2007. **5**(1): p. 12.
16. Gomes, M., et al., *Methods For Covariate Adjustment In Cost-Effectiveness Analysis That Use Cluster Randomised Trials.* Health Economics, 2012. **21**(9): p. 1101-1118.
17. Gomes, M., et al., *Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials.* Medical Decision Making, 2012. **32**(2): p. 350-361.
18. Brouwer, W.B.F., et al., *Welfarism vs. extra-welfarism.* Journal of health economics, 2008. **27**(2): p. 325-338.
19. Sen, A., *Choice, welfare and measurement*. 1982, Cambridge, MA: Harvard University Press.
20. Nussbaum, M., *The quality of life*. 1993, Oxford: Clarendon Press.
21. Robeyns, I., *The capability approach: a theoretical survey.* Journal of Human Development, 2005(6): p. 93-114.

22. Ormel, J., et al., *Quality of life and social production functions: A framework for understanding health effects.* Social science & medicine, 1997. **45**(7): p. 1051-1063.
23. Gullone, E. and R.A. Cummins, *Universality of subjective wellbeing indicators*. Social indicators reasearch series. 2002, Dordrecht: Kluwer. 196.
24. Cookson, R., *QALYs and the capability approach.* Health economics, 2005. **14**(8): p. 817-829.
25. Coast, J., et al., *Investigating Choice Experiments for Preferences of Older People (ICEPOP): evaluative spaces in health economics.* Journal of health services research & policy, 2008. **13 Suppl 3**(Journal Article): p. 31-37.
26. Brazier, J., et al., *Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale.* The European Journal of Health Economics, 2012. **13**(5): p. 575-587.
27. Forder, J.E. and J. Caiels, *Measuring the outcomes of long-term care.* Social science & medicine, 2011. **73**(12): p. 1766-1774.
28. NICE. *The social care guidance manual* 2013 [cited 2013 11.11]; Available from: http://publications.nice.org.uk/the-social-care-guidance-manual-pmg10/incorporating-economic-evaluation.
29. Coast, J., R. Smith, and P. Lorgelly, *Should the capability approach be applied in health economics?* Health economics, 2008. **17**(6): p. 667-670.

Summary

Given that the number of elderly is expected to grow in the coming decades, leading to an increased pressure on health budgets and health care professionals, the question of which healthcare interventions can be funded in health and social care for elderly becomes important. Economic evaluations can assist in answering this question and thus in resource allocation within the health care sector. Such evaluations are performed relatively routinely for curative health interventions (especially medicines), but less common in the context of elderly populations receiving social care alongside curative care. An important reason for this is the lack of appropriate outcome measures which capture the effect of social services. This risks misrepresentation of the benefits of such care and consequently suboptimal decision making. More recently, wellbeing instruments have been developed, aimed at measuring and valuing the broader benefits related to interventions in elderly care. An important example is the ICECAP-O, which is a central measure in this thesis. The ICECAP-O measures capabilities (what a person is able to do if he or she so desires). The aim of this thesis is to address some issues related to outcome measurement in economic evaluations in elderly populations consuming health and social care. More specifically, we present (i) a cost-effectiveness analysis using traditional QALYs in social care, (ii) a review of wellbeing instruments for health and social care for older people, (iii) validation studies of the ICECAP-O in different settings and groups of respondents, and (iv) an economic evaluation using the ICECAP-O.

Chapter 2 presented an example of an economic evaluation of an intervention in social care using traditional 'health QALYs' as outcome measure. Even though the intervention turned out to be highly effective and the prevalence of pressure ulcers decreased, this did not translate into significant differences in health utilities between patients in the Quality Improvement Collaborative (QIC) and those receiving standard care. As a result, the collaborative was more costly and slightly more effective than standard care. It however did not show a high probability of cost-effectiveness, even when assuming a high societal willingness to pay for QALY gains. In addition, the expected long-term effects of the intervention were highly sensitive to the sustained effectiveness of the QIC care. Hence, in order to give more definite estimates of cost-effectiveness, a longer follow-up period should be considered.

Chapter 3 showed the results of a review aimed to retrieve potentially useful instruments for economic evaluations in elderly populations using health and social care services. Our systematic

search uncovered the ICECAP-O and the ASCOT as potentially useful preference-based instruments to measure the broad benefits of health and social care in economic evaluations of interventions in elderly care. However, both instruments lack thorough validation and their ability to fully capture health benefits remains unclear. Therefore, it was recommended to use the ICECAP-O and the ASCOT *alongside* more conventional (health-related) outcome measures, such as the EQ-5D or the SF-6D, in economic evaluations of interventions aimed at (also) producing broader well-being benefits, at least until the properties of these instruments are well established.

Chapters 4 to 6 presented the results of validation studies of the ICECAP-O in different settings. The ICECAP-O appeared to be a valid measure of capability-wellbeing in Dutch post-hospitalized- frail elderly, as well as in patients in psycho-geriatric nursing homes in the Netherlands and Germany. In all three studies, the ICECAP-O was closely related to health-related quality of life measures and important measures of physical functioning, indicating convergent validity. The ICECAP-O thus appeared to be able to also measure health outcomes (at least partially). Additionally, in a Dutch nursing home, ICECAP-O scores were shown to be related to the Care Dependency Scale. The ICECAP-O was moreover associated with wellbeing measures, as expected. Therefore, the ICECAP-O also indeed appears to measure wellbeing. Furthermore, the ICECAP-O was found to discriminate between groups based on the presence of multi-morbidity, depression and limitations in social activity. It also discriminated between elderly with and without constraints in psycho-geriatric nursing homes. Furthermore, in a German nursing home setting, the ICECAP-O was shown to be related to the ADRQL, and discriminated between groups with different care levels and dementia severities. Hence, the ICECAP-O appeared able to discriminate between elderly with different health status as well as between elderly with different levels of wellbeing.

Chapters 4 to 6 also shed some light on differences in response patterns between various groups of responders to the ICECAP-O. On average, the ICECAP-O scores (based on the existing tariffs) of elderly with dementia (obtained in proxies) were 0.20 points lower than the scores provided by frail, community dwelling elderly without dementia (obtained through self-completion). We observed differences between different kinds of proxy respondents. In the Dutch study in a psycho-geriatric nursing home, family respondents to the ICECAP-O did not provide different scores for restrained versus non-restrained elderly, while nursing professionals did provide different scores for the two groups. Additionally, in Germany, nursing proxy gender, and work experience was shown to influence responses. Therefore, the choice of proxy respondent can have important consequences

for an (economic) evaluation, and therefore remains a highly relevant topic in the context of care provided to elderly.

In chapter 7, we compared an instrument which was designed to integrally measure the benefits of both health and social care (the ICECAP-O) to an instrument that measures health-relatedbenefits only (the EQ-5D) within an economic evaluation. Using QALYs as outcomes, we found little difference between capability QALYs based on the ICECAP-O scores and health QALYs based on the EQ-5D scores. After adjusting for baseline covariates, we found a small additional benefit of the investigated intervention using capability QALYs. This was not observed when using health QALYs. Although it is difficult to draw firm conclusions on the basis of one study, the study demonstrated the feasibility of performing an economic evaluation using the ICECAP-O. Moreover, the instrument appears to have the potential to measure broader outcomes than health alone, and thus may be more sensitive to differences between the intervention and comparator groups in case of effective interventions, especially when the gains relate to non-health benefits.

Chapter 8 answered the main research questions of the thesis, discussed the limitations, gave some theoretical implications, offered potential policy implications, and drew some conclusions. Without suitable outcome measures, capturing all relevant benefits of elderly care interventions, it is difficult if not impossible to perform a full economic evaluation. The ICECAP-O seems to be a relatively suitable instrument for economic evaluations in interventions where social care plays an important role. Still, despite encouraging findings, more research is encouraged, also because the research presented in this thesis had a number of limitations. Measuring appropriate outcomes is just the first step leading to routine economic evaluations of health and social care interventions targeted at elderly; feasibility and design issues have to be dealt with as well. It needs noting that the ICECAP-O is just one possible application of the capability theory in healthcare, and other applications are also possible.

The main policy implication of the current research is that wellbeing instruments should be investigated and used more broadly in economic evaluations. In addition, more research is necessary to justify a choice for basing resource allocation decisions in health care on wellbeing measures in general and capability measures (rather than functionings) in particular. Overall, using wellbeing instruments in economic evaluations is an underexplored topic. Using wellbeing

instruments can lead to a transparent method of resource allocation in forms of health care and social care, considering all relevant benefits.

Samenvatting

In de komende jaren zal het aantal ouderen naar verwachting toenemen en dit zal in Europa leiden tot een verhoogde druk op gezondheidsuitgaven en de gezondheidszorg. Daardoor wordt de beleidsvraag welke interventies (collectief) gefinancierd kunnen worden in de gezondheidszorg en de langdurige zorg voor ouderen belangrijk. Economische evaluaties kunnen helpen bij het beantwoorden van deze vraag en dus bij de toewijzing van middelen binnen de zorg. Dergelijke evaluaties zijn gebruikelijk als het gaat om curatieve interventies (met name geneesmiddelen), maar worden minder vaak uitgevoerd in de context van zorg voor ouderen, die naast curatieve zorg ook langdurige zorg ontvangen. Een belangrijke reden hiervoor is het gebrek aan geschikte uitkomstmaten om de effecten van langdurige zorg te meten. Wat langdurige zorg oplevert wordt daardoor soms verkeerd voorgesteld, hetgeen tot suboptimale besluitvorming kan leiden.

Traditioneel worden binnen economische evaluaties zogenaamde gezondheidsheidsgerelateerde kwaliteit van leven maten gebruikt (‘gezondheids QALYs’). Recentelijk zijn een aantal instrumenten ontwikkeld om uitkomsten van ouderenzorg beter te kunnen meten. Deze instrumenten meten en waarderen baten van interventies in de ouderenzorg op het gebied van welzijn van ouderen in bredere zin. Een belangrijke nieuwe maat voor welzijn van ouderen is de ICECAP-O. De ICECAP-O meet welzijn van ouderen door na te gaan wat een persoon kan doen als hij of zij dat wenst, met andere woorden zijn of haar *capabilities*. Dit proefschrift analyseert een aantal vragen rondom het meten van uitkomsten in economische evaluaties van interventies gericht op oudere populaties die zowel gezondheidszorg als langdurige zorg ontvangen. Met het oog op deze vraagstelling komen de volgende vier onderwerpen aan de orde: (i) een economische evaluatie met behulp van traditionele ‘gezondheids QALYs’ in de langdurige zorg, (ii) een overzicht van welzijnsmaten voor economische evaluaties in oudere populaties die zowel gezondheidszorg als langdurige zorg ontvangen, (iii) toepassing en validatie van de ICECAP-O in verschillende contexten en (iv) toepassing van de ICECAP-O in een economische evaluatie van een ketenzorginterventie bij kwetsbare ouderen.

Hoofdstuk 2 beschrijft een economische evaluatie van een interventie om decubitus terug te dringen in de langdurige zorg waarbij traditionele ‘gezondheids QALYs’ werden gebruikt als uitkomstmaat. Hoewel de interventie effectief bleek te zijn en de prevalentie van decubitus terugbracht, vertaalde dit zich niet in significante verschillen tussen patiënten in de Quality Improvement Collaborative (QIC) en standaard zorg in termen van QALYs. Uit de resultaten bleek dat het QIC duurder en iets meer effectief was dan standaard zorg. Uit deze studie bleek tevens dat er een lage kans was dat een dergelijke interventie tot kosteneffectief is, zelfs bij een hoge

maatschappelijke bereidheid om te betalen voor QALY winsten. Daarnaast waren de verwachte effecten van de interventie op lange termijn zeer gevoelig voor het al of niet voortduren van de nieuwe werkwijzen en resultaten, oftewel het borgen van de QIC zorg. Daarom moet een langere follow-up periode worden overwogen om een meer definitieve raming van de kosteneffectiviteit te geven.

Hoofdstuk 3 toont de resultaten van een review van mogelijk bruikbare instrumenten voor economische evaluaties in populaties van ouderen die zowel gezondheidszorg als langdurige zorg ontvangen. Uit deze systematische review kwam naar voren dat de ICECAP-O en de ASCOT mogelijk nuttige instrumenten zijn om in economische evaluaties de brede uitkomsten te meten. Voor beide instrumenten zijn preferentie-wegingen beschikbaar. Voor beide instrumenten ontbreken grondige validatie studies tot dusver echter. Daarnaast bleek het onduidelijk te zijn in hoeverre deze instrumenten de volledige gezondheidsbaten vast kunnen leggen. Dit is een belangrijk nadeel ten opzichte van bestaande, meer conventionele (gezondheidsgerelateerde) meetinstrumenten voor gezondheid zoals EQ-5D of de SF6D. Daarom is het verstandig om in economische evaluaties van interventies die (ook) gericht zijn op verbetering van welzijn, de ICECAP-O en de ASCOT te gebruiken in combinatie met conventionele (gezondheidsgerelateerde) meetinstrumenten. Zowel gezondheids- als welzijnsuitkomstmaten gebruiken in economische evaluaties lijkt wenselijk totdat de eigenschappen van laatstgenoemde instrumenten uitvoeriger onderzocht zijn.

Hoofdstukken 4 tot 6 betreffen validatie studies van de ICECAP-O in verschillende settings. De ICECAP-O bleek een valide meetinstrument te zijn voor *capability*–welzijn, zowel wanneer toegepast in thuiswonende kwetsbare ouderen na een ziekenhuisopname, als bij ouderen in psychogeriatrische verpleeghuizen in Nederland en Duitsland. In alle drie studies hingen de scores op de ICECAP-O nauw samen met scores op gezondheidsgerelateerde uitkomstmaten, zowel gezondheidsgerelateerde kwaliteit van leven als belangrijke uitkomstmaten voor lichamelijk functioneren. Dit duidt op convergente validiteit, met andere woorden: de ICECAP–O geeft (gedeeltelijk) een goede indicatie van belangrijke gezondheidsdimensies. Daarnaast was, zoals we hadden verwacht, de ICECAP-O geassocieerd met welzijnsmaten. Daarom is het aannemelijk dat de ICECAP-O inderdaad ook welzijnsdimensies meet. Bovendien bleek de ICECAP-O ook gerelateerd te zijn aan de Care Dependency Scale binnen een Nederlandse verpleeghuis setting. Een tweede doel van de analyse was om te onderzoeken in welke mate de ICECAP-O in voldoende mate gevoelig is voor verschillende doelgroepen. Uit de resultaten bleek dat de ICECAP-O onderscheidend werkt bij de aanwezigheid van multimorbiditeit, depressie en beperkingen in sociale activiteit. Daarnaast liet de ICECAP-O onderscheid zien tussen ouderen met en zonder onrustbanden die in

psychogeriatrische verpleeghuizen wonen. Verder bleek uit een de studie onder ouderen in een Duits psychogeriatrisch verpleeghuis dat de ICECAP-O ook samenhing met de ADRQL, een dementie-specifiek kwaliteit van leven instrument. De ICECAP-O was tevens gevoelig voor verschillen in zorgniveau en ernst van dementie. Concluderend kan worden gesteld dat de ICECAP-O in de hier uitgevoerde studies een valide meetinstrument bleek voor uitkomsten op het gebied van welzijn, in staat om onderscheid te maken tussen ouderen met een verschillende gezondheidsstatus en tussen ouderen met verschillende welzijnsniveaus.

Naast samenhang en onderscheidend vermogen, blijkt uit hoofdstukken 4 tot 6 ook dat de antwoordpatronen van respondenten op de ICECAP-O kunnen verschillen afhankelijk van de manier van invullen: door cliënten zelf of door een andere zogenaamde 'proxy respondent'. Gemiddeld scoren ouderen met dementie gemeten via proxies 0.20 punten (op een schaal van 0 tot 1) lager op de ICECAP-O dan thuiswonende ouderen die zelf de vragenlijst hebben ingevuld. We vonden ook verschillen tussen proxy-respondenten. In de studie binnen een Nederlands psycho-geriatrisch verpleeghuis beoordeelden verzorgenden de capabilities van ouderen in een onrustband op basis van de ICECAP-O slechter dan van ouderen zonder een onrustband, terwijl in de scores door familieleden op de ICECAP-O geen verschillen werden gevonden tussen de twee groepen. Bovendien leken in de studie uitgevoerd in Duitsland geslacht en werkervaring van de proxy respondent de scores te beïnvloeden. Samenvattend, uit deze studies is gebleken dat de keuze van de proxy respondent belangrijke gevolgen kan hebben voor de verkregen ICECAP-O scores. De keuze voor een proxy blijkt dus een zeer relevant onderwerp in de context van de zorg voor ouderen.

Hoofdstuk 7 beschrijft een economische evaluatie waarin de ICECAP-O, een instrument dat is ontwikkeld om breder welzijn te meten, werd vergeleken met de EQ-5D, een instrument dat gezondheidsgerelateerde kwaliteit van leven meet. We vonden weinig verschil tussen *capability* QALY's op basis van de ICECAP-O scores en *gezondheids*QALY's op basis van de EQ-5D scores. Na aanpassing voor achtergrond-kenmerken zoals geslacht vonden we een kleine winst in de interventiegroep in termen van *capability* QALY's. Dit werd niet waargenomen bij het gebruik van de gezondheidsQALY's. Hoewel het voorbarig is om harde conclusies te trekken op basis van deze enkele studie, lijkt het haalbaar om een economische evaluatie uit te voeren met behulp van de ICECAP-O. Een belangrijk voordeel van de ICECAP-O is dat dit instrument de potentie heeft om bredere effecten te meten dan alleen gezondheidseffecten. Hierdoor kan bij effectieve interventies de ICECAP-O gevoeliger zijn voor de verschillen tussen de interventie en vergelijkingsgroepen dan

gezondheirdsgerelateerde uitkomstmaten, vooral wanneer het gaat om niet gezondheids-gerelateerde welzijnsbaten.

Hoofdstuk 8 bevat de discussie van dit proefschrift. Hierin werden de belangrijkste bevindingen naar aanleiding van het proefschrift bediscussieerd, alsook beperkingen van de studies aangegeven. Tevens werd ingegaan op een aantal mogelijke implicaties voor de toewijzing van middelen in de ouderenzorg. Een belangrijke conclusie van dit proefschrift is dat het moeilijk is, zo niet onmogelijk, om een volledige economische evaluatie uit te voeren zonder geschikte uitkomstmaten die alle relevante baten van interventies binnen de ouderenzorg meten. Tevens blijkt uit dit proefschfit dat de ICECAP-O een relatief geschikt instrument lijkt voor economische evaluaties in interventies waar de langdurige zorg een belangrijke rol speelt. Toch, ondanks bemoedigende bevindingen, is meer onderzoek noodzakelijk, aangezien het in dit proefschrift beschreven onderzoek een aantal beperkingen had. Het meten van geschikte uitkomsten is slechts de eerste stap die leidt tot routinematig uitvoeren van economische evaluaties van interventies gericht op ouderen binnen de gezondheidszorg en langdurige zorg. Kwesties van haalbaarheid en design moeten ook worden behandeld. Belangrijk op te merken is dat de ICECAP-O slechts een mogelijke toepassing is van de *capability*-theorie in de gezondheidszorg en in de toekomst worden wellicht ook andere maten ontwikkeld om welzijn in de vorm van capabilities te meten.

De bevindingen uit dit proefschrift hebben ook gevolgen voor gezondheidszorgbeleid. Een eerste beleidsimplicatie is dat welzijnsinstrumenten moeten worden onderzocht en vaker gebruikt zouden moeten worden in economische evaluaties. Daarnaast is verder onderzoek nodig om toewijzing van middelen in de gezondheidszorg te rechtvaardigen op basis van welzijnsuitkomsten in het algemeen en *capabilities* (in plaats van functionings) in het bijzonder. Over het algemeen is het gebruik van welzijns-instrumenten in economische evaluaties een onderbelicht onderwerp. Gebruik van welzijnsinstrumenten kan leiden tot een transparante methode voor de toewijzing van middelen in de vorm van gezondheidszorg en langdurige zorg, rekening houdend met alle baten.

Acknowledgements

Having finished this book is a very important personal and individual achievement; all the hard work now is done. However, many people have contributed to this book in one way or another, and I would like to thank all of them.

First I would like to thank my supervisors, who due to their varied disciplinary background helped me finish this work.

Anna, you were formerly my co-supervisor and finally my supervisor. From the first day your drive and passion for research with older people was inspiring. Thanks to your background in quantitative sociology, I learned to integrate different disciplinary approaches into my work on quality of life. Even though I feel that supervising me was a real challenge for you, but we still made it.

Marc, thanks to your background as one of the pioneers in cost-effectiveness studies you really complemented Anna's input with a keen eye for improving the analyses as well as the reporting. I've much appreciated your very efficient method of co-supervision. You consistently read my papers before a meeting, always magnifying the one or two points most in need of improvement, and allowing me to learn from these. Also thank you for taking the time on occasion to delve into the analysis; actually taking the time to sit behind a PC with me. You strengthened my analytical skills on a practical level and gave me ideas to build on when I really was lost. We did not have the chance to often work really closely, but when you were there at a discussion, you were presence was felt.

And, I'd like to thank my first supervisor, Werner. Your creativity and experience as a health economist was vital in completing this book, since applying cost-effective analysis techniques to interventions in elderly care is simply not straightforward. Werner, your close reading improved my writing, allowing me to formulate my thoughts far more clearly and readable than I previously thought possible. I also thank you for your flexibility, when making an appointment made more sense in Utrecht or in Veenendaal, you made this possible. But most importantly, that you helped to keep this book focused even though there were many opportunities to steer off course during the long journey this has been. I consider myself fortunate to have worked with you, a supervisor who radiates quality, even though I did not realize how much time and work that meant.

I also wish to thank the members of the manuscript committee, Maureen Rutten, Jan van Buschbach, and Joanna Coast. Joanna, thank you for letting me translate and use the ICECAP shortly after its development.

Franziska, I thank you for the opportunity to supervise your master thesis. I am also grateful that you allowed me to expand on and modify your work and flesh out a full article.

Many thanks to my other co-authors on the articles in this thesis, Roland, Elly (twice!), Job, Isabelle, Willemijn, René, and Eddy, and people who often commented on a manuscript or I could brainstorm with: Pepijn, Mathilde, and Erwin.

Sarah, I thank you for sharing a room with me for 3 years. Thank you for the many discussions on academic life, and reflections on the world around us in general. You always gave me a different perspective when I was dealing too deeply with an issue. And almost everything I know about teaching on an academic level I learned from you. Tineke, my other psychologist: thank you for being a listening ear whenever I needed one. Thanks to your positive insight, this always helped me to rebound. And thank you for correcting my Dutch, when such services were extensively needed. And thank you both for all the fun moments, when work was unimportant.

Joris, I thank you for taking me along for a ride on a consultancy project. Working on VEKTIS's large database was a new experience. Finishing the analyses for the policy report on a short term with a hard deadline was really an exciting challenge, even though we will never publish this work in an academic journal. But these are separate worlds, and I learned a lot about this other world through our cooperation.

I would also like to thank the other members of the Care for Better research team, Esther, Teun, Annemiek for giving me another perspective on the evaluation of complex interventions.

I'm really grateful to all the people who helped me collect the data for the thesis. First, I would like to thank all the nurses who helped me collect the data. Second, I would like to thank the project leaders in Vilans for allowing me access to the projects and the data, Marjolein van Vliet in particular. Third, I was grateful to the student assistants I worked with, Karin, Marloes, and Lindsey.

Also, I would like to thank the members of the HSMO, and all the other colleagues I had lunch with during the years at the BMG, or whom I can thank a wonderful congress experience, or gave me

some good advice over the years. In particular I want to thank Martin, Honorine, Henk, Lisbeth, Linda, Martina, Dirk, Marcello, Birghitta, Caroline, Margreet, Saskia, Kim, Saskia, Renske, Claudine, Apostolos, Ernst, Samantha, Antoinette, Hester, Jolanda, Rik, Wim, Annemarie, Jaqueline, Jane, Gerrit, Marleen, Trea, Ilaria, Richard, Jeroen. And a special thanks to the secretaries, Alexia, Anne, Lisa, and Karin, who helped me with the difficult task of making appointments with many busy people.

Many thanks my colleagues of the Geriatrics department in Nijmegen, as facilitators of this book and simply for the open and carefree environment where I could finish this book. Marcel, I am particularly thankful for your support and patience with allowing me to be a postdoc without a PhD this long, even as I was raking in BONZO points with all the publications. As for the carefree environment: I thank everyone I shared on office with over the years, Marieke, Marjolijn, Martje, Miriam, Teun, Sharon, Maurice, and beyond my office Cynthia, Lia, Bianca, Gerrita, Jennifer in particular. Not to mention the various interns, and the other participants of the talent management program and finally all the others too numerous to name.

Laszlo, although you were not involved in this thesis, I would still like to thank you for the inspiration that you gave me as a student to pursue research on health economics/health services research issues. Without your help, I would never been here today.

I would also like to thank my friends and family. Dana, Ludmilla, Alex, Helia, Gyorgyi, Tamas, Andras, and especially my father and my mother. Thank you for your help you gave in the past few years and for your advice along the way.

Kati, you were also my greatest support in this whole process. I thank you for patiently listening to the all the complaining about how long this was taking, helping me at all the difficult moments, and for taking care of Elon on the weekends when this thesis was finished.

Elon, thanks for giving daddy the necessary peace of mind during the time we were together.

Curriculum Vitae

Peter Makai was born in Budapest on September 11, 1981. He studied economics and political science at the Corvinus University of Budapest, with a specialization in Health Economics. Additionally, after moving to the Netherlands, he obtained a second Master's degree in Public Health from the University of Maastricht in 2007, with a specialization in Health Policy, Organization and Economics. After working for the Netherlands Institute of Health Services Research, Peter joined the Institute of Health Policy and Management at the Erasmus University of Rotterdam in 2008, as a junior researcher.

At the Institute of Health Policy and Management, Peter worked on a number of projects concerning elderly, mostly within the Care for Better Evaluation team. This work formed the basis of his PhD thesis, concerning economic evaluation of interventions in elderly populations, focusing on outcome measures. Peter's work was published in Social Science & Medicine, Quality of Life Research, Health and Quality of Life Outcomes, and other journals focusing on economic evaluations, quality of life and quality of health-care. Peter also supervised a master thesis and lead working groups in a number of bachelor courses.

Peter currently works as a researcher at the Department of Geriatrics at the Radboud University Medical Center in Nijmegen, focusing on projects concerning e-health, outcome measures for frail elderly, and economic modelling. There, he also participated in the Da Vinci talent development program of the Radboud University Medical Center, aimed at promising young researchers.

Peter lives together with his partner, Katalin Katona. Peter and Katalin have a two year old son, Elon.

PhD Portfolio

PhD Student Peter Makai

Department: Institute of Health Policy and Management

PhD period iBMG: 2008-2012

Promoters: Prof. dr. Werner Brouwer and Prof. dr. Anna Nieboer

Courses	Year
Cost-effectiveness modelling methods, Maastricht University	2009
Statistics, Tinbergen institute	2009
Econometrics 1 Tinbergen institute	2010
Econometrics 2 Tinbergen institute	2010
Workshop on implementation and spread of improvements in health care, Erasmus University	2011
Probleemgestuurd onderwijs [Problem-oriented teaching], Erasmus University,	2008
Basiscursus didactiek [Basic didactics], Erasmus University Rotterdam	2008
Academic writing	2010
Invited presentations	
2nd ICECAP users' group workshop University of Birmingham	2011
Conference participation	
28-th ISQUA International Conference Hong Kong, China (Presentation)	2011
8-th ECHE conference Helsinki, Finland (Presentation)	2010
26-th ISQUA international Conference Dublin, Ireland (Poster)	2009
Teaching, working groups	
Inleiding methoden en technieken van onderzoek [Introduction to Research Methods]	2008, 2009, 2011
Kwantitatief Gezondheidszorgonderzoek [Quantitative healthcare research]	2011-12
Kwaliteit en doelmatigheid [Quality and economic evaluation]	2010
Organisatiewetenschappen [Organizational Science]	2009
Thesis supervision	
Master Thesis in Health Economics, Policy and Law – Specialization Health Economics	2011

Publications by the Author

An up to date list of publications by the author can be found here:

http://scholar.google.nl/citations?user=xLDaddUAAAAJ&hl=en&oi=ao

Published/Accepted for publication

***Makai, P.** Beckenbans, F. Exel, J. WBF Brouwer QoL in German nursing homes: a further investigation of ICECAP-O *PLOS-One* (Accepted for publication)

Makai P., Melis, RJF, Olde-Rikkert, MGM. (2014) Technical Difficulties and Evaluating e-health interventions. *JAMA Internal Medicine* 174 (2) 304.

Melis, RJF., **Makai P.,** Perry, M. (2014) Letter to the editor in response to "Association of Patient Preferences for Participation in Decision Making With Length of Stay and Costs Among Hospitalized Patients" *JAMA Internal Medicine* (accepted for publication)

***Makai, P.,** Brouwer, WBF., Koopmanschap MA., Stolk, EA., Nieboer, AP. (2014) Quality of Life instruments for economic evaluations in health and social care for older people: a systematic review *Social Science & Medicine* 102 83-93

***Makai, P.**, Koopmanschap, MA., Brouwer, WBF., Nieboer, AP. (2013) A Validation of the ICECAP-O in a population of post-hospitalized older people *Health and Quality of Life Outcomes* 11(1) 57

Hofman, CS., Melis, RJF., Boter, H., Buurman, BM, de Craen, AJM., Olde-Rikkert MGM., **Makai, P.,** Donders, ART. (2013) *European Geriatric Medicine* (4) S72-S73.

Robben, SHM., **Makai, P. (2012)** Development and implementation of a new healthcare social media platform for individual care coordination for frail elderly *European Geriatric Medicine (*4) S60.

Makai, P., Cramm, MJ., van Grotel, M., Nieboer, AP. (2012) Effectiveness and sustainability of innovative projects. Forthcoming: *International Journal of Quality in Health Care.*

***Makai, P.,** Brouwer, WBF., Koopmanschap, MA., Nieboer, AP. **(2012)** Capabilities and Quality of Life in Dutch psycho-geriatric nursing homes: An exploratory study using a proxy version of the ICECAP-O *Quality of life Research* 21(5):pp. 801-12.

***Makai, P.,** Koopmanschap, MA., Bal, R., Nieboer, AP. **(2010)** Cost-effectiveness of a Quality Improvement Collaborative *Cost Effectiveness and Resource Allocation* 8:11

Duckers, M., **Makai, P.,** Vos, L., Groenewegen, P., Wagner, C. (2009) Longitudinal analysis on the development of hospital quality management systems in the Netherlands *International Journal for Quality in Health Care* 21(5), 11 pp. 330-340

Makai, P., Boncz, I, Klazinga, N.S., Wagner, C., Gulacsi, L. **(2009)** Quality management and patient safety; survey results from 102 Hungarian hospitals *Health Policy* 90, pp. 175-180

Makai, P., Patient Safety: a policy framework **(2007)** in: Nemec, J (eds) *Lessons and Recommendations for Improvement: Central and Eastern European Public Administration and Public Policy* NISPAcee Bratislava, 2007 pp. 65-86.

Under review

***Makai P.,** Looman WM, Melis, RJF, Adang, EMM., Stolk, EA., Fabriccotti, IN. (2014) Cost-efectiveness of integrated care in frail elderly: does choice of instrument matter? (*European Journal of Health Economics*)

Makai P. Perry M. Schers, HJ. Olde-Rikkert, MGM., Melis, RJF. Early evaluation of an e-health intervention in chronic care for frail elder people (Journal of Medical internet Research)

Hofman, CS., **Makai, P.,** Boter, H., Buurman, BM., de Craen, AJM, Olde-Rikkert, MGM, Donders, ART, Melis, RJF. Establishing a composite endpoint for measuring the effectiveness of Geriatric Interventions Based on Older Persons' and Informal Caregivers' Preference Weights: A vignette study (BMC Geriatrics)

Makai P. Perry, M., Schers, H. Heinen, MM. Olde-Rikkert, MGM., Melis, RJF. Which frail older patients use online health communities and why? (Medical Decision Making)

*Publications in this thesis

Professional publications in other languages:

Robben, SHM., Heinen, MM., Makai, P., Olde-Rikkert MGM., Perry, M., Schers, HJ., Melis, RJF. (2013) Reducing fragmentation in the care of frail older persons: the successful development and implementation of the Health and Welfare Information Portal Journal of Gerontology and Geriatrics (Tijdschrift voor Gerontologie en Geriatrie) 44 (2) 59-71.

Houkes, A., **Makai, P.** Rijken, M. **(2008)** Chronisch zieken en eigen betalingen, een internationalen onderzoek NiVEL SEO economisch onderzoek report nr. 2008-48.

Gulácsi L, **Makai P.**, Péter T, Jakobsen Á, Boncz I, Quality and quality improvement in Healthcare 2005 (In Hungarian**;** ***Journal of Physician Continous Education*** **(Orvostovábbképző Szemle), 2006;13:12:13-22.**

Makai, P., Gulácsi L. Patient Safety in Hungarian Hospitals: results of 103 hospitals (In Hungarian) *Health Management Review* (Egészségügyi Gazdasági szemle) 2006. 44.1:53-6

Makai, P., Gulácsi, L. Basics of Patient Safety and the Danish System II. (In Hungarian) *Informatics and Management in Healthcare* (Informatika és Menedszment az egészségügyben) 2005 4.9:15-18.

Makai, P., Gulácsi, L. Basics of Patient Safety and the Danish System I-II. (In Hungarian) *Informatics and Management in Healthcare* (Informatika és Menedszment az egészségügyben) 2005 4:.6.:22-25.

Makai, P., Gulácsi L. 2005 Patient Safety – International review of literature (In Hungarian) *Health Management Review* (Egészséggazdasági Szemle) 2005. 43:3.16-23.

Ilonszki, G., Jáger, K., **Makai, P.** Before Campaign, after Campaign: Representation, female representatives and the media (In Hungarian) Jel-Kép 2005.3: 3-21.