Development and validation of a composite endpoint as a quality indicator in elderly care

Cynthia S. Hofman
The work described in this thesis was funded by the Dutch National Care for the Elderly Programme, coordinated and sponsored by ZonMw, Organization for Health Research and Development, The Netherlands (grant number 60-6190098-302).

Financial support for the publication and distribution of this thesis was kindly provided by the Department of Geriatric Medicine, Radboud university medical center and Vilans, Centre of Expertise on Long-Term Care.

ISBN
978-90-9029698-2

Development and validation of a composite endpoint as a quality indicator in elderly care

Cynthia Sabrina Hofman
geboren op 9 juni 1981
te Bergen op Zoom

2016 Cynthia Sabrina Hofman

All rights reserved. No part of this thesis may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording or otherwise, without prior written permission of copyright owners or the author.
Development and validation of a composite endpoint as a quality indicator in elderly care

Doctoral Thesis

To obtain the degree of doctor from Radboud University Nijmegen on the authority of the Rector Magnificus, According to the decision of the Council of Deans to be defended in public on Tuesday 31 May 2016 at 12:00 hours

by

Cynthia Sabrina Hofman
Born on June 9, 1981 in Bergen op Zoom

Promotor
Prof. dr. M.G.M. Olde Rikkert

Copromotoren
Dr. R.J.F. Melis
Dr. A.R.T. Donders

Manuscriptcommissie
Prof. dr. G. Westert (voorzitter)
Prof. dr. P.L.C.M. van Riel
Prof. dr. A.P. Nieboer (Erasmus Universiteit Rotterdam)
Contents

Part 1 Introduction 9
  Chapter 1 General introduction and outline 11

Part 2 Development of TOPICS-CEP 23
  Chapter 2 Establishing TOPICS-CEP; The health state preferences of older persons and informal caregivers 25
  Chapter 3 Adjusting TOPICS-CEP; The influence of rater characteristics on TOPICS-CEP’s preference weights 43

Part 3 Relevance of TOPICS-CEP 59
  Chapter 4 Development of a method to compare one self-assessment versus multiple peer-assessments 61
  Chapter 5 Comparing self-assessment versus peer-assessments 73
  Chapter 6 Comparing health state preferences of older persons and healthcare professionals 81

Part 4 Psychometric properties of TOPICS-CEP 101
  Chapter 7 Validation of TOPICS-CEP 103

Part 5 Summary and Discussion 121
  Chapter 8 Summary and general discussion 123
  Chapter 9 Nederlandse samenvatting | Dutch Summary 149

Part 6 Supplementary 159
  Dankwoord | Acknowledgements 161
  Curriculum Vitae 167
  Curriculum Vitae in Dutch 170
  Publications 173
Part 1
General introduction
Chapter 1

General introduction and outline
1.1 Background

The mean life expectancy of the Dutch population has increased from 71 to 81 years over the past fifty years. The proportion of people aged 65 years or over in the Netherlands has increased from 771 thousand in 1950, representing 8% of the Dutch population, to 3 million (18%) in 2015, and is expected to reach 4.5 million (26%) within the next 20 years.1

The aging population has led to an increase of healthcare delivery, as older patients are more likely to suffer from multiple diseases, take more medications, undergo more procedures, and use more healthcare than younger patients. Over the next few decades the number and diversity of older people with complex healthcare needs will increase even further and will place unprecedented demands on aging services and our entire healthcare system.

To improve care, quality of life, and self-reliance among older persons, the Dutch Ministry of Health, Welfare, and Sport commissioned the National Care for the Elderly Programme.2 This programme funded over 60 research and implementation projects as well as the development of The Older Persons and Informal Caregivers Minimum Data Set (TOPICS-MDS). As part of TOPICS-MDS initiative, a uniform dataset measuring key outcomes in health, wellbeing, and health services utilization was created and administered to all older persons and caregivers participating in these studies.3

In this thesis, we describe research project funded by the National Care for the Elderly Programme, the Older People’s Relevant Outcome of Care Score (OPROCS) project which aimed to develop and validate a preference-weighted composite endpoint (CEP) for TOPICS-MDS (henceforth referred to as TOPICS-CEP) as an indicator for value in healthcare.

1.2 Quality in healthcare

In relation to healthcare there are many definitions of quality used. The Institute of Medicine (IOM) has proposed a definition that largely captures the features of many other definitions and received wide acceptance.

“Quality of care is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge”.4
For evaluating quality of healthcare the most dominant conceptual model, on which many other frameworks are based, was developed by Donabedian [Figure 1]. According to this model, information about quality of healthcare can be drawn from three categories: structure, process, and outcomes. ‘Structure’ refers the context in which healthcare is delivered, including hospital buildings, staff, financing, and equipment. ‘Process’ describes the transactions between patients and providers throughout the delivery of healthcare. Finally, ‘Outcomes’ denotes to the effects of healthcare on the health status of patients and populations. These categories are the classifications for the types of information that can be obtained in order to judge whether the quality of care is poor, fair, or good. As shown in Figure 1 the categories influence and interact with each other.

The concept of health outcomes is included in both the IOM definition of quality in healthcare as well as Donabedian’s model. This is not surprising with the transition from volume to value driven healthcare which is central to most healthcare reform efforts around the world. As firmly stated by Michael Porter,

“...In any field, improving performance and accountability depends on having a shared goal that unites the interest and activities of all stakeholders. [...] In most fields, the preeminent goal is value [...] Defining and measuring value is essential to understanding the performance of any organisation and driving continuous improvement [...] Since value depends on results, not inputs, value in healthcare is measured by the outcomes achieved [...] Achieving good patient health outcomes is the fundamental purpose of healthcare. Measuring, reporting, and comparing outcomes is perhaps the most important step toward unlocking rapid outcome improvement and making good choices about reducing costs. Outcomes are the true measures of quality in health care [...] outcome measurement is perhaps the single most powerful tool in revamping the healthcare system”.6

This backdrop underlines that we should strive to maximize the outcomes of healthcare in order to improve the quality of healthcare. We should not waste time and money on the use or implementation of interventions that are not as beneficial as other interventions. Hence, it is important to perform Comparative Effectiveness Research (CER) in which the outcomes (effectiveness, benefits, and harms) of different treatment options are compared. CER has the purpose to provide guidance and assist patients, providers, and policymakers to make more informed decisions and to improve healthcare at both individual and population levels.7

1.3 Measuring Outcomes in Healthcare

Administrative data can be a powerful source to follow patients’ paths through the healthcare system over time. This allows a focus on both costs as well as outcomes, such as death, (re)admissions, complications. Yet, the patient reported outcome measure (PROM) is a truly patient-centred data source. PROMs assess the status of a patient's health condition which comes directly from the patient without interpretation of the patient’s response by a clinician or anyone else. It builds on the idea that the outcome, as experienced by the patient, is the ‘gold standard’ by which interventions and health services have to be judged. PROMs have been around for decades and are widely used in medical science to determine whether a new treatment is more effective (have better outcomes) than another treatment. Yet, it is only recent that the Bupa hospital group in the UK decided to used PROMs for evaluating the clinical outcomes of their procedures. Rather than asking the doctor whether the operation went well, the patients who have undergone the operation were asked if the surgery relieved their pain and restored their previous physical ability.

Figure 1 The Donabedian model of measuring healthcare system performance
Before measuring the outcomes of healthcare delivery, it is important to choose or develop the right instrument or tool (e.g. PROM) for the job. In order to do so, one needs to answer the following questions: “Do you want to assess the outcome for a specific condition or for a more general concept?” and “Do you want to assess the outcome of one specific domain or for multiple domains?”.

To assess the outcome for a specific patient group one needs to select or develop a condition-specific instrument. These instruments focus on symptoms and signs that reflect the status of a given medical condition. Generic instruments, on the other hand, are more comprehensive and assess a single aspect or multiple aspects of a general concept, e.g. general wellbeing. In contrast to condition-specific instruments, generic instruments can be applied in different types of diseases, interventions, and populations.

To circumvent these problems preference-weighted composite endpoints/outcomes are established and used. Preference-weighting refers to placing value judgments on health states achieved or avoided by treatment has been suggested also in elderly populations. These weights reflect the relative importance of various domains when compared with an anchor, such as perfect health, quality of life, or general wellbeing. The fact that most definitions of quality of care consistently stress the importance of patient-centredness underlines the necessity of using preference-weights to combine multidimensional items if the aim is to measure the value of care. Hence, in order to assess quality of healthcare, the outcome measure used needs to reflect the value of the change accomplished according to the patient.

Obtaining preference-weights
An essential step involved in conducting preference weighting is the choice of judges. Although, there are no fixed rules about whose opinions should be used, it is important to keep in mind that the choice of a reference panel is crucial, because different groups may generate quite different weights. For example, patients place more weight on instrumental activities of daily life (IADL, e.g. shopping, travelling, preparing a meal), whereas professionals weight the activities of daily life (ADLs, e.g. bathing, shaving, eating) more heavily. Potential raters include the patients themselves, peers, healthcare professionals, family members, regulators, policymakers, and the general public. It seems logical to base preference-weights of a PROM on the judgements made by patients. Yet, it is interesting to compare them with preference-weights based on the judgements of others, such as healthcare professionals, as this may underline the importance of using PROMs to increase value in healthcare.

After deciding on the judges, a choice needs to be made regarding the methodology to obtain the preference-weight. Discrete choice experiments (DCE) are often used to obtain preferences. The method requires respondents to state their choice over sets of hypothetical alternatives with each alternative described by several characteristics known as attributes. The responses are used to derive the value placed on each attribute. DCE studies specifically designed for and conducted with older persons remain relatively rare compared with those designed, conducted and reported with general adult samples. Milte and colleagues sought to investigate the reliability of DCE responses from older persons with varying levels of cognition and the threshold level of cognitive ability required for an older person to reliably complete a DCE. Their results showed that the presence of mild cognitive impairment did not have a significant effect on the consistency of responses to the DCE, however, their findings provide important preliminary evidence relating to the effect of mild cognitive impairment of DCE responses for older persons.

Another option to derive preference-weights is by means of vignette studies (Box 1). Vignettes (or profiles) are short descriptions of a person or a social situation, which contain precise references to what are thought to be the most important factors in the decision- or judgment-making processes of respondents. Over the last few years, the number of vignette studies increased in various fields of application, such as psychology, sociology, marketing, education and training, and clinical practice. These kinds of studies are typically used to study the beliefs, values, or judgments of respondents. Hence, they are useful to derive preference weights for single index values. Because of the high number and interdependence of attributes, we preferred to use vignettes studies instead of DCE’s to establish TOPICS-CEP.
To establish and validate TOPICS-CEP the following research questions were defined:

1. What are the preference weights of TOPICS-CEP components based on the health valuations made by older persons, informal caregivers, and healthcare professionals, respectively?
2. Are there significant differences when the preference weights of TOPICS-CEP components based on the health valuations made by older persons, informal caregivers, and healthcare professionals are compared?
3. Which characteristics of the raters from the group of older persons and informal caregivers have a significant influence on the preference weights of TOPICS-CEP components?
4. What is the construct validity of TOPICS-CEP?
   - Are TOPICS-CEP scores related to other often used outcome measures, such as the EQ-5D utility score?
   - Can TOPICS-CEP discriminate between groups, e.g. older persons with versus without dementia?

Outline
This thesis consists of four sections: the development of TOPICS-CEP, the importance of a preference-weighted composite endpoint, the validation of TOPICS-CEP, and a general discussion.

Chapter 2 and 3 focus on the development of TOPICS-CEP. Chapter 2 presents the results of a vignette study, which was conducted to derive and compare the preference-weights based on the health state preferences of older persons and informal caregivers. In Chapter 3 the influence of our participants’ gender, age, and self-perceived health on the preference weights are examined.

Chapter 4 introduces an adapted Bland-Altman approach to evaluate possible patterns of discord between two measurement methods with an unequal number of observations per case. Chapter 5 and 6 concentrate on how the perspective of the rater (self-rating versus external rating, and perspective of lay persons versus healthcare professionals) may affect their valuation of the outcome and thus explores the importance of the establishment of TOPICS-CEP. In a letter to the editor Chapter 5 assessments made by older persons themselves are compared with those made by peers. Chapter 6 presents the results of a vignette study which was conducted to compare the preference weights based on the health state preferences of healthcare professionals with those of older persons and informal caregivers.

1.4 Aims and thesis outline

In order to fulfil the need for a valid endpoint that combines TOPICS-MDS outcomes from eight domains, we developed and validated TOPICS-CEP. As the primary aim of this project was to develop an endpoint that considers the desirability of health outcomes, we used health valuations made by older persons and informal caregivers to obtain the preference weights for the eight domains (TOPICS-CEP components). Vignette studies were conducted in which older persons, informal caregivers, and healthcare professionals were asked to value described health states. The health valuations made by healthcare professionals were used to explore the relevance of a preference-weighted composite endpoint. As a secondary aim, this project validated TOPICS-CEP.
Chapter 7 addresses the psychometric properties of TOPICS-CEP. Convergent validity, known group validity, and generalizability of TOPICS-CEP are examined by evaluating a large heterogeneous sample of older persons aged 65 years and older.

Chapter 8 provides a summary of the main findings of this thesis and discusses the implications of these results.

1.5 References


Part 2
The development of TOPICS-CEP
Chapter 2

Establishing TOPICS-CEP; The health state preferences of older persons and informal caregivers

Published as:
Establishing a composite endpoint for measuring the effectiveness of geriatric interventions based on older persons’ and informal caregivers’ preference weights: A vignette study


CHAPTER 2 HEALTH STATE PREFERENCES OF OLDER PERSONS

2.1 Background

The number of elderly is increasing worldwide, due to increasing life-expectancy. Ageing of our populations will have a major impact on the organization and delivery of health care, as healthcare systems have to meet the needs of geriatric patients, while the shortage of healthcare workers is likely to grow. To restrain healthcare spending and improve the quality of care it is necessary to measure, report, and compare outcomes in healthcare delivery. However, comparing intervention outcomes for elderly is a great challenge because their health states are complex with problems in multiple domains, e.g. morbidities and physical functioning, and interventions often target a broad range of domains. A generic measurement instrument with a composite endpoint (CEP) would, therefore, be helpful to compare the effectiveness of different geriatric interventions.

With the increasing proportion of elderly and its impact on the organization and delivery of health care in mind, the Dutch Ministry of Health, Welfare, and Sport commissioned the National Care for the Elderly Programme (NCEP) with the aim to develop a more proactive, integrated healthcare system for older patients. Over 60 scientific projects were conducted under this programme. To achieve standardized outcome measurements within the NCEP, The Older Persons and Informal Caregivers Survey Minimal Dataset (TOPICS-MDS) instrument was constructed and integrated into the research protocols. TOPICS-MDS was developed by a small working group and includes validated instruments that are frequently used in older populations. Additionally, the instrument’s content and utility was evaluated by an independent multi-disciplinary panel with expertise in gerontology, epidemiology, biostatistics and health services research and a plain language expert was commissioned to revise the instrument for clarity and readability.

Although TOPICS-MDS is used to gather uniform data of the NCEP projects in a National Database (collecting dataset of over 32,000 elderly persons), there is currently no consensus on how to combine and weight the information from multiple outcome domains into a CEP. This means that the effectiveness of the projects can only be evaluated comparing the multiple individual domains separately and not the overall outcome. Using a single TOPICS-MDS item or item subset to compare outcomes leads to confusion when competing projects demonstrate different patterns of effect, as the items or domains may not be equally important. For example, it is difficult to decide which intervention is more effective if one intervention reduces the number of functional limitations and reduces pain sensation, while another improves social functioning and emotional wellbeing. Hence, for optimal comparison of the NCEP projects’ effectiveness a CEP that accounts for the relative importance of different outcomes is required.
In this study, we explore how multidimensional TOPICS-MDS outcomes from the Care receiver questionnaire can be weighted and combined into a CEP. The relative importance of the outcomes are reflected by preference weighting of TOPICS-MDS information compared with an anchor. We opted for best and worst general wellbeing (GWB) as the anchor, because improving patients’ GWB is a goal all stakeholders share. Basically, GWB is a concept that covers a broad spectrum of health and it is influenced by various health outcome domains. Since the purpose of healthcare is to meet the needs of patients, our main focus should be on outcomes that matter to the patients. However, as relatives of elderly persons often deliver informal care and serve as proxies, e.g. when the elderly person has a low cognitive status, we are interested in the relative importance of the items according to them as well. Thus, the aim of this study is to examine the preference weights of elderly persons and informal caregivers and explore whether their preference weights differ.

2.2 Methods

Ethical approval

The Medical Ethics Committee of the Radboud University Medical Center formally stated that this study was exempt from ethical review (Radboud University Medical Center Ethical Committee review reference number: CMO: 2010/244).

Study design

This study has three components that are similar to those described in the valuation study of Brazier, Roberts, and Deverill. Firstly, TOPICS-MDS questionnaire for care receivers has been reduced in size and complexity. Secondly, a valuation study was conducted to derive the preference weights for the TOPICS-MDS outcomes. However, in contrast to the study of Brazier et al. we used a numeric rating scale to value the health states. Thirdly, the results of the valuation study were used in a model to calculate the composite endpoint for the vignette cases.

Vignette study

In our valuation study vignettes were being used. Over the last few years, the number of vignette studies increased in various fields of application, such as psychology, sociology, marketing, education and training, and clinical practice. These kinds of studies are typically used to study the beliefs, values, or judgments of respondents. Hence, they are useful to derive preference weights for single index values. Vignettes are short descriptions of a person or a social situation which contain precise references to what are thought to be the most important factors in the decision- or judgment-making processes of respondents.

Participants

A sample of 124 community dwelling elderly aged ≥ 65 years and 76 informal caregivers participated as raters. We used a rather broad definition of informal caregiver: “An informal caregiver provides voluntary and unpaid care on a structural basis to a care recipient with physical, mental or psychological limitations who is most often a relative, friend or neighbour. The provided care involves assisting the care receiver with tasks (s)he would do him-/herself in normal health” derived from the NCEP website. In this study only informal caregivers who provided care to a care receiver aged ≥ 65 years were included. The participants were eligible if they mastered the Dutch language sufficiently. This was explored by the trained research assistants during first contact with the participants. When communication in Dutch was possible (asking questions regarding marital status, living arrangements, and family) the participants were included in the study.

The participants were recruited and the data was collected by four academic centres: Radboud University Medical Center, University Medical Centre Groningen, Academic Medical Centre Amsterdam, and Leiden University Medical Centre. These centres were spread over the Netherlands, and cover both urban and more rural parts of the country. To ensure a representative sample the participants were recruited in hospital outpatient clinics, general practitioner (GP) practices, nursing homes, day care facilities, and via the internet (recruitment messages were placed online). Written informed consent was obtained from each participant before the start of the vignette study.

Material

In total 292 vignettes were constructed based on data of real persons (cases) derived from TOPICS-MDS National database. As the participants were asked to read the vignettes by themselves we used a large font size (14 points) and double spacing. In general, each vignette included 46 items and described elderly persons covering eight health domains: morbidity, functional limitations, emotional wellbeing, pain experience, cognitive functioning, social functioning, self-perceived health and self-perceived quality of life (QOL) and four demographic characteristics: gender, age, marital status, and living situation. Table 1 gives an overview of the health domains, items per domain, and levels per item which were included in the vignettes and used in the analyses.

By using empirical data only vignettes with plausible health state combinations were constructed. The cases described in the vignettes had a mean age of 81.4 years (SD 5.72) and 58.6% (N=171) was female. The majority of these cases were either married (42.8%, N=125) or their partner was deceased (42.8%, N=125), and 39.7% (N=116) lived independently with someone, e.g. a partner or family member.
Table 1 The domains, items, and aggregated items included in the vignettes with the descriptive.

<table>
<thead>
<tr>
<th>Outcome domains</th>
<th>Vignette items</th>
<th>Outcome levels</th>
<th>Descriptives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morbidity (Local and national health monitor)11</td>
<td>Presence of: Dementia; Depression; Incontinence; Stroke, CVA or TIA; Hip fracture; Panic or anxiety disorder; Dizziness with falling; Vision disorder; Asthma; Osteoporosis; Diabetes; Arthritis; Heart failure; Form of cancer; Complaints due to benign enlarged prostate; Fracture other than hip fracture; Hearing disorder</td>
<td>Number of diseases present Range: 0-17</td>
<td>Mean: 3.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD: 2.0</td>
</tr>
<tr>
<td>Functional limitations (modified KATZ-ADL Index)12</td>
<td>Needing help with: Brushing hair; Going to the toilet; Taking medication; Sitting down and getting up from chair; Getting dressed; Travelling; Handling finances; Grocery shopping; Walking about; Taking a bath or shower; Housekeeping; Preparing a meal; Eating; Using the telephone</td>
<td>Number of limitations in (I)ADL Range: 0-15</td>
<td>Mean: 3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD: 3.5</td>
</tr>
<tr>
<td>Emotional wellbeing (RAND-36, mental health subscale)13</td>
<td>Feeling down; Feeling blue; Feeling nervous; Feeling happy; Feeling calm</td>
<td>Raw mental health score Range: 5-30</td>
<td>Mean: 10.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD: 4.7</td>
</tr>
<tr>
<td>Pain experience (Single item EQ-5D+C)14</td>
<td>Pain experience</td>
<td>No</td>
<td>130 (44.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Moderate</td>
<td>127 (43.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Severe</td>
<td>34 (11.7)</td>
</tr>
<tr>
<td>Cognitive functioning (Single item EQ-5D+C)14</td>
<td>Cognitive problems</td>
<td>No</td>
<td>222 (77.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Moderate</td>
<td>64 (22.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Severe</td>
<td>2 (0.7)</td>
</tr>
<tr>
<td>Social functioning (Single item RAND-36)15</td>
<td>Social activities hampered by physical health or emotional problems</td>
<td>Never</td>
<td>217 (75.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rarely</td>
<td>19 (6.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sometimes</td>
<td>29 (9.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mostly</td>
<td>6 (2.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Continuously</td>
<td>18 (6.3)</td>
</tr>
<tr>
<td>Self-perceived health (Single item, RAND-36)15</td>
<td>Self-perceived health in general</td>
<td>Excellent</td>
<td>18 (6.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Very good</td>
<td>20 (7.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Good</td>
<td>27 (44.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reasonable</td>
<td>108 (37.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Poor</td>
<td>14 (4.8)</td>
</tr>
<tr>
<td>Satisfaction with quality of life (QOL) (Single item formed using phrasing similar to self-perceived health question, RAND-36)</td>
<td>Self-perceived QOL in general</td>
<td>Excellent</td>
<td>26 (9.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Very good</td>
<td>40 (14.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Good</td>
<td>170 (59.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reasonable</td>
<td>41 (14.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Poor</td>
<td>9 (3.1)</td>
</tr>
</tbody>
</table>
CHAPTER 2 HEALTH STATE PREFERENCES OF OLDER PERSONS

First, a mixed model with random effects was constructed to obtain the preference weights for all raters, for both elderly raters and informal caregivers (N=200). We used the GWB scores as dependent variable and the eight outcomes as independent variables (fixed factors). Then, we repeated the analysis with the variable “informal caregiver” (0/1; no/yes) as additional independent variable to explore the influence of the informal caregiver role on the preference weights using interaction effects. The participants who fulfilled the role as informal caregiver and were aged ≥ 65 years were included in the group informal caregivers.

Stage II

For the majority of the 292 vignette cases (95.5%, N=279) we were able to calculate a TOPICS-CEP score (using the unstandardized coefficients found in stage I (Table 2) as preference weights) as they had no missing data points. Among these 279 cases 86.3% (N=241) had rated their own GWB.

Differences in mean TOPICS-CEP scores between sexes and between age groups were explored using T-test and ANOVA, respectively. The same was done for the differences in mean self-assessment scores. Differences between the calculated TOPICS-CEP scores and the self-assessment scores were examined using a paired sample T-test and Pearson’s correlation.

2.3 Results

Raters

The participants included in the group elderly raters (N=124) had a mean age of 78.3 years (SD 6.70) and 62.9% (N=78) was female. The majority of these raters were married (59.7%, N=74) and 60.5% (N=75) lived independently with someone, e.g. their spouse or a relative. The elderly raters gave their own GWB a mean score of 7.7 (SD 0.92).

The 76 informal caregivers who participated in this study had a mean age of 63.0 years (SD 12.14), 72.4% (N=55) was female, and 92.1% (N=70) took care of a family member. The informal caregivers gave their own GWB a mean score of 7.2 (SD 1.15).

Completion rates

There were 2400 numerical rating scale valuations completed by the participants out of the 2400 possible (124×12 for elderly raters and 76×12 for informal caregivers). All 200 participants were capable to read the vignettes themselves and language comprehension was not an issue.
Stage I
The linear mixed regression model that combined the eight outcomes showed that p-value of the outcomes: morbidities, limitations in daily functioning, emotional wellbeing, cognitive functioning, and self-perceived health was smaller than 0.05 (Table 2).

The linear mixed regression model that combined the eight outcome and the additional variable “informal caregiver” showed that the p-value of the outcomes: morbidity, functional limitations, emotional wellbeing, cognitive functioning, and self-perceived health was smaller than 0.05. In addition, the interactions between the “informal caregiver” variable and each of the domains were not significant (p>0.05).

Examining the residuals we found no large departures from normality nor evidence for the presence of outliers. Based on the narrow confidence intervals multicollinearity between the outcome domains of the CEP is unlikely.

Stage II
Among the 282 of 292 vignette cases for whom a TOPICS-CEP could be established and who rated their own GWB, the minimum TOPICS-CEP score calculated was 4.72 and the maximum score was 8.45 [Mean (±SD): 6.95 (0.73)]. The overall distribution of the TOPICS-CEP scores was tailed to the left (not shown). The distribution of the TOPICS-CEP scores was more normalized within the age group aged at least 85 years than within the younger age groups (Figure 1). Mean TOPICS-CEP scores (±SD) significantly differed across sex and between age groups [Men: 7.10 (0.76); Women: 6.84 (0.67); p=0.00] [<80: 7.15 (0.65); 80-84: 6.90 (0.75); ≥85: 6.74 (0.75); p=0.00].

Overall, the self-assessment scores had a broader range compared to TOPICS-CEP scores. The correlation matrices indicate moderate correlation between the two scores for all age groups. Pearson correlation test on whole group (r = 0.52, p = 0.00).

Of the 249 cases who rated their own GWB the majority gave their own GBW a score of 7.0 (33.6%) or 8.0 (35.7%) (Figure 1). Mean self-assessment scores (±SD) did not significantly differed across sex and between age groups [Men: 7.46 (1.40); Women: 7.26 (1.22); p=0.25] [<80: 7.15 (1.33); 80-84: 7.29 (1.18); ≥85: 7.28 (1.38); p=0.63].

Compared to TOPICS-CEP scores, the self-assessment scores had a broader range and a significantly higher mean [Range: 1.50-10.0; Mean difference (±SD): 0.34 (1.10), p=0.00]. The two scores were moderately correlated (r=0.52, p=0.00).

### Table 2: Results of mixed linear analyses with as dependent variable the GWB scores given by all raters.

<table>
<thead>
<tr>
<th></th>
<th>Older persons and informal caregivers (N=200)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unstandardized coefficient</td>
<td>Standardized coefficient</td>
</tr>
<tr>
<td>Intercept</td>
<td>9.03</td>
<td>1.54</td>
</tr>
<tr>
<td>Morbidities</td>
<td>-0.14</td>
<td>-0.10</td>
</tr>
<tr>
<td>Functional limitations</td>
<td>-0.12</td>
<td>-0.09</td>
</tr>
<tr>
<td>Emotional wellbeing</td>
<td>-0.04</td>
<td>-0.03</td>
</tr>
<tr>
<td>Pain experience</td>
<td>-0.13</td>
<td>-0.10</td>
</tr>
<tr>
<td>Cognitive functioning</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>Social functioning</td>
<td>-0.16</td>
<td>-0.12</td>
</tr>
<tr>
<td>Self-perceived health</td>
<td>-0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>Self-perceived QOL</td>
<td>-0.03</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

Note 1: The unstandardized coefficients estimate the increase in the dependent variable (general wellbeing scores) per unit increase of the environmental predictors.

Note 2: The unstandardized coefficients are the preference weights of the components included in the composite endpoint.

*p<0.05 **p<0.001.
2.4 Discussion

Our primary findings support that a CEP for TOPICS-MDS Care receiver questionnaire can be established based on the preference weights of both elderly persons and informal caregivers, which were derived by means of our vignette study. The narrow confidence intervals of our estimated parameters suggest that there was enough information present in the dataset, hence, that the sample size was large enough.

Our secondary analysis indicates that using a CEP that can be calculated based on assessments from patients (e.g. by means of a questionnaire) is related to GWB, yet measures a different concept as the correlation is of medium strength.

In contrast to previous research, elderly persons and informal caregiver (or family members) share the same preferences when it comes to the assessment of a subjective measure such as GWB.25-28 Perhaps, the discrepancy between our findings and findings in other studies can be explained by the fact that in our study there was no personal relationship between the informal caregiver and elderly patient (cases described in the vignettes) that could influence the assessment made, e.g. response shift bias or caregiver burden.26, 29-31 We asked elderly persons and informal caregivers to assess the GWB of neutral cases, while in other studies elderly persons were asked to assess their own GWB and informal caregivers were asked to assess the GWB of their loved ones.

Our results and implications need to be interpreted in light of several limitations. First, the vignettes we used in this study were based on empirical data derived from the TOPICS-MDS National Database, which means that some combinations of the outcome domains were not represented, e.g. a case with dementia, dizziness with falling, hip fracture and fracture other than hip fracture who do not have any functional limitations. However, by using empirical data only vignettes with plausible health state combinations were constructed. Second, the distribution of marital status and living arrangement characteristics over the participants are similar to those over the Dutch population (≥ 65 years).32 However, in our study the elderly raters had a mean age of 78.3 years and 62.9% of the sample was female, while the mean age of the Dutch elderly population is 74.3 years and 56% of this population is female.32 Hence, women and elderly aged 80 years and over are overrepresented in our sample. Previous research has shown individual variation in health state preferences influenced by gender and age.33, 34 Therefore, we will explore the influence of our raters’ characteristics on the TOPICS-CEP’s preference weights in our next study.

Third, even though the most important health domains from TOPICS –MDS Care receiver questionnaire were included in the CEP there may be aspects that influence the general wellbeing of elderly that are not included in the questionnaire and the CEP, such as isolation and loneliness.
The benefits of using TOPICS-MDS and its CEP are that a range of important endpoints will be collected and incorporated in a single metric, which can index the overall impact of interventions according to elderly persons and informal caregivers in a standardized way and reduce sample size requirements. Hence, establishing the value of interventions will be easier and more objective. Similar to other composite endpoints, such as the Disease Activity Score in rheumatology, the use of TOPICS-CEP may improve analysis of clinical trials and it may even be applicable to clinical care.\textsuperscript{21, 26}

For future research we suggest to explore the responsiveness of the established CEP and its prognostic value. Also, we advise to compare the preference weights of older persons and informal caregivers derived in this study with those of healthcare providers.

\section*{2.5 Conclusions}

TOPICS-MDS has been successfully incorporated into all NCEP research projects. Until now, the effectiveness of the projects could only be compared per item, item subset, or comparing multiple endpoints. With the establishment of TOPICS-CEP for the care receiver questionnaire that accounts for the relative importance of different outcomes based on the preferences of elderly persons and informal caregivers, optimal comparison of NCEP project’s effectiveness can be realized. A syntax to calculate the TOPICS-CEP score will be available on the TOPICS-MDS website in the latter half of 2013.\textsuperscript{7}

Besides NCEP projects, other projects in the geriatric field can use the TOPICS-MDS to collect research data and the TOPICS-CEP allowing standardized assessment of patient outcomes reflecting the preferences of elderly persons and informal caregivers.\textsuperscript{7}

\section*{2.6 References}


Chapter 3

Adjusting TOPICS-CEP; The influence of rater characteristics on TOPICS-CEP’s preference weights

Published as:
The influence of age on health valuations: the older olds prefer functional independence while the younger olds prefer less morbidity


Abstract

Background
To assess the effectiveness of geriatric interventions, The Older Persons and Informal Caregivers Survey – Composite Endpoint (TOPICS-CEP) has been developed based on health valuations of older persons and informal caregivers. This study explored the influence of the raters’ age on the preference weights of TOPICS-CEP’s components.

Methods
A vignette study was conducted with 200 raters (mean age ± standard deviation: 72.5±11.8 years; 66.5% female). Profiles of older persons were used to obtain the preference weights for all TOPICS-CEP components: morbidity, functional limitations, emotional wellbeing, pain experience, cognitive functioning, social functioning, self-perceived health, and self-perceived quality of life. The raters assessed the general wellbeing of these vignettes on a 0–10 scale. Mixed linear regression analysis with interaction terms was used to explore the effects of raters’ age on the preference weights.

Results
Interaction effects between age and the TOPICS-CEP components showed that older raters gave significantly (P<0.05) more weight to functional limitations and social functioning and less to morbidities and pain experience, compared to younger raters.

Conclusion
Researchers examining effectiveness in elderly care need to consider the discrepancies between health valuations of younger olds and older olds when selecting or establishing outcome measures. In clinical decision making, health care professionals need to be aware of this discrepancy as well. For this reason we highly recommend shared decision making in geriatric care.

3.1 Introduction
Comparing intervention outcomes for older persons is challenging because their health states are complex, with problems on multiple health domains, and because interventions often target a broad range of these domains. However, comparative effectiveness research in geriatric care becomes more straightforward when clinically important outcome parameters are combined into a multidimensional preference-weighted outcome measure. Such a composite endpoint (CEP) can efficiently deal with the issue of multiplicity, eg, if more than one outcome is important for effectiveness evaluation or if an intervention has the potential to improve more than one health domain. By using a preference-weighted multifaceted outcome measure, the relative importance of the various outcomes is taken into account.

A CEP is of particular interest for The Older Persons and Informal Caregivers Survey Minimum Data Set (TOPICS-MDS), a data repository containing data from 41 research projects participating in the Dutch National Care for the Elderly Program. To promote comparability between these research studies, a preference-weighted CEP for the TOPICS-MDS for care receivers (referred to as TOPICS-CEP) was established by means of a vignette study and based on the health-state valuations of older persons and informal caregivers. Briefly, TOPICS-CEP is a preference-weighted index ranging from 0 (worst possible general wellbeing) to 10 (best possible general wellbeing). It combines 42 data points from TOPICS-MDS covered by eight components: morbidities (a list of 17 predefined conditions used in the Netherlands), functional limitations (Katz index of independence), emotional wellbeing (mental health subscale of the RAND-36), pain experience (pain dimension of the EQ-5D), cognitive problems (cognition dimension of the EQ-5D+c), social functioning (item 10 from the RAND-36), self-perceived health (item 1 from the RAND-36), self-perceived quality of life (QOL) (phrasing similar to the self-perceived health item from the RAND-36). The data points included in TOPICS-CEP regard all the variables (or items) from TOPICS-MDS for older persons that carry information relevant for understanding an individual’s outcome. This excludes demographics and health service utilization. Excluding these components was based on the rationale that demographics such as sex and age and health service utilization cannot be influenced by health care delivery. Detailed information about the development of TOPICS-CEP can be found elsewhere.

Previous research has shown that variation in health-state preferences is influenced by various characteristics such as sex, age, and current health. This variability means that the value of a particular health status depends on who served as participants.
when evaluating the various health states. The variability can be explained by, eg, one's reference point, perspective, and coping strategies. TOPICS-CEP was established based on the health-state preferences of a heterogeneous group of raters. Results from our previous study indicated that the preference weights of older persons and informal caregivers were not significantly different.

This study aimed: 1) to explore whether the age of our raters systematically influenced the preference weights of TOPICS-CEP components, and 2) to examine how age-based preference weights affected TOPICS-CEP scores. The framework we used can be found in Figure 1.

### 3.2 Methods

#### Ethical approval

The Medical Ethics Committee of the Radboud University Medical Center formally stated that this study was exempt from ethical review (Radboud University Medical Center Ethical Committee review reference number: CMO: 2010/244).

#### Study design

A vignette study was conducted to obtain the preference weights for the eight TOPICS-CEP components: morbidities, functional limitations, emotional wellbeing, pain experience, cognitive problems, social functioning, self-perceived health, and self-perceived QOL. The participants rated the general wellbeing (GWB) of case vignettes, which were short descriptions or profiles of older persons.

#### Participants

Two hundred persons who mastered the Dutch language sufficiently participated as raters. They were recruited and the data were collected at four academic centers (Radboud University Medical Center, University Medical Centre Groningen, Academic Medical Center, and Leiden University Medical Centre) spread over the Netherlands, covering both urban and more-rural parts of the country. All participants provided written informed consent.

#### Material

In total, 292 different versions of the same basic vignette (same items) were constructed to obtain a wide spectrum of health states. The vignettes were based on data of a sample of cases derived from the TOPICS-MDS national database, which consists of pooled data from various research projects that differ across study design, sampling framework, and inclusion criteria. By using empirical data, only vignettes with

---

**Figure 1** Study framework.

Note: Framework: combining eight components into one generic composite endpoint (TOPICS-CEP) and the influence of rater characteristics.
Does age affect TOPICS-CEP scores?

Based on our findings, we constructed a model that represented the health-state preferences of 65-year-old persons (TOPICS-CEP-65) and a model that represented those of 85-year-old persons (TOPICS-CEP-85), the older olds (Table 2). These models showed that on average the TOPICS-CEP-65 score decreased by 0.15 points per additional disease present in a vignette case, while the TOPICS-CEP-85 score decreased by 0.12 points. Thus, the component morbidities had a larger negative impact on the GWB scores given to the vignettes by younger persons than those given by older persons. The same applies for the component pain experience (0.09 versus 0.03). In contrast, the models showed that on average the TOPICS-CEP-65 scores decreased by 0.11 points per additional functional limitation, while TOPICS-CEP-85 scores decreased by 0.13 points. Hence, the component functional limitation has a smaller negative impact on the GWB scores given to the vignettes by younger persons than those given by older persons. The same applies for the component social functioning (0.03 versus −0.03).

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Linear mixed models with interaction terms between the raters’ age and each of the TOPICS-CEP components.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimates</td>
</tr>
<tr>
<td>Intercept</td>
<td>9.03</td>
</tr>
<tr>
<td>Age (centered)</td>
<td>-0.01</td>
</tr>
<tr>
<td>Morbidities</td>
<td>-0.14</td>
</tr>
<tr>
<td>Functional limitations</td>
<td>-0.12</td>
</tr>
<tr>
<td>Emotional wellbeing</td>
<td>-0.03</td>
</tr>
<tr>
<td>Pain experience</td>
<td>-0.05</td>
</tr>
<tr>
<td>Cognitive problems</td>
<td>-0.14</td>
</tr>
<tr>
<td>Social functioning</td>
<td>0.01</td>
</tr>
<tr>
<td>Self-perceived health</td>
<td>-0.16</td>
</tr>
<tr>
<td>Self-perceived QOL</td>
<td>-0.02</td>
</tr>
<tr>
<td>Morbidities x raters’ age</td>
<td>0.00</td>
</tr>
<tr>
<td>Functional limitations x raters’ age</td>
<td>0.00</td>
</tr>
<tr>
<td>Emotional wellbeing x raters’ age</td>
<td>0.00</td>
</tr>
<tr>
<td>Pain experience x raters; age</td>
<td>0.01</td>
</tr>
<tr>
<td>Cognitive problems x raters; age</td>
<td>0.00</td>
</tr>
<tr>
<td>Social functioning x raters’ age</td>
<td>0.00</td>
</tr>
<tr>
<td>Self-perceived health x raters’ age</td>
<td>0.00</td>
</tr>
<tr>
<td>Self-perceived QOL x raters’ age</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Notes: Dependent variable: general wellbeing score; *P<0.05; **P<0.01.
Our findings suggest that the preference weights of TOPICS-CEP components derived in our previous study were influenced by our raters’ age. On average, the older the rater 1) the greater the impact of functional limitations and social functioning was on the GWB scores given to the vignette cases; and 2) the smaller the impact of morbidities and pain experience was on the GWB scores given to the vignette cases.

Results of the paired-sample t-test suggest that age does have an effect on health state valuations. On average, TOPICS-CEP scores of the sample cases were higher when the preference weights of the older olds are used compared those used of younger olds.

Table 2 TOPICS-CEP models based on the health state preferences of 65-year-old versus 85-year-old persons.

<table>
<thead>
<tr>
<th>TOPICS-CEP-65</th>
<th>TOPICS-CEP-85</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>9.14</td>
</tr>
<tr>
<td>Morbidities</td>
<td>-0.15</td>
</tr>
<tr>
<td>Functional limitations</td>
<td>-0.11</td>
</tr>
<tr>
<td>Emotional wellbeing</td>
<td>-0.04</td>
</tr>
<tr>
<td>Pain experience</td>
<td>-0.09</td>
</tr>
<tr>
<td>Cognitive problems</td>
<td>-0.13</td>
</tr>
<tr>
<td>Social functioning</td>
<td>0.03</td>
</tr>
<tr>
<td>Self-perceived health</td>
<td>-0.14</td>
</tr>
<tr>
<td>Self-perceived QOL</td>
<td>-0.05</td>
</tr>
</tbody>
</table>

Of the 292 cases described in the vignettes, the majority (96.6%, n=282) had no missing data points for the calculation of TOPICS-CEP scores. The minimum and maximum scores for the TOPICS-CEP-65 were 3.54 and 9.97, respectively, and were 3.36 and 9.90 for the TOPICS-CEP-85. Compared to TOPICS-CEP-65 scores (mean ± SD =7.46±1.17), the TOPICS-CEP-85 scores (mean ± SD =7.57±1.23) had a significantly higher mean (mean difference ± SD =0.11±0.26; P<0.001). The two TOPICS-CEP scores were highly correlated (r=0.98, P<0.001).

The Bland–Altman plot showed consistent variability and there was no trend visible across the graph (Figure 2). The average of TOPICS-CEP-65 and TOPICS-CEP-85 scores ranged from 3.48 to 9.93 (mean ± SD =7.51±1.19), and the difference between the scores ranged from −0.72 to 0.85. The average bias was −0.11±0.26, and the limits of agreements were −0.63 and 0.41.

Figure 2 Bland-Altman plot of TOPICS-CEP scores based on preference weights of 65 versus 85 year olds.

Notes: Bland–Altman plot: difference between CEP scores of the vignette cases based on preferences of 65-year-old versus 85-year-old persons plotted against their average. Bold solid line represents the average difference between methods; dotted lines represent the 95% confidence limits.

3.4 Discussion

Our findings suggest that the preference weights of TOPICS-CEP components derived in our previous study were influenced by our raters’ age. On average, the older the rater 1) the greater the impact of functional limitations and social functioning was on the GWB scores given to the vignette cases; and 2) the smaller the impact of morbidities and pain experience was on the GWB scores given to the vignette cases.

Results of the paired-sample t-test suggest that age does have an effect on health state valuations. On average, TOPICS-CEP scores of the sample cases were higher when the preference weights of the older olds are used compared those used of younger olds.
The Bland–Altman plot shows that TOPICS-CEP scores based on the health-state preferences of 65-year-old persons are systematically different from those based on the preferences of 85-year-old persons. Even though 95% of the observations were located between the limits of agreement, it is questionable whether the range of these limits was not too wide. In our opinion, the limits of agreement was relatively big (1.04 points on a 6.61 range); therefore, we proposed to alter the limits of agreement. After alteration, fewer than 95% of the observations were located between the two levels, suggesting low agreement when TOPICS-CEP-65 and TOPICS-CEP-85 scores are compared.

Preferences are known to vary by persons’ own experiences.11-13 This could possibly compare levels, suggesting low agreement when TOPICS-CEP scores are considered. After alteration, fewer than 95% of the observations were located between the two levels, suggesting low agreement when TOPICS-CEP-65 and TOPICS-CEP-85 scores are compared.

The Bland–Altman plot shows that TOPICS-CEP scores based on the health-state preferences of 65-year-old persons are systematically different from those based on the preferences of 85-year-old persons. Even though 95% of the observations were located between the limits of agreement, it is questionable whether the range of these limits was not too wide. In our opinion, the limits of agreement was relatively big (1.04 points on a 6.61 range); therefore, we proposed to alter the limits of agreement. After alteration, fewer than 95% of the observations were located between the two levels, suggesting low agreement when TOPICS-CEP-65 and TOPICS-CEP-85 scores are compared.

Preferences are known to vary by persons’ own experiences.11-13 This could possibly explain the influence of our raters’ age on the TOPICS-CEP preference weights. It is well known that prevalence of chronic conditions and multimorbidity is higher in older age groups and that functional limitation increases with age.34, 35 Hence, the older the rater, the higher the chance this person has experience with having one or more chronic conditions and functional limitations. The discrepancy in preference weights between younger and older persons may be explained by the reference point or framework from which people think. With multimorbidity being the norm rather than the exception for older persons, they may have accepted it as part of the normal aging process and adapted to this deterioration by altering their expectations and norms, adapting to their situation, and adjusting their standards of “good” health accordingly.18 Moreover, older persons who have experienced functional decline may understand the full impact of functional limitations even better than those who have not. Functional limitations does not only mean one needs daily support, but it also means loss of independence, autonomy, and dignity in some cases.19

Benefits and limitations of the study
The benefit of this study is to give more insight into the influence of raters’ age on individual health domains as well as overall value. However, we need to contextualize the findings in light of some limitations. Even though we have explored the influence of raters’ age, other characteristics may have influenced the TOPICS-CEP preference weights. The influence of age might have been biased by the omission of other variables, such as socioeconomic status and multimorbidity.26 However, because the raters were recruited in both urban and more-rural parts of the Netherlands, and because the number of morbidities within our study sample was equally distributed, we do not expect socioeconomic status and multimorbidity to have influenced our findings. Additionally, we have explored whether observed characteristics such as raters’ role as informal caregiver (previous study) and raters’ sex, self-perceived health state, and self-perceived QOL influenced our results, which was not the case (analyses not shown). Moreover, even though we have established that the preference weights of TOPICS-CEP need to be adjusted for the age distribution of the Dutch population aged 65 years and older, further investigations are necessary to assess the reliability, validity, and generalizability of TOPICS-CEP.

Benefits and limitations of TOPICS-CEP
The benefits of using TOPICS-MDS (questionnaire) and TOPICS-CEP (scores) are that a range of important endpoints will be collected and incorporated in a single metric, which can index the overall impact of interventions in a standardized way. Consequently, establishing the value of interventions will be easier and more objective, and investigators do not need to make arbitrary choices when deciding which measure to elect as primary outcome measure. Moreover, TOPICS-CEP establishes a link between various health domains and a measurement of general wellbeing.

Benefits and limitations of TOPICS-CEP
The benefit of this study is to give more insight into the influence of raters’ age on individual health domains as well as overall value. However, we need to contextualize the findings in light of some limitations. Even though we have explored the influence of raters’ age, other characteristics may have influenced the TOPICS-CEP preference weights. The influence of age might have been biased by the omission of other variables, such as socioeconomic status and multimorbidity.26 However, because the raters were recruited in both urban and more-rural parts of the Netherlands, and because the number of morbidities within our study sample was equally distributed, we do not expect socioeconomic status and multimorbidity to have influenced our findings. Additionally, we have explored whether observed characteristics such as raters’ role as informal caregiver (previous study) and raters’ sex, self-perceived health state, and self-perceived QOL influenced our results, which was not the case (analyses not shown). Moreover, even though we have established that the preference weights of TOPICS-CEP need to be adjusted for the age distribution of the Dutch population aged 65 years and older, further investigations are necessary to assess the reliability, validity, and generalizability of TOPICS-CEP.

Benefits and limitations of TOPICS-CEP
The benefits of using TOPICS-MDS (questionnaire) and TOPICS-CEP (scores) are that a range of important endpoints will be collected and incorporated in a single metric, which can index the overall impact of interventions in a standardized way. Consequently, establishing the value of interventions will be easier and more objective, and investigators do not need to make arbitrary choices when deciding which measure to elect as primary outcome measure. Moreover, TOPICS-CEP establishes a link between various health domains and a measurement of general wellbeing.

The limitations of using TOPICS-CEP are that the scores can be difficult to interpret, as the observed effect of TOPICS-CEP does not necessarily reflect the effects of the single components. Incorrect interpretation may result in overestimation of the effects of an intervention. Therefore, we recommend reporting the effects of an intervention not only on TOPICS-CEP scores but also on the eight components separately. Further, TOPICS-CEP’s content has not yet been mapped to the International Classification of Functioning, Disability, and Health (ICF), which is a classification framework of health and health-related conditions developed by the World Health Organization that aims at providing a unified and standardized language for describing and classifying health domains and health-related states, thus providing a common framework for the development of outcome measures.21 To compare TOPICS-CEP results with results obtained by means of instruments that have been mapped to ICF categories, further investigation is necessary to link TOPICS-CEP components to ICF categories.

3.5 Summary
This current study provides evidence that the relative weights of health domains vary by the age of the raters who assessed the described health states; the older olds preferred functional independence while the younger olds preferred less morbidity. These variations imply that the preference weights that were obtained in our previous study and were used to establish TOPICS-CEP are a result of our random selection of participants. Therefore, we adjusted the preference weights of TOPICS-CEP for the age distribution of the Dutch population aged 65 years and older. These weights can be found in TOPICS-CEP guideline, which is available online.4 To our knowledge, TOPICS-CEP is the first generic composite endpoint in geriatric care based on health valuations of the Dutch older population aged 65 years and older.
CHAPTER 3 THE INFLUENCE OF AGE

3.6 Conclusion

Researchers examining the effectiveness of health care interventions in elderly care need to consider the discrepancies between the health-state preferences of younger and older persons when choosing or developing outcome measures. Failure to recognize these discrepancies may lead to incorrect interpretation of the findings and consequently the establishment of inappropriate health care policies. Furthermore, in clinical decision making, health care professionals need to be aware of this discrepancy as well. Therefore, we highly recommend shared decision making in geriatric care.

3.7 References


Part 3

The relevance of TOPICS-CEP
Chapter 4

Development of a method to compare one self-assessment versus multiple peer-assessments

Published as:
Adapted Bland-Altman method was used to compare measurement methods with unequal observations per case
C.S. Hofman, R.J.F. Melis, A.R.T. Donders
4.1 Introduction

The Bland-Altman approach is a very often used method for studies that examine the agreement between two methods of the same medical measurement or between two raters. Briefly, the Bland-Altman method calculates the mean difference between two methods of measurement (the “bias”) and 95% limits of agreement as the mean difference (1.96 standard deviation (SD)). The Bland-Altman plot displays the difference between the measurements by two methods (vertical axis) against their mean (horizontal axis), and limits of agreement can be added. Moreover, a regression line can help to detect a non-consistent systematic bias. The existence of a non-consistent bias indicates that the difference between the methods is relative to the size of the measurement. In contrast, a fixed or consistent bias indicates that the difference between the methods is equal for all sizes of the measurement.

Although agreement is often examined between individual raters, there are situations where agreement is needed between two groups of raters. For example, quality of life (QOL) is a subjective concept and should therefore be captured directly from the patient. However, proxy reporting of QOL is widespread, particularly in pediatric studies where children have been considered to be unreliable respondents and in studies where the respondents have cognitive impairments. When the agreement between two groups of raters of unequal sizes is examined, the Bland-Altman method requires adjustment. In this article, we describe how the method should be adapted when the agreement between two measurement methods with an unequal number of observations per case is examined. We shall focus on the possible patterns of discord.

As a part of the Older Persons’ Relevant Outcome of Care Score (OPROCS) project, we aim to investigate the agreement and possible patterns of discord between the general wellbeing (GWB) scores older persons (our cases) gave themselves (\( N = 1 \)) and the scores they were given by external raters (\( N \geq 1 \)), for example, informal caregivers, doctors, nurses. If we were to evaluate the agreement and possible patterns of discord between the self-assessment scores of the cases and the GWB score each case was given by one single external rater, the standard Bland-Altman approach could have been used: plotting the difference between the self-assessment scores and the external GWB scores on the vertical axis vs. the average of both scores on the horizontal axis. Visual inspection and linear regression could then be used to investigate trends and relationships between the horizontal and the vertical axis. Yet, in the OPROCS project, we are primarily interested in the patterns of discord between multiple groups with an unequal number of raters. For example, we are interested to compare the self-assessment scores of our cases with the GWB scores.
CHAPTER 4
ADAPTING THE BLAND-ALTMAN METHOD

4.2 Methods

The Bland-Altman method for the comparison of pairs of measurements \([X, Y]\) requires to plot the difference \([X - Y]\) (vertical axis) against the mean \([\frac{X + Y}{2}]\) (horizontal axis). From this plot, it is easy to assess the magnitude of disagreement, spot outliers, and see whether there is any trend, for example an increase in variance \([X - Y]\) for high values of the mean \([X + Y]/2\). The method is based on the fact that when there is not a non-consistent systematic bias between \(X\) and \(Y\), the correlation between \([X - Y]\) and \([X + Y]/2\) is zero and a horizontal regression line will fit the points in the plot.

However, when the average of a number of assessments \([X]\) is compared with a single assessment \([Y]\) or when the average scores of a group of assessors \([X]\) is compared with the average scores of another group of assessors of a different size \([Y]\), the differences \([X - Y]\) and \([X - Y]\) will be correlated with the mean of \([X]\) and \([Y]\) and the mean of \([X]\) and \([Y]\), respectively (see Box 1). This can be explained by the fact that the means of scores given by a larger group of persons tend to be less extreme than the scores given by individual persons or smaller groups of persons. In contrast, in the absence of a non-consistent systematic bias, the differences \([X - Y]\) and \([X - Y]\) will not be correlated with the mean of all assessments irrespective of who made them (see Box 1). Hence, the latter approach should be used for the valuation.

To illustrate both approaches, we use simulation data and data derived from the OPROCS project. In short, the OPROCS project aims to construct a multidimensional preference-weighted composite end point for The Older Persons and Informal Caregivers Survey Minimum Dataset (TOPICS-MDS); a large data-sharing initiative. As part of the OPROCS project, we are interested to examine possible patterns of discord between self-assessment scores and the GWB scores given by external raters; in this article, we used the GWB scores given by external raters who were aged 65 years and older. In total, 250 cases (patient profiles) were derived from TOPICS-MDS.

56 57

Box 1

Suppose the assessed persons are numbered 1 \(\ldots N\). Person \(i\) is assessed by \(K_i\) raters of type A, and this resulted in the ratings \(X_{i,1}, X_{i,2}, \ldots, X_{i,K_i}\) with average score \(X_i\).
The same person is assessed by \(M_i\) raters of type B, and this resulted in the ratings \(Y_{i,1}, Y_{i,2}, \ldots, Y_{i,M_i}\) with average score \(Y_i\).

In the absence of systematic differences, \(E[X_i] = E[Y_i] = c_i\).

Let \(\alpha\) be the variance of \(X_i\) and \(\beta\) be the variance of \(Y_i\). Therefore, the difference between the mean score of the external raters and the self-assessment score should be plotted vs. the mean of all scores independent of the rater. This is somewhat paradoxical because for the bias between the mean score given by external raters and the self-assessment score \([X - Y]\), the single score of the elderly person \(Y\) has the same weight as the mean scores of all external raters together \([X]\), whereas for the overall mean of the scores, the self-assessment score has the same weight as the score of each of the external raters \((K_i X + Y)/(K_i + 1))\); thus, its impact will be negligible when the number of external raters is high.

In this article, we will describe the original Bland-Altman approach and two possible methods of adaptation in detail. Both empirical and simulation data will be used to illustrate these methods.
in line with the Dutch grading system). Thus, we had 250 self-assessment scores and 1,000 scores given by external raters.

For the simulation, we followed the situation described previously. We first generated 250 true scores uniformly from 4.5 to 7.5. The self-assessment scores were generated by adding a normally distributed error score (mean = 0; SD = 1.5) to the true scores. Next, four external rater scores were generated by repeating this same process four times using a normal distribution for the error terms with equal mean and SD as for the self-assessment scores. Thus, self-assessment scores and external rater scores are comparable with respect to expected value and error variance (reliability). Finally, we present adapted Bland and Altman plots where we leave out the limits of agreement as the focus of this article is on biases. For the purpose of assessing non-consistent structural bias, we added a regression line to the plot. We present a traditional Bland and Altman plot based on the self-assessment scores and the average of the scores of the external raters and a Bland and Altman plot based on the suggestions made previously (ie, we use the average of all five scores for the x-axis).

### 4.3 Results

First, we present the results of the simulation study. Figure 1A shows the adapted Bland-Altman plot of the difference between the mean of the GWB scores given by external raters and the self-assessment scores of the cases (vertical axis) vs. the average of the mean scores by external raters and the self-assessment scores (horizontal axis) (thus (mean external scores + self-assessment score)/2). The correlation between the mean differences and the uncorrected average of the means is 0.40; the slope of the regression line is 0.50 (P<.001). Figure 1B shows the adapted Bland-Altman plot of the difference between the mean of the GWB scores given by external raters and the self-assessment scores of the cases (vertical axis) vs. the overall mean of the assessments irrespective of who made them (horizontal axis) (thus (mean external scores x number of external raters + self-assessment score)/(number of external raters + 1)). The correlation between the mean differences and the corrected average of the means is 0.02, and the slope of the regression line is 0.02 (P = 0.78).

Next, we present the results from the OPROCS project. A selection of the GWB scores and corresponding means can be found in Table 1. The correlation between the mean differences and the uncorrected average of the means is 0.40; the slope of the regression line is 0.52 (P < .001; Figure 2A). The correlation between the mean differences and the corrected average of the means is 0.08, and the slope of the regression line is 0.03 (P = 0.68; Figure 2B).

![Figure 1](image1.png) **Figure 1** Adapted Bland-Altman plot based on simulation data. 
A. Bland-Altman plot comparing self-assessment scores and GWB scores given by multiple raters using uncorrected (or unweighted) average of means, thus \( M = \frac{\sum x_i}{n} \); B. Bland-Altman plot comparing self-assessment scores and GWB scores given by multiple raters using corrected (or weighted) average of means, thus \( MW = \frac{\sum x_i + \sum y_i}{n + 1} = \frac{X_{1,1} + \ldots + X_{1,n} + Y_{1,1} + \ldots + Y_{1,n}}{n + 1} \); Broken line represents the correlation line.

![Figure 2](image2.png) **Figure 2** Adapted Bland-Altman plot based on OPROCS data. 
A. Bland-Altman plot comparing self-assessment scores and GWB scores given by multiple raters using uncorrected (or unweighted) average of means, thus \( M = \frac{\sum x_i}{n} \); B. Bland-Altman plot comparing self-assessment scores and GWB scores given by multiple raters using corrected (or weighted) average of means, thus \( MW = \frac{\sum x_i + \sum y_i}{n + 1} = \frac{X_{1,1} + \ldots + X_{1,n} + Y_{1,1} + \ldots + Y_{1,n}}{n + 1} \); Broken line represents the correlation line.
Hence, although Figures 1A and 2A suggest there is a non-consistent systematic bias between the scores given by external raters and cases themselves (the trend line has a nonzero slope), Figures 1B and 2B suggest there is no non-consistent bias. The diagonal streaks in the plot for the OPROCS data were caused by the fact that the raters tend to give integers and half integers only although the use of one decimal was encouraged.

### 4.4 Discussion

Assessment methods should not be compared by plotting of \([X - Y]\) vs. \([Y]\). Therefore, the Bland-Altman approach is often used to compare measurement methods. In this article, we have described an adaptation of the Bland-Altman approach to the situation in which possible patterns of discord between two measurement methods with an unequal number of observations per case are explored. Similar to Bland and Altman, we assumed that the data were approximately normally distributed, but for skewed data, a logarithmic transformation may be considered.
4.5 References


Chapter 5

Comparing self-assessment versus peer-assessments

In press as:
Examining consistency between self-assessed QOL and peer ratings in older people.

C.S, Hofman, J.E. Lutomski, M.G.M. Olde Rikkert, R.J.F. Melis, A.R.T Donders

Journal of the American Geriatrics Society
To the Editor:
The concepts ‘quality of life’ and, more specifically, ‘health-related quality of life’ (HRQOL) refer to the physical, psychological, and social domains of health. These domains can be measured in two dimensions: objective assessment and more subjective perceptions of health status. Although the objective assessment of a person’s functioning on a number of health domains is important in defining a patient’s degree of health, the patient’s subjective perceptions translate that objective assessment into the actual QOL experienced. These valuations of specific health-related outcomes are called health utilities.

To answer the question ‘What is the effect of the intervention on patients’ day-to-day QOL?’ one can ask them to simply rate their perceived QOL. However, such subjective assessment of HRQOL not only depends on the effect of the intervention, but also on other (health related) events. Moreover, subjective assessments are subject to artefacts such as response shift effects. Response shift gives patients the ability to cope with limitations and disability, which greatly affect a person’s perception of health and satisfaction with life.

Even though self-rated QOL provides useful global assessments of perceived QOL in individuals, peers may provide complementary information on the utility of health states that is of great interest to fully understand the value of a certain health outcomes. We investigated the systematic differences between these two perspectives on how QOL is perceived by older persons themselves and how their QOL is rated by peers.

5.1 Methods

Data were derived from the OPROCS Study. In total, 292 vignette cases were constructed based on data derived from The Older Persons and Informal Caregivers Survey Minimum Dataset. The cases described in the vignettes rated their own situation on a scale from zero to ten (worst to best possible situation) to one decimal place, which is in line with the Dutch grading system. In a similar manner, a panel of older adults serving as external raters (N=124) were asked to rate the situation of 10 randomly assigned cases. Consequently, each vignette case had one self-rated score and three to five scores given by external raters (peers).

An adapted Bland-Altman plot was derived to evaluate levels of agreement between self- and peer-ratings as well as identify patterns of discrepancy. For the purpose of this study, we decided a priori that a threshold of ± 0.50 (on a scale of 0-10) would be
applied for the discrepancy between self- and peer-rated scores. Significant differences in mean self-rated and peer-rated scores were determined using a t-test (SPSS version 20.0; SPSS IBM, New York, USA).

5.2 Results

The self- and peer-rated scores ranged from 1.5 to 10 and from 2.0 to 10 respectively. The differences between self- and mean peer-rated scores for the same vignette ranged from -4.6 to 3.4. The mean of self- and peer-rated scores ranged from 4.3 to 9.0. Of the 1,240 peer ratings, 24.6% (n=305) were identical to self-rated scores. Figure 1 shows the adapted Bland-Altman plot of the difference between the mean of the peer-ratings and the self-rated scores of the cases versus the combined overall mean of self-rated and peer-rated scores. There was no trend visible. Vignette scores (standard deviations) were significantly different between self- and peer-ratings [self: 7.4 (±1.3), peer: 6.9 (±0.9), p<.001]. The upper and lower levels of agreement were -1.9 and 2.9 respectively, which exceeded our predefined threshold of ± 0.50 for relevant differences. In 196 (67%) vignettes the difference exceeded the threshold.

5.3 Discussion

Given the wide levels of agreement between the self- and peer-rated scores (± 2.4), our findings suggest that these scores are not interchangeable, but rather self- and peer-rated assessments are complementary to each other. As the first gives insight into how patients experience their QOL at a given time, while the second gives insight into how the population to which the patient belongs would assess the QOL on average.

As it is impossible to continuously ask peers to evaluate a patient’s QOL we established an outcome measure based on health valuations made by Dutch older persons aged 65 years and over. The measure is multifaceted and combines outcomes collected by The Older Persons and Informal Caregivers Survey – Minimum DataSet (TOPICS-MDS).

The assessment of longitudinal change in HRQOL is a key component of many clinical and research evaluations. Peer-ratings may provide useful complementary insight into perceived QOL to understand the utility of health states relevant for older persons. Yet, it is necessary to tests systematic differences between self- and peer-assessment within various patient groups.

Figure 1
Adapted Bland-Altman plot of the difference between the mean of the proxy ratings and the self-assessed ratings of the cases (vertical axis) versus the overall mean of both sets of ratings (horizontal axis). Red bold solid line represents the average difference between self-rated and proxy-rated scores; red bold dotted lines represent the 95% levels of agreement. Black solid line represents the zero bias line; black dotted lines represent the predefined a threshold of ± 0.50.
5.4 References


Chapter 6
Comparing health state preferences of older persons and healthcare professionals

Published as:
Comparing the health state preferences of older persons, informal caregivers and healthcare professionals: A vignette study


CHAPTER 6 OLDER PERSONS VERSUS HEALTHCARE PROFESSIONALS

6.1 Background

The population is aging across the world. This demographic shift will lead to extraordinary demands on our healthcare system. With the limited financial resources and insufficient number of healthcare professionals, evaluating the effectiveness of healthcare interventions has become an integral part of health policy and decision-making. However, it is a great challenge to evaluate interventions for elderly because their health states are complex and interventions often target more than one domain.

An expert panel of The American Geriatrics Society formulated guiding principles on how clinicians should approach the care of older adults with multi-morbidity. Several steps were defined including “Consider patient preferences” and “Is relevant evidence available regarding important outcomes?”. These principles are also crucial for researchers evaluating the effectiveness of intervention in older adults. Hence, a generic measurement instrument with a composite endpoint (CEP) that is preference-based and includes important outcomes would be helpful to compare outcomes across groups, thereupon, to establish and compare the effectiveness of different geriatric interventions.

The Dutch National Care for the Elderly Programme (NCEP) was established in 2008 to promote proactive, integrated healthcare for older persons with complex healthcare needs. Within the NCEP The Older Persons and Informal Caregivers Survey Minimum Dataset (TOPICS-MDS) was developed to collect uniform information from all research project funded under this Programme. A detailed description of TOPICS-MDS has been presented elsewhere.

Briefly, TOPICS-MDS is a collection of four validated instruments which was designed to collect essential information on the physical and mental wellbeing of older persons and informal caregivers in the Netherlands. The survey was administered in multiple research settings to elicit uniform outcome data in the aim of creating a national data repository on older persons’ health. Over 60 NCEP research projects have already incorporated TOPICS-MDS in their research protocol and evaluated more than 32,000 participating elderly using the survey.

To compare the effectiveness of these projects a preference-weighted outcome measure that combined multidimensional TOPICS-MDS outcomes into a composite endpoint (TOPICS-CEP) was developed based on the health state preferences of older persons and informal caregivers.
Briefly, TOPICS-CEP is a preference-weighted index ranging from 0 (worst possible general wellbeing) to 10 (best possible general wellbeing). It combines 42 data points from TOPICS-MDS covered by eight components, such as functional limitations (Katz index of independence)\(^{12}\) and emotional wellbeing (mental health subscale of the RAND-36).\(^{13}\) The components vary in both scale range and preference weight. Raw TOPICS-CEP scores are transformed into indexed scores. More detailed information about the development of TOPICS-CEP and its scoring procedure can be found elsewhere.\(^{11,14}\)

The various stakeholders in geriatrics share a mutual goal which is to improve a person’s health and wellbeing. However, studies have shown significant differences between the perspectives of older persons and their healthcare professionals.\(^{15,16}\) Consequently, we expected that the preference weights of the TOPICS-CEP’s components would differ between those of older persons and their healthcare professionals, which could potentially lead to treatment decisions by professionals that are at odds with patient preferences and to incorrect interpretation of findings in effectiveness studies. For that reason, we explored the TOPICS-CEP components’ weights based on the health state preferences of healthcare professionals in this current study and compared them with the weights based on the health state preferences of older persons and informal caregivers found in our previous study.

In short, the primary objectives of this study were: (1) to examine the association of preference weights with the healthcare professionals’ characteristics; (2) to examine the difference between healthcare professionals’ preference weights and those of older persons and informal caregivers; and (3) to explore the influence of the cases’ gender and age on the distribution of the composite scores.

### 6.2 Materials and Methods

**Ethical approval**

The Medical Ethics Committee of the Radboud University Medical Center formally stated that this study was exempt from ethical review (Radboud University Medical Center Ethical Committee review reference number: CMO: 2010/244). Written informed consent was obtained from the older persons and informal caregivers who participated in our previous study.

**Study design**

Vignette studies were conducted to obtain the preference weights for the eight TOPICS-CEP components: morbidities (list of 17 pre-defined conditions)\(^{17}\), functional limitations (Katz index of independence)\(^{12}\), emotional wellbeing (mental health subscale of the RAND-36)\(^{13}\), pain experience (pain dimension of the EQ-5D)\(^{18}\), cognitive problems (cognition dimension of the EQ-5D+C)\(^{19}\), social functioning (item 10 from the RAND-36)\(^{20}\), self-perceived health (item 1 from the RAND-36)\(^{13}\), self-perceive QOL (phrasing similar to self-perceived health item from the RAND-36)\(^{13}\). The participants rated the general wellbeing (GWB) of case vignettes, which were short descriptions or profiles of older persons (further called: cases).

**Participants**

**First vignette study: Older persons and informal caregivers**

In the first vignette study, 124 community dwelling older persons and 76 informal caregivers participated as raters. They were recruited and their data was collected by four academic centres: Radboud University Medical Center, University Medical Centre Groningen, Academic Medical Centre, and Leiden University Medical Centre. A full report and more detailed information can be found elsewhere.\(^{11}\)

**Second vignette study: Healthcare professionals**

The 330 healthcare professionals who rated the cases in the second vignette study were recruited during two national geriatric conferences in February 2012 and October 2012, via websites of various professional associations, and via the website of NCEP. The professionals worked as physicians, nurses, welfare staff or allied health professional across the Netherlands, covering both urban and more rural parts of the country.

**Materials**

The vignettes were based on data of a sample of cases derived from TOPICS-MDS data repository, which consists of pooled data from various research projects which differ across study design, sampling framework, and inclusion criteria. In general, each vignette included 46 items covering the eight previously described TOPICS-CEP components: morbidities, functional limitations, emotional wellbeing, pain experience, cognitive problems, social functioning, self-perceived health, and self-perceive QOL. The information included in the vignettes regards all the variables (or items) from TOPICS-MDS for older persons which carry information relevant for understanding an individual’s outcome. This excludes demographics and health service utilization. Excluding these components was based on the rationale that demographics such as gender and age and health service utilization cannot be influenced by healthcare delivery.
By using empirical data, vignettes with plausible health state combinations were constructed. We made sure that the complete ranges of outcomes for the different health domains were covered. All raters evaluated a limited number of cases and we assured that all cases were rated by a sufficiently large number of raters. Since it was to be expected that some of the disciplines would consist of lower numbers of participants, we used a smaller set of cases in this present study. To guarantee that each discipline evaluated the complete range of the outcomes we chose a new set of cases for this study.

First vignette study: Older persons and informal caregivers
The cases (N=292) of whom the GWB were assessed by older persons and informal caregivers had a mean age (±SD) of 81.4 (5.72) years and 58.6% (N=171) was female. The majority of these cases were either married (42.8%, N=125) or their partner was deceased (42.8%, N=125), and 39.7% (N=116) lived independently with someone, e.g. a partner or family member.

Second vignette study: Healthcare professionals
The cases (N=161) of whom the GWB were assessed by healthcare professionals had a mean age (±SD) of 82.4 (6.5) years and 67.7% (N=109) was female. The majority of these cases were either married (28.0%, N=45) or their partner was deceased (57.1%, N=92), and 43.5% (N=70) lived in either a nursing home or a residential care facility.

An overview of the health domains, items per domain, and levels per item which were included in the vignettes and used in the statistical analyses can be found in Table 1.

Procedure
After reading each vignette, raters were asked to give a score ranging from zero to ten representing how bad or good, in their opinion, the GWB of the described case was.

First vignette study: Older persons and informal caregivers
The vignette study within the group of older persons and informal caregivers was conducted on paper. After two trial cases, which were the same for every participant within the study, the raters were asked to give scores to a random selection of ten cases. More information about the exact procedure can be found elsewhere.11

Second vignette study: Healthcare professionals
The healthcare professionals had the opportunity to evaluate the cases on paper or online via the website of QuestionPro (online survey software to create, publish, and distribute online surveys); both the hardcopy and the online survey had the same format and the participants had to follow the same procedure. After a trial case, which was the same for every participant, the raters were asked to give scores to a random selection of five cases. In addition, we asked them to answer a couple of questions regarding: age, gender, occupation, number of years in this occupation, and number of patients/clients aged ≥ 65 years per week.

Statistical analysis
The statistical procedures for both vignette studies were comparable. The analyses to derive the preference weights for TOPICS-CEP’s components based on the health state preferences of older persons and informal caregivers can be found elsewhere.11 To derive the weights for the components based on the health state preferences of the healthcare professionals five mixed linear regression models were constructed. Each model had the following structure: (1) The GWB scores were used as dependent variable; (2) The eight CEP components were used as independent variables (predictors): morbidities, functional limitations, emotional wellbeing, pain experience, cognitive functioning, social functioning, self-perceived health, and self-perceived QOL; and (3) To correct for clustering within raters a random (rater dependent) intercept was included. Furthermore, we included in each model one of the following five factors: profession (physician, nurse, welfare staff, and allied health professional), physicians’ discipline (general practitioner, nursing home physician, internist, geriatrician), years of experience, number of patients aged ≥65 years per week, or rater group (healthcare professional / older person or Informal caregiver) together with the interaction between the included factor and each of the CEP components. The parameter estimates for the eight domains represent the preference weights.

Subsequently, for the cases used in both vignette studies, we described the distribution of TOPICS-CEP scores (based on the preference weights of older persons and informal caregivers) across cases’ gender and age groups and compared them with the distribution of such a composite score when one would base it on healthcare professionals’ preferences (further referred to as: HP’s CEP). A paired sample T-test was used to examine the difference between TOPICS-CEP and the HP’s CEP. In addition, to explore the level of agreement between the two composite outcome measures a Bland-Altman plot was used.
Table 1 The domains, items, and aggregated items included in the vignettes with the descriptive.

<table>
<thead>
<tr>
<th>Outcome domains</th>
<th>Vignette items</th>
<th>Outcome levels</th>
<th>Descriptives Vignettes study: older persons and informal caregivers</th>
<th>Descriptives Vignettes study: Healthcare professionals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morbidity (Local and national health monitor)²¹</td>
<td>Presence of: Dementia; Depression; Incontinence; Stroke, CVA or TIA; Hip fracture; Panic or anxiety disorder; Dizziness with falling; Vision disorder; Asthma; Osteoporosis; Diabetes; Arthritis; Heart failure; Form of cancer; Complaints due to benign enlarged prostate; Fracture other than hip fracture; Hearing disorder</td>
<td>Number of diseases present Range: 0-17</td>
<td>Mean: 3.5 SD: 2.0</td>
<td>Mean: 3.3 SD: 2.3</td>
</tr>
<tr>
<td>Functional limitations (modified KATZ-ADL index)²²</td>
<td>Needing help with: Brushing hair; Going to the toilet; Taking medication; Sitting down and getting up from chair; Getting dressed; Travelling; Handling finances; Grocery shopping; Walking about; Taking a bath or shower; Housekeeping; Preparing a meal; Eating; Using the telephone</td>
<td>Number of limitations in IIADL Range: 0-15</td>
<td>Mean: 3.0 SD: 3.5</td>
<td>Mean: 4.4 SD: 4.2</td>
</tr>
<tr>
<td>Emotional wellbeing (RAND-36, mental health subscale)²³</td>
<td>Feeling down; Feeling blue; Feeling nervous; Feeling happy; Feeling calm</td>
<td>Raw mental health score Range: 5-30</td>
<td>Mean: 10.4 SD: 4.7</td>
<td>Mean: 9.8 SD: 4.4</td>
</tr>
<tr>
<td>Pain experience (Single item EQ-SD+C)²⁴</td>
<td>Pain experience</td>
<td>No Moderate Severe</td>
<td>130 (44.7) 127 (43.6) 34 (11.7) (N=291)</td>
<td>50 (39.4) 60 (47.2) 17 (13.4) (N=127)</td>
</tr>
<tr>
<td>Cognitive functioning (Single item EQ-SD+C)²⁴</td>
<td>Cognitive problems</td>
<td>No Moderate Severe</td>
<td>222 (77.1) 64 (22.2) 2 (0.7) (N=288)</td>
<td>74 (58.6) 37 (29.1) 16 (12.6) (N=127)</td>
</tr>
<tr>
<td>Social functioning (Single item RAND-36)²⁵</td>
<td>Social activities hampered by physical health or emotional problems</td>
<td>Never Rarely Sometimes Mostly Continuously</td>
<td>217 (75.9) 19 (6.6) 29 (9.1) 6 (2.1) 18 (6.3) (N=289)</td>
<td>90 (52.6) 25 (14.6) 22 (12.9) 14 (8.2) 20 (11.7) (N=171)</td>
</tr>
<tr>
<td>Self-perceived health (Single item, RAND-36)²⁶</td>
<td>Self-perceived health in general</td>
<td>Excellent Very good Good Reasonable Poor</td>
<td>18 (6.3) 20 (7.0) 127 (44.3) 108 (37.6) 14 (4.8) (N=287)</td>
<td>7 (4.0) 9 (5.2) 66 (38.2) 73 (42.2) 18 (10.4) (N=173)</td>
</tr>
<tr>
<td>Satisfaction with quality of life (QOL) (Single item formed using phrasing similar to self-perceived health question, RAND-36)</td>
<td>Self-perceived QOL in general</td>
<td>Excellent Very good Good Reasonable Poor</td>
<td>26 (9.1) 40 (14.0) 170 (59.4) 41 (14.4) 9 (3.1) (N=286)</td>
<td>5 (2.9) 17 (9.8) 94 (54.0) 54 (31.0) 4 (2.3) (N=174)</td>
</tr>
</tbody>
</table>
6.3 Results

The healthcare professionals who participated as raters in this study had a mean age of 43.0 years (SD 11.0) and 80.3% was female (N=265). The majority of the healthcare professionals conducted the vignette experiment online (76.7%, N=253). Additional information about the characteristics of the healthcare professionals can be found in Table 2.

<table>
<thead>
<tr>
<th>Years active in current position</th>
<th>Number of patients/clients aged 65 years and over</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Physicians</td>
<td>127</td>
</tr>
<tr>
<td>Nurses</td>
<td>102</td>
</tr>
<tr>
<td>Welfare staff</td>
<td>45</td>
</tr>
<tr>
<td>Allied health professionals</td>
<td>56</td>
</tr>
<tr>
<td>Total</td>
<td>330</td>
</tr>
</tbody>
</table>

Note: In our previous study, older persons (N=124) and informal caregivers (N=76) participated as raters. The older persons had a mean age of 78.3 years (SD: 6.7) and 62.9% (N=78) was female. The informal caregivers had a mean age of 62.9 (SD: 12.1) and 72.4 (N=55) was female.

Healthcare professionals’ characteristics and their preference weights

The models including the interaction terms between profession, physician’s discipline, or number of patients aged 65 years and older with each of the predictors showed no significant interaction effects. In contrast, the model that included the interaction terms between years of experience and each of the predictors showed a significant interaction effect between years of experience and morbidities (p = 0.02). For each additional year of experience, the preference weight of the component morbidities declined with 0.01 points. Hence, the association between the number of morbidities and GWB score became less strong.

Comparing preference weights: healthcare professionals vs. older persons and informal caregivers

For several components of TOPICS-CEP the healthcare professionals’ preference weights differed significantly from those of older persons and informal caregivers. The components’ weights based on the health state preferences of older persons

### Table 2 Distribution of the healthcare professionals (N=330).

<table>
<thead>
<tr>
<th>Years active in current position</th>
<th>Number of patients/clients aged 65 years and over</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Physicians</td>
<td>127</td>
</tr>
<tr>
<td>Nurses</td>
<td>102</td>
</tr>
<tr>
<td>Welfare staff</td>
<td>45</td>
</tr>
<tr>
<td>Allied health professionals</td>
<td>56</td>
</tr>
<tr>
<td>Total</td>
<td>330</td>
</tr>
</tbody>
</table>

Note: In our previous study, older persons (N=124) and informal caregivers (N=76) participated as raters. The older persons had a mean age of 78.3 years (SD: 6.7) and 62.9% (N=78) was female. The informal caregivers had a mean age of 62.9 (SD: 12.1) and 72.4 (N=55) was female.

### Table 3 Mixed linear regression model: Unstandardized coefficients with corresponding P-values and confidence interval, standardized coefficients, and p-value for the interaction effect between each predictor and the independent variable ‘healthcare provider’.

<table>
<thead>
<tr>
<th>Healthcare professionals (N=330)</th>
<th>Older persons and informal caregivers (N=200)</th>
<th>P-value for the interaction between predictor and healthcare provider</th>
<th>Unstandardized Coefficient B</th>
<th>Confidence interval for B</th>
<th>P-value</th>
<th>Standardized Coefficient</th>
<th>P-value</th>
<th>Dependent variable: GWB scores given by healthcare professionals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>10.47</td>
<td>10.14</td>
<td>10.80</td>
<td>0.00**</td>
<td>8.84</td>
<td>8.84</td>
<td>9.22</td>
<td>0.00**</td>
</tr>
<tr>
<td>Morbidities</td>
<td>-0.02</td>
<td>-0.06</td>
<td>0.02</td>
<td>0.28</td>
<td>-0.15</td>
<td>-0.15</td>
<td>-0.12</td>
<td>0.00**</td>
</tr>
<tr>
<td>Functional limitations</td>
<td>-0.05</td>
<td>-0.07</td>
<td>-0.03</td>
<td>0.00**</td>
<td>-0.05</td>
<td>-0.05</td>
<td>-0.03</td>
<td>0.00**</td>
</tr>
<tr>
<td>Emotional wellbeing</td>
<td>-0.07</td>
<td>-0.08</td>
<td>0.01</td>
<td>0.00**</td>
<td>-0.13</td>
<td>-0.13</td>
<td>-0.11</td>
<td>0.00**</td>
</tr>
<tr>
<td>Pain experience</td>
<td>-0.21</td>
<td>-0.22</td>
<td>0.06</td>
<td>0.00**</td>
<td>-0.22</td>
<td>-0.22</td>
<td>-0.22</td>
<td>0.00**</td>
</tr>
<tr>
<td>Cognitive problems</td>
<td>-0.23</td>
<td>-0.23</td>
<td>0.02</td>
<td>0.00**</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>0.00**</td>
</tr>
<tr>
<td>Social functioning</td>
<td>-0.13</td>
<td>-0.12</td>
<td>0.06</td>
<td>0.00**</td>
<td>-0.12</td>
<td>-0.12</td>
<td>-0.12</td>
<td>0.00**</td>
</tr>
<tr>
<td>Self-perceived health</td>
<td>-0.18</td>
<td>-0.18</td>
<td>0.04</td>
<td>0.00**</td>
<td>-0.20</td>
<td>-0.20</td>
<td>-0.20</td>
<td>0.00**</td>
</tr>
<tr>
<td>Self-perceived QOL</td>
<td>-0.28</td>
<td>-0.28</td>
<td>-0.19</td>
<td>0.00**</td>
<td>-0.07</td>
<td>-0.07</td>
<td>-0.07</td>
<td>0.00**</td>
</tr>
</tbody>
</table>

a Dependent variable: GWB scores given by healthcare professionals
b Dependent variable: GWB scores given by older persons and informal caregivers

*P < 0.05
**P < 0.001

---

90 91
and informal caregivers versus those based on the preferences of healthcare professionals can be found in Table 3.

Significant interaction effects were found between the factor healthcare professional and the outcome domains: morbidities, functional limitations, pain experience, social functioning, and self-perceived QOL (p<0.05). The estimated differences of these preference weights were: -0.12, -0.07, 0.14, 0.15 and 0.26, respectively. These estimates weights indicated that morbidities and functional limitations were given more weight by older persons and informal caregivers than by healthcare professionals, whereas the opposite was true for pain experience, social functioning, and self-perceived QOL.

The components morbidities and functional limitations had stronger associations with GWB scores given by older persons than with scores given by healthcare professionals: for every morbidity present the GWB score based on the preference weights of older persons and informal caregivers declined with 0.14 points, whereas the GWB score based on the preference weights of healthcare professionals declined with 0.02 points. These numbers were 0.12 versus 0.05 for every functional limitation, respectively. On the other hand, the components pain experience, social functioning, and self-perceived QOL had stronger associations with GWB scores given by healthcare professionals than with scores given by older persons and informal caregivers: when pain increased one point on the Likert scale (no pain, moderate pain, severe pain) the GWB score based on the preference weights of older persons and informal caregivers declined with 0.04 point, whereas the GWB score based on the preference weights of healthcare professionals declined with 0.18 points. These numbers were 0.01 versus 0.16 for social functioning, and 0.02 versus 0.28 for self-perceived QOL, respectively.

Distribution of the CEP’s

Of the 453 cases described in the vignettes from both vignette studies, the majority (84.7%) had no missing data points for the calculation of TOPICS-CEP scores or HP’s CEP. Consequently, both composite outcome measures were calculated for 384 cases.

The overall distribution for both measures were tailed to the left (not shown), though became more normalized when stratified by age (Figure 1). Mean scores (±SD) significantly differed across gender and age groups for both HP’s CEP [(Men: 7.695 (0.80); Women: 7.43 (0.76); p=0.001) (<80: 7.67 (0.79); 80-84: 7.50 (0.78); ≥85: 7.42 (0.76); p=0.039)] and TOPICS-CEP [(Men: 7.01 (0.82); Women: 6.73 (0.786; p=0.001) (<80: 7.02 (0.75); 80-84: 6.83 (0.79); ≥85: 6.64 (0.80); p=0.001)].

The minimum and maximum HP’s CEP scores calculated were 4.75 and 9.21, respectively. These scores were 4.38 and 8.42 for TOPICS-CEP. The mean HP’s composite score (±SD) differed from the mean TOPICS-CEP [HP’s CEP: 7.53 (0.78); TOPICS-CEP: 6.84 (0.79); p=0.001]. The two composite outcome measures were highly correlated (r=0.88, p<0.001).

The Bland-Altman plot showed consistent variability and there were no trends visible across the graphs (Figure 2). For the cases aged younger than 80 years, the average of HP’s CEP and TOPICS-CEP scores ranged from 4.74 to 8.80 [80-84: 4.59 to 8.64; ≥85: 5.02 to 8.64], the difference between the scores ranged from -1.44 to 0.74 [80-84: -2.00 to 0.55; ≥85: -1.98 to 0.29], the average bias was -0.64 [80-84: -0.66; ≥85: -0.78] and the level of agreements were -1.33 and 0.05 [80-84: -1.44 and 0.12; ≥85: -1.15 and 0.02].
6.4 Discussion

Our primary findings indicate that the weights of TOPICS-CEP’s components based on the health state preferences of healthcare professionals differed significantly from those based on the preferences of older persons and informal caregivers. These findings are in line with other studies exploring the discrepancies between older persons and healthcare professionals concerning health state preferences.\textsuperscript{15,16}

Our results indicate that the presence of morbidities and functional limitations in the vignette cases have a greater impact on the GWB scores given by older persons and informal caregivers than on the scores given by healthcare professionals. However, the presence of increased levels of pain experience, hampering of social functioning, and a decrease of self-perceived QoL status have a greater impact on the GWB scores given by healthcare professionals than on the scores given by older persons and informal caregivers. Furthermore, our results suggest that healthcare professionals’ number of years of experience influence the preference weight of morbidities. The higher the numbers of years of experience the lower the negative impact of the number of morbidities on GWB scores. To explore whether a change of 0.005 point per additional year is clinically relevant further research needs to be conducted.

Finally, our results show that the mean differences between HP’s CEP and TOPICS-CEP scores were not close to zero for any of the age groups, which indicates that the two composite outcome measures are systematically producing different results. Yet, to understand whether these systematic differences are clinically relevant further research needs to be conducted. Moreover, in the Bland-Altman plot there were no trends visible in any of the age groups.

Our results and implications need to be interpreted in light of several limitations. First, the vignettes used in the two studies were not the same. This means that the GWB of a vignette case was never assessed by both an older person or informal caregiver and a healthcare professional. However, all vignettes were based on empirical data derived from the TOPICS-MDS National Database and the cases were all plausible health state combinations. Consequently, none of the raters had to assess impossible health state combinations, e.g., a case that has eight morbidities and experiences severe pain, but does not have any functional limitations. With this in mind, we do not expect that the use of a different set of vignettes influenced our findings.
Second, we compared the components’ preference weights between the various professions and explored the influence of work experience on these weights. However, we have not studied the influence of personal characteristics of the professionals, such as gender and age, on the preference weights. This was a well-considered decision as the aim of our study was to establish a CEP based on the preference weights of a random sample of healthcare professionals.

6.5 Conclusion

If more than one outcome is important for effectiveness evaluation or if an intervention has the potential to improve more than one health domain, a CEP can efficiently deal with the issue of multiplicity, e.g. in elderly care. By using a preference-weighted multifaceted outcome measure, such as TOPICS-CEP, the relative importance of the various outcomes is taken into account. At the macro level, TOPICS-CEP which is based on older persons’ health state preferences may be considered as a general patient reported outcome measure to be used for evaluating healthcare interventions for (frail) older subjects.

When examining the effectiveness of healthcare interventions in elderly care need to consider the discrepancies between the health state preferences of older persons and healthcare professionals. Failure to recognize these discrepancies may lead to incorrect interpretation of the findings and the establishment of inappropriate healthcare policies. Furthermore, healthcare professionals need to keep in mind that their own health state preferences may not be the same as those of their older patients. This provides a good argument for shared decision making in healthcare.
6.6 References

Part 4

The psychometric properties of TOPICS-CEP
Chapter 7
Validation of TOPICS-CEP

In press as:
Examining the Construct and Known-group Validity of a Composite Endpoint for The Older Persons and Informal Caregivers Survey Minimum Data Set (TOPICS-MDS); A Large-Scale Data Sharing Initiative


Membership of TOPICS-MDS research consortium is provided in the Acknowledgments.

Plos One
CHAPTER 7 VALIDATION OF TOPICS-CEP

7.1 Introduction

Aging of the population has a major impact on the organization and delivery of healthcare. The shift from acute to chronic illnesses and the expected shortage of healthcare workers will be of particular importance. 1 To ensure high quality care for older persons, the evaluation and monitoring of three aspects of health care delivery need to be regularly evaluated: structure, process, and outcomes. 2 However, comparing outcomes in older persons is challenging. Firstly, the health states of older persons are complex, as older individuals often present different combinations of chronic multi-morbidity and functional limitations. 3 Secondly, interventions often influence a broad range of health domains both directly and indirectly. For example, occupational therapy aims to enable people who have physical restrictions to achieve greater independence. By engaging in meaningful social activities, health and psychological wellbeing are also indirectly and positively influenced. 4 Thus, occupational therapy can improve both physical and mental wellbeing. The two obstacles can be circumvented if the important outcome parameters are collected and combined into a preference-weighted composite endpoint (CEP). 5, 6

In 2008, the Dutch Care for the Elderly Programme was commissioned by the Ministry of Health, Welfare and Sport with the guiding principles of improving care, quality of life, and self-management among older persons. As part of this Programme, The Older Persons and Informal Caregivers Survey Minimum Data Set (TOPICS-MDS) was developed to uniform collection of outcome measures. 7 To promote comparability between research studies, a preference-weighted CEP was established for TOPICS-MDS based on the health state valuations of older persons and informal caregivers. This CEP (referred to as TOPICS-CEP) was designed as a multi-faceted outcome measure applying weights derived from older persons’ priorities for different outcomes to assist in the evaluation of interventions in older persons. 8

TOPICS-CEP has been previously developed using a vignette study in which 200 persons participated. Profiles of older persons (vignettes) were used to obtain the preference weights for TOPICS-CEP’s components. 4 The aim of this current study was to determine TOPICS-CEP’s convergent and known-groups validity in large heterogeneous samples of older persons aged 65 years and older and across general population, primary care and hospital setting.

Abstract

Background
Preference-weighted multi-faceted endpoints have the potential to facilitate comparative effectiveness research that incorporates patient preferences. The Older Persons and Informal Caregivers Survey – Composite endpoint (TOPICS-CEP) is potentially a valuable outcome measure for evaluating interventions in geriatric care as it combines multiple outcomes relevant to older persons in a single metric. The objective of this study was to validate TOPICS-CEP across different study settings (general population, primary care and hospital).

Methods
Data were extracted from TOPICS Minimum Dataset (MDS), a pooled public-access national database with information on older persons throughout the Netherlands. Data of 17,603 older persons were used. Meta-correlations were performed between TOPICS-CEP indexed scores, EuroQol5-D utility scores and Cantril’s ladder life satisfaction scores. Mixed linear regression analyses were performed to compare TOPICS-CEP indexed scores between known groups, e.g. persons with versus without depression.

Results
In the complete sample and when stratified by study setting TOPICS-CEP and Cantril’s ladder were moderately correlated, whereas TOPICS-CEP and EQ-5D were highly correlated. Higher mean TOPICS-CEP scores were found in persons who were: married, lived independently and had an education at university level. Moreover, higher mean TOPICS-CEP scores were found in persons without dementia, depression, and dizziness with falls, respectively. Similar results were found when stratified by subgroup.

Conclusion
This study supports that TOPICS-CEP is a robust measure which can be used in broad settings to identify the effect of intervention or of prevention in elderly care.
7.2 Methods

Data source
Data were derived from TOPICS-MDS (www.topics-mds.eu), which is a public data repository designed to capture essential information on the physical and mental wellbeing of older persons and informal caregivers in the Netherlands. A detailed description of TOPICS-MDS has been presented elsewhere. Briefly, TOPICS-MDS consists of pooled data from various research projects which differ across study design, sampling framework, and inclusion criteria. All data were cleaned locally using a standardized protocol. Anonymized individual-level data were then submitted to a central institution (Radboud University Medical Center, Nijmegen, the Netherlands) for further validation checks and creation of the pooled dataset. Since various research projects submit information to TOPICS-MDS, the database is dynamic in nature and thus regularly updated with new observations.

Our present analysis uses the first version of the dataset available as of January 2013 and is based on 41 studies with data available on 32,310 older persons. Studies which omitted TOPICS-CEP data points by design were excluded from this study. This, resulted in a final study sample of 17,603 older persons.

TOPICS-MDS is a fully anonymized dataset available for public access, and therefore this analysis was exempt from ethical review (Radboud University Medical Center Ethical Committee review reference number: CMO: 2012/120).

Measures

TOPICS-CEP score is a preference-weighted index ranging from 0 (worst possible state) to 10 (best possible state) that combines 42 data points representing eight domains: morbidities (list of 17 pre-defined conditions widely used in the Netherlands), functional limitations (Katz index of independence), emotional wellbeing (mental health subscale of the RAND-36), pain experience (pain dimension of the EQ-5D), cognitive problems (cognition dimension of the EQ-SD+C), social functioning (item 10 from the RAND-36), self-perceived health (item from the RAND-36), and self-perceived quality of life (phrasing similar to self-perceived health item from the RAND-36). The components vary in scale range and preference weight. More detailed information about TOPICS-CEP, including a description of the data points, can be found elsewhere. Briefly, TOPICS-CEP score is calculated in four steps. Firstly, data points are coded in the same direction by means of reversed scoring. Secondly, all items that belong to the same health domain are aggregated into one component. Thus, 17 morbidity items are combined into the component number of morbidities, 15 items regarding functional limitations into number of functional limitations, and 5 emotional wellbeing items into raw emotional wellbeing score. Thirdly, a raw TOPICS-CEP score is calculated by means of applying the preference weights for the Dutch population aged 65 years and over.

\[
\text{Raw TOPICS-CEP score} = 9.00 \text{ (Intercept)} - [0.13 \times \text{morbidities}] - [0.12 \times \text{functional limitations}] - [0.03 \times \text{emotional wellbeing}] - [0.03 \times \text{pain experience}] - [0.14 \times \text{cognitive problems}] - [0.01 \times \text{social functioning}] - [0.17 \times \text{self-perceived health}] - [0.02 \times \text{self-perceived quality of life}].
\]

Finally, the raw TOPICS-CEP score is transformed into an indexed score (referred to as TOPICS-CEP score) ranging 0 to 10.

\[
\text{TOPICS-CEP score} = \frac{\text{[raw TOPICS-CEP score} - \text{minimum raw TOPICS-CEP score})/\text{raw score range}} \times 10.
\]

In this current study, only missing data points were allowed for the aggregated TOPICS-CEP components morbidities, functional limitations and emotional wellbeing. The thresholds used were less than 5 missing values for morbidities and functional limitations respectively, and less than 2 missing values for emotional wellbeing. Estimation for these data points was done by pro-rating the score. For instance, the component functional limitations includes 15 items and the scale range is 0 to 15; when 12 items are answered and the sum of the answered items is 6, then score pro-rating = [(6/12) x 15] = 7.5.

Other measures
The Cantril’s life satisfaction score is a one-dimensional index ranging from 0 (completely unsatisfied with life) to 10 (completely satisfied with life) and measures self-perceived general QOL. We used a modified version of Cantril’s self anchoring ladder where respondents were asked to rate their present life on a scale between zero and ten.

The EuroQol-5D (EQ-5D) utility score measures health related QOL (HRQOL). 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) are combined into one utility score by means of applying the scoring values for the Dutch population. The EQ-5D utility score ranges from -0.33 to 1.00 where a score of less than zero is indicative of a health state worse than death.
Socio-demographic characteristics included in our analyses were marital status, living arrangements, and education level. Included clinical data points were dementia, depression, and dizziness with falls.

Convergent validity
Convergent validity refers to how closely a measure is related to other measure of the same construct. We examined convergent validity of TOPICS-CEP score with the Cantril’s life satisfaction score and the EQ-5D utility score respectively. Convergent validity is determined by the correlation between the outcome measures.

Analyses
Feasibility was assessed by calculating the number of missing values for TOPICS-CEP. Floor and ceiling effects were assessed by reporting the proportion of respondents with minimum and maximum TOPICS-CEP scores, respectively. A floor or ceiling effect of 15% was considered the maximum acceptable.

Since TOPICS-MDS is a pooled dataset, we applied meta-analytical techniques to account for clustering within individual research projects. Pearson's correlations were used to examine convergent validity between TOPICS-CEP, Cantril’s life satisfaction scale, and EQ-5D utility score within each study. To calculate the pooled correlation coefficients random effects meta-correlations were performed. Correlations below 0.3 were referred to as weak, between 0.3 and 0.5 as moderate, and above 0.5 as strong.

Known group validity was examined by determining significant differences in mean TOPICS-CEP index scores. Mixed linear regression analyses were used to compare the scores between groups and to examine whether differences between groups were still present when adjusted for age and gender. To account for clustering within individual research projects the models included random intercepts for project. The models were constructed based on a priori expectations. Differences between parameter estimates smaller than 15% were considered to be acceptable. Analyses were performed using SPSS version 20.0 (SPSS IBM, New York, USA) and the Meta package in R (Foundation for statistical computing, Vienna, Austria).

7.3 Results
Sample characteristics
Data from 17,603 older persons from 28 projects were included in this study. The majority of the study sample were women (N= 10,817, 61.5%) and the mean (±SD) age was 79 (7) years. Overall, the sample consisted of 7,849 (44.9%) subjects living independently with others, 8,187 (46.7%) were married or cohabiting, and 7,965 (46.7%) had a secondary education level. The conditions dementia, depression, and dizziness with falls were present in; 962 (5.6%), 1,558 (9.1%), and 2,495 (14.6%) subjects of the study sample respectively. The socio-demographic distribution within...
the subgroups (primary care (N=11,892), general population (N=3,331), and hospital (N=1,534)) were similar to the combined sample.

Outcomes
Of the 17,603 participants, the majority had no missing data points for TOPICS-CEP: 88.7% (N=15,612), Cantril’s ladder: 91.9% (N=16,178) and EQ-5D: 96.6% (N=17,006). The means (±SD; minimum and maximum scores achieved) were TOPICS-CEP: 7.37 (1.23; 1.88 – 10.0); Cantril’s ladder: 7.12 (1.40; 0.0 – 10.0); and EQ-5D: 0.63 (0.29; -0.33 – 1.0). Table 1 gives an overview of the mean (±SD) scores and floor and ceiling effects for the complete sample and stratified by subgroup. The highest values possible for TOPICS-CEP, Cantril’s ladder, and EQ-5D was reported for 18 (0.1%), 379 (2.2%), and 2,009 (11.4%) older persons respectively. For each outcome measure, the lowest value possible was calculated for less than 1% of the subjects.

Table 1 The mean (±SD) scores and floor and ceiling effects for the complete sample and stratified by subgroup.

<table>
<thead>
<tr>
<th>Subgroups by study setting</th>
<th>Mean (SD)</th>
<th>Floor N (%)</th>
<th>Ceiling N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Complete study sample (N=17,603)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOPICS-CEP</td>
<td>7.37 (1.23)</td>
<td>0 (0.0)</td>
<td>18 (0.1)</td>
</tr>
<tr>
<td>Cantril’s ladder</td>
<td>7.12 (1.40)</td>
<td>22 (0.1)</td>
<td>379 (2.2)</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>0.63 (0.29)</td>
<td>0 (0.0)</td>
<td>2009 (11.4)</td>
</tr>
<tr>
<td><strong>Subgroups by study setting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Primary care setting (N=11,892)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOPICS-CEP</td>
<td>7.44 (1.15)</td>
<td>0 (0.0)</td>
<td>12 (0.1)</td>
</tr>
<tr>
<td>Cantril’s ladder</td>
<td>7.11 (1.42)</td>
<td>17 (0.1)</td>
<td>257 (2.2)</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>0.61 (0.28)</td>
<td>0 (0.0)</td>
<td>1100 (9.2)</td>
</tr>
<tr>
<td><strong>General population (N=3,331)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOPICS-CEP</td>
<td>7.37 (1.40)</td>
<td>0 (0.0)</td>
<td>5 (0.2)</td>
</tr>
<tr>
<td>Cantril’s ladder</td>
<td>7.07 (1.34)</td>
<td>4 (0.1)</td>
<td>47 (1.4)</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>0.72 (0.26)</td>
<td>0 (0.0)</td>
<td>653 (19.6)</td>
</tr>
<tr>
<td><strong>Hospital (N=1,534)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOPICS-CEP</td>
<td>7.48 (1.20)</td>
<td>0 (0.0)</td>
<td>1 (0.1)</td>
</tr>
<tr>
<td>Cantril’s ladder</td>
<td>7.36 (1.35)</td>
<td>1 (0.1)</td>
<td>49 (3.2)</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>0.61 (0.30)</td>
<td>0 (0.0)</td>
<td>210 (13.7)</td>
</tr>
</tbody>
</table>

When stratified by subgroup the mean (±SD) scores showed similar patterns. For each outcome measure the lowest value possible was achieved by less than 1% of the older persons whereas the highest possible value for EQ-5D was calculated for 19.6% (N=653) and 13.7% (N=210) of the older persons sampled from the general population and hospital respectively.

Convergent validity
Table 2 gives an overview of the meta-correlation coefficients and the 95% CI. Expectedly, TOPICS-CEP and Cantril’s ladder were moderately correlated in the overall sample and subgroups Complete sample: r=0.43; Primary care: r=0.41; General population: r=0.50; Hospital: r=0.43. In comparison, TOPICS-CEP and the EQ-5D were highly correlated (Complete sample: r=0.63; Primary care: r=0.60; General population: r=0.71; Hospital: r=0.57).

Table 2 Meta-correlation coefficients and the 95% CI of the outcome measures TOPICS-CEP, Cantril’s ladder, and EQ-5D utility score for the complete study sample and stratified by subgroup.

<table>
<thead>
<tr>
<th>Subgroups by study setting</th>
<th>TOPICS-CEP</th>
<th>Cantril’s ladder</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Complete study sample (N=17,603)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cantril’s ladder</td>
<td>0.43 (0.39 - 0.48)</td>
<td>0.34 (0.28 - 0.40)</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>0.63 (0.58 - 0.67)</td>
<td>0.34 (0.28 - 0.40)</td>
</tr>
<tr>
<td><strong>Subgroups by study setting</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Primary care (N=11,892)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cantril’s ladder</td>
<td>0.41 (0.33 - 0.48)</td>
<td>0.31 (0.21 - 0.41)</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>0.60 (0.52 - 0.67)</td>
<td>0.31 (0.21 - 0.41)</td>
</tr>
<tr>
<td><strong>General population (N=2,221)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cantril’s ladder</td>
<td>0.53 (0.51 - 0.56)</td>
<td>0.43 (0.35 - 0.50)</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>0.71 (0.68 - 0.74)</td>
<td>0.43 (0.35 - 0.50)</td>
</tr>
<tr>
<td><strong>Hospital (N=1,534)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cantril’s ladder</td>
<td>0.43 (0.35 - 0.51)</td>
<td>0.29 (0.25 - 0.34)</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>0.57 (0.51 - 0.62)</td>
<td>0.29 (0.25 - 0.34)</td>
</tr>
</tbody>
</table>

Known group validity
Table 3 illustrates the association between TOPICS-CEP scores and sample characteristics. In line with our expectations, higher mean TOPICS-CEP scores were found in older adults who were married, lived independently and had a higher education level, respectively. Moreover, the mean TOPICS-CEP scores were higher in the persons...
Table 3 The associations between TOPICS-CEP scores and sample characteristics for the complete study sample and stratified by subgroup.

<table>
<thead>
<tr>
<th></th>
<th>Complete sample</th>
<th>Subgroups by study setting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjusted for gender and age</td>
<td>Primary care</td>
</tr>
<tr>
<td></td>
<td>Estimates</td>
<td>Estimates</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>95% CI</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>78 year old¹</td>
<td>7.27 [7.05 - (7.49)]</td>
<td>7.29 [6.99 - (7.57)]</td>
</tr>
<tr>
<td>Per additional year</td>
<td>0.03 [0.02 - (0.04)]</td>
<td>0.04 [0.04 - (0.04)]</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male¹</td>
<td>7.56 [7.32 - (7.8)]</td>
<td>7.51 [7.21 - (7.82)]</td>
</tr>
<tr>
<td>Female</td>
<td>-0.39 [-0.43 - (-0.36)]</td>
<td>-0.34 [-0.38 - (-0.3)]</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married / Cohabiting¹</td>
<td>7.50 [7.26 - (7.74)]</td>
<td>7.19 [6.97 - (7.42)]</td>
</tr>
<tr>
<td>Partner deceased</td>
<td>-0.37 [-0.41 - (-0.34)]</td>
<td>-0.08 [-0.13 - (-0.04)]</td>
</tr>
<tr>
<td>Other</td>
<td>-0.22 [-0.28 - (-0.16)]</td>
<td>-0.12 [-0.18 - (-0.06)]</td>
</tr>
<tr>
<td>Living arrangements</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent alone¹</td>
<td>7.37 [7.14 - (7.59)]</td>
<td>7.27 [7.05 - (7.49)]</td>
</tr>
<tr>
<td>Independent with others</td>
<td>0.19 [0.16 - (0.23)]</td>
<td>0.01 [-0.03 - (0.05)]</td>
</tr>
<tr>
<td>Dependent</td>
<td>-1.01 [-1.08 - (-0.95)]</td>
<td>-0.86 [-0.94 - (-0.82)]</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school³</td>
<td>7.13 [6.88 - (7.37)]</td>
<td>7.02 [6.8 - (7.25)]</td>
</tr>
<tr>
<td>Secondary school</td>
<td>0.27 [0.23 - (0.31)]</td>
<td>0.17 [0.13 - (0.21)]</td>
</tr>
<tr>
<td>University</td>
<td>0.44 [0.39 - (0.5)]</td>
<td>0.30 [0.24 - (0.35)]</td>
</tr>
<tr>
<td>Dementia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No¹</td>
<td>7.43 [7.1 - (7.77)]</td>
<td>7.23 [6.93 - (7.54)]</td>
</tr>
<tr>
<td>Yes</td>
<td>-1.13 [-1.21 - (-1.04)]</td>
<td>-1.04 [-1.13 - (-0.95)]</td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No¹</td>
<td>7.42 [7.15 - (7.71)]</td>
<td>7.25 [6.99 - (7.51)]</td>
</tr>
<tr>
<td>Yes</td>
<td>-1.16 [-1.21 - (-1.1)]</td>
<td>-1.16 [-1.22 - (-1.11)]</td>
</tr>
<tr>
<td>Dizziness with Falls</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No¹</td>
<td>7.49 [7.2 - (7.75)]</td>
<td>7.29 [7.04 - (7.54)]</td>
</tr>
<tr>
<td>Yes</td>
<td>-1.07 [-1.11 - (-1.01)]</td>
<td>-1.01 [-1.05 - (-0.96)]</td>
</tr>
</tbody>
</table>

¹Reference, for example: Mean TOPICS-CEP Male = 7.56, Female = 7.17
without dementia, depression and dizziness with falls, respectively. Further, Table 3 illustrates the relationships between TOPICS-CEP scores and sample characteristics adjusted for gender and age. The parameter estimates of marital status and education level remained significant (P-values < 0.05) after adjustments; however, these exceeded the 15% threshold of change. Thus, for example the average difference between TOPICS-CEP scores of persons who were married or cohabiting versus those who had a deceased partner was still significantly different, however the difference between the scores decreased from 0.37 to 0.08. Furthermore, the parameter estimate of living independently with others was no longer significant after adjustment for gender and age. Without adjustment, the average difference TOPICS-CEP scores of persons living independently alone versus living independently with others were 0.19 points and with the adjustment the difference was 0.01 point. When stratified by subgroup similar results were found (data not shown).

7.4 Discussion

The purpose of this study was to determine convergent and known group validity of TOPICS-CEP in a large and heterogeneous sample of persons aged 65 years and older. Preference-weighted composite endpoints such as TOPICS-CEP have the potential to facilitate comparative effectiveness research, thus it is important to establish the validity of these kinds of endpoints prior to their use in the population of interest.

In this current study, TOPICS-CEP was able to accurately represent the heterogeneous composition of the overall study population. TOPICS-CEP scores obtained covered most of the entire current score range of the index and there were no floor or ceiling effects found in the total sample nor in the subsample taken from general population, primary care or hospital settings. This is important for its performance as an outcome measure. At the same time, the EQ-SD utility scores showed considerably larger ceiling effects in the general population sample. The most plausible reason why this specific subgroup exhibited this effect would be that the persons from the general population sample were less frail compared to those from the primary care and hospital sample.

Our correlation analyses revealed significant associations between TOPICS-CEP score versus Cantril’s ladder and EQ-SD utility score. The stronger correlation between TOPICS-CEP and EQ-SD indicates that the TOPICS-CEP measures important aspects of health. As expected, the correlation between TOPICS-CEP and Cantril’s ladder was moderate because the two outcomes measure different concepts. Moreover, our findings supported our hypothesis that there would be a strong correlation between TOPICS-CEP components and the EQ-SD dimensions.

TOPICS-CEP scores adhered expected patterns across marital status, living arrangements, and education level. Additionally, TOPICS-CEP was able to distinguish subjects who had dementia, depression, and dizziness with falls even when adjusted for age and gender. These findings further support the overall validity of the tool.

Our results indicate that there were no floor or ceiling effects for TOPICS-CEP in the different settings. However, similar to other studies we found a ceiling effect for the EQ-SD utility score as the percentage of persons with the highest possible EQ-SD utility score of 1.00 exceeded the 15% threshold. These ceiling effects may be due to a small range of responses (3 levels per item).24,25

There are several limitations to consider. Even though a large heterogeneous sample was used to validate TOPICS-CEP, institutionalized older persons were underrepresented in our sample. Secondly, additional research is required to examine other important properties of TOPICS-CEP, such as minimal clinically important difference and the sensitivity to detect change. For these reasons, longitudinal validation would be beneficial.

In conclusion, preference-weighted multi-faceted endpoints have the potential to facilitate comparative effectiveness research that incorporates patient preferences. This study supports that TOPICS-CEP is a good option for researchers who need an outcome measure to assess important outcomes for older persons even when it is across a range of differently functioning subpopulations. TOPICS-CEP is a robust measure which can be used in broad settings to identify the effect of intervention or of prevention in elderly care. It deserves further spread as the various outcome domains included in the measure are of great importance to the older population.
Acknowledgements
The authors thank TOPICS-MDS Consortium members.

TOPICS-MDS Consortium:
Project Group N Bleijenberg (Julius Center for Health Sciences and Primary Care, UMC Utrecht, Netherlands); JW Blom (Dept. of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands); GJM Kempen (CAPHRI School for Public Health and Primary Care, Dept. of Health Services Research, Maastricht University, Netherlands); PFM Krabbe (Dept. of Epidemiology, University of Groningen, University Medical Center Groningen, Netherlands); RJF Melis (Radboud Institute for Health Sciences, Dept. of Geriatric Medicine, Radboud University Medical Center, Nijmegen, Netherlands); EP Moll van Charante (Dept. of Internal Medicine and Geriatrics, University Medical Centre Utrecht, Netherlands); MJ Schuurmans (Dept. of Rehabilitation, Nursing Science & Sports, University Medical Center Utrecht, Netherlands); MJG Schols (Dept. of Family Medicine and Dept. of Health Services Research, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, Netherlands); MJ Schuurmans (Dept. of Rehabilitation, Nursing Science & Sports, University Medical Center Utrecht, Netherlands); DA Smilde (Stichting GENERO, Rotterdam, Netherlands).

Steering Committee BM Buurman (Dept. of Internal Medicine and Geriatrics, Academic Medical Center, Amsterdam, Netherlands); J Gusselklo (Dept. of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands); HE van der Horst (Dept. of General Practice and Elderly Care Medicine/EMGO+ Institute for Health and Care Research, VU University Medical Centre Amsterdam, Netherlands); MGM Olde Rikkert (Donders Institute for Brain, Cognition and Behaviour, Dept. of Geriatric Medicine, Radboud University Medical Center, Nijmegen, Netherlands); SEJA de Rooij (University Center for Geriatric Medicine, University Medical Centre Groningen, Netherlands); IMGA Schols (Dept. of Family Medicine and Dept. of Health Services Research, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, Netherlands); MJ Schuurmans (Dept. of Rehabilitation, Nursing Science & Sports, University Medical Center Utrecht, Netherlands); DA Smilde (Stichting GENERO, Rotterdam, Netherlands).

Working group D van den Brink (Radboud Institute for Health Sciences, Dept. of Geriatric Medicine, Radboud University Medical Center, Nijmegen, Netherlands); JE Lutomski (Radboud Institute for Health Sciences, Dept. of Geriatric Medicine, Radboud University Medical Center, Nijmegen, Netherlands); L Qin (Radboud Institute for Health Sciences, Dept. of Geriatric Medicine, Radboud University Medical Center, Nijmegen, Netherlands).

7.5 References


Part 5
Summary and Discussion
Chapter 8

Summary and general discussion
With the increasing proportion of elderly and its impact on the organization and delivery of healthcare, the Dutch Ministry of Health, Welfare, and Sport commissioned the National Care for the Elderly Programme (NCEP) with the aim to develop a more proactive, integrated healthcare system for older patients. This programme funded over 60 research and implementation projects as well as the development of The Older Persons and Informal Caregivers Minimum Data Set (TOPICS-MDS). As part of the TOPICS-MDS initiative, a uniform dataset measuring key outcomes in health, wellbeing, and health services utilization was created and administered to all older persons and caregivers participating in these studies.¹

Despite the collection of standardized outcome measures in TOPICS-MDS, there remained the inherent need to create a composite endpoint (CEP) to facilitate the objective evaluation of interventions. Thus, the general aim of this thesis was to develop and validate a CEP for TOPICS-MDS to measure HR-QOL in older persons. This CEP, henceforth referred to as TOPICS-CEP, integrates 42 TOPICS-MDS items from eight domains (chronic conditions, functional limitations, emotional wellbeing, pain experience, cognitive functioning, social functioning, self-perceived health and self-perceived QOL) into a single preference-weighed outcome measure based on the health state preferences of older persons and informal caregivers.

This chapter summarizes the main findings of each study (described in Chapters 2 to 4) and reflects on methodological issues. Furthermore, implications for future research and clinical practice are provided. With the strengths and limitations of the different studies being discussed in the preceding chapters, this chapter focuses on general considerations.

8.1 Summary of the main findings

This thesis describes the development and validation of TOPICS-CEP. Since person-centred care is becoming increasingly important in modern medicine, Part 2 focuses on the health state preferences of older persons as well as their informal caregivers, who often serve as proxies. After establishing preference weights for TOPICS-CEP, the relevance of these preference-weights for the valuation of health outcomes was examined in Part 3 by comparing health state preferences of older persons, informal caregivers and healthcare professionals. Lastly, TOPICS-CEP’s psychometric properties were studied in Part 4 to validate TOPICS-CEP for HR-QOL.
In Chapter 1, the rationale for the development of TOPICS-CEP are described. Briefly, an aging population has led to an increased strain on healthcare services. Nowadays, outcomes are considered to be a necessity in determining the quality of healthcare. Comparative effectiveness research in elderly care is challenging, because the health states of older persons are complex and interventions often influence a broad range of domains. To circumvent these obstacles, TOPICS-CEP a preference-weighted multifaceted CEP, was established.

Part 2 describes the development of TOPICS-CEP. Ageing of our population has a major impact on the organisation and delivery of healthcare. Healthcare systems have to meet the needs of geriatric patients while the shortage of healthcare workers is likely to grow resulting in an increase in the number of informal caregivers. Informal caregivers often serve as proxies which means they are appointed to make healthcare decisions on behalf of the older person in case he or she is incapable of making his or her wishes known, e.g. when the patient has a low cognitive status. Therefore, the preference weights for the eight TOPICS-CEP domains were obtained based on the health state preferences of both older persons and informal caregivers. A vignette study was used to retrieve the weights and the results are described in Chapter 2. A sample of 124 community dwelling elderly aged 65 years and older and 76 informal caregivers participated. When the preference weights given by older persons were compared with those given by informal caregivers no significant differences were found. Consequently, TOPICS-CEP was established based on the health state preferences of both groups.

Previous research has shown that variation in health state preferences is influenced by various characteristics such as sex, age, and current health.2-4 This variability can further be explained by personal perspective and coping strategies. TOPICS-CEP was established based on the health state preferences of a heterogeneous group of raters. Therefore, we studied whether the characteristics of our raters systematically influenced the preference weights of TOPICS-CEP’s components and how this may affect TOPICS-CEP scores. Chapter 3 describes the results of this study. Our findings suggest that the preference weights of TOPICS-CEP’s components (described in Chapter 2a) were influenced by our raters’ age. Moreover, when we used weights based on the preferences of 65 year olds and those based on 85 year olds to evaluate the situation of a sample of cases results show low agreement between the two outcome measures. Consequently, since the age distribution of our study sample differed from the age distribution of the Dutch population aged 65 years and over, the preference weights of TOPICS-CEP had to be adjusted. With these adjustments, TOPICS-CEP may, on average, better reflect the health state valuation of the older Dutch population.

Part 3 provides evidence to support the relevance of a preference-weighted CEP based on patients’ health valuations, such as TOPICS-CEP. There were several reasons to establish TOPICS-CEP which combines objective outcomes and subjective preferences into one multifaceted outcome measure. Even though simply assessing the value of healthcare by asking patients to rate their own situation on a scale from 0 to 10 is common, the approach may result in a different valuation of a certain health state than utility measures based on external raters. This, as self-assessments are prone to adaptation and response shift. Such phenomena may result in over- or underestimation of the magnitude of the change in health status. Secondly, relying solely on objective measures, such as physicians’ reports of chronic conditions or functional limitations omits relevant factors, primarily the patient’s preferences.

Instead of using self-ratings in our vignette study, we used the ratings of multiple peers to obtain the preference weights for TOPICS-CEP components. The Bland-Altman plot, the most preferred method to investigate agreement, was originally designed to evaluate the agreement of two measurement methods with two single ratings per case. Yet, as we want to be able to compare one self-assessed rating versus three to five peer ratings per case, an adapted Bland-Altman plot was derived. The adjustment of the Bland-Altman approach to evaluate possible patterns of discord between two measurement methods with an unequal number of observations per case is described in Chapter 4.

Self-rated QOL provides useful global assessments of perceived QOL, however, peers may provide complementary information that are of great interest. Chapter 5 describes the study in which we investigated the agreement between self-rated versus peer-rated scores. As expected, the findings suggest that self-rated and peer-rated scores are not interchangeable. However, in our opinion self- and peer-rated assessments are complementary to each other. As the first gives insight into how patients experience their QOL at a given time, while the second gives insight into how the population to which the patient belongs would assess the QOL on average. With TOPICS-CEP established, based on the health valuations made by peers, we have the ability to translate objective assessments into perceived QOL to evaluate the effect of interventions on average.

Even though various stakeholders in geriatrics share a mutual goal which is to improve a person’s health and wellbeing, studies have shown significant differences between the perspectives of older persons and their healthcare professionals.7-8 Such disparities can lead to treatment decisions by professionals that are at odds with patient preferences and biased interpretations in effectiveness studies. Consequently, we expected that the preference weights of the TOPICS-CEP’s components would differ between those of older persons and their healthcare
professionals. For that reason, we explored the TOPICS-CEP components’ weights based on the health state preferences of healthcare professionals and compared them with the weights based on the health state preferences of older persons and informal caregivers obtained in our previous study (Chapter 2). The results of this study are described in Chapter 6. Additional to the sample of 124 community dwelling elderly aged 65 years and older and 76 informal caregivers who participated in our previous study, data were obtained from 330 healthcare professionals (physicians, nurses, welfare staff and allied healthcare professionals). Comparing the preference weights of the older persons and informal caregivers versus those of healthcare professionals showed significant differences, underscoring why outcomes that are based on patients’ preferences are critical. Morbidities and functional limitations were given more weight by older persons and informal caregivers than by healthcare professionals, while the opposite was true for pain experience, social functioning, and self-perceived QOL. Hence, we concluded that it is important to recognize the discrepancies between the health state preferences of various stakeholders to: (1) correctly interpret results when studying the effectiveness of interventions in elderly care; and (2) establish appropriate healthcare policies.

Preference-weighted CEP, such as TOPICS-CEP, have the potential to facilitate comparative effectiveness research. It is important to establish the validity of these kinds of endpoints prior to their use in the population of interest. Therefore, we examined TOPICS-CEP’s construct validity in large heterogeneous samples of older persons aged 65 years and older. Chapter 7 describes the results of this study which aimed to: (1) study how closely TOPICS-CEP scores were related with Cantril’s life satisfaction scores and EQ-5D utility scores; (2) assess whether TOPICS-CEP scores were significantly different between groups known to differ in HR-QOL; and (3) examine whether the validation results for TOPICS-CEP were generalizable across different settings (general population, primary care and hospital). Data from 17,603 older persons obtained from TOPICS-MDS national dataset were included in this study. Firstly, our meta-analyses revealed significant associations: a strong positive correlation between TOPICS-CEP and EQ-5D and only a moderate positive correlation between TOPICS-CEP and Cantril’s ladder (i.e. an overall QOL measure). Secondly, based on multivariable mixed models, we found that TOPICS-CEP scores adhered expected patterns across known-groups. Finally, our results indicate that there were no floor or ceiling effects for TOPICS-CEP in the different settings whilst a ceiling effect was found for the EQ-5D utility score in the general population. The TOPICS-CEP behaved in accordance with a priori hypotheses, thus supporting construct and known-group validity. Hence, TOPICS-CEP is a robust generic measure which can be used in broad settings to establish health outcomes in elderly care.

8.2 General discussion

In today’s healthcare environment, it is increasingly important to be able to quantify the amount of change (ultimately: the added value) associated with a given intervention. The aim of this thesis was to develop and validate TOPICS-CEP; a preference-weighted multifaceted outcome measure.

In this section, the relevance of quality indicators and the possible role of TOPICS-CEP are considered. We discuss the practical, theoretical and empirical challenges and opportunities we encountered while carrying out the research project reported in this thesis. The benefits and limitations of TOPICS-CEP are discussed. Finally, implications and recommendations are made for future research and clinical practice.

From TOPICS-MDS with single outcomes towards a Composite EndPoint

TOPICS-MDS was created to record relevant outcomes. The current healthcare model emerged from the need to treat acute or singular chronic conditions in younger populations. Consequently, with our aging society, healthcare professionals are now working with a quality of care paradigm and therefore in an healthcare environment that is incompatible with the needs of older persons with increasingly complex health states. Their symptoms often reflect an interaction between multiple chronic conditions and functional limitations which influence the QOL. For these persons, the incoherent treatment and monitoring of single conditions without taking into account general wellbeing does not promote healthy aging, not even when the single diseases are fully taken care of in line with the leading guidelines and quality indicators in the specific fields. Instead, such a narrow focus neglects additional gains in physical, mental, and social health that are often found when a more integrated approach to healthcare is taken.

Because of this lack of appropriate outcome measures TOPICS-MDS was created. It was designed not only to record relevant morbidity outcomes, but also to provide insight into physical, mental, and social wellbeing of older persons and their caregivers. In brief, for older persons, information is collected on demographics, morbidity, QOL, functional limitations, mental health, social functioning and health service utilisation. For informal caregivers, information is collected on demographics, hours of informal care and QOL.

The importance of TOPICS-CEP

Outcomes collected by means of TOPICS-MDS can be analysed separately (e.g. per item or item subset). However, using a single TOPICS-MDS item or item subset to compare outcomes has the disadvantage that it may lead to confusion when
competing interventions demonstrate different patterns of effect. Additionally, researchers and healthcare professionals have the tendency to evaluate merely the outcome they are directly accountable for, which makes it even more difficult to compare the effects of competing interventions. Therefore, TOPICS-CEP was developed to facilitate comparative effectiveness research and provide users of TOPICS-MDS, for older persons, with the option of a preference-weighted measure which combines the outcomes of eight domains.

To our knowledge, there is no other preference-weighted quality indicator available specifically designed to assess and compare the outcomes of interventions in elderly care prior to the development of TOPICS-CEP. Previous studies have used generic instruments such as the Medical Outcomes Study 36-item Short-Form survey (MOS SF-36), the Patient-Reported Outcomes Measurement Information System 29-item general health form (PROMIS-29), and the EuroQol-5 Dimensions (EQ-5D) to assess the effect of interventions on health outcomes. However, these are generic health status measurement scales, which all use a number of items that are not appropriate for older subjects (e.g. work performance), while specific elements most relevant for older persons are not addressed at all (e.g. cognitive decline and loss of autonomy).

Moreover, there is no consensus on how to combine the (subscales of the) SF-36 items into one CEP. Hence, SF-36 consists of eight scaled scores for the following outcome domains: vitality, physical functioning, bodily pain, general health perceptions, physical role functioning, emotional role functioning, social role functioning, and mental health. The scores are the sum of all items within each of the domains transformed into a score ranging from 0 to 100. This method is based on the assumption that all items within a domain are equally relevant. This is in contrast to the new paradigm of health proposed by Machteld Huber et al. and her empirical work supporting this. In their quest toward operationalising the new concept of health which is “the ability to adapt and self-manage, in the face of social, psychical, and emotional challenges” the importance of the following six health dimensions were explored: bodily functions, mental functions and perception, spiritual / existential dimension, QOL, social and societal participation, and daily functioning as viewed by various stakeholders. Their results suggest that the dimensions are not considered to be equally important by any of the stakeholder groups. Thus, a CEP needs to be preference-weighted based on the health state preferences of the specific target group for whom the instrument is developed.

PROMIS-29 consists of seven raw scores and T-scores for the outcomes: physical functioning, anxiety, depression, fatigue, sleep disturbance, satisfaction with social role, and pain interference plus a single pain intensity rating. With the standardized T-score statements can be made with regard to the score of an individual in comparison to the mean of a certain population whose norm scores are available. However, similar to the SF-36 no statements can be made regarding the overall change (effect of an intervention on all domains) as valued by the Dutch older population aged 65 years and older.

EQ-5D consists of five dimensions: mobility, self-care, daily activities, pain and discomfort, anxiety and depression. Each dimension has three levels: no problems, moderate problems, extreme problems. Data are combined into one utility score by means of applying the scoring values for the Dutch population. However, a limitation of the EQ-5D utility score is the ceiling effect reported in various studies including our own validation study. The most plausible reason why in our study the ceiling effect was present in the subgroup general population is that the persons from this sample were less frail compared to those from the primary care and hospital sample. Similar drawbacks are met in using the EQ-5D+C, which is the EQ-5D with the addition of one cognition item. Weighting of the EQ-5D items is done by translating the health states into utilities, which is a distinct method of weighting. Often these weights are based on a population that is different from the targeted population, which also is a limitation of this measure.

Based on these limitations of other generic tools and the results found in our validation study, TOPICS-CEP may be of great value for quality improvement in the elderly care. By using preference-weighted outcome measures the desirability of health outcomes are considered. These kind of measures are distinct from health status instruments, because they characterize how health outcomes are valued as a whole based on the values of relevant respondents. TOPICS-CEP reflects on average the value of interventions according to the Dutch population aged 65 years and older. By reporting these values, quality may be well monitored and quality improvement driven.

Determining perceived importance of the outcomes
There are methodological challenges in deriving valuations of older participants due to reduced cognitive capacity and the multifaceted nature of geriatric care. To obtain the preference weights for TOPICS-CEP we considered two often used methods: vignette studies and discrete choice experiments (DCE’s). In vignette studies respondents are asked to value descriptions of situations (scenarios, e.g. health states) of persons (profiles, cases) to obtain the value placed on each component (item), included in the vignette. In DCE’s respondents are asked to state their choice over sets of hypothetical alternatives to derive the value placed on each component (item) included in the alternatives.
At the beginning of our project, we explored the feasibility of the two different methods in a small pilot (not described in this thesis). All six participants were able to understand both tasks. When they were asked which method they preferred they all chose the vignettes. A DCE was regarded as more difficult task to complete since they could not always easily discern a preferable health state. In contrast, rating a single health status, described in a vignette, was perceived as an easy task to complete. This is in line with the literature which has found that DCE’s with too many items (attributes) negatively affect the practical feasibility by imposing a significant cognitive burden on respondents.

Further, in the design of a DCE hypothetical alternatives are generated and combined to create choice sets. For example, 10 items with each two levels leads to $2^{10} = 1024$ alternatives. In our study we had 46 items (covering 8 domains) with the number of levels ranging from 2 to 18. Thus, there would be too many alternatives to combine and choose from to perform a DCE. Besides this practical limitation, we found in the literature that the inclusion of too many attributes with conceptual overlap between two or more attributes further affects the statistical quality of a DCE. This would prevent the accurate estimation of the main effect (the direct independent effect) of a single attribute on the dependent variable due to interaction effects.

In light of these marked disadvantages, the use of vignettes was the preferable alternative to deriving preference weights for TOPICS-CEP. This method allowed for the inclusion of a relatively large number of items without comprising statistical integrity. By using empirical data, derived from TOPICS-MDS National Database, vignettes with plausible health states (based on real older persons) were constructed. Moreover, none of the participants had any problems with the execution of the assignment, which was to rate the general wellbeing a random sample of older persons who were described in vignettes.

However, the vignette approach does have one important disadvantage. Throughout the literature concerns are raised that vignettes do not accurately reflect ‘real world’ phenomena, and that this artificiality affects the validity of results and conclusions of these studies. Although respondents are asked to imagine themselves in certain hypothetical situation, they are reporting about someone else. Respondents who are actually in the described situation may respond quite differently. Therefore, we have explored the influence of our raters’ characteristics on health state preferences. Our findings suggested that the preference weights of TOPICS-CEP components were influenced by our raters’ age but not by other characteristics, such as functional limitations or chronic conditions (only the first was reported in our publication). Thus, in our opinion, when vignettes are used appropriately they can be a flexible, practical, and powerful tool for studying how various factors influence judgments and decisions. This conclusion has not changed since the start of this study, which warrants further use in future research.

Motivations to use TOPICS-CEP

General wellbeing or HR-QOL include various domains such as chronic conditions, functional limitations, and emotional wellbeing. When a researcher wishes to measure the effect of an intervention on improving patients health states, TOPICS-CEP can be used as it combines items or composites into one single measure and it takes into account the perceived importance of the various domains according to older persons and informal caregivers.

Using TOPICS-CEP to measure, report, and compare outcomes in health services also confers several key advantages. First, it can drive improvement in healthcare by giving providers a simple tool to benchmark performance measures for individual patients (to be combined in the plan-do-check-act cycle of quality improvement). Thus, by measuring TOPICS-CEP score over time, providers are able to make adjustments in the care they provide, share their successes, and explore for causes when progress is lacking or stays behind. Second, by making performance transparent (e.g. publicly reporting TOPICS-CEP scores), patients have the opportunity to be well informed. Well informed patients are better able to assess quality of healthcare for themselves (e.g. by means of using TOPICS-CEP as a PROM), and may use this information to make choices, ask questions, and advocate for good healthcare. Third, assessing outcomes enables the shift in payment systems for healthcare from volume to value based healthcare reimbursement. Increasingly, worldwide private and public payers use outcomes such as patient reported outcomes (PROs) as preconditions for payment and targets for bonuses.

Besides these theoretical motivations for the use of TOPICS-CEP there are several practical motivations. Similar to other often used continuous composite endpoints producing a single score, such as the Disease Activity Score with 28-Joint Counts (DAS28), TOPICS-CEP has the advantage over the interpretation of individual components as it provides a meaningful estimate of HR-QOL with interpretation of multiple data points simultaneously and with a more responsiveness to change than single items. Additionally, continuous CEP’s are less affected to selection bias related to the reporting of a single measurement and they are preferable for statistical analysis in studies. Combining single items or components into a single score increases the event rate, which can be translated into a reduction in the minimum number of patients required for the clinical trial. If larger effect sizes are of interest, then a trial using TOPICS-CEP can gauge the effect of an intervention on
HR-QOL using a smaller sample size compared to single items (only when everything else remains equal).

A well-designed clinical trial that prospectively embeds a CEP into its primary analysis plan is empowered to measure smaller effects, while using single endpoints could miss these effects due to the narrow focus. Assessments of single endpoints reveals that the isolated interpretation of a single component or health domain can be misleading. For example, consider the vignette / profile of Jeremy introduced in Chapter 1 [Box 1]. Jeremy was experiencing moderate pain, had no problems with his memory, attention and thinking, and was capable to handle his own finances without any assistance. If a researcher is interested in the effect of pain medication on the single outcome domain pain experience and Jeremy were to report, after six weeks, that he does not experience any pain, but does have moderate problems with his memory, attention and thinking, and is no longer capable to handle his own finances, the intervention would be labelled as successful. This, as the focus would only be on the fact that Jeremy is no longer experiencing any pain. However, if the researcher were to use TOPICS-CEP for a more holistic approach, the intervention would not be labelled as successful: TOPICS-CEP score before the intervention was 7.3, and after the intervention 6.7. Furthermore, if we use the case of Jeremy and calculate a composite score based on the health state preferences of healthcare professionals the intervention would be labelled as neutral: Composite score before and after the intervention was 8.2. This example underlines the importance of using generic preference-weighted outcomes to assess the effects of interventions on health states.

8.3 Limitations of TOPICS-CEP

Compromise between completeness and practical feasibility
Before measuring quality of health care, it is important to choose the right tools for the job. Each measurement tool assesses performance from a specific angle, e.g. focussing on process, structure or outcome. Moreover, a tool may have been developed for a specific group of patients (e.g. the Hip disability and Osteoarthritis Outcome score is a condition-specific tool) or it may have been developed for a generic group (e.g. EuroQol-5D is a generic tool). Hence, when selecting or developing a tool the included items are of great importance.

TOPICS-MDS was developed as part of a National initiative to create a minimum data set on older persons’ health outcomes. Key domains and items for the standardized questionnaire were outlined by a working group; the instrument’s content and utility evaluated by an independent multidisciplinary panel with expertise in gerontology, epidemiology, biostatistics and health services research. Despite this rigorous approach, some may question the exclusion of certain items, such as an indicator for loneliness. Yet, as described by Lutomski et al. (2013) the use of a minimal dataset is by definition a compromise between maximum desired completeness and practical feasibility. Given this limitation, it is not the intention to suggest that TOPICS-MDS

Box 1 Vignette / profile of Jeremy

Baseline:
Jeremy is a 75 year old male and lives independently with his wife Nora.
In the last 12 months, Jeremy was diagnosed with diabetes and complained about dizziness and falling.
Jeremy needs help with 4 out of 6 self-care tasks:
He need assistance with: bathing, getting dressed, walking, and getting up out of a chair.
He does not need any assistance with: eating, combing his hair, and going to the toilet.
Jeremy needs help with 4 out of 7 other daily tasks:
He need assistance with: travelling, grocery shopping, preparing a meal, and performing household tasks.
He does not need any assistance with: using the telephone, taking his medication, and dealing with finances.
According to Jeremy, the last month he was always happy and calm, almost never sad or nervous, and he never felt like he could not be cheered up. Moreover, his social life has not been influenced at all by his emotional wellbeing over the last four weeks.
Today, Jeremy is experiencing moderate pain. He has no problems with his memory, attention and thinking.
According to Jeremy his health is good and his quality of life is excellent.

Reported change during Follow-up:
Jeremy is no longer experiencing any pain. However, he has moderate problems with his memory, attention and thinking, and is no longer capable to handle his own finances.

<table>
<thead>
<tr>
<th>TOPICS-CEP score:</th>
<th>Score based on health state preferences of healthcare professionals:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline = 7.3</td>
<td>Baseline = 8.2</td>
</tr>
<tr>
<td>Follow-up = 6.7</td>
<td>Follow-up = 8.2</td>
</tr>
</tbody>
</table>
should take the place of all pre-existing outcomes or descriptive instruments. TOPICS-MDS was designed to be used as either a primary or complementary survey. Its primary focus was to ensure that critical core outcomes are uniformly collected. This flexible design feature can, in turn, be viewed as a major strength of the survey.13

Further, results from our meta-analyses support that TOPICS-CEP is a good option for researchers who need an outcome measure to assess important outcomes for older persons, even when it is across a range of differently functioning subpopulations. This means that TOPICS-MDS includes sufficient items to identify the effect of intervention or prevention on desirable health outcomes in elderly care.

Relevance of the components cannot be compared
To establish TOPICS-CEP we conducted two vignette studies to obtain the preference weights for the eight TOPICS-CEP components based on health valuations made by older person and informal caregivers. Additionally, we obtained the preference weights for TOPICS-CEP components based on health valuations made by healthcare professionals to explore the relevance of patient-focused CEPs. Unfortunately, with the preference weights obtained we cannot make any comparisons between the components, thus we cannot make statements regarding which domain was considered most relevant by the older persons, informal caregivers, and healthcare professionals, respectively. However, this also was not the purpose of this project, which was to develop a multifaceted endpoint that combines TOPICS-MDS outcomes and considers the desirability of health outcomes.

The eight TOPICS-CEP components have different number of levels, e.g. self-perceived QOL has a five-level response option ranging ‘poor’ to ‘excellent’ and cognitive problems has a three-level response option ranging from ‘no problems’ to ‘severe problems’. Because of this variety and the interdependence of the components the preference weights of the eight TOPICS-CEP components cannot be compared, e.g. even though the component self-perceived QOL has a preference weight of 0.02 and the component cognitive problems has a preference weight of 0.14 this does not imply that the older persons and informal caregivers find cognitive problems seven times more important than self-perceived health. Hence, such direct comparisons cannot be drawn.

However, the preference weights can be used to calculate TOPICS-CEP scores and a composite score based on health state preferences of healthcare professionals. When we calculated these two scores for 384 cases and compared them, we found mean differences that were not close to zero. Thus, the two outcomes were systematically producing different results, which suggests that on average health valuations made by older persons and informal caregivers differ from those made by healthcare professionals. These disparities found support the importance of patient-focused outcomes such as TOPICS-CEP.

Research versus clinical practice
The use of a CEP can simplify and streamline the general evaluation of new interventions. However, in clinical practice it is important to recognize that two individuals with identical self-assessed health status as measured by a generic instrument, such as TOPICS-MDS, may value their general wellbeing or HR-QOL differently. This can be explained by the fact that valuations are influenced by personality traits, e.g. neuroticism are correlated with lower valuation of one’s general wellbeing, while extraversion, conscientiousness, agreeableness, and openness to experience are correlated with higher valuation of one’s general wellbeing.14

Moreover, the two different valuations could be explained by a phenomenon called response shift. Response shift is defined as an adaptation to changing health and refers to a change in the meaning of QOL over time.15 It is beneficial for patients as it can help them in adapting to a new situation. There are three forms of response shift: (1) a change in someone’s internal standards (e.g. recalibration); (2) a change in the importance attributed to health domains (i.e. change in values or reprioritization); and (3) a change in the definition of the concept of QOL (i.e. reconceptualization).

Consider two persons with equal personality traits and the exact same health status. However, one of the individuals was born disabled and uses a wheelchair all of his life, while the other just had an accident a week ago and uses a wheelchair ever since. Based on our coping strategies it very likely that, even though both individuals are disabled, the first person is more content with his HR-QOL than the second person as this has been his situation this whole life, while the second person needs to adjust to this new situation. With the concept of response shift in mind, we established TOPICS-CEP based on the health valuations made by older persons who rated the health states of their peers.

Thus, when TOPICS-MDS instrument for older persons is used in clinical practice it is of great importance to ask the individual patients which outcomes (health domains) they find relevant. Yet, TOPICS-CEP is a valuable outcome for comparative effectiveness research, as more general statements can be made regarding the change of TOPICS-CEP scores which on average reflect added value according to the Dutch population aged 65 years and over.
8.4 Future directions of TOPICS-MDS and TOPICS-CEP

Future role: PROMs in clinical practice
TOPICS-MDS contains core outcomes which validly measure health and wellbeing of older persons in the Netherlands. TOPICS-CEP combines these relevant outcomes into a multifaceted outcome measure and reflects health valuations according to Dutch persons aged 65 years and older. Ideally, TOPICS-MDS and its corresponding CEP will continue to grow, promote standardized data collection, and stimulate comparing the effectiveness of interventions in both academics as well as clinical practice. The first initial steps have been made, as stated by Jennifer Lutomski.

“The Dutch Society for Clinical Geriatrics (NVKG) has recently adopted TOPICS-MDS to monitor patient wellbeing. Thus, TOPICS-MDS will have the dual purpose of collecting pertinent research data as well as assisting clinical evaluation by providing the patient’s perspective on quality of care. The embracement of TOPICS-MDS by the Dutch Society for Clinical Geriatrics reflects the growing international interest in using Patient Reported Outcome Measures (PROMs) to measure quality of care”.

In contrast to the typical way in defining successful treatments by clinician-oriented measures, e.g. the presence or absence of disease, PROMs are rooted in the perspective of the patient. A PROM is an instrument, scale, or single-item measure used to assess the concept Patient Reported Outcome (PRO) as perceived by the patient, obtained by directly asking the patient to self-report. The insight PROMs add into patients’ QOL cannot be readily extracted from electronic patient records. Thus, investigating how TOPICS-MDS and TOPICS-CEP preform as a PROMs is highly relevant for the healthcare sector.

The outcomes for patients can be improved by the use of PROMs as clinical tools. PROMs are promoted to support and enhance collaborative models of patient-clinician interaction and shared decision making. Moreover, they help ensure that patients’ wishes, values and preferences. To build on this previous work, additional research to examine the responsiveness (does the instrument detect changes over time that matter to patients) of TOPICS-MDS as a PROM is currently underway in cooperation with ZonMw (The Dutch Organisation for Health Research and Development). The responsiveness of TOPICS-CEP as a composite PROM has not yet been explored. However, with the individual TOPICS-CEP components (e.g. EQ-5D utility score and KATZ index) often being used in trials to indicate the effect of interventions, we expect TOPICS-CEP to be responsive enough to reflect change over time.

Health insurers and various healthcare facilities have made it clear that they are interested in TOPICS-CEP. They want to use TOPICS-CEP scores to interpret and compare the effectiveness of clinical interventions and assess care management. Since 2006, the Dutch healthcare system has changed towards a demand-driven system of managed competition to contain costs and improve efficiency and quality of care. A precondition for managed competition is the freedom for insurers to contract and negotiate. Hence, since the introduction of the Health Insurance Act, insurers are free to decide where and by whom the benefits in the basic package are provided. This selective contracting is based on clear quality criteria. Consequently, both health insurers and healthcare providers are more interested in measuring, reporting and comparing quality of healthcare. Yet, there is insufficient data available to measure key outcomes in health and wellbeing to effectively compare the quality of care provided.

At the moment, uniform information is usually drawn from various sources, including existing clinical and administrative records and patients’ own assessments through Consumer Quality Index (CQ-index) surveys. With the growing demand for patient-centred care and for the transparency and accountability of health services, client surveys such as the CQ-index are increasingly mandatory. The CQ-index is a National standard to measure healthcare quality from the perspective of patients and is based on the Consumer Assessment of Healthcare Providers and Systems (CAHPS) from the US and the Dutch Quality Of care Through the patient’s Eyes (QUOTE) instruments. It measures what patients find important in healthcare, what their actual experiences are, and how they rate the overall quality of care. CQ-index surveys can be used for the purpose of public reporting, quality assurance and governance, however, one great disadvantage of the surveys is their lengthiness.

Priorities for future research and development
During the validation of TOPICS-CEP we found a clear disparity between the valuations and health state preferences of older persons and healthcare professionals. These findings emphasise the necessity of outcome measures that are based on patients’ wishes, values and preferences. To build on this previous work, additional research to examine the responsiveness (does the instrument detect changes over time that matter to patients) of TOPICS-MDS as a PROM is currently underway in cooperation with ZonMw (The Dutch Organisation for Health Research and Development). The responsiveness of TOPICS-CEP as a composite PROM has not yet been explored. However, with the individual TOPICS-CEP components (e.g. EQ-5D utility score and KATZ index) often being used in trials to indicate the effect of interventions, we expect TOPICS-CEP to be responsive enough to reflect change over time.

Health insurers and various healthcare facilities have made it clear that they are interested in TOPICS-CEP. They want to use TOPICS-CEP scores to interpret and compare the effectiveness of clinical interventions and assess care management. Since 2006, the Dutch healthcare system has changed towards a demand-driven system of managed competition to contain costs and improve efficiency and quality of care. A precondition for managed competition is the freedom for insurers to contract and negotiate. Hence, since the introduction of the Health Insurance Act, insurers are free to decide where and by whom the benefits in the basic package are provided. This selective contracting is based on clear quality criteria. Consequently, both health insurers and healthcare providers are more interested in measuring, reporting and comparing quality of healthcare. Yet, there is insufficient data available to measure key outcomes in health and wellbeing to effectively compare the quality of care provided.

At the moment, uniform information is usually drawn from various sources, including existing clinical and administrative records and patients’ own assessments through Consumer Quality Index (CQ-index) surveys. With the growing demand for patient-centred care and for the transparency and accountability of health services, client surveys such as the CQ-index are increasingly mandatory. The CQ-index is a National standard to measure healthcare quality from the perspective of patients and is based on the Consumer Assessment of Healthcare Providers and Systems (CAHPS) from the US and the Dutch Quality Of care Through the patient’s Eyes (QUOTE) instruments. It measures what patients find important in healthcare, what their actual experiences are, and how they rate the overall quality of care. CQ-index surveys can be used for the purpose of public reporting, quality assurance and governance, however, one great disadvantage of the surveys is their lengthiness.
Insurers and health organisations who want to use TOPICS-CEP have some concerns regarding the feasibility of TOPICS-CEP. To motivate the use of TOPICS-CEP in clinical practice, further research needs to be done to examine whether TOPICS-CEP can be calculated with fewer TOPICS-MDS items without reducing the reliability and validity of TOPICS-CEP. In addition to item reduction, TOPICS-MDS can be made even more user friendly by designing it as an online tool that can calculate TOPICS-CEP at the press of a button. Finally, as stated before two individuals with identical self-assessed health status as reported by TOPICS-MDS may value their general wellbeing differently whilst their TOPICS-CEP scores are the same. Hence, the tool should not only be used as a measurement instrument, but also to facilitate patient-clinician interaction and to promote patient-centred care and shared decision making; understanding what the wishes, values, and preferences are of the patient who is sitting in front of you.

Other recommendations
In this section we would like to elaborate on the importance of person-centred care and give our recommendations. Person-centred (or patient-centred) care is one of the overarching goals of health advocacy and it is defined as “Providing care that is respectful of and responsive to individual patient preferences, needs and values, and ensuring that patient values guide all clinical decisions”.20

To establish person-centred care, an integrated approach is necessary. A white paper published by Vilans (Centre of Expertise on Long-Term Care) describes the House of Person-Centred Care (Huis van Persoonsgerichte zorg),21 which is based on the “Co-creating health” approach and “House of Care” model from the UK.22, 23 Components of this framework are similar to the “Chronic Care Model” and “Guided Care model”, but differ because the components are structured in the shape of a house to emphasise the importance and balance needed between all components. Also, the “The House of Person-Centred Care” is meant for all people with long term care needs, and is not limited to people with chronic diseases. The model can act as a checklist (highlighting what needs to be in place), a metaphor (emphasising that person-centred care is complex and that all the components need to be in place to make it a success), and a flexible framework (guiding each healthcare professional to build a stable house designed round the needs of individual patients).

The House of Person-Centred care consists of five building blocks.21 To make this house stable the building blocks need to be secured. The left wall consists of well-informed patients and informal caregivers. The right wall are coaching healthcare professionals. The roof are organisations that facilitate and stimulate person-centred care. The foundation are context and healthcare procurement. In the heart of the house healthcare professionals and patients establish person-centred care by jointly (1) setting the agenda, (2) defining goals and planning action, and (3) evaluating and adjusting the actions when necessary during follow-up. TOPICS-CEP could fulfill its role as a PROM in the heart of the house by evaluating the actions (interventions).

Vilans, being an independent non-commercial knowledge institute in health care, might support further spread of TOPIC-CEP. Vilans operates at the interface between policy, science, and practice and helps professionals to improve care for the chronically ill, frail older people, and people with disabilities. Determining where improvements need to be made (agenda setting), which improvements need to be made (defining goals), how to achieve them (action planning and adjusting after follow-up) is done in partnership with managers, patients / clients / citizens, and policymakers in the heart of the house of care. To embed the person-centred approach, we strive to include one or more representatives of the target population within the organisation and within the projects Vilans facilitates.

Ideally, TOPICS-MDS and TOPICS-CEP are used to promote the person-centred care approach. For example, take the case of Jeremy [Box 1]. Before the consultation the...
doctor emails TOPICS-MDS, for older persons, or a selection of TOPICS-MDS items that are included in TOPICS-CEP to Jeremy ('agenda setting'). Jeremy fills in the survey and the doctor receives a notification. During the consultation Jeremy's health outcomes are discussed and the doctor asks "Jeremy, what do you need? What are your wishes?" ('coaching healthcare professionals'). Perhaps Jeremy says he just want to live without any pain or he could ask the doctor for advise on how he can become more independent again ('defining goals'). After listening to Jeremy's desires the doctor informs Jeremy about different possibilities with their effects and side effects ('well informed patients'). "I can give you pain medication to reduce the pain and make you more mobile, however, one of the possible side effects is memory loss.

Another option is occupational therapy to enable you to participate in activities of everyday life, however, not every health insurer covers these costs". After having discussed various options with their benefits and harms Jeremy and the doctor make a joint decision about the treatment ('action planning'). Finally, after a set time, Jeremy fills in the survey again. During the next consultation the doctor and Jeremy discuss the results and adjust the treatment if necessary ('follow-up'). Additionally, TOPICS-CEP scores before and after the treatment can be used for research purposes to monitor the average value of the intervention according to the older Dutch population.

Advice for patients
To successfully implement person-centred healthcare systems certain barriers need to be brought down. Creating person-centred healthcare systems requires a change in roles, behaviours and mind sets from healthcare professionals and patients. Historically, in many cultures the doctor-patient relationship was one of paternalism. In this relationship the doctor genuinely wants the best for the patients, but believes that the patient often needs to be firmly guided through the decision making process as they do not always know what is best for them. Nowadays "doctor knows best" is no longer reliable for predicting the best outcome for patients, thus it can no longer serve for planning the future. However, some people still cannot let go of the idea that "doctor knows best". Perhaps this is driven by patients' lack of knowledge, skills, and confidence to effectively manage their own health and care. Consequently, person-centred care should support individuals to self-manage, offer personalised care planning and enable people to share in decision making. Thus, the limited time patients spend with their healthcare professional should be used to coach and enable patients to become knowledgeable, confident managers of their own health and support them to make informed decisions and successfully manage their own health and care.

Moreover, patients should be supported to stop acting as passive recipients and start acting as active participants or partners. Patients can be trained, guided, and encouraged to tell their healthcare professional what their wishes, desires, values and goals are when it comes to their health, wellbeing and QOL. Additionally, they can be coached to ask questions and offer alternative suggestions when it comes to treatments. As such professionals can help transforming their collaboration with patients by mutual communication and coaching on shared goal setting, shared decision making, and shared outcome measurements (i.e. TOPICS-CEP). By shifting in this direction patients may become better informed; knowing more alternatives and gain improved understanding of their proposed treatments’ advantages and disadvantages.

Advice for healthcare professionals
The beliefs, attitudes, and behaviours ingrained in healthcare professionals could interfere with their ability to provide effective, person-centred care. Healthcare professionals may think they already use a truly person-centred approach, because there is a lack of a clear shared understanding about what effective person-centred care looks like. Moreover, some healthcare professionals may not realise the value of involving patients in co-creating health.

By considering patients as partner rather than passive recipients, healthcare professionals gain insight in the motivations and challenges that each individual patient faces in selecting an intervention or adopting healthy behaviours; not just the biomedical challenges, but the psychological and social challenges as well. This is in line with the growing interest in shared decision making. Additionally, this fundamental shift in healthcare delivery emphasise the transition from focusing on improving clinical outcomes to QOL, from providing specialist treatment to generic support, and from a system that reacts to people’s ill health to one that focuses proactively on enabling people to live as healthy as possible. Therefore, we recommend to involve patients when providing training courses for healthcare professionals. In person-centred care medical knowledge (knowing which interventions are optional) and patients' experiences (understanding what these interventions mean in real life) are equally important. By giving patients the role of co-trainers they can bring in their own personal experiences, which may lead to eye openers.

Advice for managers
Healthcare professionals could fear that a person-centred approach would lengthen the time of consultations and they don't get enough time from their managers which is a legitimate concern. However, Robert Kaplan and Derek Haas state in their article "How not to cut healthcare costs" that one of five counterproductive mistakes that
8.5 Final remarks

Complex healthcare needs are common among older persons, who more and more emphasise the importance of individually tailored support. Consequently, healthcare professionals should always strive to see the person behind the patient and offer person-centred care meaning that clinical decisions are guided by patient’s wishes and values. Moreover, in health services research outcome measures need to take patient preferences into account. Following the work presented in this thesis, TOPICS-CEP can serve as a reliable and valid preference-weighted outcome measure, which is based on the health state preferences of older persons. This is necessary, as we provided evidence that the health state preferences of older persons and healthcare professionals are not the same. I hope this thesis presents sufficient valid data to warrant further spread of TOPICS-CEP as a relevant quality outcome indicator for healthcare in older persons. Put in more general terms, and finalising this thesis, I would like to stress the importance of good communication between clinicians, patients, and informal caregivers with regard to clinical decision making and strong patient involvement when setting the agenda and developing quality indicators for health services research.

healthcare providers often make is trying to reduce costs by pressuring healthcare professionals to maximize the number of patients they see and minimize the time they spend with each patient.\textsuperscript{23} Many examples can be found in the literature with regard to how doctors treating chronic conditions could offer better advice and achieve better treatment compliance if they had more time to spend with their patients.\textsuperscript{23} The costs of this extra time would be compensated through fewer future complications, e.g. in emergency rooms and intensive care units. Thus, managers need to be aware that spending extra time with the patient is a good investment to insure delivering higher value care; better outcomes for lower costs.

Advice for researchers

A more active role for patients is not only relevant for co-creating health, but also for setting priorities for health service research. The James Lind Alliance is a non-profit making initiative which brings together patients, informal caregivers, and healthcare professionals in Priority setting Partnerships to identify and prioritise the effects of treatments that they agree upon are most relevant. The aim of this alliance is to help ensure that those who fund health service research are aware of what matters to both patients and clinicians. As stated in The James Lind Alliance guidebook, "Focusing research where it will be most useful requires finding out what patients and clinicians want to know from research, and which of their requests for research are most important and urgent. Patients and clinicians have ideas about which new technologies they would like to be fully tested, which current treatments warrant further testing, and which criteria they would use to judge success or failure".\textsuperscript{27}

As TOPICS-CEP is a preference-weighted outcome to evaluate HR-QOL, it could be used as a criteria to judge success or failure of a treatment.

Furthermore, research by INVOLVE (Public INVOLVEment in NHS, public health and social care research) found that involving patients in clinical research is of great value as it ensures acceptability of trails, and participants feel them to be ethical, well-designed and have relevant outcome measures.\textsuperscript{28} It also found that public involvement was reported to help increase recruitment to all types of research. Hence, it is of great importance to involve patients and healthcare professionals when setting the agenda for health service research.
8.6 References


Chapter 9

Nederlandse samenvatting | Dutch Summary
Nederlandse samenvatting

In 2008 is het Nationaal Programma Ouderenzorg (NPO) in opdracht van het Ministerie van Volksgezondheid, Welzijn en Sport gestart. Het NPO verbetert de zorg voor ouderen met complexe hulpvragen. Talrijke organisaties slaan daarvoor landelijk en regionaal de handen ineen. Het doel is een samenhangend zorgaanbod dat beter is afgestemd op de individuele behoeften van ouderen. Voor ouderen leidt deze kwaliteitsslag tot meer zelfredzaamheid, meer functiebehoud, minder terugvallen op de zorg en minder kans op zorg en behandelingen die onnodig belastend zijn.

Binnen het NPO zijn meer dan 125 vernieuwende projecten uitgevoerd. Ook is ‘The Older Persons and Informal Caregivers Survey – Minimum DataSet’ (TOPICS-MDS) ontwikkeld. TOPICS-MDS bestaat uit vragenlijsten voor de oudere zorgvrager en mantelzorger en een database waarin verzamelde gegevens zijn samengevoegd. De TOPICS-MDS vragenlijsten zijn in alle projecten die door het NPO gefinancierd worden ingezet om effecten van de studies te kunnen meten. De vragenlijsten zijn aan de start van het NPO samengesteld door experts uit de regionale netwerken ouderenzorg. Met hen zijn afspraken gemaakt over de te gebruiken uitkomstmaten in de projecten. Dit resulteerde in vragenlijsten die bestaan uit gangbare gevalideerde instrumenten die lichamelijke, psychische en sociale gezondheid en welzijn meten bij ouderen en mantelzorgers.

De data die vervolgens zijn verzameld binnen alle NPO projecten, zijn centraal bijeengebracht in de TOPICS-MDS database. Daarmee is een unieke bron met waardevolle gegevens over gezondheid en welzijn van een groot aantal kwetsbare ouderen en mantelzorgers verspreid over Nederland ontstaan. Deze database bevat momenteel gegevens van meer dan 43.000 ouderen en 9.000 mantelzorgers van circa 60 verschillende projecten. Baseline gegevens van de projecten zijn op dit moment verzameld in de database en ook is van een aantal projecten follow-up data aanwezig in de database.

Dit proefschrift beschrijft de ontwikkeling en validering van een samengestelde uitkomstmaat (in het Engels: Composite Endpoint (CEP)) voor de TOPICS-MDS vragenlijst voor zorgontvangers. Deze uitkomstmaat TOPICS-CEP genoemd, kan dienen als kwaliteitsindicator in de ouderenzorg. Hieronder worden de belangrijkste bevindingen kort samengeteld.

Deel 1 geeft een algemene inleiding. De wereldpopulatie is aan het vergrijzen. Hierdoor komt er grote druk te staan op de zorg. Vanwege beperkte middelen, zowel financieel als personeel, is het belangrijk om de meerwaarde (ofwel het effect) van interventies voor de vast te kunnen stellen. Het is echter niet eenvoudig om in de
ouderenzorg te bepalen welke interventies de grootste meerwaarde hebben. Dit komt doordat de gezondheidstoestanden van ouderen vaak complex zijn. Daarnaast hebben interventies vaak invloed op meerdere levensdomeinen tegelijkertijd. Een samengestelde uitkomstmaat die rekening houdt met de onderlinge relevante van de uitkomstdomeinen (concepten waaruit de samengestelde uitkomstmaat bestaat) kan een mogelijke oplossing zijn voor dit probleem. Het belang van kwaliteitsindicatoren in het algemeen en van samengestelde uitkomstmaten in het bijzonder wordt beschreven in hoofdstuk 1.

Deel 2 beschrijft de ontwikkeling van TOPICS-CEP welke werd gebaseerd op de voorkeuren van ouderen en mantelzorgers. In toenemende mate verwacht de overheid een grotere verantwoordelijkheid van de burgers als het gaat om zorg en ondersteuning voor elkaar. Door de vergrijzing en het tekort aan personeel in de zorg zien we steeds vaker dat familieleden de rol van mantelzorger krijgen toegewezen. In sommige gevallen, bijvoorbeeld wanneer een ouder lijdt aan cognitieve stoornissen, treden naasten op als proxies, d.w.z. dat meningen, opvattingen en ervaringen van personen uit de omgeving van de patiënt worden gebruikt als vervanging van die van de patiënt. Dit was voor ons een belangrijke reden om te onderzoeken of ouderen en mantelzorgers de verschillende gezondheidsdomeinen: chronische aandoeningen, functionele beperkingen, emotioneel welbevinden, pijnpervaring, cognitieve problemen, sociaal functioneren, zelf ervaren gezondheid en zelf ervaren kwaliteiten van leven (acht componenten verkregen uit de TOPICS-MDS vragenlijst) op dezelfde manier beoordelen. Dit was voor ons een belangrijke reden om te onderzoeken of de mantelzorger zijn/haar algemeen welbevinden geen lager rapportcijfer geven vanwege de rol als mantelzorger. Dit was voor ons een belangrijke reden om te onderzoeken of de mantelzorger zijn/haar algemeen welbevinden geen lager rapportcijfer geven vanwege de rol als mantelzorger. Daarnaast hebben we, gebaseerd op de voorkeuren van ouderen en mantelzorgers de verschillende gezondheidsdomeinen: chronische aandoeningen, functionele beperkingen, emotioneel welbevinden, pijnpervaring, cognitieve problemen, sociaal functioneren, zelf ervaren gezondheid en zelf ervaren kwaliteiten van leven (acht componenten verkregen uit de TOPICS-MDS vragenlijst) op dezelfde manier beoordelen, door gebruik te maken van vignetten, wat korte omschrijvingen zijn van personen, hebben wij de gewichten voor de verschillende componenten berekend en vergeleken. De gewichten, ook wel preferentiegewichten genoemd, reflecteren hoe belangrijk de ouderen en mantelzorgers de verschillende domeinen vinden. De resultaten van deze vignettenstudie worden in hoofdstuk 2 beschreven. In totaal werkten 124 ouderen en 76 mantelzorgers mee aan deze studie. Wanneer we de gewichten van deze twee groepen vergeleken, bleek dat er geen verschillen waren. Om deze reden hebben wij TOPICS-CEP ontwikkeld, een gewogen samengestelde uitkomstmaat, gebaseerd op de voorkeuren van zowel ouderen als mantelzorgers.

Verschillende studies hebben laten zien dat eigenschappen van personen invloed hebben op hoe zij oordelen vellen over bijvoorbeeld kwaliteit van leven en ervaren gezondheid. Daarom hebben wij, op basis van gegevens die wij in de eerste vignettenstudie hebben verzameld, onderzocht of en op welke manier de eigenschappen van onze deelnemers invloed hebben gehad op de preferentiegewichten die zijn opgenomen in TOPICS-CEP. De resultaten van deze studie worden in Hoofdstuk 3 beschreven. We hebben in deze studie onderzocht of het geslacht, de leeftijd en zelf ervaren gezondheid van onze deelnemers van invloed zijn geweest op de preferentiegewichten van de acht componenten waaruit TOPICS-CEP bestaat. Hierbij hebben we gecontroleerd voor de invloed van het al dan niet mantelzorger zijn. De rol als mantelzorger blijkt echter geen invloed te hebben gehad op onze bevindingen. Onze resultaten laten verder zien dat de leeftijd en ervaren gezondheid van onze deelnemers van invloed zijn geweest op de preferentiegewichten van TOPICS-CEP. Echter vervult de invloed van ervaren gezondheid wanneer er wordt gecorrigeerd voor leeftijd. Hieruit kunnen we opmaken dat de leeftijdssamenstelling van onze deelnemers invloed heeft gehad op de preferentiegewichten van de TOPICS-CEP. Deze bevinding heeft ertoe geleid dat we in de richtlijnen voor het gebruik van TOPICS-CEP (www.topics-mds.eu) duidelijk hebben aangeven op welke steekproef de preferentiegewichten gebaseerd zijn. Daarnaast hebben we, gebaseerd op de leeftijdsverdeling van ouderen (65-plussers) in Nederland, de gewichten aangepast zodat deze de preferenties van de Nederlandse populatie van 65 jaar en ouder reflecteert.

Deel 3 beschrijft de relevante van TOPICS-CEP. Het vaststellen van lange termijn veranderingen in het algemeen welbevinden of kwaliteit van leven is een belangrijk onderdeel van klinisch onderzoek en evaluaties. Vaak wordt er aan patiënten of cliënten zelf gevraagd wat zij van hun algemeen welbevinden of kwaliteit van leven vinden, bijvoorbeeld door er een rapportcijfer aan te geven. Hoewel dit een relevante vraag is en nuttige informatie kan opleveren, is de beoordeling niet alleen afhankelijk van het effect van de interventie maar ook van bijvoorbeeld “adaptatie”. Adaptatie is een natuurlijk proces waarbij men leert omgaan met tegenslagen. Door een persoon bijvoorbeeld beperkingen in het dagelijks leven heeft geaccepteerd, zal hij/zij zijn/haar algemeen welbevinden geen lager rapportcijfer geven vanwege functionele beperkingen. Daarnaast zijn mensen geneigd om hun eigen leven te vergelijken met anderen die het nog slechter hebben, dit wordt “framing” genoemd. Door gebruik te maken van een samengestelde uitkomstmaat wordt het mogelijk om een subjectief concept zoals algemeen welbevinden op een objectieve manier te benaderen maar hierbij nog wel rekening te houden met wat ouderen over het algemeen relevant vinden.

Hoofdstuk 5 beschrijft de studie waarbij wij hebben onderzocht in welke mate de beoordeling door externe ouderen (participanten uit onze eerste vignettenstudie) overeenkomen met de rapportcijfers die de casus omschreven in de vignet zichzelf had gegeven. Normaliter is de Bland-Altman methode de meest gekozen methode om de mate van overeenstemming (“level of agreement”) tussen twee beoordelaars te onderzoeken. Deze methode krijgt de voorkeur boven een scatterplot of correlatietest, omdat deze laatste twee geen rekening houden met systematische
verschillen. De Bland-Altman methode is echter in zijn huidige vorm niet te gebruiken om de overeenstemming tussen onze individuele interne beoordelaars (de casus) en onze groep externe beoordelaars te bestuderen. **Hoofdstuk 4** beschrijft op welke manier de Bland-Altman methode moet worden aangepast, zodat deze wel kan worden gebruikt om de mate van overeenstemming tussen twee groepen van ongrijpe grootte te bepalen; dus ook wanneer de ene groep bestaat uit één beoordelaar, terwijl de ander groep bestaat uit 2 of meer beoordelaars. Gebruik makend van de aangepaste Bland-Altman methode laat de plot in **Hoofdstuk 5** zien dat er systematische verschillen bestaan tussen de beoordelingen van externe beoordelaars en de zelf gegeven rapportcijfers (interne beoordeling door de casus). Deze consistente bias en de richting daarvan komt overeen met wat wij hadden verwacht hadden vanwege o.a. adaptatie en framing. Concreet betekent dit dat interne en externe beoordelingen niet identiek zijn en daardoor niet verwisselbaar zijn. Zij zijn echter wel complementair en we adviseren dan ook om beiden te gebruiken. Het zelf oordeel geeft inzicht in hoe een patiënt of cliënt diens eigen kwaliteit van leven beoordeelt op een bepaald moment, terwijl de externe beoordelingen inzicht geven in hoe de populatie waartoe de patiënt of cliënt behoort de kwaliteit van leven over het algemeen zou beoordelen.

Omdat zorgverleners vaak zelf een inschatting maken of een bepaalde interventie nut heeft gehad wilden wij onderzoeken of zij gezondheidsstoestanden op dezelfde manier evalueren als ouderen en mantelzorgers. Met andere woorden: we hebben gekeken of de gewichten van de TOPICS-CEP componenten gebaseerd op de preferenties van ouderen en mantelzorgers overeenkwamen met de gewichten gebaseerd op de preferenties van zorgverleners. Om dit te onderzoeken hebben we een tweede vignettenstudie uitgevoerd waarbij 330 zorgverleners het algeheel welbevinden van ouderen omschreven in vignetten een rapportcijfer hebben gegeven. De resultaten van deze vignettenstudie worden in **Hoofdstuk 6** beschreven. In totaal hebben 127 artsen, 102 verpleegkundigen, 45 welzijnsmedewerkers en 56 paramedici deelgenomen aan deze studie. Het aantal jaar werkervaring is voor verschillende ouderen populaties.

Deze consistente bias en de richting daarvan komt overeen met wat wij hadden verwacht vanwege o.a. adaptatie en framing. Concreet betekent dit dat interne en externe beoordelingen niet identiek zijn en daardoor niet verwisselbaar zijn. Zij zijn echter wel complementair en we adviseren dan ook om beiden te gebruiken. Het zelf oordeel geeft inzicht in hoe een patiënt of cliënt diens eigen kwaliteit van leven beoordeelt op een bepaald moment, terwijl de externe beoordelingen inzicht geven in hoe de populatie waartoe de patiënt of cliënt behoort de kwaliteit van leven over het algemeen zou beoordelen.

Omdat zorgverleners vaak zelf een inschatting maken of een bepaalde interventie nut heeft gehad wilden wij onderzoeken of zij gezondheidsstoestanden op dezelfde manier evalueren als ouderen en mantelzorgers. Met andere woorden: we hebben gekeken of de gewichten van de TOPICS-CEP componenten gebaseerd op de preferenties van ouderen en mantelzorgers overeenkwamen met de gewichten gebaseerd op de preferenties van zorgverleners. Om dit te onderzoeken hebben we een tweede vignettenstudie uitgevoerd waarbij 330 zorgverleners het algeheel welbevinden van ouderen omschreven in vignetten een rapportcijfer hebben gegeven. De resultaten van deze vignettenstudie worden in **Hoofdstuk 6** beschreven. In totaal hebben 127 artsen, 102 verpleegkundigen, 45 welzijnsmedewerkers en 56 paramedici deelgenomen aan deze studie. Het aantal jaar werkervaring is voor verschillende ouderen populaties.

Laatste opmerkingen
Complex zorgvragen komen vaak voor bij ouderen. Idealiter vraagt dit om individueel afgestemde ondersteuning. Hiervoor is het van belang dat zorgverleners de persoon achter de patiënt zien, zodat zij persoonsgerichte zorg kunnen bieden. Dit betekent dat zij de wensen en behoeften van de patiënt leidend laten zijn bij het kiezen van een interventie. Daarnaast dient in onderzoek naar de meerwaarde van aangeboden zorg uitkomsten te worden gebruikt die rekening houden met de wensen en behoeften (voorkeuren) van patiënten.
Naar aanleiding van de studies omschreven in dit proefschrift stellen we dat TOPICS-CEP kan dienen als een betrouwbaar en valide uitkomstmaat die rekening houdt met de voorkeuren van ouderen. Dat dit nodig is blijkt uit onze resultaten die laten zien dat wanneer de beoordelingen van gezondheidstoestanden worden vergeleken tussen ouderen en zorgverleners de twee groepen andere voorkeuren hebben. Hierover is een infographic gepubliceerd in de telegraaf van 6 juni 2015.

Tenslotte eindigen we dit proefschrift met het benadrukken van het belang van goede communicatie tussen clinici, patiënten en mantelzorgers met betrekking tot het nemen van klinische beslissingen. Wij adviseren zorgverleners en onderzoekers om patiënten en/of cliënten te betrekken bij het vaststellen van de agenda (waar moet over worden nagedacht) en bij de ontwikkeling van kwaliteitsindicatoren in de zorg. Het is belangrijk om de persoon achter de patiënt te zien en te blijven zien!
Part 6
Supplementary

Extras, Additional, Supplementary, More, Plus...
Dankwoord | Acknowledgements
Dankwoord (Acknowledgements)

"Het leed wat promoveren heet", maar dankzij verschillende mensen heb ik de eindstreep nu dan toch gehaald.

Allereerst wil ik alle ouderen, mantelzorgers en zorgverleners bedanken die hebben deelgenomen aan dit onderzoek. Zonder jullie was dit proefschrift er niet geweest en had er geen TOPICS-CEP bestaan. Hartelijk dank voor jullie inzet.

Natuurlijk wil ik mijn promotor, voormalig promotor en copromotoren bedanken. Marcel, jouw feedback op mijn stukken was altijd beknopt en helder, dus dat ga ik nu ook zijn. Bedankt voor jouw snelle reacties op mijn mails (zelfs tot laat in de avond) en voor het open staan voor mijn soms wat "vreemde" (animal assisted therapy) ideeën. George, helaas heb ik de eindstreep zonder jou moeten bereiken maar wat een mooie sprint hebben wij aan het begin gemaakt. Het was erg fijn om met je samen te werken en ik vond het dan ook verschrikkelijk toen ik hoorde dat je ermee ging stoppen. Rogier, jouw taak was om George op te volgen binnen ons team en dat heb je echt super gedaan. We hebben waardevolle overleggen gepleegd en mede dankzij jouw heldere uitleg over de statistiek heb ik zelf alle analyses kunnen uitvoeren, mijn dank daarvoor. René, last but not least. Waar moet ik beginnen? Bedankt voor het feit dat jij (en George) mij de kans hebt gegeven om zelfstandig onderzoek te doen; bedankt voor de feedback op al mijn stukken; bedankt voor jouw gezelschap in de trein naar Groningen, Leiden, Amsterdam en Utrecht; en bedankt dat je het met mij hebt uitgehouden.

Daarnaast gaat mijn dank uit naar de leden van de manuscriptcommissie: Prof. dr. Gert Westert, Prof. dr. Piet van Riel en Prof. dr. Anna Nieboer. Hartelijk dank voor jullie bereidheid om in de commissie plaats te nemen en dit proefschrift te beoordelen.

Verder wil ik ook het OPROCS-project team bedanken. Bianca Buurman, Han Boter en Ton de Craen fijn dat jullie onderdeel wilden zijn van ons team. Bedankt voor jullie inzet en commentaar op de papers. Peter, voor jou geldt dat je het team pas in een later stadium kwam versterken. Je hebt mij door de soms wat moeilijke dagen gesleept en geholpen wanneer dit nodig was. Je weet niet half hoe waardevol je bent geweest. Thanks!

De data had ik niet kunnen verzamelen zonder de hulp van Lieke Rijnen, Mieke Luten, Frank Schalkwijk, Marga de Fretes, Loes Kramer, Maureen van den Boom-Kalker en Michelle Langeveld-Kolman. Jullie allen heel erg bedankt.
En dan hebben we natuurlijk ook nog Wim Lemmens! Wim, bedankt voor je technische ondersteuning bij het ontwikkelen van de vignetten, maar in het bijzonder bedankt ik dat ik zomaar 10 minuten bij je kon binnenvallen voor de gezellige familie praatjes.

Het secretariaat wil ik bedanken. Lieve Gemma, Nora, Maja en Hanna, fijn dat jullie ruimtes voor de afdelingslunches wilden reserveren (ook al konden jullie zelf niet altijd aanschuiven). En bedankt voor het uitzoeken van allerlei dingetjes die weer verkeerd gingen of onduidelijk waren, met name binnen MijnTijd en MijnOntwikkeling en tijdens de implementatie van Werkplek 2.0. Ook na mijn vertrek op de afdeling kon ik altijd met vragen bij jullie terecht. Super!

Collega-onderzoekers bedankt voor de gezellige pauzes. In de afgelopen vier en een half jaar hebben heel veel mensen de afdeling verlaten en ook weer bezet, daarom ga ik er niet aan beginnen om jullie allemaal bij naam te noemen. Maar een aantal van jullie hebben toch een belangrijke rol gespeeld in de afgelopen jaren. Allereerst mijn kamergenoten waarmee ik goede en minder goede tijden mee heb mogen delen: Sarah, Petra, Marieke, Joep, Jaap, Gerrita en Kim. Bedankt dat jullie er waren. Van Kempen, jou wil ik nog even apart noemen. Wat had ik een geluk met jou als kamergenoot voor drie lange, maar geweldige jaren. En wat fijn dat René en Gerda ervoor hebben gezorgd dat wij (met onze autistiche trekken) ook op de derde verdieping een kamer mochten blijven delen. Bedankt voor het aanhoren van mijn gezeur (van tijd tot tijd) en voor je nuchtere kijk op dingen!

Daarnaast wil ik ook in het bijzonder twee TOPICS-MDS mensen bedanken: Lia en Jenn. Girls thanks for your positive outlook on life and your enthusiasm when it comes to research. I used to be a bit uncertain about my own competence, but you were there to make me believe and achieve. I have always been aware of my own weaknesses and now I am also aware of my own strengths! Thank you both and cool beans :)!

Wat is een mens zonder vrienden? Daarom wil ik ook hen bedanken voor de steun en afleiding wanneer dit nodig was. Lieve Liza, Nieky en Linda, we zien elkaar veel minder dan jaren terug, maar ik blijf ons contact bijzonder en waardevol vinden. Ongeacht hoe lang we elkaar niet hebben gesproken, wanneer we elkaar weer zien is het direct als vanouds. Bedankt dat jullie zijn wie jullie zijn! Don't ever change!

En nu komt het zware werk, nu mag ik mijn familie bedanken. Lieve Pama (mijn ouders), waar moet ik beginnen? Wat zijn jullie heerlijke mensen en wat hebben jullie mij goed opgevoed 😊 Hoewel ik eerst dacht dat vriendelijkheid en respect voor anderen (voor ouderen) normaal was, ben ik er achter genomen dat het iets is wat ik van huis uit heb meegekregen. Het is niet iets wat je kunt “faken”. Doen alsof is niet voldoende, je moet daadwerkelijk de respect hebben voor anderen om ze het gevoel te kunnen geven dat ze gerespecteerd en gehoord worden; dus bedankt voor jullie goede opvoeding. En wat heb ik een mazzel gehad dat ik altijd heb mogen doen wat ik zelf wilde, ook al kostte het meer tijd dan gepland. Ik weet dat jullie trots op me zijn, ook al weten jullie nooit precies wat ik aan het doen ben of heb gedaan. Ik houd van jullie en bedankt dat jullie nog altijd en voor altijd achter mij staan.

Lieve Ingrid en Jan, wat fijn dat toen ik Guido leerde kennen ik er gelijk een lieve familie bij kreeg. De vakanties die mij weerhielden van gek worden en wellicht de handdoek in de ring te gooien waren niet mogelijk zonder jullie. Dankzij jullie hoefde ik geen zorgen te maken om Bilbo; ik wist en weet nog steeds dat hij bij jullie in goede handen is. Bedankt voor jullie goede zorgen. En in jullie oprechte interesse in wat ik doe.

Wen en Il, zelfs op afstand hebben jullie een goede invloed op mij. Gewoon door mail contact, vakanties in Oostenrijk en bezoekjes aan jullie in Zwitserland geeft mij energie. Heerlijk! Bedankt! En tja, de jongens… Joshua en Ayden, jullie hebben van mij een trotste en vrolijke Tante Sin gemaakt. Ongeacht mijn humeur een foto of filmpjes van jullie via de WhatsApp tovert altijd weer een glimlach op mijn gezicht. Gracias!

Tot slot mijn eigen Meneer (Guido) onbegrijpelijk hoe jij het met me hebt uitgehouden. Mijn humeurige avonden, druk bezette agenda en luie weekenden, moeten niet altijd even makkelijk voor jou zijn geweest. Bedankt voor jouw geduld, voor jouw goede zorgen voor mij, Bilbo en tot vorig jaar ook voor Pluto én voor de heerlijke zomervakanties. Zonder jou had ik nooit de wereld ontdekt en had ik deze eindstreep nooit gehaald!

Veel liefs,
Sin
Curriculum Vitae
Curriculum Vitae

Cynthia Hofman was born in Bergen op Zoom (the Netherlands) on June 9th 1981. In 1998 she graduated from higher general secondary education and in 2000 from pre-university education, Mollerlyceum in Bergen op Zoom. She subsequently received a Bachelor’s degree in Psychology and a Master’s degree in Neuro- en Rehabilitation psychology at the Radboud University Nijmegen. In 2008 she worked at the Psychiatry department of the Radboud university medical center for her research internship, where PhD candidate Martine Hoogman introduced her with conducting scientific research. Cynthia wrote her Master thesis “Behaviour regulation in adults with ADHD: Attention, mental flexibility, and response inhibition” under the supervision of Paul Eling. After her internship and while writing her thesis she worked as a student-assistant at ParkinsonNet, a National project that was launched by the Neurology department of Radboud university medical center. After her graduation in 2010 she was offered a PhD position at the department of Geriatric medicine and the department for Health Evidence at the Radboud university medical center. She worked under the supervision of Marcel Olde Rikkert, René Melis, and Rogier Donders. Finally, her research resulted in this thesis. During her PhD project, Cynthia was a board member of the National Association “De Zonnebloem”. In addition, she was temporarily the webmaster of “De Zonnebloem” website (region De Liemers) and chair of the workgroup “New media”. Since July 2014 she works at Vilans (centre of expertise on long-term care) as a consultant and researcher. Her primary focus is on Frail elderly and dementia, Patient centred care, and Shared Decision Making.
Curriculum Vitae in Dutch

Publications

Journal articles (this thesis)


Hofman CS, Lutomski JE, Olde Rikkert MGM, Melis RJF, Donders ART. Letter to the editor: Examining consistency between self-assessed and peer ratings in older people. Journal of the American Geriatrics Society. [In press]


Hofman CS, Lutomski JE, Boter H, Buurman BM, de Craen AJM, Donders ART, Olde Rikkert MGM, Makai P, Melis RJF. Examining the Construct and Known-group Validity of a Composite Endpoint for The Older Persons and Informal Caregivers Survey Minimum Data Set (TOPICS-MDS); A Large-Scale Data Sharing Initiative. PlosOne. [In press]

Other publications

Book chapters


Ouderen bepalen zelf de meerwaarde van hun zorg

Beschrijving
Hofman deed onderzoek naar de zorgvoorkeur van ouderen en mantelzorgers. In het onderzoek werd een samengestelde uitkomstmaat ontwikkeld, die rekening houdt met wat ouderen en mantelzorgers belangrijk vinden voor het algeheel welbevinden. De relevantie van zo’n uitkomstmaat blijkt onder andere uit het verschillende belang dat ouderen, hun mantelzorgers en de zorgverleners (artsen, verpleegkundigen en paramedici) toekennen aan relevante uitkomstdomeinen voor de zorg.

Hofman stelt dan ook, dat ouderen en zorgverleners verschillende opvattingen hebben over welke levensdomeinen het belangrijkst (meest relevant) zijn. Zo hechten ouderen meer waarde aan zelfstandig functioneren dan zorgverleners, terwijl zorgverleners pijnreductie belangrijker vinden dan de ouderen zelf.

Hofman adviseert om ouderen zelf de meerwaarde van de ouderenzorg te laten bepalen. Daarnaast pleit ze voor gesprekken tussen ouderen, mantelzorgers en zorgverleners voor het stellen van doelen en het samen beslissen over het te volgen zorgpad.

Biografie