Going beyond health-related quality of life for outcome measurement in economic evaluation

by

Lidia Engel

M.Sc., Maastricht University, 2013

Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Doctor of Philosophy Program
Faculty of Health Sciences

SIMON FRASER UNIVERSITY

Summer 2017

Approval

Name:	**Lidia Engel**
Degree:	**Doctor of Philosophy**
Title:	***Going beyond health-related quality of life for outcome measurement in economic evaluation***

Examining Committee: **Chair:** Frank Lee
Associate Professor

Dr. David GT Whitehurst
Senior Supervisor
Assistant Professor

Dr. Scott A Lear
Co-supervisor
Professor

Dr. Stirling Bryan
Supervisor
Professor, School of Population and Public Health, University of British Columbia

Dr. Bohdan Nosyk
Internal Examiner
Associate Professor

Dr. Paula Lorgelly
External Examiner
Visiting Professor, King's College, London, United Kingdom

Date Defended/Approved: July 4, 2017

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Background: The quality-adjusted life year (QALY) framework has been challenged for use in economic evaluation because of its narrow focus on health-related aspects of quality of life, thus ignoring potential 'non-health' benefits associated with treatments and interventions. With the development of new preference-based measures, such as the ICEpop CAPability (ICECAP) instruments that adopt a broader evaluative space, the aim of this thesis was to examine methodological considerations and applied implications for outcome measurement in health economics when applying measures that extend beyond health.

Methods: A narrative review provides an overview of challenges involved for broadening the evaluative space of the QALY and the progress that has been made in this area. A critical interpretive synthesis (CIS) is then presented that conceptualized benefits beyond the *health-related* QALY, followed by three empirical analyses, each using a different dataset: (i) regression analyses testing the complementarity of a preference-based health-related quality of life (HRQoL) measure, the EQ-5D-5L, and a measure of capability wellbeing for older adults, the ICECAP-O, within the context of public health; (ii) exploratory factor analyses investigating the overlap between the ICECAP-A and five preference-based HRQoL measures; and (iii) path analyses to further explore the relationship between two HRQoL measures (EQ-5D-5L and AQoL-8D) and two wellbeing measures (ICECAP-A and subjective wellbeing).

Results: The CIS conceptualized non-health benefits into four themes: (i) benefits affecting a person's wellbeing (psychological wellbeing, subjective wellbeing, empowerment, and capability wellbeing); (ii) benefits derived from the process of health care delivery; (iii) benefits beyond the affected individual; and (iv) benefits beyond the health care sector. Three key findings were made from the empirical analyses that further explored wellbeing measures. Firstly, the ICECAP-O is more sensitive to environmental features (i.e., social cohesion and street connectivity) when compared with the EQ-5D-5L; secondly, the ICECAP-A contains domains in its descriptive system that are not measured by most HRQoL measures, except for the AQoL-8D; and thirdly, HRQoL and wellbeing measures are affected in a different way by different secondary health conditions but a similar relationship was found between the ICECAP-A and AQoL-8D.

Conclusion: The thesis concludes that the application of wellbeing measures in economic evaluations requires careful consideration due to the risk of double counting. The capability approach has the potential to extend the QALY but the operationalization of this approach – and other non-health benefits within or outside the QALY framework – requires further research.

Keywords: Quality-adjusted life year; non-health benefits; economic evaluation; outcome measurement; capability wellbeing; health-related quality of life.

To my husband

Acknowledgements

I would like to express my sincere gratitude to my senior supervisor, Dr. David Whitehurst, for the continuous support of my PhD studies, for his patience, motivation, and invaluable knowledge. His advice regarding both my research and my career has allowed me to grow as a researcher. I could not have imagined having a better advisor and mentor for my PhD studies.

I would also like to thank my committee members, Professor Stirling Bryan and Professor Scott Lear, for their insightful comments and for their critiques, which encouraged me to widen my research from various perspectives. I am thankful for the support received from staff members at the Centre for Clinical Epidemiology and Evaluation (C2E2) and am particularly grateful to Dr. Penny Brasher for her statistical advice. I would also like to thank Drs. Anna Chudyk, Maureen Ashe, and Heather McKay for their support for the work reported in Chapter 4.

My sincere thanks, furthermore, goes to Dr. Duncan Mortimer for enabling me to conduct part of my PhD research at the Centre for Health Economics at Monash University. His guidance, encouragement and insights were invaluable. I extend my thanks to other members of the Centre, especially to Professor Jeff Richardson for pushing my thinking further in our conversations, which helped to shape my PhD research. I am also thankful to Deakin Health Economics for their support during the final stages of this journey.

Finally, I would also like to take this opportunity to express my appreciation to my family for offering to help in any way they could despite the distance. A million thanks also to my friends from all over the world, who helped me maintain an optimistic outlook over the last few years as I strived towards my goal. Most of all, I want to thank my husband Dimitri for his love, support and belief in me – words cannot express how grateful I am for all the sacrifices he has made for me.

Table of Contents

List of Tables

List of Figures

List of Abbreviations / Naming Convention Descriptions

15D	15-dimension health-related quality of life questionnaire
AQoL-4D	Assessment of Quality of Life 4-dimension questionnaire
AQoL-8D	Assessment of Quality of Life 8-dimension questionnaire
ASCOT	Adult Social Care Outcomes Toolkit
CADTH	Canadian Agency for Drugs and Technologies in Health
CBA	Cost-benefit analysis
CCA	Cost-consequences analysis
CEA	Cost-effectiveness analysis
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CI	Confidence interval
CIS	Critical interpretive synthesis
CMA	Cost-minimization analysis
CUA	Cost-utility analysis
CV	Contingent valuation
CVZ	College van Zorgverzekeringen
DCE	Discrete choice experiment
EFA	Exploratory factor analysis
EQ-5D-3L	3-level questionnaire developed by the EuroQol Group
EQ-5D-5L	5-level questionnaire developed by the EuroQol Group
EUT	Expected utility theory
HRQoL	Health-related quality of life
HTA	Health technology assessment
HUI	Health Utilities Index
HUI-2	Health Utilities Index Mark 2
HUI-3	Health Utilities Index Mark 3
ICECAP	ICEpop CAPability
ICECAP-A	ICEpop CAPability measure for Adults
ICECAP-CPM	ICEpop CAPability Close Person Measure
ICECAP-O	ICEpop CAPability measure for Older adults
ICECAP-SCM	ICEpop CAPability Supportive Care Measure

MAUI	Multi-attribute utility instrument
MAUT	Multi-attribute utility theory
MCDA	Multi-criteria decision analysis
MIC	Multi Instrument Comparison
NEWS-A	Neighbourhood Environment Walkability Scale – Abbreviated
NICE	National Institute for Health and Care Excellence
OCAP-18	18-item capability measure
OxCAP-MH	Oxford CAPabilities questionnaire-Mental Health
QALY	Quality-adjusted life year
QoL	Quality of life
QWB-SA	Quality of Wellbeing Self-Administered
RMSEA	Root mean square error of approximation
SAFER	Shelter Aid for Elderly Renters
SC-5PT	Sampson's 5-item measure of collective efficacy
SCI	Spinal cord injury
SCRQoL	Social care-related quality of life
SD	Standard deviation
SF-36	Medical Outcomes Study 36-item Short Form Health Survey
SF-6D	Classification system for describing health derived from SF-36 items
SG	Standard gamble
SHC	Secondary health condition
SSWS	Street Smart Walk Score
SWB	Subjective wellbeing
TLI	Tucker-Lewis index
TTO	Time trade-off
TUI	Temporary Utility Index
UK	United Kingdom
USA	United States of America
VAS	Visual analogue scale
WHO	World Health Organization
WLSMV	Weighted least square means and variance adjusted
WTP	Willingness to pay
WTT	Walk the Talk

Related publications and conference presentations

Publications in peer-reviewed journals:

1. Engel L, Mortimer D, Bryan S, Lear S, Whitehurst DGT (2017): An investigation of the overlap between the ICECAP-A and five preference-based health-related quality of life instruments. *Pharmacoeconomics*, 35(7): 741-753.
2. Engel L, Chudyk AM, Ashe MC, McKay HA, Whitehurst DGT, Bryan S (2016): Older adults' quality of life – exploring the role of the built environment and social cohesion in community-dwelling seniors on low income. *Social Science & Medicine*, 164: 1-11.

Related conference presentations and invited seminars:

1. How well does EQ-5D-5L, ICECAP-A and a subjective wellbeing rating scale capture the effects of secondary health conditions in individuals living with spinal cord injury? A path analysis approach. *34th EuroQol Plenary Meeting*, Barcelona (Spain), 20-23 September 2017.
2. To what extent does the EQ-5D-5L capture aspects of capability wellbeing? *33rd EuroQol Plenary Meeting*, Berlin (Germany), 14-16 September 2016.
3. Conceptualizing benefits beyond health within the quality-adjusted life year (QALY) framework: a critical interpretive synthesis. *Priorities*, Birmingham (UK), 7-9 September 2016.
4. Same, same-same or different? Exploring overlap between the ICECAP-A and six preference-based HRQoL instruments. *Priorities*, Birmingham (UK), 7-9 September 2016.
5. Secondary health conditions in spinal cord injury – investigating the impact on health, capability and subjective wellbeing (an economic perspective). *ICORD Seminar*, Vancouver (Canada), June 28th, 2016.

6. Same, same-same or different? To what extent do preference-based health-related quality of life instruments capture aspects of capability wellbeing? *Vancouver Health Economics Methodology (VAN-HEM) Meeting*, Vancouver (Canada), June 10th, 2016.
7. Conceptualizing benefits beyond health within the context of quality-adjusted life years (QALYs): a critical interpretive synthesis. *SFU Health Research Day*, Surrey (Canada), May 4th, 2016.
8. Going beyond *health-related* quality of life within the quality-adjusted life year (QALY) framework. *Centre for Health Economics Research and Evaluation, University of Technology Sydney*, Sydney (Australia), March 29th, 2016.
9. Exploring the impact of the built environment on older adults' quality of life. *ISOQOL 22nd Annual Conference,* Vancouver (Canada), 21-24 October 2015.
10. The association between the built and social environment and older adults' quality of life in Metro Vancouver – Results from the Walk The Talk study. *Centre for Clinical Epidemiology and Evaluation*, Vancouver (Canada), September 1st, 2015.
11. Challenges associated in going beyond health-related quality of life in cost-utility analysis. *iHEA 11th World Congress*, Milan (Italy), 12-15 July 2015.
12. Exploring the impact of the built environment on older adults' quality of life. *16th International Medical Geography Symposium*, Vancouver (Canada), 5-10 July 2015.
13. Exploring the impact of the built environment on older adults' quality of life. *Vancouver Health Economics Methodology (VAN-HEM) Meeting*, Vancouver (Canada), June 26th, 2015.
14. Challenges associated in going beyond health-related quality of life in cost-utility analysis. *CADTH Symposium 2015*, Saskatoon (Canada), April 14th, 2015.

Chapter 1. Thesis overview

1.1. Introduction to the thesis

This thesis contributes to an emerging field of research concerned with the evaluative space of the quality-adjusted life year (QALY). QALYs have been used in the assessment of health interventions primarily to adjust someone's life expectancy based on levels of health-related quality of life (HRQoL) (Sassi, 2006; Weinstein *et al.*, 2009). The underlying idea of the QALY is that a year of life spent in full health is worth one QALY and a year of life lived in a state less than full health is worth less than one (Brazier *et al.*, 2017). In order to determine QALYs, numerous preference-based HRQoL measures exist, which all provide estimates on the relative value that people place on living in particular health states (Neumann *et al.*, 2000).

The starting point for this thesis was the widely recognized concern that the evaluative space of the QALY focuses too narrowly on *health-related* aspects of quality of life (QoL) (Al-Janabi *et al.*, 2012; Goranitis *et al.*, 2017; Payne *et al.*, 2013). The evaluative space is generally concerned with the identification of the 'objects of value' (Sen, 1993). Applied to HRQoL, the evaluative space depends on the scope of the evaluative factors concerning a person's life as defined by the dimensions of HRQoL included in preference-based HRQoL measures. It is argued that adopting an evaluative framework that is concerned with health (and only health) when conducting economic evaluations may result in suboptimal resource allocation decisions because other 'non-health' benefits of health care interventions are not considered (Coast *et al.*, 2008a). As a response to this concern, new measures have been developed that offer a broader evaluative space, i.e., one that is not limited to health alone. This shift from preference-based HRQoL measures towards a broader set of outcomes in economic evaluation provides an opportunity for further examination of methodological considerations and applied implications for outcome measurement in health economics.

This thesis explores potential challenges associated with broadening the evaluative space of the QALY and provides an overview of broader non-health benefits that currently do not fit within the QALY framework. Based on three empirical studies conducted as part of this thesis, further evidence is provided in terms of the additional information offered by newly developed measures that adopt a broader evaluative space, when compared with preference-based HRQoL measures that are currently used to derive QALYs. Given that preference-based HRQoL instruments differ greatly in their coverage of dimensions, the comparative analyses in this thesis comprise a number of different HRQoL measures and a wide range of methods are employed to shed more light on the relationship between these alternative measures. Finally, this thesis discusses practical implications for the field of health economics and economic evaluation as well as areas for further research.

1.2. Economic evaluation of health care interventions

Health care resources are scarce relative to needs, which requires some form of rationing. An aging population and expensive technological advances have increased the cost pressures in health care to the extent that explicit rationing of services is inevitable (Coast & Donovan, 1996). In the absence of effective markets for most health care services, it is necessary to make decisions regarding the allocation of resources across different kinds of services (Jan, 1998). Health care is an economic good in the sense that delivering more of one type of care (from a fixed budget) displaces care elsewhere in the system. This reflects the concept of opportunity cost, which is the value of the consequences forgone by choosing to deploy resources in one way rather than in their best alternative use (Mooney & Drummond, 1982).

Over recent decades, attempts have been made in many countries to improve the allocation of health care resources using economic evaluation. An economic evaluation can help to inform the process of shifting resources to where they are most efficiently allocated. An economic evaluation is the comparative assessment of the costs and benefits of alternative health care interventions (Drummond *et al.*, 2015). The term

'benefits' is also referred to as 'consequences', 'outcomes', or 'effects' within the economic evaluation literature. Evaluations that consider both costs and benefits can be considered 'full' economic evaluations, of which there are three distinct types: cost-effectiveness analysis (CEA), cost-utility analysis (CUA), and cost-benefit analysis (CBA) (Drummond *et al.*, 2015). Within these three types of evaluation, the unit for measuring the benefits of health care is the key distinguishing feature. The scope of benefits within the three types of evaluation moves from uni-dimensional outcomes measured in natural units in CEA (e.g., life years or depression-free days), to a two-dimensional unit capturing HRQoL and length of life as measured by QALYs in CUA, to benefits measured in monetary terms in CBA (Lancsar & Louviere, 2008). Two other types of economic evaluation exist in the literature: the cost-consequences analysis (CCA) and the cost-minimization analysis (CMA). In a CMA, equivalence in benefits is assumed in comparative groups. A CCA provides a multi-dimensional listing of costs and benefits in a disaggregated form. Both types of evaluation have faced criticism; CMA for the ambiguous evidence of clinical equivalence and CCA for the absence of a full synthesis of benefits and costs. As such, strictly speaking, both types cannot be considered as a full economic evaluation (Drummond *et al.*, 2015). However, while CMAs are rarely conducted nowadays (Briggs & O'Brien, 2001), the CCA is gaining more attention in health care policy (further details to follow) (Coast, 2004).

Despite the existence of many types of outcome measures in economic evaluation, the QALY has become the dominant approach to evaluate the cost-effectiveness of health care interventions (Payne & Thompson, 2013). While CEA can limit comparability across clinical contexts due to the use of condition-specific measures, shortcomings of the CBA include the difficulty in assigning monetary values to health care benefits (Coast *et al.*, 2008d). The use of an outcome metric that is capable of allowing comparisons across clinical areas has led to the widespread adoption of the QALY (Round, 2012). CUA, based on QALYs, is currently the recommended approach by numerous health technology assessment (HTA) agencies, including the Canadian Agency for Drugs and Technologies in Health (CADTH) in Canada (CADTH, 2017) and the National Institute for Health and Care Excellence (NICE) in the UK (NICE, 2013).

1.3. Preferences and health state valuations

Within the CUA framework, preferences play a key role and form a fundamental component of the QALY calculation. In order to assess the value of a range of interventions, the impact of the intervention on an individual's health state is measured and valued, where more preferred health states receive greater weights (Whitehead & Ali, 2010). Preference weights for health states are anchored on a 0-1 scale, with 0 being the value of 'dead' and 1 being the value of 'full health', with negative scores possible, i.e., states worse than dead (Neumann *et al.*, 2000). A year of 'full health' (i.e., a health state equal to one throughout a 12-month period) is, therefore, equivalent to one QALY.

To measure the value of a particular health state, a number of preference-elicitation techniques have been developed. Three common approaches are the visual analogues scale (VAS), time trade-off (TTO), and standard gamble (SG) (Whitehead & Ali, 2010). The VAS is a form of a rating scale and is seen as the simplest direct elicitation technique, where the upper end of the scale indicates the 'best imaginable health' and the lower end of the scale the 'worst imaginable health'. While the EuroQol Group traditionally utilized these anchor points, other end-points can be applied such as 'full health' and 'dead' (Brazier *et al.*, 2017). Individuals are then asked where on the scale they would place the health state that needs to be valued. The TTO technique, on the other hand, provides individuals with two alternative options. One option describes a life in an impaired health state for a certain time period and the second option describes a life in full health for a shorter period of time. This shorter period of time in full health is varied until the individual is indifferent between those two alternatives. The value of the health states is then represented by the ratio of the two periods, i.e., the number of years in full health divided by the number of years in the impaired health state (Attema *et al.*, 2013). The standard gamble also offers individuals two alternatives, where the choice is between a continuation of life in the current health state to be valued and a gamble. The gamble offers the possibility of full health or immediate death, which are the anchor points required for use within the QALY framework. The probabilities associated with the gamble are varied until the individual is indifferent between the two alternatives,

where the indifference point becomes the value of the health state (Gafni, 1994). Although all three techniques are used to elicit preferences for health states, only SG provides utilities based on the expected utility theory (EUT) of von Neumann and Morgenstern, which requires choices to be made under uncertainty (von Neumann & Morgenstern, 1944). The VAS and TTO provide 'values', which are obtained under conditions of certainty (Drummond *et al.*, 2015). For consistency reasons, the term utility will be used throughout this thesis, referring to the numerical representation of a person's choice behaviour (Gafni & Birch, 1995).

In clinical research, the three approaches described above (also referred to as *direct* preference elicitation approaches) require many resources and are not always practical, as patient preferences would need to be elicited each time a study is carried out (Arnold *et al.*, 2009). For these reasons, numerous *indirect* approaches have been developed, which provide so-called 'off-the-shelf' values, using standardized HRQoL questionnaires (Neumann *et al.*, 2000). Here, values are available for every health state defined by the descriptive classification system (i.e., the items and response levels in the questionnaire). Examples of such descriptive classification systems include the 15D (Sintonen, 2001), Assessment of Quality of Life 8-dimension (AQoL-8D) (Richardson *et al.*, 2009), EQ-5D (3 level (EQ-5D-3L) or 5 level (EQ-5D-5L)) (Herdman *et al.*, 2011; Rabin & de Charro, 2001), Health Utilities Index Mark 2 (HUI-2) or Mark 3 (HUI-3) (Feeny *et al.*, 2002), Quality of Well-Being Scale Self-Administered (QWB-SA) (Seiber *et al.*, 2008), and SF-6D (Brazier *et al.*, 2002; Brazier & Roberts, 2004).

In order to apply these questionnaires across different disease groups, they need to be generic and, as such, the descriptive classification systems must capture a broad range of health dimensions to reflect people's multidimensional health states. For that reason, these standardized questionnaires are often called multi-attribute utility instruments (MAUIs), utility measures, or preference-based HRQoL instruments. The latter term will be used throughout this thesis. The weights to score these standardized health state classifications are obtained using one (or a combination) of direct elicitation techniques. As such, preference-based HRQoL measures consist of a descriptive system (referring to

the questionnaire) and a valuation system, which scores each health state as defined by the questionnaire and provides a single value (Brazier *et al.*, 1999). A large number of valuation studies have been conducted in many countries, with the aim of developing national valuation sets, also known as social tariffs (Engel *et al.*, 2016).

An important question in such valuation tasks is who should value health states. Different sources are possible including patients, their carers, health professionals and the general public (Brazier *et al.*, 2017; Dolan, 1999). Currently, preferences for health states of indirect preference-based HRQoL measures, such as the EQ-5D, are obtained from the general population because of individuals' status as potential patients and payers of publicly funded health care (Versteegh & Brouwer, 2016). However, evidence has shown that there are discrepancies in health state values between the general public and patients (Peeters & Stiggelbout, 2010; Ratcliffe *et al.*, 2007). While a distinction is usually made between the sources of preferences (patients versus general population), another consideration is made in terms of what is valued. Different scenarios are identified in Table 1.1. The general public is generally asked to value a hypothetical health state by imagining what the state would be like (cell A). The description of health states is usually derived from existing classification systems, such as the EQ-5D-5L. Patient preferences refer to patients valuing their own health state, which they are currently experiencing (cell D) with the use of direct elicitation techniques such as SG or TTO. While cell B (patients valuing hypothetical health states) and cell C (members of the general public who have experiences with specific health states) are also possible, it is important to understand that there are fundamental differences between preferences in cell A and D. Further discussion on this topic will follow in Chapter 6. International health economic guidelines vary between jurisdictions, where preferences from the general public are explicitly required in the UK and the Netherlands, while other countries, like Sweden, prefer patient preferences (Versteegh & Brouwer, 2016).

Table 1.1. Classification of preferences by population and what is valued

	General public	Patients
Hypothetical health state	A	B
Experienced (own) health state	C	D

1.4. The evaluative space of the QALY

QALYs are seen as a measure of health and the descriptive systems of preference-based measures from which QALYs are derived typically focus only on outcomes affecting a person's health or HRQoL. The World Health Organization (WHO) defines health as "*a state of complete physical, mental and social well-being, and not merely the absence of disease and infirmity*" (WHO, 1948). More recently, the WHO described health in terms of impairment, disability, and participation (WHO, 2001). The most widely used preference-based HRQoL measures (i.e., EQ-5D-3L, HUI-2 and HUI-3, and SF-6D (Richardson *et al.*, 2015c)) focus on these aspects of health, although differences exist in terms of how these aspects are incorporated into the respective descriptive classification system (Brazier *et al.*, 2007). For example, the HUI can be described as a measure of the 'within-skin aspects of health' (Furlong *et al.*, 2001), focusing primarily on impairment and disability (vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain). Alternatively, the EQ-5D (3 level or 5 level) contains aspects of impairment (pain/discomfort, and anxiety/depression), disability (mobility and self-care), and participation (usual activities). These three aspects can also be found in the SF-6D through the incorporation of pain and mental health (impairment), physical functioning and vitality (disability) and role limitation and social functioning (participation).

Figure 1.1 provides a graphical presentation of the term HRQoL as reflected by existing preference-based HRQoL. Using an example of a person who suffers from rheumatism, the Figure shows to what extent preference-based HRQoL measures combine aspects of WHO's definition of health in their respective descriptive classification systems. The figure also shows that although health care can impact a

person's broader QoL, when it comes to the evaluation of health care interventions using preference-based HRQoL measures, the focus is restricted to HRQoL. The term HRQoL refers in this context to aspects of QoL that can be affected by health (i.e., aspects affected by the presence of disease and treatment) (Brazier *et al.*, 2007).

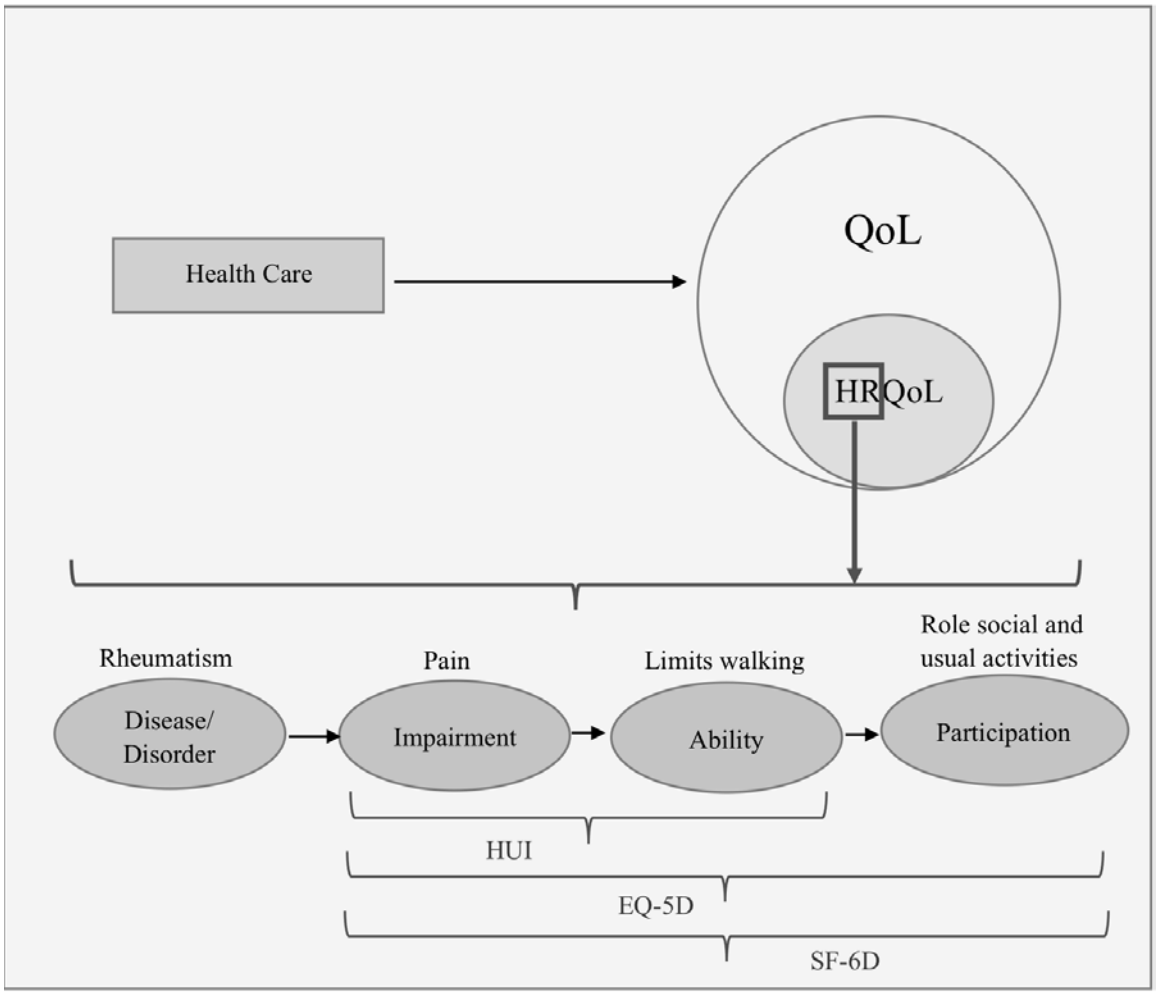


Figure 1.1. Graphical presentation of the term HRQoL as reflected by existing preference-based HRQoL measures

Besides the focus on impairment, ability and participation, preference-based HRQoL measures differ also in their scope of the evaluative factors concerning a person's life, as defined by the dimensions of HRQoL included in preference-based HRQoL measures. Table 1.2 provides an overview of dimensions included in six preference-based HRQoL measures. It shows that different HRQoL measures contain different physical and psychosocial dimensions. Given that preference-based HRQoL measures are used to derive utility weights for the QALY calculation, these measures, ultimately, determine the evaluative space of the QALY (Whitehurst & Engel, 2017).

Table 1.2. Dimensions of six preference-based HRQoL measures (derived from Richardson *et al.* 2015a) [a]

Dimension	EQ-5D-5L	SF-6D	HUI-3	15D	AQoL-8D	QWB
Physical						
Physical ability/ mobility/ vitality/ coping/ control	*	*	**	**	**	**
Bodily function/self-care	*			***	*	
Pain/discomfort	*	*	*	*	**	
Senses			**	**	**	
Usual activities/ work	*	*		*	**	
Communication			*	*	*	
Psychosocial						
Sleeping				*	**	
Depression/anxiety/anger	*	*	*	***	*** *** *	
General satisfaction					*** *	
Self-esteem					*** **	
Cognition/memory ability			*			
Social function/relationships		*			*** **	*
(Family) Role		*			*	
Intimacy/sexual relationships				*	*	
Total items/symptoms	5	6	8	15	35	[b]

[a] The number of asterisks (*) indicates the number of items measuring the dimension in the respective instrument.
[b] QWB has 3 items relating to mobility and physical and social health, plus 27 symptom groups.

1.4.1. Perceived limitations of the health-related QALY

It has been claimed that focusing on *health-related* aspects of QoL only for health care allocation decisions may result in suboptimal decisions because it does not capture broader benefits that may result from an intervention (Coast *et al.*, 2008a). Methods for the economic evaluation of health care interventions have existed for a number of years but these have mainly been applied to more narrowly defined 'clinical' interventions, such as drugs, devices and medical procedures. The attempt to evaluate interventions that are more complex has faced challenges, mainly because benefits cannot be simply measured in terms of health (Byford & Sefton, 2003; Husereau *et al.*, 2014; Payne *et al.*, 2013). Complex interventions are distinguished from simple interventions, such as pharmaceuticals, in that they often involve several interacting components, difficult behaviours in those delivering (or receiving) the intervention, more than one group or organizational levels targeted by the intervention, and numerous outcomes (Craig *et al.*, 2008). Treatments for drug users can be considered as a complex intervention, which not only result in health benefits for the affected individuals, but also for society in terms of the behaviour change of the drug user, which leads, for example, to crime reduction (Payne *et al.*, 2013). Similar in genetic services, which have important benefits for individuals and their families regarding knowledge about the diagnosis, prognosis and risk of having a condition (Payne *et al.*, 2013). In addition to the increased complexity of interventions, it can be observed that health care is shifting from *curing* towards *caring*. While cure is concerned with improvements in health functionings, care often helps sick people to maintain dignity (Dolan & Olsen, 2002).

With a shift towards care, services without measurable impact on health status are increasingly demanded (Kotzian, 2009). The provision of social care services is one such area. Social care is usually provided to individuals with physical or sensory impairments, learning difficulties and mental health problems (Netten *et al.*, 2012). For individuals with such conditions, the distinction between 'health needs' and 'social service needs' is often unclear and a shift towards an integrated approach to health and social care is observable in countries such as the UK (Mason *et al.*, 2015). The aim of social care services is to compensate a person for their lost functional ability (rather than try to

restore it), since most people using social care services have conditions that involve a permanent or declining loss of functional ability (Netten *et al.*, 2012). As such, social care services may not contribute to health *per se* but to QoL more generally (Grewal *et al.*, 2006), helping individuals with the types of tasks and activities associated with their daily activities.

Another area that often provides benefits beyond health emerges from the public health field. It has been observed that in the public health sector the use of preference-based HRQoL measures neglects relevant benefits of interventions (Goebbels *et al.*, 2012; Lorgelly *et al.*, 2008). Health promotion, for example, aims to improve an individual's ability to make informed decisions and increases control over his/her personal life. The evaluation of complex interventions, such as public health programs, is particularly challenging as they not only combine educational, social and political strategies but also because benefits are often occurring outside the health care sector (Greco *et al.*, 2016). With the increase in evidence showing that existing measures are not sufficient to capture such 'non-health' benefits (Goebbels *et al.*, 2012; Greco *et al.*, 2016), NICE guidelines in the UK have recognized that the evaluation of public health and social care interventions requires different approaches (further discussion will follow in Chapter 4) (NICE, 2012).

Traditionally, contingent valuation (CV) has been proposed as an alternative measurement approach to overcome the narrow evaluative space of the QALY. CV methods value outcomes in monetary terms based on individuals' willingness to pay (WTP) within a CBA framework (Gafni, 2006). WTP can be seen as superior to QALYs in that it does not impose restrictions on which dimensions of benefits respondents are permitted to consider in order to express their preferences (Olsen & Smith, 2001). However, this approach has some limitations, since evidence has shown that people's WTP is partly determined by their ability to pay, where the allocation of health care resources could be skewed towards the wealthy (Olsen & Smith, 2001). In addition, people are in general extremely uncomfortable valuing life and QoL in monetary terms and often overstate their real WTP (Coast *et al.*, 2008d). In practice, CBA has not been

widely accepted and current policy makers have a preference for CUA based on QALYs (Payne & Thompson, 2013).

Another approach to overcome the narrow focus on HRQoL within the QALY framework is to broaden the evaluative space. *This approach is the focus of the research conducted within this thesis*. Recent efforts to overcome the perceived limitations of the evaluative space of the QALY have led to the development of alternative preference-based instruments that are suitable for use in economic evaluation. Three such measures have resulted from the Investigating Choice Experiments for the Preferences of Older People (ICEPOP) project: namely ICEpop CAPability measure for older adults (ICECAP-O) (Coast *et al.*, 2008c), adults (ICECAP-A) (Al-Janabi *et al.*, 2012), and individuals at the end of life (ICECAP Supportive Care Measure (ICECAP-SCM)) (Sutton & Coast, 2014). In 2016, a fourth instrument became available, which captures the benefits to those close to the dying person (ICECAP-Close Person Measure, ICECAP-CPM) (Canaway *et al.*, 2016). Outside the ICECAP family of instruments, another measure was developed for the social care context, the Adult Social Care Outcomes Toolkit (ASCOT), which is a measure of a person's social care-related QoL (Netten *et al.*, 2012). ICECAP and ASCOT measures have only a few or no items directly related to physical and mental health, and are more concerned about a person's autonomy, enjoyment, control, or dignity. Table 1.3 provides an overview of the domains in the ICECAP-A and ASCOT.

ICECAP and ASCOT measures do not only have a broader evaluative space, these measures are also based on the same theoretical foundation set out by Amartya Sen's capability approach (Sen, 1993). Briefly, Sen's capability approach distinguishes between 'functionings' and 'capabilities'. Functionings are an individual's set of achieved doings and beings, or what a person manages to do or be. Capabilities are an individual's potential to achieve certain functionings, or the various combinations of functionings of what a person can do or be (Sen, 1993). Applied to the health care context, this approach implies that the evaluation of health care interventions should distinguish between 'observable health achievements' and the 'capability to achieve good

health'. These issues are discussed further later in the thesis (see Chapters 4-6). The ICECAP and ASCOT measures have been validated (although more validation work is still warranted) and have been used in a number of countries, including the UK, Australia, Canada, and the Netherlands (Coast *et al.*, 2008c; Couzner *et al.*, 2013b; Davis *et al.*, 2013; Makai *et al.*, 2014a; Makai *et al.*, 2013; van Leeuwen *et al.*, 2015a; van Leeuwen *et al.*, 2015b).

Table 1.3. Overview of the domains in the ICECAP-A and ASCOT [a]

Measure	Domains
ICECAP-A	Stability
	Attachment
	Autonomy
	Achievement
	Enjoyment
ASCOT	Control over daily life
	Personal cleanliness and comfort
	Food and drink
	Personal safety
	Social participation and involvement
	Occupation
	Accommodation cleanliness and comfort
	Dignity

[a] The ICECAP-A is used as an example here. Other measures of the ICECAP family comprise the ICECAP-O, ICECAP-SCM, and ICECAP-CPM.

1.5. Thesis aim and outline

With the development of outcome measures in economic evaluation that go beyond health, understanding the relationships between different approaches is an important area of research. The aim of this thesis is to examine methodological considerations when using outcome measures in health economic evaluation that extend beyond HRQoL and to examine implications for policy and practice when broadening the evaluative space of the QALY. The empirical investigations within this thesis focus primarily on the relationships between preference-based HRQoL measures and the adult (ICECAP-A) and older adult (ICECAP-O) ICECAP instruments. The main objective of this thesis is to examine the scope and additional information provided by the ICECAP measures when compared with conventional preference-based HRQoL measures. The thesis is constructed around three research questions:

- What are the potential challenges in broadening the evaluative space of the QALY?
- What is understood by 'benefits beyond health'?
- What information is captured by ICECAP measures over and above the information garnered by existing preference-based HRQoL measures?

To achieve these aims, a number of different approaches were undertaken, starting with a narrative review, followed by a literature review and three empirical analyses. A brief outline of Chapters 2-7 now follows.

Chapter 2 is a narrative review of challenges associated in broadening the evaluative space of the QALY. Theoretical issues associated with the welfarism and extra-welfarism approaches are examined, in addition to conceptual and normative considerations regarding the extent to which the current maximization principle in health care decision-making may be justified. Finally, important methodological and practical issues are discussed, including the measurement and valuation of broader benefits, as well as the necessity of pragmatism in economic evaluation.

Chapter 3 reports a critical interpretive synthesis of articles that discusses the notion of benefits beyond health within the context of QALYs. After examining how the literature to date has communicated such benefits, non-health benefits are organized into a thematic framework. The final output of this chapter is a concept map that displays the relationships between non-health benefits.

Chapter 4 is an empirical analysis of outcome measures within a public health context. This chapter reports findings from a study examining the role of the built environment and social cohesion for older adults' QoL, where the ICECAP-O is compared with the EQ-5D-5L. This investigation provides evidence in terms of the extent to which these instruments are able to capture broader benefits outside the health care context.

Chapter 5 investigates the extent of overlap between the ICECAP-A and five preference-based HRQoL measures in an exploratory factor analysis. The 'added value' of the ICECAP-A instrument in economic evaluation is discussed, along with practical implications regarding the use of measures of capability wellbeing.

Chapter 6 extends the comparative analysis from Chapter 5 through a path analysis, providing empirical analysis regarding the relationship between HRQoL, capability wellbeing and subjective wellbeing. Using data from individuals living with spinal cord injury who suffer from secondary health conditions, this analysis examines the effect of secondary health conditions simultaneously on HRQoL, capability wellbeing, and subjective wellbeing through direct and mediated pathways. It provides further practical considerations and implications for outcome measurement in economic evaluation.

Chapter 7 comprises a discussion of the results of each chapter and their implications for health care decision-making. The chapter outlines the degree to which previous research gaps have been addressed, as well as the limitations of this work and directions for future research.

Chapter 2. Challenges associated in broadening the evaluative space of the QALY

2.1. Introduction

The introduction of QALYs as an integral part of health care decision-making in the 1980s came along with criticisms (Harris, 1987; Rawles, 1989), where in particular the evaluative space of the QALY framework has been challenged (Mooney, 1989). Mooney cast doubt that health services are about health only and, as such, using cost per QALY as a decision rule for resource allocation purposes may be inadequate. In his view, maximizing QALYs cannot be equated with maximizing utility since there are other non-health outputs that can be considered as benefits from the health care system (Mooney, 1989). Chapter 1 provided examples of areas for which non-health outcomes are non-trivial, such as social care, public health, and other complex interventions. Evidence also exists about broader benefits that individuals derive from the actual processes of providing care, as opposed to achieving some desired level of health, referred to 'process utility' (Brennan & Dixon, 2013). Individuals may, for example, prefer easier access to certain facilities, or are willing to trade health gain against time spent being counselled. Other examples of process-related aspects of health care may also include 'being treated with dignity' or 'being treated in a pleasant way' (Brennan & Dixon, 2013). In particular, for end of life treatments, where health production may not be the focus of the care provided, process utilities can play an important role (Round, 2012).

Concerns around the evaluative space of the QALY have not just emerged in recent years, as discussions around non-health benefits have existed in the literature since the late 1980s (Mooney, 1989; Rawles, 1989). It appears that this topic has faced numerous challenges that require further consideration. This chapter outlines the theoretical, normative, conceptual, methodological, and practical challenges associated

with broadening the evaluative space of the QALY. It explores to what extent challenges have been already addressed and highlight existing challenges that require further research.

2.2. Theoretical challenges

2.2.1. Welfarism versus extra-welfarism

In standard welfare economics the overall welfare of the society is a function of individual utilities, and individual utilities are a function of the goods and services consumed by the individuals themselves. It is based on the consumer sovereignty theory (Brouwer *et al.*, 2008), which implies that individuals are themselves the best judges of *what* contributes most to their utility and *how much* that contribution is (Birch & Donaldson, 2003). Utilities here can be understood as the value of a function that represents an ordering, in particular, a preference ordering of different combinations of goods and services consumed (Gafni & Birch, 1995). In other words, utility is a numerical representation of a person's choice behaviour. In welfare economics, a social utility function is interpreted as an aggregate of individual utilities and the ultimate goal of any resource allocation scheme is to maximize the social utility function (Coast, 2009).

To determine whether an improvement in social welfare has taken place or not, the 'Pareto principle' is used. A Pareto improvement would occur if a policy in health care makes one or more persons better off (i.e., increases their utility) without making another person worse off (i.e., decreases their utility) (Coast, 2009). In practice, however, allocation of resources cannot produce only winners but also involves losers (i.e., resources devoted to the health care of one person will deny health care to another person who might have benefited). A less strict approach, called the 'potential Pareto principle' (or Kaldor-Hicks criterion (Nicholas, 1939)), allows for the possibility of the winners compensating the losers. Under this approach, society as a whole has benefited from a particular allocation decision if the winners could, in theory, compensate the losers and still remain better off than they were before the decision (Coast, 2009). Important to note

is that since individuals can derive utility from different sources, such as from the consumption of health services, education or chocolate bars, individuals not receiving health can be compensated by enhancing their utility from other sources. Health is, therefore, seen as an intermediate stage that contributes to a person's utility from the consumption of health services (Coast, 2009).

In contrast, the 'extra-welfarism' approach, sometimes called the 'non-welfarist' or 'decision-maker' approach, replaces utilities with health as the primary outcome of interest for evaluation (Coast *et al.*, 2008d). This approach tries to maximize health as opposed to overall welfare. Originally, extra-welfarism was derived from Sen's capability approach (Sen, 1993). Sen rejected the exclusive focus on individual's utilities and suggested to replace it by a broader perspective, considering the quality of utility and individual's capabilities rather than the emotional reaction of individuals to the possession of goods and capabilities. Building upon Sen's capability approach, Culyer introduced in the 1980s a theoretical framework around the extra-welfarism approach within health economics (Culyer, 1989). It differs from the welfarist perspective in four important aspects: i) it permits the use of outcomes other than utility; ii) it permits the use of sources of valuation other than the affected individuals; iii) it permits the weighting of outcomes (whether utility or other) according to principles that do not need to be preference-based; and iv) it permits interpersonal comparisons of wellbeing in a variety of dimensions, thus enabling movement beyond welfare economics (Brouwer *et al.*, 2008). These four important differences are discussed in further detail below.

While the welfarism approach exclusively focuses on individual utilities, the 'extra' in extra-welfarism is that it broadens the evaluative space to include other non-utility information. This may include, for example, equity weights and capabilities (Brouwer *et al.*, 2008). However, although the extra-welfarism approach allows taking measures other than individual utilities into account, Culyer's expression of extra-welfarism is limited in that it focuses only on health (Coast *et al.*, 2008d). Given that decision makers are mainly interested in health, health has become the primary outcome of interest for evaluation (Culyer & Evans, 1996). Whereas in the welfarism approach

health is taken into account insofar as it enables utility to be derived from the consumption of health care, the extra-welfarism approach suggests that the output of health care should not be judged in terms of preferences of health against other goods, but rather to its *contribution to health*. Culyer's operationalization of the extra-welfarism approach aims to maximize health (in terms of QALYs) rather than overall welfare, and ignores individuals' trade-offs between health and other commodities and/or characteristics (Birch & Donaldson, 2003). Even though utility theory is used in the derivation of QALYs, it is used solely to measure people's health rather than the utility they drive from it (Wagstaff, 1991).

The second characteristic of the extra-welfarism approach allows judgments to be made on behalf of, rather than by, affected individuals. In other words, while decisions in the welfarism approach are being made from the individual perspective because an individual is the best source of their own health state, the extra-welfarism approach allows external judgments that replace or supplement the subjective utility numbers in the social welfare function. It is argued that individuals need to be protected from their own foolishness, which on the flip side may override their own best interests (Birch & Donaldson, 2003). The extra-welfarism approach assumes that a health state has the same effect on all individuals. As such, two individuals with the same health states would be treated equally according to the extra-welfarism perspective, regardless of whether one of them is happier or copes well with their health state. The welfarism approach would, in contrast, encapsulate these coping differences (Gyrd-Hansen, 2005). As a result, individual preferences might be overlooked within the extra-welfarism approach, as it is the aim to maximize the overall health in a society. In the extra-welfarism approach, any number of stakeholders can make judgments in regards to the different values for the different entities, and how they should to be traded off against one another and compared interpersonally (Brouwer *et al.*, 2008). In practice, there is still a debate about the exact source of value and the question of whose preferences should count (e.g., patients versus public) (Dolan, 1999; Versteegh & Brouwer, 2016).

The weighting of outcomes is another key distinction between the two approaches, associated with the assignment of priority levels according to principles that need not be preference-based. In order to weight benefits, such as health, certain characteristics are required. These characteristics may imply that, for example, more weight should be attached to health gained by the young, or those with poor levels of initial health, or those who stand to lose a large proportion of remaining health. These weights can be used to address equity concerns within the extra-welfarism framework. Although these weights do not need to be preference-based, extra-welfarists strongly advocate the use of preference-based measures (Brouwer *et al.*, 2008). Other authors argue that distributional weights are not exclusively for the extra-welfarism approach and can be also found in the welfarist theory (Birch & Donaldson, 2003). It is argued that if individuals were concerned about distributional issues, either directly or indirectly, this would be reflected in their utility function.

It was previously mentioned that under the extra-welfarism perspective health is not treated as an intermediate stage on the way to producing utility but valued on its own right. In view of the fact that the allocation of health resources happens simultaneously with the production of health, health cannot be transferred from one individual to another once produced (Coast, 2009). It follows that the potential Pareto principle no longer provide a sufficient theoretical basis for health maximization because the production and distribution of health in extra-welfarism is in theory not separable. Therefore, a movement beyond welfare economics is acceptable by allowing *interpersonal comparisons* of wellbeing in a variety of dimensions. Interpersonal comparisons refer to the ability to compare utility functions of different individuals that are measured on the same yardstick. From a welfarist perspective, interpersonal comparisons appear impossible or meaningless since neither the Paretian principle, nor the compensation (potential Paretian) principle require interpersonal comparison. Under these two approaches, what matters is whether at least one person's utility is increased while no other individual's utility is reduced, or the minimum compensation it would take to make sure the loser is no worse off but leaves the gainer better off (Birch & Donaldson, 2003). Within the extra-welfarism approach, however, interpersonal comparisons are needed to

determine which groups or persons to give priority in allocating health care resources (Brouwer & Koopmanschap, 2000). Using health-related outcome measures, such as QALYs, enables the comparison of individuals within a health domain. Under this approach, it is irrelevant if an individual might be worse-off since the only thing that matters are the total QALYs, which are used as a summary measure for the value of an intervention and as a basis of comparisons between alternative resource allocations (Birch & Donaldson, 2003).

2.2.2. Incorporation of non-health benefits under the welfarism and extra-welfarism approach

The welfarism approach is currently the underlying approach of a CBA based on WTP. In welfare economics the measurement of social welfare is the wellbeing of individuals as assessed by themselves. This implies that in economic evaluation, benefit assessment for a given intervention needs to be based on the extent to which individuals are affected and how they value this impact (Brazier *et al.*, 2007). A CBA aims, therefore, to assess how people value changes in their own health. In this context, both health outcomes and non-health outcomes can be included as long as individuals derive utility from it. Here, health is only taken into account insofar as it enables utility to be derived from the consumption of health care.

Even though broader outcomes are considered under the welfarist approach, it does not account for process utilities, as from the traditional welfare economic view it is assumed that health care has no value in use (Mooney, 2009). Based on Grossman's model of the derived demand for health, individuals pay for health care only because of the benefits derived from it, through expected health gains. The welfarism perspective is based on 'consequentialism', which implies that utility can only be derived from the *outcomes* of behaviour and processes rather than the processes themselves or intentions that led to the outcomes (Brouwer *et al.*, 2008). In other words, processes and procedures would only be considered if they have utility consequences. However, as Donaldson and Shackley have argued, there are other elements in the consumer's utility function – not

only consequences and outcomes (Donaldson & Shackley, 1997). There can be utility in the process of health care as well as in its outcome.

Whereas CBA falls under the welfarist approach, the QALY has no foundation in economic welfare theory (Birch & Donaldson, 2003). Firstly, in welfare economics the affected group of individuals is the primary source of valuation. This is contrary to the QALY concept, which is often not based on an individual's own valuation of their health but rather societal preferences using preference-based HRQoL measures (e.g., EQ-5D-5L (Herdman *et al.*, 2011)). Secondly, health is not the single component of utility. Beside health, there are other items captured in terms of individual utility, like processes, institutions, equity, quality of relationships, and social norms. If health were indeed the single component of an individual's utility function, then the QALY concept would fit within the welfarism approach. But if other non-health aspects would also contribute to an individual's wellbeing, this does not hold. It follows that the QALY suits better the extra-welfarism approach. In fact, when the QALY was introduced, some health economists criticized its fit within the welfarist approach and the extra-welfarism approach can be seen as an attempt to justify health maximization and, in turn, QALY maximization in the health care objective function (Mooney, 2009). Recent research has explored the WTP for a QALY, which can be interpreted as a CBA since the non-monetary consequences in a CEA can be translated into monetary terms (Gyrd-Hansen, 2005). Although this approach could be interpreted within the welfare theoretical framework, it is not being used to date.

From an extra-welfarist approach, based on CUA using QALYs, neither process utility nor non-health benefits are considered, as can be observed in existing national guidelines (CADTH, 2017; NICE, 2013). Given that the objective of publicly-funded health care is to improve population health, the extra-welfarism approach focuses on health outcomes only and ignores the individual's potential willingness to trade between health and other aspects that may yield wellbeing. Consequently, the current extra-welfarism approach would only consider non-health outcomes or process utility if they contribute to health.

Other authors have argued that if decision makers would accept a broader concept of wellbeing than that typically adopted within extra-welfarism, process utilities can be added into the QALY calculation (Brennan & Dixon, 2013). In fact, there is no conceptual reason to exclude process utility and non-health outcomes within the extra-welfarism approach (Round, 2012). In theory, therefore, non-health outcomes can be included within the QALY framework. This reflects also the viewpoint of Brouwer and colleagues who have provided another interpretation of the extra-welfarism approach that is more aligned with Sen's capability approach (Brouwer *et al.*, 2008). Although Culyer argued that his extra-welfarism framework was based on Sen's capability approach, the extent of Sen's influence is limited due to three reasons. Firstly, Culyer's expression of extra-welfarism relies purely on health as an outcome, whereas the capability approach is multi-dimensional as can be seen in Nussbaum's list of ten central human capabilities (Nussbaum, 2003). Secondly, Sen makes a clear distinction between capabilities and functionings, which is not reflected in Culyer's extra-welfarism approach. Finally, whereas Sen's capability approach moves away from a maximization principle towards equity and poverty reduction, Culyer's expression of extra-welfarism is ultimately to maximize health in terms of QALYs (Coast *et al.*, 2008d). Therefore, accepting a broader extra-welfarism perspective as proposed by Brouwer *et al.* compared with the narrow explanation introduced by Culyer would allow for the incorporation of benefits beyond health (e.g., non-health outcomes and process utility) if a decision maker with authority decides to do so (Brouwer *et al.*, 2008).

Table 2.1 compares the welfarism approach with the two extra-welfarism frameworks; one framework that adopts a 'narrow' extra-welfarism view based on Culyer's interpretation and one 'broader' extra-welfarism view based on Brouwer and colleagues. Table 2.1 also summarizes to what extent non-health benefits can be incorporated into these theoretical frameworks.

Table 2.1. Key characteristics of the welfarism approach compared with two extra-welfarism frameworks [a]

	Welfarism	'Narrow' Extra-welfarism (Culyer 1989)	'Broad' Extra-welfarism (Brouwer *et al.* 2008)
Relevant outcomes	• Focus on individual utilities • Social welfare is a function of individual welfares • Consequentialism	• Focus on health	• Focus on outcomes other than utility • Selection of outcomes is context-dependent
Source of valuation of relevant outcomes	• The affected individual (individual sovereignty)	• Permits the use of sources of valuation other than the affected individual	• Permits the use of sources of valuation other than the affected individual
Weighting of relevant outcomes	• Weighting of outcomes reflected in individual utilities	• Permits the weighting of outcomes	• Permits the weighting of outcomes
Interpersonal comparison	• Individual utilities are not comparable	• Permits interpersonal comparisons	• Permits interpersonal comparisons
Maximization rule	• Maximize social welfare	• Maximize health (expressed in QALYs)	• Maximization rule is context-dependent
Incorporation of non-health outcomes	• Yes, if individuals derive utility from it (utility can only be derived from outcomes and not from processes)	• Only if they contribute to health	• Yes, if a decision-maker with authority decides to include them

[a] Information provided in this Table were partly derived from Brouwer *et al.* (2008).

2.3. Normative challenges

The extra-welfarism approach is based on the premise that the aim of an economic evaluation is to maximize whatever the decision maker wants to maximize. The question is then: what should be maximized? This is an important normative question because only those benefits that the decision maker considers to be relevant will be followed by analysts and included in the primary analysis. Within the context of health care, the focus is exclusively on health, as it is argued that individuals prefer more health to less health, which is why health is the appropriate maximand to be used in comparing resource allocations (Culyer & Evans, 1996). For many years, the emphasis in health economics has been almost exclusively on health (Mooney, 2009). Grossman's work on the demand for health in the early 1970s defines health as illness-free days, which indicates the ability to work (Grossman, 1972). The healthier a person, the more healthy time is available either for work or for leisure activities. Grossman's model showed that individuals do not demand health care for its own sake but because they have a demand for health. Thus, the demand for health care is a *derived* demand (Grossman, 1972). Given this emphasis on health, as opposed to health care, it is no surprise that health is treated as an end goal.

The focus on health was further strengthened with the introduction of QALYs, which are mainly concerned about a person's morbidity and mortality (Mooney, 2009). Changes in other outcomes, such as patient autonomy, respect, or information are not measured. Although many would argue that health is the key consideration for assessing the benefits of medical interventions, it is less clear how non-health benefits should be dealt with. Particularly, when comparisons are made across a broad range of interventions (e.g., drugs and public health services), interventions that result in non-health benefits would be disadvantaged in allocation decisions if such broader benefits were not captured. Ryan and Shackley provided examples of such benefits, including information, autonomy, dignity, the process of treatment, and non-medical reasons for visiting the doctor (Ryan & Shackley, 1995).

Not taking such evidence into account can potentially lead to misallocation of resources because the inclusion of broader outcomes can have significant effects on the conclusion of an economic evaluation, as a recent study demonstrated (Makai *et al.*, 2015). In an economic evaluation of an integrated care model for frail seniors, Makai and colleagues found that the intervention had a higher probability of being cost-effective when using the ICECAP-O, which is a broader measure of a person's capability wellbeing (see Chapter 4.2.2), when compared with use of the EQ-5D-3L. A second example came from a study that evaluated the cost-effectiveness of two psychological interventions for drug addictions (Goranitis *et al.*, 2017). This study applied the ICECAP-A and EQ-5D-5L and concluded that under the health maximization principle based on the EQ-5D-5L, the results yielded different treatment recommendations than applying a 'sufficient capability' approach (Mitchell *et al.*, 2015b) using the ICECAP-A. Similar implications can be found for process utility within the QALY framework. McNamee and Seymour have shown that process values produced a different number of QALYs and QALY gains compared with those derived from health outcome values (McNamee & Seymour, 2008). As such, the authors argue that the estimation of process utility provides additional information to policy makers in judgments over the cost-effectiveness of health care interventions and offers a promising alternative to standard cost-per-QALY estimation.

An important question is how far we should go with the evaluative space of outcome measurement in economic evaluation. As Donaldson and Shackley discussed in regards to process utilities, the possibility of trading non-health attributes for health attributes may mean that the provision of television sets in hospital could be recommended over some health-enhancing action if the utility gain from the former is greater (Donaldson & Shackley, 1997). This concern brings back into question whether going beyond health represents 'good value for money'. Currently, 'value' in health care can be understood as health outcomes achieved per dollar spent (Porter, 2010). Whether this definition is an accurate reflection of societal preferences can be challenged. It is likely that citizens demand health care services that might have little or no impact on health status. As such, it is important to know what society considers as 'good value'. For

example, as people get richer, there is a change in what they expect from their health care services (Kotzian, 2009), meaning that increases in wealth may cause a change in how health care is valued. The production of non-health outputs does not necessarily make the health care system inefficient. In fact, not considering these preferences can result in a misinterpretation of health care systems being inefficient (Kotzian, 2009).

When it comes to the evaluation of health care systems, there is the need to take societal preferences into account. To date, evidence in terms of whether the society is willing to give up health gains for non-health benefits remains scarce. A recent study by Bansback and colleagues explored which aspects in health care priority setting members of the public value most (Bansback *et al.*, 2014). An analytic hierarchy process was applied in order to elicit preferences for six different attributes from a representative sample in British Columbia, Canada (health benefit, condition severity, prevention, fairness, environmental, and non-health benefits). In this study, non-health benefits were defined as greater convenience or comfort, as well as increased confidence or autonomy. The study showed that study participants valued health benefits most (33%) and non-health benefits least (7%). While this study provided important empirical evidence, a potential limitation concerns framing effects. As participants were asked which attributes should be given higher priority in *health care*, it is maybe not surprising to see that they attached a low value to non-health outcomes. The problem here arises also in how we label these outcomes, which leads to further conceptual challenges around this topic.

2.4. Conceptual challenges

The discussion about health and non-health outcomes for use in economic evaluation is difficult due to the lack of a clear definition of such outcomes. The WHO defines health "*as a state of complete physical, mental and social well-being, and not merely the absence of disease*" (WHO, 1948). However, in health economics and in the development of preference-based HRQoL measures, the WHO definition of health has influenced this field only to a limited extent. In a recent paper, de Vries and colleagues have argued that existing scales (and corresponding QALYs) may not provide a

comprehensive picture of the effectiveness of an intervention for a patient's health as defined by the WHO (de Vries *et al.*, 2016). They argue that while QALYs are currently derived from health measures that focus primarily on physical and mental functioning, they ignore social wellbeing.

While the WHO definition of health is still most commonly adopted in practice, others have argued that this definition is no longer fit for purpose. Primarily, criticism has been focused on the word 'complete' in relation to wellbeing. Particularly for chronic diseases and disabilities, the WHO definition declares these individuals as ill and minimizes the role of the human capacity to cope autonomously with life's changing physical, emotional, and social challenges, and to function with fulfilment and feeling of wellbeing. Consequently, Huber *et al*. define health as "*the ability to adapt and to self-manage"* and argue that this is more meaningful for health policy where "*health gain in survival years may be less relevant than societal participation, and an increase in coping capacity may be more relevant and realistic than complete recovery*" (Huber *et al.*, 2011, p.2).

Having a clear concept of health would enable a better understanding of non-health outcomes, but the absence of such a definition means the term 'non-health outcomes' remains unclear. In a Dutch study that tried to identify important non-health outcomes for health promotion, Goebbels *et al.* used a working definition that categorized non-health outcomes as all outcomes not covered by the EQ-5D measure, which contains the dimensions mobility, self-care, usual activities, pain/discomfort, and anxiety/depression (Goebbels *et al.*, 2012). At the same time, the authors acknowledged that the dividing line between the non-health and health outcomes was not always a clear-cut. Conceptual challenges arise also due to the interchangeable use of the terms health, HRQoL, and QoL in the economic evaluation literature (Karimi & Brazier, 2016). Given that HRQoL is sometimes equated with health status and sometimes with QoL, it appears to be impossible to draw a line between health outcomes and non-health outcomes. Outside the heath economics literature, more consistent definitions are observed, where HRQoL focuses on the measurement of symptoms and functions, whereas QoL is not

only influenced by disease and treatment but also by personality, economic status, environment, social relationship and culture (Doward & McKenna, 2004).

There is a general consensus that, in an economic evaluation, health outcomes are defined in a narrow sense, i.e., the effects of health care on life years and on *functional* aspects of HRQoL, including aspects related to morbidity and disease, such as mobility, disability, pain and distress (Donaldson & Shackley, 1997). While it can be argued that this concept might miss important attributes of health care, such as reassurance, dignity and information, it has been argued that these attributes can be attributed to health because they reflect peoples' mental health state (Donaldson & Shackley, 1997). Process utility can also be construed as health if the process of care were to affect a person's mental health. Some authors have stated that process utility should be measured as part of a wider attempt to maximize wellbeing in accordance with the definition of health given by WHO (Higgins *et al.*, 2014). However, a further issue with process utility is that is remains unknown what aspects of care fall into the process utility category. In a review by Opmeer *et al.*, process-related preferences were considered as one type of non-health outcomes, including aspects such as discomfort, duration, or health care professional, whereas intervention-related outcomes (e.g. invasiveness) and patient-related outcomes (e.g. uncertainty or embarrassment) were seen as separate types of non-health outcomes (Opmeer *et al.*, 2010).

The absence of a clear definition of what constitutes an outcome and what is regarded as a process is associated with further challenges. In a study by Donaldson and Shackley, which examined the existence of process utilities in cholecystectomy patients using WTP, process utilities were defined as anything that takes place during surgery and while recovering from surgery, including faster recovery, quicker return to work, shorter hospital stay, and less pain. Outcomes, on the other hand, were defined as a state in which the patient is left after they have recovered (Donaldson & Shackley, 1997). While such definitions provide further clarity within the context of an individual study, Ryan & Shackley stress that a clear definition is not needed. What matters more is the fact that

patients may derive utility (whether called process or outcome) from aspects of health care other than health outcome (Ryan & Shackley, 1995).

While the usefulness of clear definitions of non-health outcomes remains debatable, current practice still requires a distinction to be made. For example, the recent NICE guidelines state that for some interventions (e.g., interventions with a social care focus), the intended outcomes of interventions are broader than improvements in health status and, as such, broader preference-based measures of outcome are more appropriate (NICE, 2014). The guidelines continue saying that if an intervention is associated with both health and non-health-related outcomes, it may be helpful to present these elements separately in a CCA. However, in order to justify the use of another type of economic evaluation, a better understanding of such benefits is needed. The fact that NICE do not specify what constitutes a health and non-health outcome means that further conceptual examinations are warranted.

2.5. Methodological challenges

Outcome assessment in economic evaluation is divided into three stages: the identification stage, the measurement stage, and the valuation stage. The examination of methodological challenges associated with broadening the evaluative space of the QALY framework, are assessed for each stage separately.

2.5.1. Identification stage

Under Culyer's extra-welfarism approach, having a definition of non-health outcomes is inevitable. Given that non-health outcomes are outside the evaluative space, such outcomes would need to be excluded unless they contribute to health. In contrast, adopting a welfarist perspective does not require a strict definition of health and non-health outcomes because the goal would be to maximize an individual's utility, regardless whether health outcomes or non-health outcomes enter an individual's utility function. In current practice, where decisions are made on behalf of the affected individual, decision

makers determine the appropriate maximand. Given the maximand, a clear scope needs to be formulated in terms of what should be measured.

A crucial role in determining the relevant outcomes is defined by the perspective of the economic evaluation (Claxton *et al.*, 2010; Gold *et al.*, 1996). The perspective not only determines the evaluative scope (i.e., who is affected by the intervention and who should be included in the analysis) but also the evaluative space (i.e., which outcomes should be measured). In 1996, the US Panel on Cost-effectiveness in Health and Medicine explicitly recommended the societal perspective as the reference case (i.e., the set of preferred methods that an analyst should follow when conducting the base case analysis in an economic evaluation). The societal perspective implies that all costs and benefits should be taken into account, regardless of where these costs and benefits occur (Gold *et al.*, 1996). Given the societal perspective, an outcome measure would capture everything that matters to society and would need to reflect what decision makers would like to accomplish in the public interest. Although the US Panel acknowledges that QALYs do not fully reflect this, QALYs still took the role of the main outcome given the fact that other important public values cannot be incorporated in a useful way (Gold *et al.*, 1996). In an update to the 1996 recommendations, the second Panel on Cost-effectiveness in Health and Medicine reflected on their initial recommendations and stated that since the original publication, many cost-effectiveness analyses have not used a societal perspective (Sanders *et al.*, 2016). In addition, the Panel highlighted that a number of HTA bodies in Europe, Australia and Canada focus on a health system perspective. New recommendations for cost-effectiveness analyses include now two reference case analyses: one based on a health care sector perspective and another based on a societal perspective (Sanders *et al.*, 2016).

It can be argued that identifying and measuring all relevant benefits will lead to more efficient resource allocation decisions and will enhance the legitimacy and acceptability of such decisions. Yet, only a few countries adopt a societal perspective, such as the Netherlands or Denmark (Mathes *et al.*, 2013). Given the current focus on the health care system in many countries, societal preferences are only taken into account

when eliciting preferences for health states. In general, decision makers have little information about the preferences of the general public towards other benefits that are currently not captured by existing preference-based HRQoL instruments. More research is needed to examine what else the general public values in health care besides health.

2.5.2. Measurement stage

Once identified, outcomes beyond health need to be measured. It can be assumed that using existing preference-based HRQoL will not capture important non-health outcomes of health interventions, as they were developed to measure explicitly health-related aspects of QoL. However, the term 'health-related' seems to be vague in this context, as some of the existing preference-based measures contain dimensions in their descriptive systems that are affected by aspects other than health alone (Brazier *et al.*, 2007). The SF-6D, for example, asks respondents about their role limitations or social activities, which are not only influenced by health (Brazier *et al.*, 2002). Even more such dimensions can be found in the AQoL-8D, including a person's role in the community, relationships with family and friends, happiness, or control of one's life (Richardson *et al.*, 2011). Therefore, when measuring non-health outcomes, existing HRQoL instruments should be examined more closely in terms of the extent to which they already incorporate broader benefits beyond health. This will be examined in more detail in subsequent Chapters 5 and 6 of this thesis.

2.5.2.1. Developing new measures

When developing new measures of non-health and process utility for use within the QALY framework, it is essential to provide evidence that these types of outcomes are currently not captured by existing preference-based measures (Brazier & Tsuchiya, 2015; Brennan & Dixon, 2013). For example, if a new measure of non-health outcomes will be developed that includes social activities as one of the dimensions in its descriptive system, double-counting can occur if such a measure will be applied alongside the SF-6D. On the other hand, it can be argued that double counting is already evident in existing preference-based HRQoL measures, since diseases that impact, for example, on pain are

likely to impact on multiple dimensions, such as usual activities (Brazier *et al.*, 2007). This implies that respondents completing the EQ-5D-5L, which includes pain and usual activities in its descriptive system, may be overvaluing the impact of pain on QoL. The issue of double counting is particularly important for comprehensive measures like the AQoL-8D (see Table 1.2).

Developing a new instrument that captures non-health benefits from health care interventions requires some additional considerations. The selection of domains for the descriptive systems of many existing preference-based measures were derived from literature reviews and expert judgments with little input from the patients (Stevens, 2016). Some authors have suggested that in order to improve content validity, it is important to ask patients to reflect on what matters to them. This is particularly the case when examining non-health outcomes and process attributes, as patients are a better source for identifying important dimensions based on their own experiences. Patients should, therefore, be involved in the development and testing phase of measures (Coast *et al.*, 2008a; Stevens & Palfreyman, 2012). For generic measures, the challenge lies in the selection of respondents. If the aim is to develop a generic measure to be applicable across a number of patient groups, a purposive sampling of respondents is needed in the development of instruments and their evaluation in specific populations groups (Michel *et al.*, 2016).

2.5.2.2. Developing bolt-ons

The addition or extension of dimensions of existing measures to cover the dimensions deemed to be missing from the measure is another way to capture non-health benefits. Examples of so called ‘bolt-ons’ have been developed for vision, hearing and tiredness (Longworth *et al.*, 2014), cognition (Krabbe *et al.*), sleep (Yang *et al.*, 2014), satisfaction (Dolan *et al.*, 2013), dignity (Dixon *et al.*, 2011), and for patients with psoriasis (Swinburn *et al.*, 2013). The key challenge with bolt-ons is that the addition of another dimension to an existing measure has consequences for the valuation of health states defined by the new instruments, meaning that the use of bolt-ons would require the re-valuation of existing measures with the bolt-on. A second challenge is to identify

specific dimensions that are 'missing' from the respective descriptive system. Existing *generic* instruments also require the addition of *generic* dimensions. However, Chapter 1 outlined that benefits beyond health are particularly important for certain conditions and disease areas, and also examples described above of previously developed bolt-ons are disease-specific. As such, it can be argued that the discussion around the inclusion of non-health benefits is not different than the discussion around disease-specific measures. The inappropriateness, or rather the limited sensitivity of some generic preference-based measures for certain clinical conditions has been highlighted in the literature a number of times, including asthma (Flood *et al.*, 2006), diabetes (Sundaram *et al.*, 2010), or epilepsy (Mulhern *et al.*, 2012). As a response, many disease-specific preference-based measures have been developed for use in economic evaluation (Versteegh *et al.*, 2012), although the usefulness of specific measures remains limited with regard to comparisons across disease and population groups (Brazier & Tsuchiya, 2010).

2.5.2.3. Using existing measures that adopt a broader evaluative space

The final approach to measure broader benefits would be the use of existing measures, such as the ICECAP measures (Al-Janabi *et al.*, 2012; Canaway *et al.*, 2016; Coast *et al.*, 2008a; Sutton & Coast, 2014) or the ASCOT (Netten *et al.*, 2012). These measures were developed with the aim to broaden the evaluative space and to capture benefits beyond health. Whereas the ICECAP measures were designed to assess a person's capability wellbeing, the ASCOT measure is designed to capture information about an individual's social care-related quality of life (SCRQoL). The ICECAP measures and the ASCOT are based on Sen's capability approach and the distinction between capabilities and functionings is of central importance to the development of both measures (Karimi *et al.*, 2016). However, the use of such measures is associated with additional challenges, which are outlined below.

The main challenge in using these instruments relate to the way they measure capabilities, which is in tension with Sen's capability approach. Sen distinguishes between functionings, which are the things someone can do and be, and capabilities, which are the combinations of functionings available to a person, also referred to a

person's capability set. A common example provided in the capability literature relates to the nutritional state of a person. In this context, a person who is fasting has the capability and opportunity to eat compared with a person who is starving. As such, capability reflects the intrinsic value of having a choice and the opportunity for a better combination of functionings. In an attempt to operationalize the capability approach for the purposes of economic evaluation, the ICECAP developers phrased the questions in terms of 'are you able to' or 'can you' in order to explicitly shift the focus away from functionings (which are concerned with what a person 'is' and 'does'). This operationalization is different to the ASCOT measure. The ASCOT measure contains four response options (i.e., ideal state, no needs, some needs, high needs) for each domain. While the lower three levels of each domain reflect levels of basic functioning (or needs), the highest level refers to capabilities. The developers state that "*once needs are met, it is also essential to identify capabilities: whether or not people are able to achieve their desired situation*" (Netten *et al.*, 2012, p.x).

In a recent article, Karimi *et al.* state that existing capability measures developed for use in economic evaluation have two problems: (i) they measure each domain independently of other domains and, therefore, ignore the value of choice in the capability set, and (ii) they do not provide an accurate description of an individual's entire capability set (Karimi *et al.*, 2016). Although Sen advocated that interpersonal comparisons could be made in the space of capabilities, Sen avoided the provision of a list of capabilities because a person's capability set may vary. The list of capabilities is context-dependent, in terms of both the geographical area to which it applies as well as the sort of evaluation that is done (Sen, 1993). It is likely that Sen would argue that reducing people's capability achievements to a number of attributes to measure capability wellbeing, as done by existing measures, is inappropriate as these vary across individuals in quality and quantities. Three other problems with existing capability measures relate to: (i) collapsing multidimensional capabilities information into a single index score using preference-based valuation techniques; (ii) measuring perceived capabilities rather than objective capabilities, where problems with adaptation might occur (i.e., individuals may not recognise their own lack of capability because they have adapted to their

situation); and (iii) ignoring a person's agency goals, which refers to goals other than one's wellbeing (e.g., wellbeing of others or commitments outside the person). Before using existing measures to capture broader benefits of health care interventions, these challenges for the capability approach need to be resolved.

2.5.3. Valuation stage

The valuation of non-health outcomes or process attributes faces additional challenges. There are contradictory views about the use of conventional valuation techniques, such as TTO, SG or VAS for the valuation of non-health outcomes. On the one side, it was argued that given the amount of risk one would accept or the idea of giving up life-years for attributes such as the provision of information or other process-related aspects may appear unrealistic and could involve respondents dealing in very small risks or amounts of time (Donaldson & Shackley, 1997). Nevertheless, these techniques and others were used to measure process utility in previous studies (Brennan & Dixon, 2013). Opmeer and colleagues identified nine other preference-elicitation methods that were used in the literature to value non-health outcomes, such as ranking, recommendation to someone else, or WTP (Opmeer *et al.*, 2010). Also Donaldson and Shackley suggested using WTP as a measure to estimate such benefits (Donaldson & Shackley, 1997), whereas Ryan has proposed the use of conjoint analysis to take account of factors beyond health outcomes (Ryan, 1999). Conjoint analysis is based on the premise that any good or service can be described by its characteristics, also referred to as attributes, and the extent to which an individual values a good or service depends on the levels of these characteristics (Lancaster, 1966). Types on conjoint analysis include ranking, rating, and discrete-choice experiments (DCEs). In particular, DCEs are gaining attention in the health economics literature. While initially designed to measure non-health outcomes (Ryan, 1999), DCEs are now being used in the valuation of health outcomes. Examples can be found in a number of countries, where DCEs were used to produce QALY weights for the EQ-5D-5L, such as Australia (Viney *et al.*, 2014) and Canada (Bansback *et al.*, 2012). However, the use of DCEs within the QALY framework is only possible if a 'life year' attribute is added to the DCE task (Bansback *et al.*, 2012),

which would provide values anchored on the dead-full health utility scale. Another anchoring approach was proposed in a study by Brazier *et al.*, which included using the state 'dead' in the DCE to rescale regression coefficients (Brazier *et al.*, 2012). Rowen and colleagues compared a number of other methods for converting DCE values onto the dead-full health QALY scale (Rowen *et al.*, 2015). Although some issues for consideration remain in using DCEs to estimate utility values (Flynn, 2010a), the DCE can be considered as a promising technique to value non-health outcomes within the QALY framework.

Some have argued that the DCE tasks are less cognitively challenging than the conventional elicitation techniques (Bansback *et al.*, 2012), although it can be still complex for certain population groups (Ratcliffe *et al.*, 2011). Best-worst scaling (BWS) has been proposed as an alternative to DCE tasks (Flynn *et al.*, 2007). Rather than making choices between states, this method asks participants to choose the best and the worst attribute (dimension) defining the (health) state (Coast *et al.*, 2008b). Different types of BWS exist (Flynn, 2010b), which have recently been applied to measure process-related utility of treatments for Parkinson disease (Weernink *et al.*, 2016). BWS techniques are particularly useful if there is an interest in knowing the absolute impact (utilities) of attributes, such as testing the hypotheses that 'waiting time is more important than continuity of care' (Flynn *et al.*, 2007).

When using traditional techniques such as SG or TTO, the valuation of broader benefits requires additional consideration of the duration of the health state in the valuation task. Whereas, usually, valuation measures were used to value *chronic* health states, within the context of non-health outcomes, health states are *temporarily* and not chronic. Although Torrance originally developed a temporary health state approach based on TTO (Torrance, 1986), others have argued that this method is not useable in a study of process utility, for which outcomes may be in terms of hours or days (Swan *et al.*, 2003). A review of methods for measuring temporary health states for cost-utility analyses identified numerous other methods (Wright *et al.*, 2009). A temporary health state was defined as a transient health state lasting < 1 year that may impact some discomfort or

temporary reduction in QoL. The methods identified by the study included: TTO with specified duration of the health state, TTO with a lifespan modification, waiting trade-off, chained approaches for TTO and SG, and sleep trade-off. Wright and colleagues concluded that advantages and disadvantages vary by method and that no gold standard method merged from their analysis.

Besides the different techniques that can be applied to value important non-health outcomes and process attributes, one important question still remains in terms whose values should be considered in the valuation process – those of the general population, health professionals, or patients. The ongoing discussion within health outcomes research is not different to non-health outcomes. It appears that as long as national HTA guidelines favour societal values in economic evaluation, it is more appropriate to consider the general population the primary source of values in this process.

Finally, the valuation of a new measure of broader benefits could also be conducted using statistical mapping techniques. Mapping has been proposed as a technique to transfer a non-preference based measure into utility data, where an algorithm is estimated between the non-preference based measure and a target measure that provides utility, such as the EQ-5D-5L (Chuang & Whitehead, 2012). However, since the mapping function relies on statistical association, mapping would not be appropriate if the two measures have no conceptual overlap (Brazier & Tsuchiya, 2015), which is likely to be the case between a health measure and a non-health measure.

2.6. Practical challenges

The HTA process contains two stages, the assessment stage and the appraisal stage. In the assessment stage, clinical and cost-effectiveness evidence is gathered based on expert evaluation of the quality and meaning of scientific evidence submitted by manufacturers or other means, such as a systematic review of literature and advice from clinical experts. The appraisal stage derives recommendations about which intervention(s) ought to be recommended, for whom, and under what circumstance

(Morgan *et al.*, 2006). Although the focus of this thesis is on the incorporation of non-health outcomes within the assessment stage, outcomes beyond health can also be explored separately in the appraisal stage of health care decision making (Goebbels *et al.*, 2012). Where decisions need to be made that are based on more than one criterion, the concept of multi-criteria decision analysis (MCDA) has been introduced, which refers to a set of methods and approaches to aid decision-making. MCDA provides a systematic process for clarifying what is being taken into account (criteria), how each of these criteria is to be measured, and how much importance (weight) to put on each criterion (Devlin & Sussex, 2011). Rather than incorporating non-health outcomes into economic evaluation, non-health outcomes could be added to a MCDA exercise, which would provide decision makers with important additional information in the appraisal stage.

The extent to which the information on non-health outcomes would enter the decision-making process under the MCDA framework is unclear. Additional criteria for the appraisal stage already exist but have been used infrequently. Within the UK context, the 'social value judgments' proposed by NICE describe that for an incremental cost-effectiveness ratio (ICER) > £20,000 per QALY gained, other criteria will be considered, such as (i) the degree of certainty around the ICER, (ii) whether the assessment of the change in QoL has been adequately captured, and (iii) whether the intervention is an innovation (NICE, 2008). The need for accurate information of QoL benefits is an indication that broader benefits, which are not necessarily captured by existing measures, are already taken into account during the appraisal stage – yet, empirical evidence indicates the opposite. As one assessment of the impact of NICE's criteria on its decision-making practice has shown, cost-effectiveness is the principle determinant of most NICE decisions, predicting 82% of decisions, and the probability of rejection increases significantly with an increasing ICER (Dakin *et al.*, 2013). It follows that the provision of such additional criteria in the appraisal stage is useful but the process to weight them against each other requires further consideration.

It seems that the incorporation of non-health benefits in the assessment stage will find a greater role in health care decision-making than in the appraisal stage, although

decision makers may find it challenging to compare different health-based QALYs and non-health based QALYs. With the introduction of the ASCOT instrument, NICE social care guidance now refer to the social-care QALY (NICE, 2016), but there is also the possibility of a capability-QALY (Cookson, 2005b), a process-adjusted QALY, or a 'Super QALY' (Buxton, 2008). In addition, Brazier and Tsuchiya have recently introduced the concept of wellbeing-adjusted life years (WELBYs) (Brazier & Tsuchiya, 2015). As the name suggests, a WELBY is similar to a QALY in that it combines quantity and QoL into a single measure but the descriptive system used to derive WELBYs focuses on wellbeing rather than health alone. Given these many different ways of weighting time spent in varying QoL states, the original idea of the QALY as a comparable measure across all clinical areas is unlikely to hold (i.e., one WELBY ≠ one QALY ≠ one Super QALY etc.). Also, while countries currently use implicit or explicit threshold values for a cost per QALY gain (Cleemput *et al.*, 2011), new threshold values would need to be determined for the different outcomes described above.

It is already known that different preference-based HRQoL measures produce different health state valuations (Brazier *et al.*, 2004; McDonough & Tosteson, 2007; Whitehurst & Bryan, 2011; Whitehurst *et al.*, 2014a). The introduction of additional non-health measures to generate QALYs will increase the complexity of health-care decision-making. The fact that NICE guidance requires the use of the EQ-5D instrument, as the measurement of outcomes for economic evaluation, provides an important example for this discussion (NICE, 2013). It follows that in health care decision-making simplicity and consistency seem to play an important role. Pragmatism is a further central feature in health care decision-making, as can be seen by the current use of QALYs. Despite criticism of the QALY, its strengths have led to its widespread adoption. It is often argued that because of the lack of a better measure as an alternative, the QALY is an indispensable tool (Johnson, 2009) and should be used as a reference method in order to make comparisons across diseases and interventions (Drummond *et al.*, 2009). Although some improvements to QALY measurement have been proposed, these ideas resulted only in journal publications and were not implemented by decision makers (Johnson, 2009). Evidence also indicates that decision makers already find the current concepts

behind QALYs difficult to understand and show limited knowledge about formal methodology (Coast, 2004). It can be assumed that the consideration of non-health benefits in economic evaluation will add another layer of complexity that may face decision makers' rejection. Nevertheless, as shown by a recent qualitative study with 22 experts in the field of health economics and/or public health, the importance of non-health outcomes and the need to measure them are widely accepted (van Mastrigt *et al.*, 2015). While different methods are applied across countries, including the use of different perspectives, countries that adopt a narrow health care sector perspective may encounter more challenges in going beyond health than countries that already adopt a societal perspective.

2.7. Conclusion

This chapter provided an overview of theoretical, normative, conceptual, methodological, and practical challenges in broadening the evaluative space of the QALY in economic evaluation. It can be concluded that from a theoretical point of view, non-health outcomes can be incorporated into the QALY framework if a broader extra-welfarist view is adopted. There are also strong arguments for including such broader benefits, as this approach may provide a better representation of what constitutes 'value' from the deployment of scarce health care resources. A number of different approaches have been proposed for the measurement and valuation of broader benefits, and with the development of the ICECAP and ASCOT instruments, this topic has gained attention among decision makers. Although certain practical challenges remain, the fact that some HTA bodies (e.g., NICE and College van Zorgverzekeringen (CVZ)) now embed the ICECAP and ASCOT in their guidelines is a big step forward (NICE, 2013; Zorginstituut Nederland, 2016). A major challenge remains on the conceptual level. Given the existing inconsistencies in the literature in terms of what constitutes a health outcome, a non-health outcome, or a process, more clarification is needed. The next chapter will focus on this issue and will conceptualize benefits beyond the health-related QALY.

Chapter 3. Conceptualizing benefits beyond the health-related QALY – a critical interpretive synthesis

3.1. Introduction

Chapter 2 showed that one of the main challenges for the topic of interest remains at the conceptual level (i.e., what is understood as a non-health benefit). To shed more light on benefits beyond the *health-related* QALY, understanding the underlying idea of the 'QALY approach' or the 'QALY framework' in the health economics literature is crucial. It is noticeable that there are at least three interpretations of the QALY: (i) the QALY as a measure of utility that captures preferences over health states multiplied by a person's life expectancy; (ii) the QALY as a measure of health functionings over length of life, using measures such as the EQ-5D-3L; and (iii) the QALY that adjusts length of life by indices of QoL, referring to the 'Q' in the QALY. Although QoL is so broad that it could, in principle, incorporate a range of outcomes, the QALY currently has a strong focus on health, which implies that the assessment of benefits of health care interventions is either judged in terms of changes in health status or extension of life.

The terms used to describe the Q in the QALY can be vague in the sense that QoL, HRQoL and health have been used interchangeably in the health economics literature. Karimi and Brazier recently provided an overview of these terms (Karimi & Brazier, 2016). The authors state that while it is easiest to distinguish between health and QoL, the term HRQoL is more complex. The definitions of HRQoL identified by Karimi and Brazier resemble either health status or QoL and included: (i) HRQoL as functioning (an individual's ability to carry out activities) and wellbeing (an individual's subjective feelings); (ii) HRQoL as the health aspect of QoL, where non-health aspects such as economic circumstances are not included in HRQoL; (iii) HRQoL referring to aspects of QoL that can be affected by health (i.e., aspects affected by the presence of disease and

treatment); and (iv) HRQoL referring to the value of health and the corresponding values assigned to different health states (Karimi & Brazier, 2016). These varied definitions and their interchangeable use, have, to some degree, led to the different interpretations of the QALY framework.

Given that many interpretations of the QALY currently exist, it can be also expected that many diverse understandings will be identified of so-called 'non-health benefits', 'broader benefits', 'benefits beyond health', or any other nomenclature. While some descriptions of non-health benefits may have resulted from a misinterpretation of the QALY framework in the first instance, it is crucial to understand which broader benefits are thought to be outside the QALY. To gain a better understanding of such non-health benefits, an examination of how previous literature has described such broader benefits is warranted. While recent discussions have contributed to a better understanding of what is included in the QALY, benefits that are currently not captured by the QALY have received less attention. In light of recent calls to go beyond the *health-related* QALY (Brazier & Tsuchiya, 2015; Payne & Thompson, 2013), a comprehensive overview of benefits that currently do not fit within the QALY framework is warranted. The objectives of this chapter are to critically review and synthesize existing literature regarding benefits beyond health within the context of QALYs, and to develop a concept map that shows the relationship between different aspects of non-health benefits.

3.2. Methods

3.2.1. Critical Interpretive Synthesis

A critical interpretive synthesis (CIS) was undertaken. CIS is similar to meta-ethnography (i.e., systematic analysis and synthesis of qualitative research), with the difference that the synthesis is not of qualitative studies but of methodologically diverse literature (Dixon-Woods *et al.*, 2006). CIS builds interpretation by, firstly, identifying interpretations offered by the original studies and, secondly, enabling the development of new interpretations that go beyond those offered in primary studies (Campbell *et al.*,

2011). Dixon-Woods *et al.* also described these interpretations as *synthetic constructs* (i.e., broad theoretical categories), which are the result of a transformation of the underlying evidence into a new conceptual form (Dixon-Woods *et al.*, 2006). Table 3.1 provides a comparison of key characteristics of a systematic review and CIS. The term 'critical' in CIS relates to the fact that judgments need to be made about how the evidence contributes to the development of the new concept. CIS is distinguished from critical appraisal in that CIS treats literature as an object of inquiry. This critical scrutiny of the literature applies to the entire process (i.e., sampling, selection of publications, and data synthesis). The final output of a CIS is the *synthesizing argument*, which integrates the evidence from across the studies into a coherent thematic framework. It can consist of synthetic constructs but also original interpretations reported in the literature. The process of CIS is characterized by its iterative, interactive, dynamic, and recursive nature, where often searching, sampling, critique and analysis go hand in hand.

3.2.2. Literature Search

The literature search was undertaken in Web of Science in October 2015, using the 'citation pearl growing' method (Hartley *et al.*, 1990). Firstly, key 'pearls' or key papers were identified that were relevant for the topic of interest. These key papers, which were known to the doctoral candidate and supervisory committee, were used to search for papers that cited the key papers in order to retrieve the first 'wave of pearls'. After screening these new papers for eligibility (see section 3.2.3), papers that were included were used to generate the second wave of pearls, i.e., identifying papers that have cited the wave one inclusions. These steps were repeated and after each wave of pearls a test of saturation was applied that determined the point at which to stop searching for new literature. The saturation point was determined by answering after each wave of searching whether the literature retrieved adds anything new to the understanding, and whether further searching is likely to add new knowledge (further details are provided in the next section). The final step comprised a reference list search of included papers. The process of the citation pearl growing method is illustrated in Figure 3.1.

Table 3.1. Comparison of key characteristics of a systematic review and CIS [a]

	Systematic review [b]	**Critical interpretive synthesis**
Purpose	• To systematically summarize data	• To develop an understanding of topics, concepts and theories grounded in evidence
Process	• Structured process of searching, appraising, and synthesizing findings	• Iterative, interactive, dynamic and recursive • Searching, sampling, and critique happen concurrently • There is a need for constant reflexivity
Research question	• Defined *a priori* and does not change	• Developed at the outset but can be refined
Searching	• Explicit and exhaustive search strategies in bibliographic databases	• Can include bibliographic searches but also purposive selection of relevant literature from various sources • Sampling until theoretical saturation is reached
Selection	• *A priori* defined inclusion and exclusion criteria aiming for comprehensive identification and inclusion of all relevant literature	• Selective, purposive and informed by the emerging theoretical framework
Quality appraisal	• Quality checklists applied to all articles included	• Quality judged in terms of relevance and rigour
Synthesis	• Data are extracted in a systematic way and synthesized quantitatively (e.g., meta-analysis) or qualitatively based on findings from studies included	• Data extraction useful but not essential; the aim is to develop a *synthesizing argument*, linking existing constructs from the findings to *synthetic constructs* (new constructs developed during synthesis)
Reproducibility	• Results need to be reproducible	• Authorial voice is noticeable and results may not be reproducible

[a] This table is a summary of previous comparisons (Dixon-Woods *et al.*, 2006; Entwistle *et al.*, 2012; Wilson *et al.*, 2014).
[b] Most but not necessary all points typically apply to a systematic review.

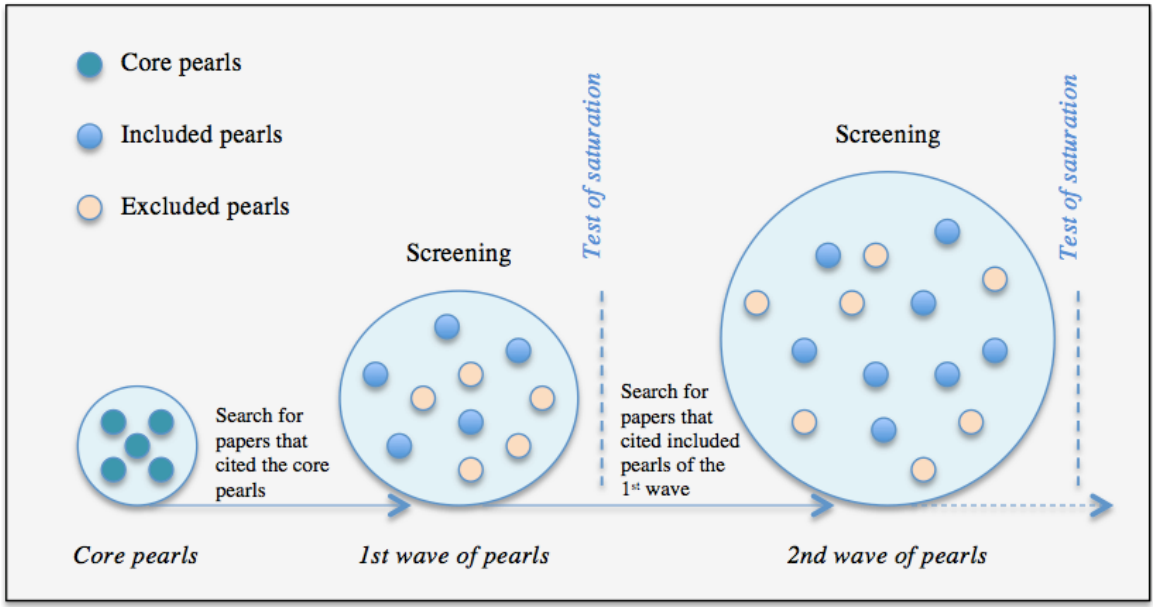


Figure 3.1. Graphical illustration of the citation pearl growing method

The idea of applying a citation pearl growing method is to find papers similar to the key set of papers. Compared with a conventional database search, a citation search can identify publications that would not be found via standard database searches because there is no constraint by the vocabulary of a search strategy. After each wave, full-texts of all identified articles were screened to determine eligibility given that titles and abstracts were not sufficient to make this judgment. Full-texts were appraised in a systematic way by the doctoral candidate only. In a first step, the paragraph was reviewed where the reference that identified the paper was cited. In a second step, a search was undertaken for keywords, such as 'non-health', 'broader', 'beyond', 'QALY', 'cost-utility', and 'utility'. The selection of papers was driven a by a list of pre-specified selection criteria.

3.2.3. Selection criteria

The selection process in a CIS can be generally described as subjective and is typically informed by the emerging theoretical framework (Dixon-Woods *et al.*, 2006). In order to be included in the analysis, papers had to be journal articles written in English.

An additional inclusion criterion was the explicit demonstration of a link between the non-health benefit and the QALY framework. This included papers that used different terminology, such as 'non-health outcomes', or discussed any kinds of benefits that currently lie outside the QALY framework (according to the author(s) of the respective study). Given that QALY weights are often derived from standardized preference-based measures that have been found to lack sensitivity for certain disease areas or population groups (Brazier & Tsuchiya, 2010), papers that focused on the development of a specific preference-based instrument (e.g., the development of an asthma-specific measure) were excluded. The rationale for this was the focus on generic measures and the identification of broad constructs of non-health benefits that are outside the QALY framework. Given that the aim of a CIS is not to identify *all* relevant references but literature that contributes something *new* to the topic of study, papers were also excluded if they did not provide further understanding beyond the general critique, or if they cited already included studies. For example, if a paper discussed a previously identified non-health benefit without providing a new/alternative definition, or further descriptive attributes that would offer more clarity, it did not contribute to a better knowledge of the topic of study and was not included. There were no constraints on publication dates.

Different types of evidence were eligible to be included in the CIS (empirical, qualitative, commentaries, etc.). Papers were not judged by their quality in view of the fact that methodologically weak papers may still prove conceptually insightful. It was more important to judge quality in terms of relevance and rigor in how the paper subsequently informed the test of saturation. Relevance indicates whether the paper addresses the concept of interest, while rigor refers to whether the paper has sufficient weight to make a contribution. Quality judgement in terms of relevance and rigor has been generally suggested for qualitative reviews (Pawson *et al.*, 2005). The selection process was conducted with the support using EndNote X6.

3.2.4. Data extraction and synthesis

Multiple approaches were used for the data extraction process. Firstly, key characteristics of the included papers were extracted in a standardized table (see

Appendix 3.1). Furthermore, a second table documented all non-health benefits offered by the original studies. While non-health benefits were being added to this table throughout the identification phase, a thematic analysis was applied at the same time, identifying constructs of similar non-health benefits. This analytic step generated the synthetic constructs, which is the result of a transformation of the underlying evidence into a new conceptual form. Constructs were not defined *a priori* but were specified in an iterative approach, i.e., they were modified in response to search results and findings from retrieved studies. Initial constructs were developed by the doctoral candidate and verified by the senior supervisor (DGTW). These constructs were then organized into higher-level themes, capturing the broad phenomena described.

In a last step, a synthesis argument was developed that integrated evidence from studies and synthetic constructs into a coherent and explanatory framework. In this way, contradictions in the evidence when producing the synthesis argument were considered. Since the synthesis argument describes the complete phenomenon under review, while considering the synthetic constructs as well as the relationships between them, this step subsequently informed the organization of the concept map. The concept map displays the relationships between concepts and the potential overlap between concepts. A concept map does not provide causal relationships but rather a better understanding of a phenomenon (i.e., interpretative approach). In the following section, the results of the CIS are presented by themes derived from the thematic framework analysis. Constructs within a theme are then discussed and presented in a concept map. Attributes, derived from existing literature, that describe each constructs are presented in Appendix 3.2 and purposively reported in the results section (an appropriate method of the CIS).

3.3. Results

The CIS of the literature was based on 109 papers. Figure 3.2 shows the flow diagram of the literature search. Firstly, 21 core references were identified from the authors' own collections and were included in the review (see Appendix 3.1). The first wave of searching that cited these 21 core references retrieved 274 unique references and

47 references were included after full-text screening. These 47 references were cited by 523 additional unique references, of which 31 were included. Although these 31 papers were relevant and were included in the review, they did not yield new constructs but rather provided further understanding of existing constructs. Since it was not anticipated that further searching is likely to add new knowledge, the rule of saturation was applied after the 2nd wave of search.

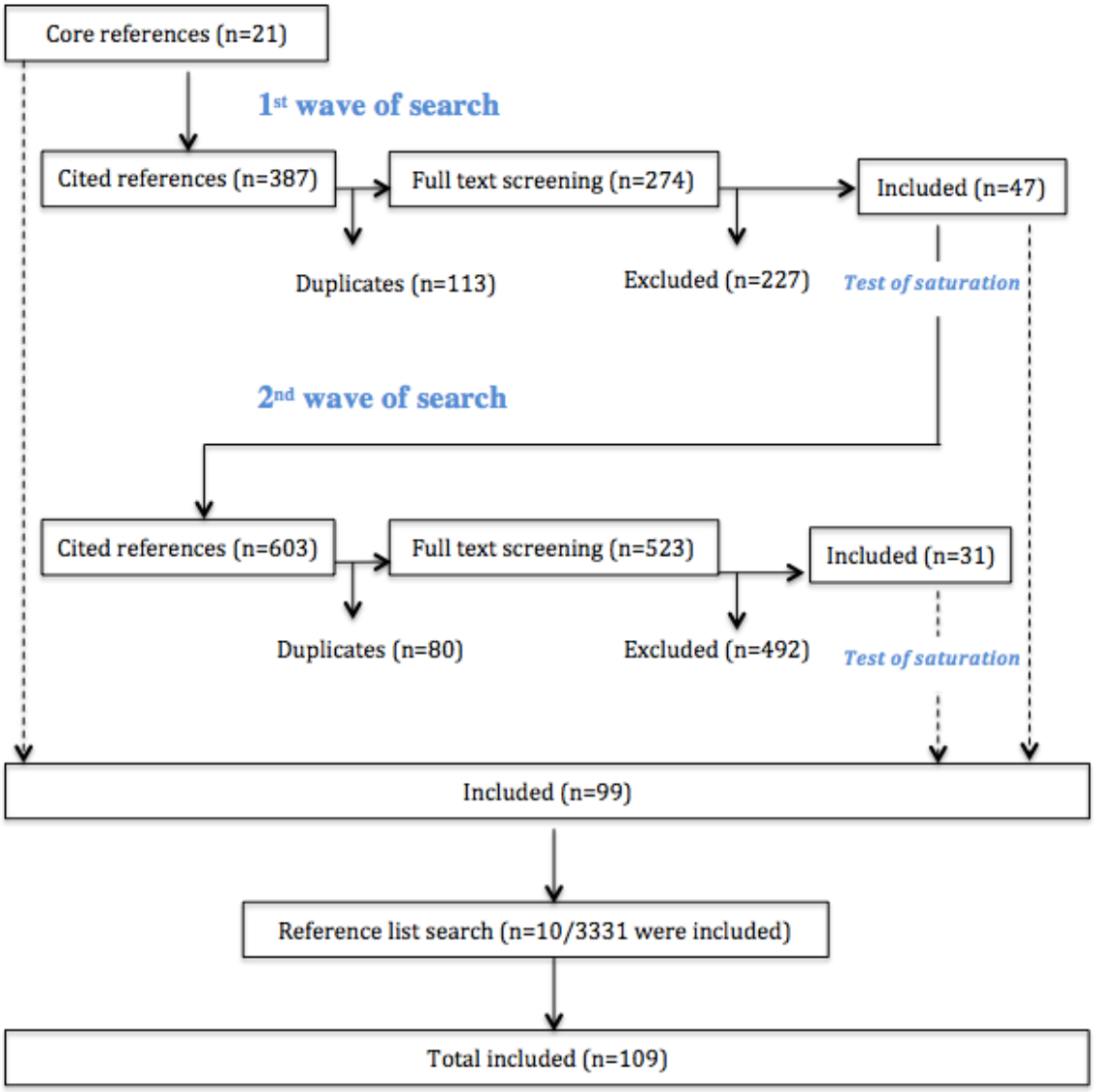


Figure 3.2. Flow diagram of the literature search

In the final step, the reference list search of 99 papers retrieved additional 10 references that were included in the analysis. These 10 additional references were identified through a process of title and abstract screening, followed by a full-text review. Characteristics of all included studies are provided in Appendix 3.1.

The synthetic constructs were based around four themes: (i) benefits affecting wellbeing (subjective wellbeing, psychological wellbeing, capability wellbeing, and empowerment); (ii) benefits derived from the process of health care delivery (process utility); (iii) benefits beyond the affected individual (e.g., spillover effects, externalities, option value, and distributional benefits); and (iv) benefits outside the health care sector. Table 3.2 provides an overview of the main constructs as well as the higher-level themes. Additionally, Appendix 3.2 provides a list of all benefits beyond the health-related QALY identified from the literature and applied to the thematic framework. What follows is a detailed discussion of the results by theme.

Table 3.2. Thematic framework analysis of benefits beyond the health-related QALY

Theme	Construct
Wellbeing	Subjective wellbeing
	Psychological wellbeing
	Capability (wellbeing)
	Empowerment
Process utility	Intervention characteristics
	Provider characteristics
	Structural characteristics
	Process characteristics
Benefits beyond the affected individual	Spillover effects
	Externalities
	Option value
	Distributional benefits (equity & need)
Benefits outside the health care sector	Other sectors

3.3.1. Benefits affecting wellbeing

Benefits that extend beyond the narrowly defined health-related evaluative space of the QALY were often linked to the concept of wellbeing or a general definition of QoL. Makai and colleagues proposed two concepts of wellbeing (Figure 3.3) (Makai *et al.*, 2014b). Firstly, wellbeing can be interpreted as an inherently subjective concept, which implies that it does not contain health dimensions. The overarching concept of QoL then constitutes functional HRQoL dimensions and subjective wellbeing. This is illustrated in Figure 3.3 in the left part, labeled 'A'. The second concept of wellbeing (see Figure 3.3, labeled 'B') treats wellbeing as representing individuals' welfare, which is dependent on individuals' functioning, thus encompassing HRQoL dimensions.

While Makai and colleagues have offered diverse interpretations of the concept of wellbeing, the thematic analysis retrieved four general categories of benefits that affect a person's wellbeing: (i) subjective wellbeing, (ii) psychological wellbeing, (iii) capability wellbeing, and (iv) empowerment (see Table 3.2).

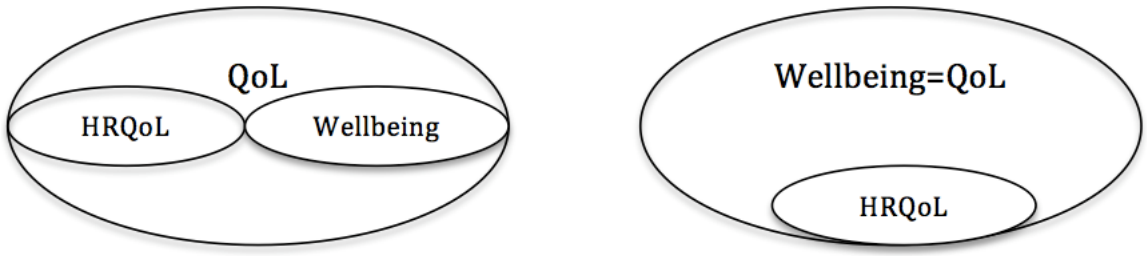


Figure 3.3. Two concepts of wellbeing based on Makai *et al.* (2014b)

Subjective wellbeing is associated with positive affects or negative affects (i.e., basic experiences of the ongoing events in people's lives), satisfaction with life, and domain satisfaction. For example, Van Mastrigt *et al.* identified emotions (i.e., a sense of satisfaction or enjoyment) as a non-health outcome relevant for economic evaluations of

public health interventions (van Mastrigt *et al.*, 2015). Negative feelings were also considered in the literature, such as regret or disappointment (Salkeld, 1998; Salkeld *et al.*, 2004), and burden (i.e., not feeling oneself to be a burden on others) (Coast, 2014; Coast *et al.*, 2008a). The second component of subjective wellbeing identified from the literature referred to (life) satisfaction, which can be understood as the global judgment about the quality of a person's life (Schulz *et al.*, 2014; Stanczyk *et al.*, 2014). This was usually discussed within the broader concept of QoL, general wellbeing, and view of life (Benning *et al.*, 2015; Davis *et al.*, 2015; Gandjour, 2001; Hoefman *et al.*, 2013; Thorn *et al.*, 2014). Lastly, domain satisfaction reflects a person's evaluation of the specific domains in his or her life, such as satisfaction with leisure and work activity (Connell *et al.*, 2014; Cookson, 2005b), usual or outside activities (Al-Janabi *et al.*, 2011; Davis *et al.*, 2013), and creativity and play (Benning *et al.*, 2015; Makai *et al.*, 2014b). Activities relating to family (Cookson, 2005b; van Mastrigt *et al.*, 2015) and social interactions/relationships formed two other domains in the literature (Al-Janabi *et al.*, 2011; Chisholm *et al.*, 1997; Coast *et al.*, 2008a; Davis *et al.*, 2013; Goebbels *et al.*, 2012; Hausman, 2012; Lorgelly *et al.*, 2010; Makai *et al.*, 2014b; Netten *et al.*, 2012; van Mastrigt *et al.*, 2015). These activities were closely related to the concept of (social) support (i.e., feeling supported) (Benning *et al.*, 2015). A number of other domains were discussed, including security and safety (Ali & Ronaldson, 2012; Makai *et al.*, 2014b), environment/community/neighbourhood (Makai *et al.*, 2014b; Ong *et al.*, 2009; van Mastrigt *et al.*, 2015), and activities relating to religion or spirituality (Coast, 2014; Coast *et al.*, 2008a; Cookson, 2005b; Ong *et al.*, 2009; Round, 2012). Further examples for the concept of subjective wellbeing can be obtained from Appendix 3.2.

While subjective wellbeing refers to hedonic wellbeing in terms of pleasure, enjoyment and satisfaction, a second concept of wellbeing was identified, referring to psychological wellbeing. Psychological wellbeing, or eudaimonic wellbeing, includes aspects of flourishing, where wellbeing increases when an individual more closely fulfills their nature as a human being (i.e., he or she flourishes). Psychological wellbeing includes benefits such as autonomy (i.e., having one's autonomy respected) (Mooney, 1998; Ryan & Shackley, 1995), freedom (Makai *et al.*, 2014b), independence (Chisholm

et al., 1997; Makai *et al.*, 2014b; Makai *et al.*, 2013), and control (i.e., perceived life control, and the ability to keep things in perspective) (Benning *et al.*, 2015; van Mastrigt *et al.*, 2015). Self-perception, such as self-confidence (Benning *et al.*, 2015; Borghi & Jan, 2008; van Mastrigt *et al.*, 2015), self-esteem (Borghi & Jan, 2008; Cookson, 2005b; Hausman, 2012), self-efficacy (Borghi & Jan, 2008), and self-respect (Coast *et al.*, 2008a) were also discussed in the literature, which all relate to psychological wellbeing. Other examples categorized under psychological wellbeing included goals (i.e., to set more and realistic goals) (Goebbels *et al.*, 2012), hope (i.e., having goals and aspirations, and being involved in activities that were fulfilling and had meaning and purpose) (Connell *et al.*, 2014; Salkeld, 1998), and the ability to cope (Payne & Thompson, 2013).

Another construct of wellbeing refers to individuals' capabilities. The capability approach was introduced by Amartya Sen who argued that capabilities (things that people are free to do or be) should be included in the overall assessment of a person's wellbeing and not only outcomes (functional utilities) (Simon *et al.*, 2013). Here, the focus is on choice and control, and to measure what people can do, rather than what they actually do (Netten *et al.*, 2012). Capabilities may be seen as a conceptualization of wellbeing in a broader sense, where capability wellbeing captures a variety of health and non-health outcomes (Al-Janabi *et al.*, 2012; Makai *et al.*, 2013). While some of the identified studies referred to the capability approach in general (Buchanan & Wordsworth, 2015; Crosignani *et al.*, 2015; Lorgelly *et al.*, 2015; Thorn *et al.*, 2014), others have identified specific capabilities and developed instruments to measure them. Examples include the Oxford CAPabilities questionnaire-Mental Health (OxCAP-MH), for use in mental health (Simon *et al.*, 2013), and the OCAP-18, which was developed for public health (Lorgelly *et al.*, 2015). Both instruments consist of similar capabilities, which were modified to improve the application for the respective field of study. Another set of capability measures was developed by the ICECAP group, which consists of the ICECAP-A (dimensions: stability, attachment, autonomy, achievement, and enjoyment) (Al-Janabi *et al.*, 2012; Al-Janabi *et al.*, 2013b), ICECAP-O (dimensions: attachment, security, role, enjoyment, and control) (Coast *et al.*, 2008a), and ICECAP-SCM (dimensions: choice, love & affection, physical & emotional suffering, dignity, being supported, and

preparation) (Sutton & Coast, 2014). Similar capabilities were included in the capability measure for those who experience chronic pain, developed by Kinghorn *et al.,* although this is not part of the ICECAP family (Kinghorn *et al.*, 2015). Most of the instruments mentioned above use particular wording to measure a person's capabilities, such as what a person 'can do' or 'is able to do'. A slightly different approach to measuring capabilities was provided by Netten *et al.* with the development of the Adult Social Care Outcomes Toolkit (ASCOT). The ASCOT approach combines both functionings and capabilities concerning food and drink, personal care, safety, social participation and involvement, control over daily living, accommodation, cleanliness and comfort, occupation and dignity. Each attribute contains four response options. While the lower three options of each attribute reflect levels of basic functioning or needs, the highest level refers to capabilities. The authors state that "*once needs are met, it is also essential to identify capabilities: whether or not people are able to achieve their desired situation*" (Netten *et al.*, 2012, p.x).

The final construct of wellbeing identified by the thematic analysis is the concept of empowerment (McAllister *et al.*, 2012). The definition of empowerment provided by McAllister describes patients as "*self-determining agents with some control over their own health and healthcare, rather than as passive recipients of healthcare*" (McAllister *et al.*, 2012, p.2). Using a case study of clinical genetics, McAllister and colleagues developed a model that conceptualized empowerment as a multi-dimensional construct including: cognitive control, decisional control, behavioural control, emotional regulation, and hope for the future. Empowerment was also discussed by others (Alayli-Goebbels *et al.*, 2013; Coulter *et al.*, 2013; Lorgelly, 2015; Lorgelly *et al.*, 2010; Payne *et al.*, 2013; Schulz *et al.*, 2014) in relation to consciousness, self-esteem, and life skills, such as health literacy, problems solving and communication skills, stress management, and skills to cope with emotions (Alayli-Goebbels *et al.*, 2013). Closely related is the concept of decision-making (i.e., being able to make informed decisions) (Goebbels *et al.*, 2012; Payne *et al.*, 2013; Ryan *et al.*, 2014), where control (i.e., control over decision-making) (Goebbels *et al.*, 2012; Petrou & Wolstenholme, 2000), provision of information and knowledge sharing (Benning *et al.*, 2015; Borghi & Jan, 2008; Dowie,

2001; Grosse *et al.*, 2008; Mooney, 1998; Payne *et al.*, 2013; Ryan, 1999; Ryan & Shackley, 1995; Salkeld *et al.*, 2004), and health literacy (Benning *et al.*, 2015) play an important role. Finally, behavioural change (Essink-Bot *et al.*, 2003; Lorgelly, 2015; van Mastrigt *et al.*, 2015) was discussed as a non-health outcome that can be categorized under the umbrella term of empowerment. Behavioural change can be linked to awareness of health risks and overcoming addictions (Goebbels *et al.*, 2012), self-management capacities (Benning *et al.*, 2015), and effort to change lifestyle behaviour (Goebbels *et al.*, 2012).

3.3.2. Benefits derived from the process

This concept of benefits denotes that the focus should not only be of *what* is achieved in a health care system but *how* it is achieved. The narrow consequentialist view of the QALY only considers consequences and outcomes but others have argued that there are other elements in the consumer's utility function (Donaldson & Shackley, 1997). Therefore, when consuming health care services individuals may have preferences over not only health consequences but also the circumstances associated with those consequences (Birch & Donaldson, 2003). This utility or disutility derived from the actual processes can be labeled under the umbrella term 'process utility'.

It has been argued that ignoring process utility in the overall valuation of utility from health care could lead to a sub-optimal provision of health care (Higgins *et al.*, 2014), which is particularly important when treatment alternatives produce similar health outcomes but differ in terms of process-of-care factors (McNamee & Seymour, 2008). Others have noted that process utility is a term used as a synonym for 'patient experiences' and 'satisfaction' from the process (Annemans *et al.*, 2013; Buchanan & Wordsworth, 2015; Mooney, 1998; Payne *et al.*, 2013; Ryan *et al.*, 2014; Swan *et al.*, 2016; Tinelli *et al.*, 2013), which can be linked to the quality of care provided. Dirksen discussed the role of 'patient preferences' that reflect patient's evaluation of the process of health care (Dirksen, 2014). Given that the QALY framework is based on the mean public or patient values, the treatment recommended to an individual patient (e.g., through the process of shared decision making) may differ from what is being

recommended at a system-wide level. Such mean preferences can potentially override preferences of the patient or affected individual. In view of the fact that individuals' preferences are influenced by factors relating to the health care process, it is argued that incorporating patient preferences into the QALY framework will lead to improved satisfaction with the process of care and better health outcomes (Brazier *et al.*, 2009).

Opmeer and colleagues previously discussed non-health outcomes of health care interventions related to process utility, and distinguished between features of the intervention, features of the health care process, or subjective/perceived impact on patients (Opmeer *et al.*, 2010). Similar concepts were also used in the thematic analyses of this CIS to summarize aspects of process utility, which comprise: (i) intervention characteristics, (ii) provider characteristics, (iii) structural characteristics, and (iv) process characteristics. These concepts are reported in Appendix 3.2 and summarized below.

Several papers identified non-health benefits that can be derived from certain features of the health care intervention itself. Such features include, for example, the accuracy of a test (Payne *et al.*, 2013), mode and frequency of administration (Ali & Ronaldson, 2012; Brouwer, 2008), invasiveness (intensity) of the intervention (Alayli-Goebbels *et al.*, 2013; Opmeer *et al.*, 2010), and complications or morbidity related to the invasiveness, such as the pain of testing or postoperative bleeding (Swan *et al.*, 2016; Swan *et al.*, 2003; Swan *et al.*, 2010). The extent to which the intervention is tailored to the individual, group, or family has also been associated with non-health benefits (Alayli-Goebbels *et al.*, 2013; Grosse *et al.*, 2008).

Numerous structural characteristics that refer to attributes of the setting in which care is provided were discussed. Access to health care was described within the context of the location of care and distance to the health facility (including travel time) (Ali & Ronaldson, 2012; Benning *et al.*, 2015; Birch *et al.*, 2003; Dowie, 2001; Grosse *et al.*, 2008; Payne *et al.*, 2013; Petrou & Wolstenholme, 2000; Ryan, 1999; Ryan *et al.*, 2014; Shackley *et al.*, 2001; Swan *et al.*, 2016; Swan *et al.*, 2010). Furthermore, the coordination of services (Swan *et al.*, 2016), availability of non-health related services (Ali & Ronaldson, 2012), and the type of care delivery including follow-up care

(continuity of care) (Ryan, 1999; Shackley *et al.*, 2001) have been considered as characteristics of the structure that provide utility. Lastly, the treatment environment itself, such as continuity of staff (Shackley *et al.*, 2001), comfortable accommodation (Kotzian, 2009), and cleanliness of facilities (Ali & Ronaldson, 2012; Swan *et al.*, 2016) were stated as important characteristics that can lead to utility gain.

Provider characteristics, which include the attitude of the health care staff (Ryan, 1999) and the relationship to the health care provider (Coulter *et al.*, 2013), were discussed as additional aspects of process utility. This contains the quality of the relationship (Payne *et al.*, 2013), as well as the extent of physicians' involvement during the process (Alayli-Goebbels *et al.*, 2013). The nature of the consultation formed an important component of process utility, which refers to prompt and clear communication, and time to talk (Grosse *et al.*, 2008). In this context, individuals may derive benefits during the process if the provider acts in a respectful manner, demonstrates competency (Swan *et al.*, 2016), and is able to provide recommendations or advice in terms of the treatment options (Annemans *et al.*, 2013). Ryan also stated that there might be non-medical reasons to see the doctor (Ryan & Shackley, 1995), which closely relates to Borghi and Jan's argument that the demand for some health care interventions may be a derived demand not for health but, for example, for information (Borghi & Jan, 2008).

The provision of information and the 'value of knowing' were discussed as key characteristics of the process of health care delivery (Donaldson & Shackley, 1997; Ryan & Shackley, 1995; Towse & Garrison, 2013). This goes beyond the 'value of information' for decision-making and includes the value of information even when the information does not affect treatment (Neumann *et al.*, 2012). Specifically, the information produced by health care intervention was discussed (Birch *et al.*, 2003), often within the context of genetic testing (Bajaj & Veenstra, 2013; Buchanan *et al.*, 2013; Eden *et al.*, 2013; Grosse *et al.*, 2009; Lu & Cohen, 2015). Bajaj *et al.* listed some consequences of knowing, such as anxiety (or anxiety relief) or the implications of test results for other family members (Bajaj & Veenstra, 2013). Grosse suggested that the utility of genomic information could be considered from three perspectives: the public

health perspective (i.e., health improvements), the clinical perspective (i.e., diagnostic thinking and therapeutic choice), and the personal perspective, where information can have a value *per se* (Grosse *et al.*, 2009). Veenstra and colleagues used the term 'health-related utility' that combines the attributes of clinical utility and social utility that are associated with influences on clinical outcomes or QoL (Veenstra *et al.*, 2010). Buchanan *et al.*, however, distinguished between 'personal utility' and 'clinical utility', where the value of personal genetic information may not be captured by clinical utility but by personal utility, which describes benefits and harms that are manifested outside of medical contexts (Buchanan *et al.*, 2013). This could include, for example, improved certainty of knowing and, hence, individuals' sense of control, self-identity and autonomy (Buchanan *et al.*, 2013). Several studies stated that information is important for the reassurance it can provide (Buchanan *et al.*, 2013; Cookson, 2005b; Donaldson & Shackley, 1997; Grosse *et al.*, 2008; Howard *et al.*, 2008; Ryan, 1999; Ryan & Shackley, 1995; Salkeld *et al.*, 2004; Swan *et al.*, 2003).

The concept of process utility was also used as a collective term for all activities that occur during health care delivery (Dirksen, 2014). This includes whether or not a patient is treated with dignity (Al-Janabi *et al.*, 2011; Ali & Ronaldson, 2012; Coast, 2014; Coast *et al.*, 2008a; Coast *et al.*, 2008d; Cookson, 2005b; Donaldson & Shackley, 1997; Makai *et al.*, 2014b; Makai *et al.*, 2013; Mooney, 1998; Ryan, 1999; Ryan & Shackley, 1995; Schlander *et al.*, 2014; Torgerson & Raftery, 1999), experiences fear/anxiety or uncertainty (Borghi & Jan, 2008; Opmeer *et al.*, 2010; Salkeld, 1998; Swan *et al.*, 2010), comfort/discomfort (Al-Janabi *et al.*, 2012; Cookson, 2005b; Makai *et al.*, 2014b; Makai *et al.*, 2013; Opmeer *et al.*, 2010), stress or convenience/inconvenience (Ding *et al.*, 2011; Higgins *et al.*, 2014), being treated in a pleasant/unpleasant way (Brennan & Dixon, 2013; Brouwer *et al.*, 2008; Donaldson & Shackley, 1997), and embarrassment or cosmetic issues (Borghi & Jan, 2008; Opmeer *et al.*, 2010; Swan *et al.*, 2016). Privacy (Ryan *et al.*, 2014), duration of treatment, and other time-related aspects of the process of care (e.g., time to talk, waiting time, time to complete a test and knowing the results, length of stay, and convenience of appointment times) were also associated with additional benefits (Dowie, 2001; Grosse *et al.*, 2008; Opmeer *et al.*,

2010; Payne *et al.*, 2013; Ryan, 1999; Ryan *et al.*, 2014; Shackley *et al.*, 2001; Swan *et al.*, 2016; Swan *et al.*, 2010). The importance of choice (e.g., choice of medical staff, service level, and type of provider) and the act of choosing may also lead to an additional effect on utility (Ali & Ronaldson, 2012; Kotzian, 2009; Ryan & Shackley, 1995). Brouwer *et al.* discussed the choice of becoming a carer within the context of informal care and the associated burden of care (Brouwer *et al.*, 2005). In general, care (i.e., being cared for, receiving care, and providing care) was mentioned as an important characteristic of the process from which individuals may derive utility (Brouwer *et al.*, 2005; Davidson & Levin, 2010; Mooney, 1998). Finally, end-of-life aspects, such as being prepared for death and the quality of dying (i.e., emotional and physical suffering) were highlighted as important process characteristics (van Mastrigt *et al.*, 2015).

3.3.3. Benefits beyond the affected individual

A number of benefits were discussed in the literature that occur beyond the affected individual, which refers to the 'evaluative scope' of the QALY (i.e., who is affected by the intervention and who should be included in the analysis). The current evaluative scope of the QALY contains only patient QALYs and ignores QALY gains in significant others (Al-Janabi *et al.*, 2011). Significant others may include family members and friends (Ali & Ronaldson, 2012; Brouwer *et al.*, 2006; Coast, 2014) or informal caregivers (Al-Janabi *et al.*, 2011; Hoefman *et al.*, 2013). Such spillover effects refer either to the 'family effect', where family members may experience health changes because someone in their social environment is ill (effects of caring *about* other people) or 'caregiver effect', which is the effect on health induced by the burden of informal care (effects of caring *for* someone who is ill) (Bobinac *et al.*, 2010; Bobinac *et al.*, 2011; Brouwer, 2008). Providing informal care is, however, not necessarily associated with negative effects, as there are also positive effects, such as feelings of being appreciated by the cared-for patient or spending time together (Davidson & Levin, 2010). Brouwer and colleagues referred to this as the process utility from providing informal care (Brouwer *et al.*, 2005). A few studies noted that the benefits and harms of genetic testing

include not just the impact on the individual being tested but the health effects for family members (Eden *et al.*, 2013; Grosse *et al.*, 2008; Rogowski *et al.*, 2014).

Spillover effects can be seen as a form of externalities, where a general distinction can be made between: (i) caring externalities, which refer to the utility people derive from the knowledge that other people are receiving health care, or (ii) selfish externalities where an individual cares about others' consumption of health care because this consumption affects their own health (Donaldson & Shackley, 1997; Labelle & Hurley, 1992; Olsen & Smith, 2001; Ryan & Shackley, 1995). The former is sometimes described as altruism (Jacobsson *et al.*, 2005), which is a type of non-use value (Olsen & Smith, 2001; Sach *et al.*, 2007). Another type of non-use value refers to the concept of 'option value', which derives from uncertainty regarding future need (Labelle & Hurley, 1992). Here, option value describes the utility obtained from having the option to use health care services at some point in the future (Donaldson & Shackley, 1997; Eden *et al.*, 2013; Olsen & Smith, 2001; Ryan & Shackley, 1995). Since not only users of the services derive utility but also current non-users, who may or may not use the service in the future (Ryan & Shackley, 1995), it represents a benefit that extends beyond the affected individual. Similarly, Borghi & Jan referred to 'passive use value' within the context of health promotion programs, which refers to the indirect learning from others or reassurance from knowing that others have learned (Borghi & Jan, 2008). Although it has been previously acknowledged that benefits of a program can accrue to the broader community, such externalities lie outside QALY framework, which could lead to a non-optimal allocation of resources (Labelle & Hurley, 1992).

Generally, the three concepts described above (spillover effects, externalities and option value) refer to non-health benefits outside the affected individual. A fourth non-health benefit that is not directly relevant for the affected individual but the society overall relates to distributional benefits, such as equity and need. A number of papers noted that applying the QALY maximization rule leads to sub-optimal allocation of resources because distributional issues are ignored (Baker *et al.*, 2010; Schlander *et al.*, 2014). Under the QALY framework, the overall health benefits of an intervention are

calculated by aggregating the QALY gains accruing to individuals using a simple, unweighted summation, which is known as the 'QALY maximization' rule. Since concerns about how benefits are distributed across individuals are ignored (distributive neutrality), all QALYs are of equal social value regardless to whom they accrue and the context in which they are enjoyed ('QALY egalitarianism'). In other words, a 'QALY is a QALY is a QALY' under all circumstances and regardless of the characteristics of recipients (Singer *et al.*, 1995; Williams, 1997). Although a theoretical basis for equity weighting is provided by the extra-welfarism approach, until today no such weights are used in practice. Different approaches have been put forward to incorporate equity considerations into the QALY framework, such as distributional weights based on age, severity, lifestyle, socioeconomic status, and having dependents, etc. (Baker *et al.*, 2010). Herlitz & Horan emphasize the importance of health 'needs' (i.e., the severity of pre-treatment health states) that can be seen as an independent objective for allocation decisions, which is different but not mutually exclusive to health maximization or equity concerns (Herlitz & Horan, 2016). Allocation based on a need-based prioritization framework was also discussed for genetic tests (Rogowski & Schleidgen, 2015).

3.3.4. Benefits outside the health care sector

A number of papers discussed that the allocation of resources in health care will affect other sectors, such as education, criminal justice, environment, housing, and labour (Ali & Ronaldson, 2012; Bayoumi, 2004; Beale *et al.*, 2012; Benning *et al.*, 2015; Coast *et al.*, 2008d; Edwards *et al.*, 2015; Edwards *et al.*, 2013; Gandjour, 2001; Hausman, 2012; Laxminarayan *et al.*, 2014; Lorgelly *et al.*, 2010; Lu & Cohen, 2015; van Mastrigt *et al.*, 2015). At the same time, it has been cautioned that the inclusion of such benefits on the effect side can lead to double counting if the effect is already captured by the cost side of an economic evaluation (which has been the typical approach to capture such broader impacts) (Bayoumi, 2004). Yet, controversial views still exist around this topic in the literature. For example, decreased criminal activity may be associated with QoL effects. Therefore, treating such outcomes only as costs runs the risk of undervaluing the associated QoL benefits (Bayoumi, 2004). Particularly within the context of health

promotion, Benning and colleagues have highlighted the importance of reductions in criminal behaviour, as well as better educational achievements, increased labour participation and work productivity (Benning *et al.*, 2015).

With regard to labour and productivity gains, the Washington Panel's 1996 guidelines for economic evaluation of health care interventions initially proposed that productivity costs should be incorporated as health effects in the denominator of the cost-effectiveness ration (rather than in numerator) (Brouwer *et al.*, 1997). The rationale for this was based on the definition of productivity cost by the Washington Panel, which also accounted for the lost or impaired ability to engage in leisure activities. However, although lost or impaired ability to engage in leisure activities should be reflected in QALYs, it has been highlighted that current preference-based HRQoL instruments used to calculate QALYs do not include such benefits of survivor consumption and leisure forgone (Liljas, 2011; Nyman, 2011). Besides the lost or impaired ability to engage in leisure activities that affect QoL, it can also be argued that lost or impaired ability to work can also influence an individual's QoL. As such, particularly within a cost-utility framework, it has been cautioned to include productivity loss in the denominator of the ICER because the QALY also reflects the influence of work loss, which may lead to double counting (i.e., productivity loss being included in both the numerator and denominator of the ICER).

Despite these controversies, the Second Panel on Cost-Effectiveness in Health and Medicine now recommends that productivity loss should be considered in the cost side in an economic evaluation (Sanders *et al.*, 2016). Currently, there is no clear approach as to how to incorporate other benefits outside the health care sector, such as reductions in criminal behaviour. An alternative method was suggested to incorporate a multi-sectorial approach, or an inter-sectoral compensation test, where inter-sectoral costs and benefits are simultaneously captured and adjusted for budgets and resources that are allocated by different ministries (van Mastrigt *et al.*, 2015). However, there are no simple ways of integrating costs and benefits in other sectors within existing decision rules and the proposed methods are still at a theoretical stage of development.

3.3.5. Concept map

The concept map shown in Figure 3.4 displays the relationships between different non-health benefits identified by the CIS. The four higher-order themes discussed above were further broken down into two groups that are indicators of whether benefits accrued within or outside the affected individual. Benefits outside the affected individual comprise (i) benefits beyond the affected individual and (ii) benefits outside the health care sector. Benefits affecting wellbeing and benefits derived from the process form a second group that is concerned with the affected individual. While a distinction has been made between process and outcome, it needs to be noted that it was not always possible to make a clear cut between what constitutes an outcome and what is process-related. This is illustrated by the dashed line in the concept map.

With regard to wellbeing, the four concepts described above overlap conceptually, which is also reflected in the concept map. For example, positive relationships can be considered as a domain of psychological wellbeing, but social interactions and being satisfied with social relationships is also an important component of subjective wellbeing. Likewise, certain capabilities such as enjoyment or autonomy conceptually overlap with psychological and subjective wellbeing. The concept of empowerment was previously described as another conceptualization of Sen's capability approach (McAllister *et al.*, 2012; Payne *et al.*, 2013) and, in addition, is closely related to psychological wellbeing. Moreover, since empowerment can be seen as a process by which individuals gain control over their lives, it can be also linked to the concept of process utility.

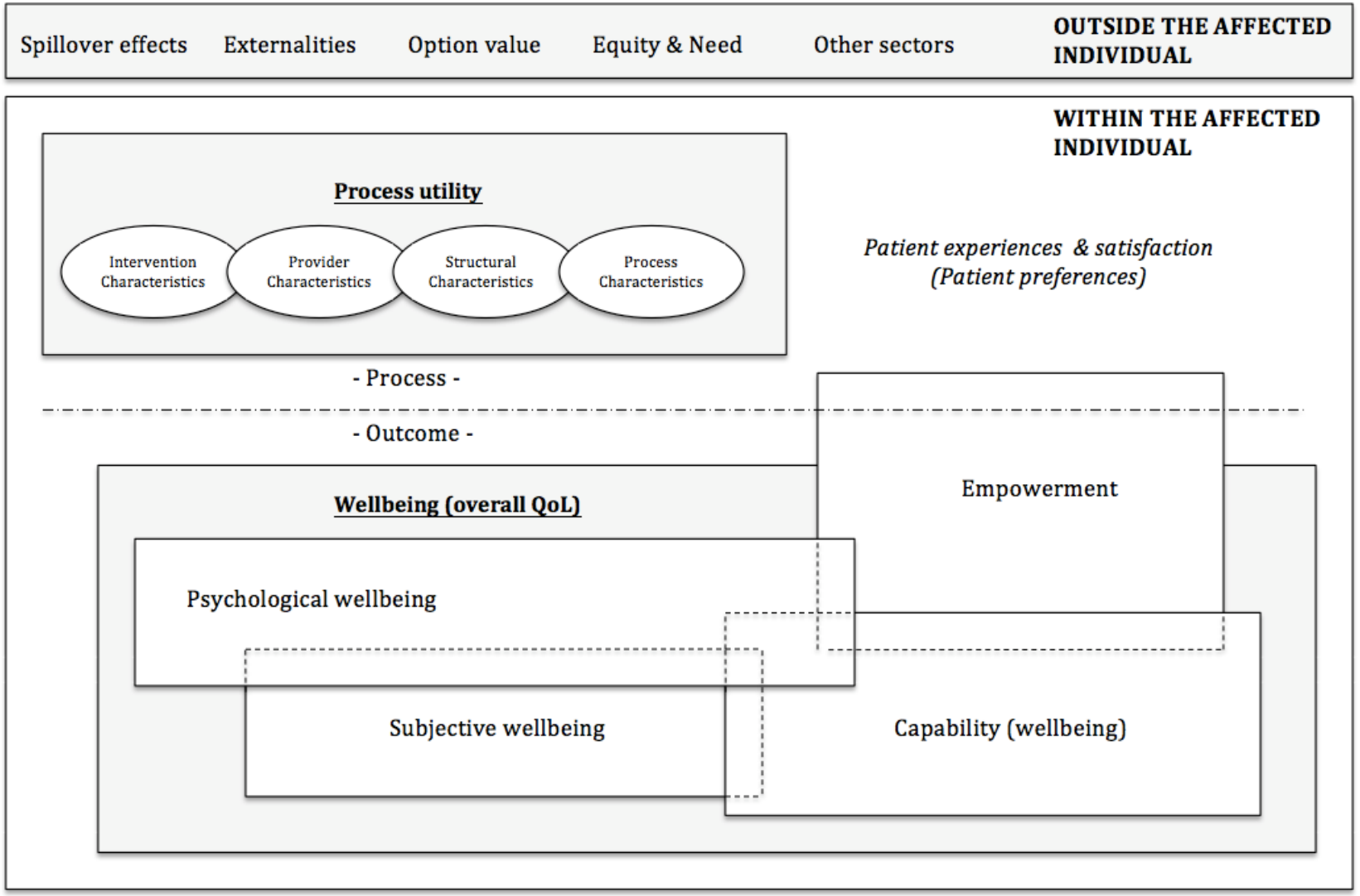


Figure 3.4. Concept map of benefits outside the *health-related* QALY based on literature identified through the CIS

3.4. Discussion

The identification of benefits of health care interventions and their consideration in health economic evaluation remains a challenging task. While QALYs have been most frequently used in the assessment of health interventions, there are increasing concerns that the QALY framework does not capture all benefits of health care interventions. The aim of this chapter was to provide a better understanding of benefits beyond the *health-related* QALY. This work critically reviewed existing evidence and identified four broad types of non-health benefits that are currently considered outside the QALY framework, which include: (i) benefits affecting wellbeing; (ii) benefits derived from the process of health care delivery; (iii) benefits beyond the affected individual; and (iv) benefits outside the health care sector. Given these findings, the future course of action concerning such non-health benefits remains unclear. Possible choices include ignoring non-health benefits, abandoning the QALY for something else, or incorporating non-health benefits into the QALY framework. These three options are discussed in the next sections.

3.4.1. Ignoring non-health benefits within the QALY framework

Given that non-health benefits have played a minor role in health care resource allocation decisions to date, it can be argued that they should be ignored within the QALY framework. However, countries that currently base their reimbursement decisions on QALY gains are recognizing the potential importance of non-health benefits in economic evaluation, as reflected in recent guidelines. With the introduction of the ASCOT and ICECAP measures in the NICE guidelines, broader measures of wellbeing are gaining significant attention in health care decision-making (NICE, 2014). In addition, NICE now also allows consideration of evidence on the 'process characteristics' of health care technologies that have a value to people independently of any direct effect on health, such as improving convenience in delivery and administration of care, and the level of information provided to patients (NICE, 2013). A similar development can be also observed in Canada, where recent guidelines proposed the incorporation of non-

health effects using a CCA to complement the health effects captured in a CUA (CADTH, 2017). An important statement made in the CADTH guidelines is that the value of non-health effects should be based on being traded-off against health. In this context, the guidelines refer to societal values that should be reflected in the values of non-health benefits. As such, to answer the question whether or not non-health benefits should be ignored, more studies are needed to explore whether society is willing to sacrifice health gains in order to achieve a range of distributional, wellbeing, and process-related objectives.

Williams noted that if, besides health, priority should be given to aspects related to 'access to health care', it is important to understand that this would imply that health care resources could be taken away from improving people's health and to devote them to reinforcing people's sense of security about care being available to them when they need it (Williams, 1996). Whether this is a reflection of what society wants from the health care system is unclear. One ethical objection to incorporating non-health benefits is if someone who is in medical need will be not treated because health care resources were spent for non-health benefits (Broome, 2002). Getting more views from citizens would help to inform this debate and enable policy makers to identify an appropriate maximand.

In search of a maximand, there is, however, possibly a tension between what patients want versus what the society and the decision maker regards as relevant. It was previously argued that non-health benefits are particularly valued by patients, whereas decision makers or tax-payers might value a more limited, health-focused set of attributes (Bryan & Dolan, 2004). Similar views have been provided by others, suggesting that compared with patients, citizens may be reluctant to fund health care that provides non-health benefits but does not affect health status (Clark & Olsen, 1994). While health is considered unique in the sense that it enables people to 'flourish', process utility does not have necessarily this feature and, thus, it has been argued that individuals will not feel an obligation to subsidize their fellow citizens' use of health care resources that are not health enhancing (Clark & Olsen, 1994). A contrary view was offered by Mooney, who questioned the role of decision makers to define what is important for a patient's utility

function, especially the legitimacy of non-health outcomes (Mooney, 1994). In view of the fact that patients may derive utility from aspects of health care other than health outcome, ignoring such broader benefits seems to be an incorrect approach to proceed with. The problem appears to be in the limited ability of the QALY to incorporate non-health benefits. As such, a better solution could be the identification of another approach that allows the incorporation of non-health benefits and to abandon the QALY for something else.

3.4.2. Abandoning the QALY

Alan Williams once said about the QALY: "*there is quite a lot I don't like about it, but it's not so bad when you consider the alternative*" (Williams, 1996, p.1803). The list of alternative approaches of resource allocation decisions in the health economics literature is long and comprises, for example, CBA, CCA, or MCDA. A recent qualitative study has offered some other alternatives that were proposed by experts, such as the use of the Happiness Index, Wellbeing Index, capability indices, a CUA using a broader QALY concept, or CUA/CEA using the multi-sectorial approach (van Mastrigt *et al.*, 2015). Nevertheless, these alternatives are also not free of problems as the authors underscore. Drawbacks of CBA include the use of individual preferences, the difficulty to express benefits in monetary terms, the lack of standardised methods, and the need for a tailored CBA for every specific intervention. The main limitation of the CCA, on the other hand, is the difficulty to determine the overall impact of the intervention, where different outcomes under consideration could potentially indicate a different course of action. Due to the lack of a standardized approach for MCDA and for the multi-sectorial approach, these two alternatives can also not be considered as superior to the QALY. Questionable is also the use of bolt-ons to derive a 'broader QALY'. Finally, given the difficulty to measure capabilities and the uncertainty about the society's acceptance of wellbeing as a metric to inform policy decisions, these alternatives need to be explored further before their application in practice.

Whether the QALY should be abandoned for something else needs also to be examined in light of the role of the QALY in the decision-making process. QALYs were

introduced as a tool to inform decision makers as explicitly as possible about the interpersonal comparisons of utility (Mooney, 1989). Importantly, QALYs do not inform decision makers what is the 'best' option in terms of overall social value. This is because apart from improving health outcomes, decision-makers usually consider other criteria in resource allocation decisions (Shah *et al.*, 2012). A frequent problem is that most decision-making agencies lack explicit and transparent frameworks for incorporating the multiple criteria considered when prioritising health technologies (Hansen, 2012). Given the absence of the relative importance of additional criteria, and a clear maximand for resource allocation decision-making, it appears that the exclusion of benefits beyond health does not appear to be problematic for the QALY metric *per se* but rather a wider issue for the decision-making.

In addition, while the QALY can be seen as a framework that informs resource allocation decisions in health care, the instruments used to derive QALYs play an important role in determining the evaluative space of the QALY. As such, it was not surprising to see that authors of identified papers discussed the limitation of the evaluative space of the QALY with a particular linkage to an instrument – in particular, the most frequently used instrument, the EQ-5D-3L. Two Dutch studies, for example, have studied the importance of non-health outcomes and defined such broader outcomes as "*all outcomes not incorporated in EQ-5D based QALYs*" (Goebbels *et al.*, 2012, p.180; van Mastrigt *et al.*, 2015). While such working definitions most likely stem from the fact that the EQ-5D is also the preferred instrument by some national HTA agencies, such as NICE in the UK (NICE, 2013) or CVZ in the Netherlands (Zorginstituut Nederland, 2016), it raises the question as to what extent this particular preference for an instrument is the cause of the problem itself i.e., driving the discussions around broader non-health benefits.

Brazier and colleagues noted that the key criticism of most generic preference-based HRQoL measures is that researchers have designed them with little or no input from people with relevant health problems (Brazier *et al.*, 2014). The development of broader instruments (e.g., ICECAP and ASCOT), on the other hand, was shaped by

findings from qualitative interviews with affected population groups. Given that, it remains unclear whether existing preference-based HRQoL measures deviate fundamentally from broader instruments or whether important aspects of the impact of impairments on people's lives have been missed due to the failure to include end users in the development phase. From a theoretical point of view, the QALY measure can embrace any HRQoL characteristics that are important to people (Williams, 1996). As Chapter 2 outlined, under the extra-welfarism approach there are no theoretical reasons to limit the QALY to health outcomes only. Yet, in practice, it seems that QALYs are currently defined by what is measured rather than the conceptual origin. It follows that although non-health benefits have been largely ignored when estimating QALYs, one should not dismiss the QALY framework but rather develop appropriate instruments that capture such broader benefits (Round, 2012). Besides this focus on instruments to broaden the evaluative space of the QALY, a number of methodological advances have been proposed in the literature that would allow the incorporation of non-health benefits into the QALY framework.

3.4.3. Incorporating non-health benefits into the QALY framework

The CIS identified four broader types on non-health benefits: (i) benefits affecting wellbeing; (ii) benefits derived from the process of health care delivery; (iii) benefits beyond the affected individual; and (iv) benefits outside the health care sector. This section challenges whether these constructs truly represent non-health benefits and discusses the extent to which they are already considered within the QALY framework or can be incorporated into the QALY framework

A number of benefits affecting wellbeing were identified, such as subjective wellbeing, psychological wellbeing, capability wellbeing, and empowerment. It has been argued that existing preference-based HRQoL measures that are used to generate QALYs do not account for wellbeing, which has resulted in the development of broader outcome measures, such as the ASCOT and the ICECAP instruments. However, it needs to be noted that several wellbeing attributes identified through the CIS can be found in existing preference-based HRQoL measures. For example, 'happiness' in the HUI or 'control',

'pleasure', and 'coping' in the AQoL-8D. This indicates that concepts of wellbeing are not excluded from the QALY framework *per se*. Instead, the absence of such attributes is largely a reflection of the dimensions that are included in the descriptive systems of preference-based HRQoL measures. Accordingly, there is the need for further investigation in terms of the extent to which aspects of wellbeing are currently captured by existing preference-based HRQoL measures. This research question is the focus of Chapter 5. It is also important to note that even if newly developed measures provide additional information beyond that of existing preference-based HRQoL measures, their use within the QALY framework is not straightforward. While the ASCOT can be used to derive 'social-care QALYs' as it is anchored on a scale of 0 'being dead' to 1 'ideal SCRQoL state' (Netten *et al.*, 2012), the ICECAP instruments do not have QALY properties, given that they were anchored on a 'full capability' to 'no capability' scale (Flynn *et al.*, 2015). This means that only the ASCOT can be used to calculate QALYs, while ICECAP measures just provide sets of preference-weighted index values. Other capability measures identified in this chapter (e.g., OxCAP-MH) are also not suitable for the use within the QALY context, and some researchers abstain from collapsing multidimensional capabilities information into a single index score because it is in tension with the original capability approach (Simon *et al.*, 2013).

Although process attributes are not directly included in existing preference-based HRQoL measures, there are still similar concerns with regard to double counting. Brennan and Dixon highlight that the extent to which process utility is not captured by existing preference-based HRQoL measures is currently unknown (Brennan & Dixon, 2013). In order to avoid double counting, this information is required before any separate estimates of process utility are added into cost-effectiveness analyses. Ryan and Shackley provided an example on this issue within the context of screening programmes (Ryan & Shackley, 1995). Here, double counting could occur because screening programmes provide the opportunity for reassurance but may provoke anxiety in patients, where anxiety is already included in many preference-based HRQoL measures (Ryan & Shackley, 1995). While further research is needed to explore the extent to which process-related aspects are captured by existing preference-based HRQoL measures, newly

developed preference-based measures already pay close attention to process-related aspects, which can be observed in the inclusion of a dignity question (i.e., the impact of the care process on the person's sense of self-worth) in the ASCOT instrument or a question regarding choice (i.e., having a say in the decision-making) in the ICECAP-SCM.

Brennan and Dixon mention that there is currently no sound methodological framework for the incorporation of 'process' into QALY frameworks and suggest the exploration of alternative methodologies, such as the use of psychometric approaches or the use of a process-related 'bolt-on' (Brennan & Dixon, 2013). However, the use of bolt-ons is again not straightforward, given that new valuation studies would be required in order to capture the relative impact of the added dimension on heath state valuations. An attempt to incorporate process preferences directly within the QALY framework was demonstrated by McNamee and Seymour through the use of TTO and SG tasks (McNamee & Seymour, 2008). Alternatively, Swan and colleagues have developed a MAUI to capture the effects of testing and screening on QoL, called the Temporary Utility Index (TUI) (Swan *et al.*, 2010). It is also possible to complement QALYs with a patient-reported experience measure (PREM) that accounts for patient satisfaction with the treatment (Olsen, 2017). Although no perfect methods exist to capture process utility within the QALY framework, these studies have demonstrated that the incorporation of process utility within the QALY framework is not impossible.

With regard to benefits beyond the affected individual, recent efforts showed that it is feasible to include family spillovers in economic evaluations within the extra-welfarism framework (Al-Janabi *et al.*, 2015; Al-Janabi *et al.*, 2016). Al-Janabi and colleagues have developed a framework that adapts the conventional cost-effectiveness decision rule to include two multiplier effects – one specified for the health benefit generated, and one for the health benefit displaced by funding the intervention (Al-Janabi *et al.*, 2015). This allows for consideration of health spillovers, as these multiplier effects express the ratio of total health effects (for patients *and* their family networks) to patient health effects. However, the authors highlight that accounting for health spillovers (or

health externalities) in resources allocation decisions may result in greater funding for services for people with dependents, at the expense of those without (Al-Janabi *et al.*, 2015). Inevitably, the ethical implications of such evaluations will require further consideration.

While other equity issues were identified in this chapter around the QALY framework, Mooney urged caution that QALYs *per se* do not create any ethical problems but they may if health is assumed the only output of health services. He stressed that, in practice, equity problems can be addressed by attaching empirical weights (Mooney, 1989). A theoretical basis for equity weighting is also provided by the extra-welfarism approach, which in fact does no longer separate between equity and efficiency concerns that would usually allow economist to take on a neutral position towards the distributional issues (Coast, 2009). However, the main problem is to identify such weightings. An important step to incorporate equity concerns for priority setting in health care has been made in the Netherlands, with the introduction of the concept of proportional shortfall (van de Wetering *et al.*, 2013). This concept combines elements of fair innings and severity-of-illness, where priority is given to those patients who lose the greatest proportion of their remaining health expectancy due to illness if this illness remains untreated. In other words, proportional shortfall looks at the fraction of QALYs that people lose relative to their remaining life expectancy (Stolk *et al.*, 2004). The proportional shortfall is measured on a scale from 0 (no health loss) to 1 (complete loss of remaining health) (van de Wetering *et al.*, 2013). The advantage of this approach is that equity concerns can be addressed in an explicit and transparent way. However, others provided a contrary view and argued that incorporating equity aspects into the QALY framework is not straightforward (Wailoo *et al.*, 2009), which could potentially overload the framework and it becomes even more difficult to understand the key concepts behind it (Nord *et al.*, 2009).

Finally, benefits outside the health care sector have also gained attention in recent years. Brazier and Tsuchiya have provided ten possible approaches for extending the *health-related* QALY using wellbeing and monetary-based methods (Brazier & Tsuchiya,

2015). In particular, the proposal of a WELBY is noteworthy. The WELBY is the same as a QALY, except that the descriptive system of a standardized instrument would be concerned with wellbeing rather than HRQoL. Whether the WELBY is the answer for improving cross-sector comparisons remains unknown. Other 'minor adjustments' to the QALY that would allow for the comparison of outcomes across sectors is the use of mapping, bolt-ons, or the valuation of measures on a common scale using preferences. However, the authors conclude that any choice between these approaches involves political decisions about what counts in measuring benefits of interventions (Brazier & Tsuchiya, 2015). Additionally, there is a risk of double counting given that outcomes across sectors are not unrelated. Here, double counting relates to the fact that the dimensions in one measure may be captured to some extent by dimensions in the other measure. As such, cross-sector comparisons require an overall measure of benefit whilst avoiding double counting. Another form of double counting that requires further research relates to the issue whether benefits outside the health care sector should be captured in the denominator or the numerator of the ICER. This is particularly an issue for reduction in criminal behaviour, better educational achievements, as well as increased labour participation and work productivity.

It can be concluded that non-health benefits identified are to some extent already included in the QALY framework or methods have been developed to incorporate those. Generally, part of the discussion and the way these broader benefits have been labeled is largely driven by the theoretical foundation of the QALY. The QALY was introduced as a measure of health outcome and as such the focus moved away from focusing on individuals' utility. Accounting for utility associated with process of care, as well as wider effects on SWB is considered outside the QALY. Also the incorporation of spillover effects falls under the utilitarian idea of aggregating all utility impact in all affected individuals and not only the individual. However, even if the QALY as a measure of health outcome appears to be a simple metric, health is complex and identifying, measuring and valuing different types of health improvements is challenging. For that reasons, adapting a broader extra-welfarism approach proposed by Brouwer and colleagues is a move towards a more precise QALY measure that is more of significance

to the decision maker involved or the public served by the decision-making authority (Brouwer *et al.*, 2008). Therefore, if a decision maker with authority decided that such broader benefits are of relevance for the allocation of health care resources (which can be observed in countries like the UK, where the focus on wellbeing and process-attributes is growing), it is an acknowledgment that these benefits are associated with health improvements and as such should not be labelled as 'non-health benefits'.

3.4.4. Strengths and limitations

A number of strengths and limitations of this work are worth noting. With regard to the chosen method, the conduct of a CIS using a citation pearl growing methods was considered as a strength of this work, which allowed for the identification of papers in this research area, and a more appropriate synthesis of the findings. For example, the lack of consistent indexing and terminology across papers would make it difficult to develop a search strategy for bibliographic databases, as it is usually done in conventional systematic reviews. It was also expected that discussions around benefits beyond health are not always included in the abstracts, meaning that relevant papers could be missed by a bibliographic database search. The citation pearl growing method has been previously applied within the context of health economics (Dolan *et al.*, 2005; Mitchell *et al.*, 2017a) and recommended for searching within the social sciences literature (Papaioannou *et al.*, 2010). A second strength of the CIS can be seen in the way evidence was appraised and synthesized. While conventional systematic reviews aim to summarize data and require a basic of comparability between concepts, this work involved different forms of evidence (with incompatible aims) and concepts could not be well specified *a priori*. The use of more qualitative approaches to synthesize evidence that allowed for the development of own interpretations can be seen as another strength of this work.

With regard to limitations, it needs to be acknowledged that not all relevant papers on this topic were included in the analyses. Firstly, the search did not consider grey literature and, secondly, with regard to the selection criteria, papers that cited already included papers were excluded, as it was assumed that such papers would not identify any new concepts that provide a deeper understanding of the topic. However, it is

worth highlighting that it was not the aim of this chapter to generate an exhaustive list of papers that discuss broader benefits within the QALY framework but rather to provide a better understanding of the concepts describing non-health benefits, which is in line with the CIS approach. Finally, in contrast to conventional systematic review methodology, the findings produced by the CIS may not be reproducible due to approaches applied for the literature search and synthesis. With regard to the literature search, it is possible that a different researcher would have used a different core set of papers, which could have resulted in the inclusion of different papers. However, this is not to say that a different thematic framework would have emerged. While the production of the synthetic constructs and the synthesizing argument is an interpretive process that allows for going beyond interpretations provided by the original studies, deviations are more expected in the classification of different non-health attributes (i.e., referring to the list in Appendix 3.2), rather than the synthetic constructs or the broader themes identified.

3.5. Conclusion

This chapter has synthesised existing literature of broader benefits that are currently outside the QALY framework. While different alternatives were discussed in terms of dealing with non-health benefits in future, it can be concluded more effort is needed to incorporate non-health benefits into the QALY framework. It appears that the QALY framework does not exclude such broader benefits *per se* but rather the narrow focus of preference-based instruments used to derive QALYs and the absence of a clear maximand for resource allocation decision-making. With a range of methods introduced to extend the *health-related* QALY, these approaches need to be tested empirically and explored further in practice to rule out double counting.

Whether or not the evaluative space of the QALY is limited to health alone will be explored further in the following chapters. This will be done through the direct comparisons of preference-based HRQoL measures that are used to derive QALY weights with outcome measures that adopt a broader evaluative space, particularly wellbeing measures that relate to the concept of capability and subjective wellbeing. The

next chapter reports findings from a study that applied the EQ-5D-5L and the ICECAP-O, a measure of capability wellbeing for older adults, within the context of public health. Results of this work aim to inform the current research gap in terms of the extent to which both instruments are able to capture broader benefits outside the health care context. The subsequent chapters will then explore the extent of double counting, and further relationships between wellbeing measures and other preference-based HRQoL measures.

Chapter 4. Complementarity of the ICECAP-O and EQ-5D-5L in the context of public health

4.1. Introduction

Methods for economic evaluation of clinical interventions have become well established and the use of CUA has become the dominant form (CADTH, 2017). However, as highlighted in the previous chapter, the application of conventional health economic evaluation methods to *public health* interventions faces some methodological challenges. Public health interventions comprise a wide range, from screening and immunizations through to the promotion of healthy eating, physical activity and wellbeing (Drummond *et al.*, 2007). Public health interventions are often considered as 'complex' (Craig *et al.*, 2008) because they not only combine educational, social and political strategies but also because the benefits of the interventions are often occurring outside the affected individual and outside the health care sector (Greco *et al.*, 2016). Drummond *et al.* have identified four main challenges in conducting an economic evaluation of public health interventions: (i) attributing outcomes to interventions, (ii) measuring and valuing outcomes, (iii) incorporating equity considerations, and (iv) and identifying intersectoral costs and consequences (Drummond *et al.*, 2007). Chalkidou *et al.* identified four further challenges that concern (i) the analytical perspective, (ii) the extrapolation of results to the appropriate time horizon, (iii) the quality of evidence, and (iv) the cost-effectiveness threshold (Chalkidou *et al.*, 2008).

Given that one of the methodological challenges for public health is the measurement and valuation of benefits, the exploration of the field of public health is of particular relevance to the topic of this thesis. It was previously reported by Lorgelly and colleagues that benefits of public health interventions impact broader aspects of QoL and not just health, which is why using QALYs as the measure of outcome may result in

underestimation of the relative benefits (Lorgelly *et al.*, 2010). The evaluation of public health interventions requires instead multiple or broader outcome measures (Greco *et al.*, 2016). Although public health guidance by NICE now places more emphasis on CCA and CBA, CUA is still required because it provides a single yardstick or currency for measuring the impact of interventions on health (NICE, 2012). Likewise, the Canadian guidelines state that a CCA may be helpful to present results for interventions involving public health (CADTH, 2017). Despite the proposal of other methods that could be employed to measure and value a broader set of benefits generated by public health interventions (Banke-Thomas *et al.*, 2015; Claxton *et al.*, 2007; Greco *et al.*, 2016; Huter *et al.*, 2016; Trueman & Anokye, 2013), the use of the capability approach is gaining particularly attention in the public health field (Lorgelly, 2015; Marsh *et al.*, 2012; van Mastrigt *et al.*, 2015). Under the capability approach, the evaluation of public health interventions would then be based on their impact on a person's capability to function in a particular way. At this current stage, little is known about the application of the capability approach to outcome measurement in public health, more specifically, how well capability measures perform when compared with preference-based HRQoL measures. The aim of this chapter is to investigate the complementarity of a capability measure, the ICECAP-O, and a preference-based HRQoL measure, the EQ-5D-5L, in a public health context. Looking at a specific population group consisting of older adults, this chapter will provide further evidence about the extent to which these two measures are able to pick up benefits of different 'characteristics of the environment' that play an important role in public health research.

4.1.1. The role of the environment in older age

Characteristics of the environment can be broken down into two components: social and physical characteristics. The latter refers to the natural environment (e.g., plants, water, earth, air quality, climate) and the built environment. Features of the built environment comprise the *urban design* (the design of the city and the elements within it), *land-use* (the distribution of activities across space), and the *transportation system* (including the physical infrastructure of roads, sidewalks, bike paths, railroad tracks,

bridges and services provided) (Handy *et al.*, 2002). The social environment, on the other hand, encompasses *interpersonal relationships* (e.g., social support and social networks), *social inequalities* (e.g., socioeconomic position and income inequality, racial discrimination etc.), and *neighbourhood and community characteristics* (e.g., social cohesion and social capital, and neighbourhood factors) (McNeill *et al.*, 2006a). The social environment comprises broader factors that could affect large groups or entire communities such as culture, norms, indicators of social disorder, and is associated with place attachment that can be understood as a sense of belonging to the neighbourhood (Barnett & Casper, 2001; Rowles, 1983; Sallis, 2009).

The built environment and social environment are of particular importance for older adults. The extent to which older adults engage with their environments is different compared with most other people, mainly because as individuals age, their environment tends to shrink to the locale of their home or immediate neighbourhood (King, 2008). The environmental press theory by Lawton suggests that the environment places a certain degree of 'press' or stress on individuals (Lawton, 1977). How well individuals function in their environment is a reflection of the degree to which individuals' competence meets the press imposed by the environment (Lawton, 1977). As physical health declines, older adults are less able to function within their surroundings and, therefore, are more vulnerable to the forces within their environment (Noreau & Boschen, 2010). Environmental challenges include, for example, uneven sidewalks, high curbs, increased traffic, or short timing for crosswalks. A combination of physical impairments and lower neighbourhood walkability presents challenges to moving about, which may lead to loss of independence, social isolation, and the inability to remain in a familiar social environment (Hanson *et al.*, 2013). Social isolation, in turn, can lead to depression and other adverse mental health outcomes (Rosso *et al.*, 2011). A supportive environment with fewer barriers, on the other hand, can promote physical activity (Morris *et al.*, 2008) as well as social interaction (Day, 2008), and is associated with better perceived QoL (Rantakokko *et al.*, 2010).

Although previous conceptual frameworks have emphasized the importance of environmental characteristics on QoL (Ferrans *et al.*, 2005; WHO, 2002b; Wilson & Cleary, 1995), a recent review has identified only ten studies that explored this association in an older adult population (Garin *et al.*, 2014). It was shown that accessibility, residential satisfaction, home size, housing type, heavy traffic, higher usability, exterior environment, interior environment, street noise, and safety from traffic were associated with QoL (Garin *et al.*, 2014). However, the authors stress the difficulty to hypothesize on the nature of relationships due to the different use of environmental variables that may have resulted in conflicting findings.

With regard to the social environment, an increase in social cohesion has the potential to positively influence older adults' QoL. Social cohesion can be understood as the extent of connectedness and solidarity among groups in society (Kawachi & Berkman, 2000). Previous literature indicated that neighbourhood cohesion was predictive of good health, wellbeing and QoL (Elliott *et al.*, 2014; Friedman *et al.*, 2012; Gale *et al.*, 2011). However, none of this research has examined these relationships using preference-based measures of HRQoL or capability wellbeing and further investigation of the association between the built environment and social cohesion with older adults' QoL is required. In consideration of this existing research gap, the objectives of this chapter are to explore the association between the built environment and social cohesion with older adults' perceived HRQoL and capability wellbeing, and to shed more light on the extent to which the EQ-5D-5L and ICECAP-O are influenced by characteristics of the environment.

4.2. Methods

4.2.1. Data source

The data used for this secondary analysis are from the 'Walk the Talk (WTT): Transforming the Built Environment to Enhance Mobility in Seniors' project. This study was approved by the University of British Columbia's Clinical Research Ethics Board

(certificate: H10-02913) and detailed information on study design and methods have been published elsewhere (Chudyk *et al.*, 2014). Briefly, this cross-sectional study consists of qualitative and quantitative components that investigate the impact of the built environment on the mobility and health of older adults living on low income in Vancouver, Canada.

Study participants were older adults that were in receipt of a Shelter Aid for Elderly Renters (SAFER) rental subsidy from BC Housing (BC Housing, n.d.). The rental subsidy is available to residents of British Columbia aged ≥ 60 years who pay more than 30% of their gross monthly household income towards the rent of their residence. This, however, does not imply that SAFER recipients necessarily reside in low-income neighbourhoods. Individuals were eligible if (i) they were current SAFER recipients aged ≥ 65 years (according to WHO's definition of an older or elderly person (WHO, 2002a)), (ii) resided in one of eight select cities within Metro Vancouver, (iii) self-reported that they were able to participate in a mobility assessment that involved a 4-meter walk, (iv) were able to walk ≥ 10 meters with or without mobility aid, (v) had left their home to go into the community at least once in a typical week, (vi) had not been diagnosed with dementia, and (vii) spoke and understood the English language. In total, 5806 households in eight select cities within Metro Vancouver were sampled using a stratified design. Following this, 200 households were randomly selected from within each decile of Walk Score® (n=2,000) to ensure diversity across the built environment. Five households were excluded because they were contacted prior for participation in the pilot study. Out of 1995 mailed invitations, 161 participants agreed to participate in the WTT study. Participants took part in a two-hour measurement session where they completed questionnaires about their neighbourhood built environment, social cohesion, health (including QoL) and physical activity, and underwent performance-based measures of cognition and lower-extremity function.

4.2.2. Outcome variables

The two primary outcome measures for QoL were the EQ-5D-5L and the ICECAP-O. While the EQ-5D-5L is considered a measure of an individual's HRQoL, the

ICECAP-O is a measure of person's capability wellbeing, which accounts for the fact that a person's capabilities (what a person *can* do) may differ from their functionings (what a person actually *does*) (Al-Janabi *et al.*, 2012). The ICECAP-O is considered to go beyond health-related aspects of QoL and into a broader measure of a person's capability wellbeing, and was specifically designed for older adults (Grewal *et al.*, 2006). It covers five attributes – attachment, security, role, enjoyment, and control – with scores ranging from zero (no capability) to 1 (full capability). The UK tariff was used to calculate the ICECAP-O index scores (Coast *et al.*, 2008b), as no value set is available yet for the Canadian population. To ensure comparability, the crosswalk value set for the EQ-5D-5L that reflected the values of the UK population was also used (van Hout *et al.*, 2012). The psychometric properties of the EQ-5D-5L are well established across many countries and population groups (Alvarado-Bolanos *et al.*, 2015; Feng *et al.*, 2015; Golicki *et al.*, 2015; Herdman *et al.*, 2011). Also the ICECAP-O has been validated in a number of studies (in the UK (Coast *et al.*, 2008c), Germany (Makai *et al.*, 2014a), Australia (Couzner *et al.*, 2013a) and the Netherlands (Makai *et al.*, 2013)) and has consistently been found to be a valid instrument to measure capability wellbeing.

4.2.3. Explanatory variables

The perceived built environment was measured using the Neighbourhood Environment Walkability Scale – Abbreviated (NEWS-A) (Cerin *et al.*, 2006). This instrument contains twelve subscales (A-L) and describes individuals' perceptions of their neighbourhood built environment features (for reference, labels A-L are used later in the chapter). Subscales are scored such that higher scores indicate higher walkability, except for subscale G, H, K, and L where higher scores indicate lower walkability. Scores for ten subscales range from 1-4, while subscale A is anchored on a 173-865 scale (residential density items were weighted relative to the average density) and subscale B on a 1-5 scale. For the WTT study, the NEWS-A was modified (items were dropped or added) in order to make it more suitable for older adults. As a result of these changes, subscale E (infrastructure and safety) could not be calculated for this dataset because of a dropped item. Items that were added to the NEWS-A were not used in the analyses. The

validity of the unmodified NEWS-A was previously established in older adults residing in the United States (Starnes *et al.*, 2014).

Next, the perceived social cohesion was measured using Sampson's 5-item measure of collective efficacy (SC-5PT) (Sampson *et al.*, 1997). This measure asks respondents to indicate the extent to which they agree or disagree (on a 5-point scale) that 'this is a close-knit neighbourhood', 'people around here are willing to help', 'people in this neighbourhood can be trusted', 'people in this neighbourhood generally do not get along with other', and 'people in this neighbourhood do not share the same values'. After reversing the coding for the last two statements, items are averaged to reflect levels of perceived social cohesion (average scores ranging from 1 to 5, where higher scores indicate higher collective efficacy). The SC-5PT showed being a reliable and valid measure in previous literature (Lochner *et al.*, 1999; Sampson *et al.*, 1997).

Finally, the Street Smart Walk Score ® (SSWS) was used as an objective measure of the built environment. Previous evidence has shown differences between objective and perceived measures of the environment (Gebel *et al.*, 2011; Leslie *et al.*, 2005; McGinn *et al.*, 2007), which suggest that these measurement types are likely capturing different constructs and both should be considered (Rosso *et al.*, 2011). The SSWS is based on the Walk Score® (www.walkscore.com), which is a publicly available index that measures the walkability of an address based on distances to nearby destinations (e.g., grocery stores, restaurant, shops). Road connectivity characteristics, such as intersection density and block length, are included in the score. Walk Score® uses data from Google, Education.com, Open Street Map, Localeze, and places added by the Walk Score user community. The SSWS uses an updated methodology that better reflects the pedestrian-walking experience (Frank, 2013), with scores ranging from zero to 100, where higher scores denote higher walkability.

4.2.4. Covariates

Previous literature has identified a number of covariates relevant for studying the relation of interest, including sociodemographic characteristics, health status, personality

traits, emotional stability, and social support (Elliott *et al.*, 2014; Friedman *et al.*, 2012; Gale *et al.*, 2011). For the analyses in this work, age, sex, living arrangements, physical function (number of comorbidities and gait speed), and psychosocial measures (loneliness and self-efficacy for walking) were included as precision variables and potential confounders - although no distinction is being made between these two in the statistical model. A self-report questionnaire was used to gather participants' sociodemographic information. Number of comorbidities was measured with the Functional Comorbidity Index, a self-report measure of the presence of eighteen comorbid diseases (Groll *et al.*, 2005). Gait speed (usual pace) was assessed as part of the 4-meter walk component of the Short Physical Performance Battery (Guralnik *et al.*, 1994). Loneliness was measured with 11 items drawn from the Revised UCLA Loneliness Scale (R-UCLA) (Russell *et al.*, 1980) – a measure of social isolation, loneliness, and dissatisfaction with one's social interactions. Finally, self-efficacy for walking was evaluated with the Ambulatory Self Confidence Questionnaire, which assesses perceived self-efficacy to walk in twenty-two different environment situations (Asano *et al.*, 2007).

4.2.5. Statistical analyses

Pearson's correlation analyses were conducted for environmental features and outcome measures; correlations were interpreted as weak (0.10 to 0.30), moderate (0.30 to 0.50), or strong (greater than 0.50) (Cohen, 1988). To examine the associations between the built environment and social cohesion with QoL while adjusting for covariates, separate stepwise backward Tobit regression models for the EQ-5D-5L and ICECAP-O index scores were used. The Tobit model is the recommended model for QoL measures that are subject to upper censoring (Austin *et al.*, 2000). Some have argued that the use of the Tobit model or the censored least absolute deviations (CLAD) model, which both treat the distribution of index scores as censored at 1, are not appropriate within the context of QALYs (Pullenayegum *et al.*, 2010). Pullenayegum and colleagues have argued that in an economic evaluation the outcome of interest should be utility (anchored on a 0-1 QALY scale) rather than HRQoL (which is a broader, more abstract

construct). Sullivan has taken a contrary view and argues that utilities are not bounded at 1.0 because regardless how 'full health' is defined by descriptive systems, there will be individuals who exceed the definition (Sullivan, 2011). As such, the basic characteristics of utilities require that scores exceed 1.0. This viewpoint finds further support with the development of the new Canadian tariff for the EQ-5D-5L, where the upper anchor was set at 0.949 rather than 1, which challenged the idea that no problems in the five dimensions of the EQ-5D-5L can be equated to full health (i.e., a health state =1) (Xie *et al.*, 2016). As such, analyses by Pullenayegum and colleagues showing that the Tobit and CLAD models lead to bias when the utility was bounded at 1 may not provide reasonable proof against the use of such alternative models.

All covariates (sex, age, living arrangement, number of comorbidities, gait speed, loneliness, and self-efficacy for walking), NEWS-A subscales, SC-5PT, and SSWS were entered into the model. Backward stepwise elimination was used to remove variables relating to the built environment and social cohesion (i.e., exploratory variables); a *p*-value of greater than 0.1 was the criterion for removal. Covariates were not eligible for removal to explicitly control for them when exploring the associations. To examine the associations with each environmental feature separately, while controlling for covariates, another set of Tobit models for the EQ-5D-5L and ICECAP-O index scores was fitted. Prior to running the regression models, missing value analysis was carried out to explore the patterns of missing data and to determine whether data were missing at random. Missing data were imputed using multiple imputation by chained equations (MICE) and combined estimates and variances were obtained using Rubin's rule (Rubin, 2004). Backward stepwise Tobit regression analyses were also repeated within a 'complete case' sample and compared with the results from a backward stepwise OLS regression. Appropriate tests for multicollinearity (VIF and tolerance test), homoscedasticity (White's test), and normality (Shapiro-Wilk test) indicated that assumptions were not violated. Descriptive statistics are presented as mean (SD), number (%), median, and interquartile range (IQR). All statistical tests performed were two-sided with $p \leq 0.05$ considered to be statistically significant. STATA version 10 (StataCorp, 2007) was used for all analyses and Microsoft Excel for Mac 2011 for graphical presentations.

4.3. Results

Characteristics of the study population are reported in Table 4.1. In total, 160 individuals (women n=102) aged ≥ 65 years were included in this study. Participants were mostly white, Canadian-born and lived alone. Although this population is of low socioeconomic status as measured by household income, the percentage of those individuals who completed some high school or less was only 12.5%, i.e., the majority was well educated. Table 4.1 also presents the characteristics of the study population used for the complete case analysis when applying the EQ-5D-5L and ICECAP-A.

Table 4.1. Characteristics of the study population (values are numbers (percentages) unless states otherwise)

	Total (n=160)	Complete case EQ-5D-5L (n=137) [a]	Complete case ICECAP-O (n=135) [a]
Mean Age (SD)	74.3 (6.3)	73.8 (6.1)	73.8 (6.1)
Gender			
Females	102 (63.8)	87 (63.5)	87 (64.4)
City of residence			
Vancouver	58 (36.3)	50 (36.5)	49 (36.3)
Surrey	38 (23.8)	31 (22.6)	31 (23.0)
West Vancouver	14 (8.8)	12 (8.8)	12 (8.9)
Burnaby	14 (8.8)	10 (7.3)	10 (7.4)
Richmond	12 (7.5)	11 (8.0)	10 (7.4)
North Vancouver	10 (6.3)	9 (6.6)	9 (6.7)
New Westminster	9 (5.6)	9 (6.6)	9 (6.7)
White Rock	5 (3.1)	5 (3.6)	5 (3.7)
Born in Canada			
Yes	87 (54.4)	80 (58.4)	79 (58.5)
Mean (SD) years in Canada	43.9 (15.7)	44.4 (15.9)	44.0 (15.8)
Ethnicity			
White	125 (78.1)	110 (80.3)	109 (80.7)
Other	16 (10.0)	13 (9.5)	13 (9.6)

Chinese	7 (4.4)	6 (4.4)	5 (3.7)
South Asian	3 (1.9)	2 (1.5)	2 (1.5)
Black	2 (1.3)	1 (0.7)	1 (0.7)
Latin American	2 (1.3)	2 (1.5)	2 (1.5)
Southeast Asian	2 (1.3)	1 (0.7)	1 (0.7)
Japanese	2 (1.3)	2 (1.5)	2 (1.5)
Filipino	1 (0.6)	0 (0)	0 (0)
Marital Status			
Divorced	60 (37.5)	55 (40.1)	55 (40.7)
Widowed	50 (31.3)	40 (29.2)	39 (28.9)
Single (never married)	25 (15.6)	21 (15.3)	20 (14.8)
Married (or common law)	14 (8.8)	11 (8.0)	11 (8.1)
Separated	11 (6.9)	10 (7.3)	10 (7.4)
Education Level			
University	55 (34.4)	47 (34.3)	46 (34.1)
Trade/ College diploma	53 (33.1)	46 (33.6)	46 (34.1)
High School	32 (20.0)	29 (21.2)	28 (20.7)
Some High School or less	20 (12.5)	15 (10.9)	15 (11.1)
Mean (SD) years lived in current residence	9.4 (9.5)	9.6 (9.8)	9.6 (9.9)
Living Arrangements			
Alone	130 (81.3)	111 (81.0)	110 (81.5)
With a spouse or partner	16 (10.0)	13 (9.5)	13 (9.6)
With another family member	10 (6.3)	9 (6.6)	8 (5.9)
With a friend or roommate	3 (1.9)	3 (2.2)	3 (2.2)
Other	1 (0.6)	1 (0.7)	1 (0.7)

SD=standard deviation.

[a] Complete case refers to the number of cases with complete data on the outcome variables, explanatory variables and covariates.

Histograms of the EQ-5D-5L and ICECAP-O index scores, presented in Appendices 4.1 and 4.2, demonstrate that both measures show a certain degree of ceiling effects; such findings confirmed the appropriateness of the Tobit model. Appendices 4.3 and 4.4 indicate some strong correlations between environmental features, while no or only weak correlations were found between environmental features and the two outcome measures. Based on the missing value analysis, one participant was excluded from further analyses due to language barriers that resulted in missing data (i.e., this formed a total sample size of 160 individuals). The missing data analysis also showed a high degree of missing values (17%) for the NEWS-A subscale I (parking) for the total sample. Eighty-two percent of individuals who did not respond to this question reported that they had no vehicle at their disposal in the past seven days. This indicated that the subscale was not applicable for this study population and was dropped from further analyses. Missing values for the remaining variables ranged from 0-5% (see Appendix 4.3). Each variable was imputed using a regression model conditional on all other variables used for the analyses. In total, 15 datasets were generated that reflected the percentage of missing values considering all variables (16% overall) and combined estimates and variances were calculated. Descriptive statistics for selected instruments are reported in Table 4.2. Mean (SD) scores for the EQ-5D-5L and the ICECAP-O were 0.794 (0.15) and 0.843 (0.11), respectively.

Figure 4.1 and Figure 4.2 show the percentage of responses at each level across all dimensions for the EQ-5D-5L and ICECAP-O. None of the participants reported the lowest level (extreme problems) on any of the EQ-5D-5L dimensions. Furthermore, with the exception of the 'pain or discomfort' dimension, the majority of this study population reported 'no problems' on all EQ-5D-5L dimensions. For the ICECAP-O, Figure 4.2 indicates that this study population was very independent, with 62% reporting the highest level (i.e., greatest capability).

Table 4.2. Descriptive statistics for instruments under consideration

	N	Mean	SD	Median	IQR	Min	Max
Environmental Instruments							
NEWS-A[a]							
Scale A: Residential density	157	331.27	158.35	265	247	173	792
Scale B: Land use mix-diversity	158	2.81	0.87	2.75	1.33	1.18	4.82
Scale C: Land use mix-access	158	3.39	0.76	3.67	1.00	1.00	4.00
Scale D: Street connectivity	156	3.07	0.77	3.00	1.00	1.00	4.00
Scale F: Aesthetics	160	3.16	0.72	3.25	1.00	1.25	4.00
Scale G: Traffic hazards [b]	155	2.57	0.57	2.67	1.00	1.00	4.00
Scale H: Crime [b]	152	1.69	0.69	1.67	1.00	1.00	4.00
Scale J: Lack of cul-de-sacs	159	2.98	1.02	3.00	2.00	1.00	4.00
Scale K: Hilliness [b]	159	2.02	1.05	2.00	2.00	1.00	4.00
Scale L: Physical barriers [b]	159	1.42	0.82	1.00	0	1.00	4.00
SC-5PT	156	3.45	0.70	3.60	1.00	1.80	5.00
Street Smart Walk Score	160	72.04	24.94	81.00	37.00	0.00	100
Quality of Life Instruments							
EQ-5D-5L	160	0.79	0.15	0.80	0.17	0.23	1.00
ICECAP-O	158	0.84	0.11	0.87	0.14	0.48	1.00
Other Instruments							
Functional comorbidity Index	158	2.90	2.12	3.00	3.00	0	9.00
Loneliness questionnaire	160	1.56	0.43	1.45	0.73	1.00	3.00
Gait Speed	160	1.00	0.26	0.98	0.31	0.45	1.91
Ambulatory questionnaire	160	8.36	1.67	8.91	2.29	3.27	10.00

SD=standard deviation, IQR= interquartile range, Min=minimum, Max=maximum, NEWS-A= Neighbourhood Environment Walkability Scale – Abbreviated, SC-5PT= Sampson's 5-item measure of collective efficacy.

[a] NEWS-A Scale E (Infrastructure and Safety) cannot be calculated because of a dropped item; NEWS-A Scale I (Parking) removed due to missing values.

[b] Reverse coding, where higher score indicate 'less pedestrian friendly'.

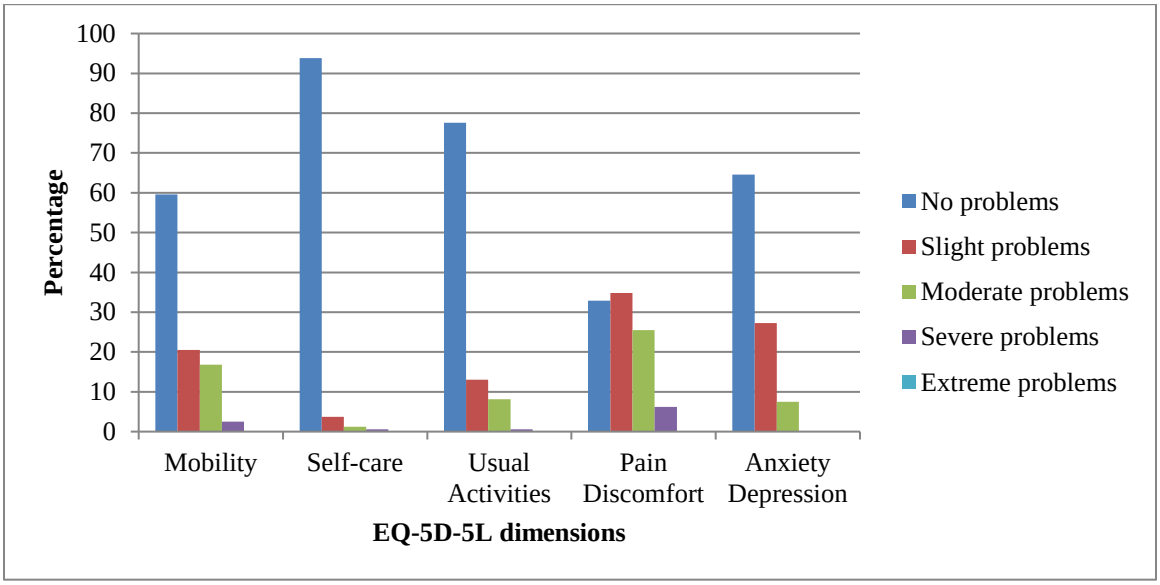


Figure 4.1. Percentage of responses at each level across all EQ-5D-5L dimensions (n=160)

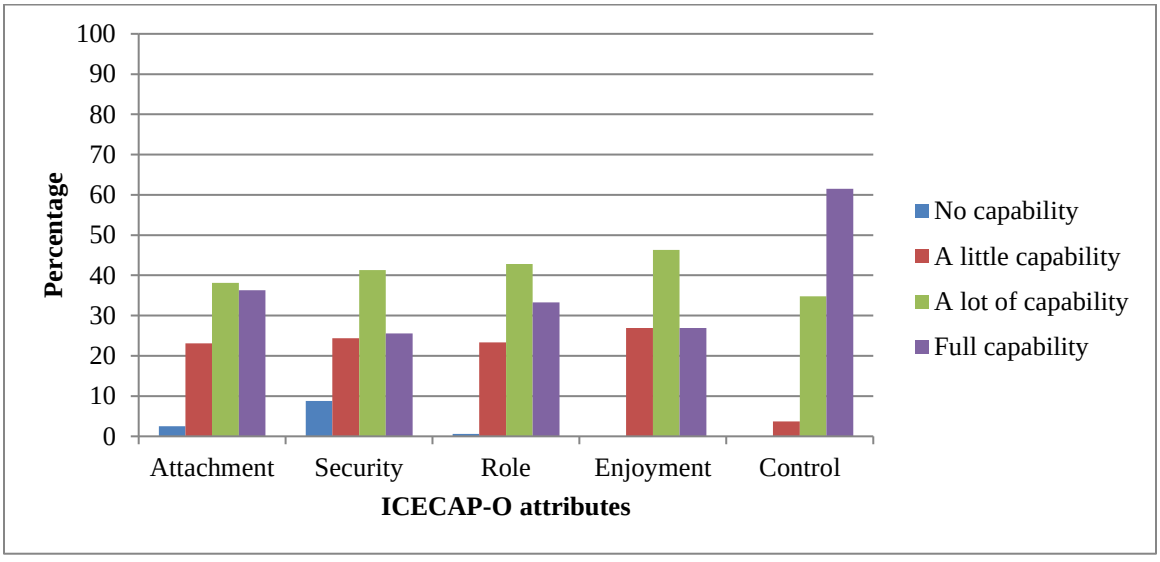


Figure 4.2. Percentage of responses at each level across all ICECAP-O dimensions (n=158)

Table 4.3 (showing the results of the backward stepwise Tobit regression analysis) and Table 4.4 (showing the results examining each environmental feature separately) indicate that no statistically significant variables were found for older adults' HRQoL using the EQ-5D-5L. However, when measuring QoL using the ICECAP-O, better social cohesion (SC-5PT) was associated with higher ICECAP-O index scores. In addition, a negative association between NEWS-A subscale D (street connectivity) and ICECAP-O was found, i.e., short distances between intersections and many alternative routes were associated with lower capability wellbeing. Overall, associations were small; an increase in street connectivity by one unit decreased ICECAP-O index scores by 0.028 (0.024 when built environment features were analyzed separately; Table 4.4). For social cohesion, a 0.029 (0.024 when analyzed separately) increase in ICECAP-O index scores was found when social cohesion was increased by one unit.

Results for the complete case analyses are provided in Appendix 4.6 and 4.7. Backward stepwise OLS and Tobit regression analysis yielded similar results and showed that after controlling for covariates, the EQ-5D-5L was associated with land-use (access), aesthetics, and cul-de-sacs, while the ICECAP-O was associated with aesthetics and social cohesion. However, BIC, AIC and LR chi^2 indicated a better model fit for the Tobit model when compared with simple OLS, which supports the use of the Tobit model in the main analyses.

Table 4.3. Backward stepwise Tobit regression analysis

	EQ-5D-5L (n=160)			ICECAP-O (n=160)		
	β (SE)	95% CI	*p*	*β* (SE)	95% CI	*p*
Constant	0.219 (0.189)	(-0.151, 0.590)	*0.248*	0.829 (0.119)	(0.596, 1.062)	*<0.001*
Covariates						
Sex (ref. Female)	-0.040 (0.026)	(-0.091, 0.010)	*0.119*	0.008 (0.016)	(-0.023, 0.039)	*0.622*
Age	0.005 (0.002)	(0.002, 0.010)	*0.006*	0.000 (0.001)	(-0.002, 0.003)	*0.680*
Living with someone (ref. Yes)	0.033 (0.031)	(-0.027, 0.092)	*0.288*	0.032 (0.019)	(-0.005, 0.068)	*0.089*
Functional Comorbidity Index	0.027 (0.006)	(-0.038, -0.016)	*<0.001*	<0.001 (0.003)	(-0.006, 0.008)	*0.792*
Loneliness	-0.057 (0.027)	(-0.110, -0.003)	*0.039*	-0.143 (0.017)	(-0.177, -0.108)	*<0.001*
Gait Speed	-0.019 (0.054)	(-0.124, 0.086)	*0.728*	0.037 (0.033)	(-0.027, 0.102)	*0.253*
Ambulatory self-efficacy	0.048 (0.008)	(0.032, 0.064)	*<0.001*	0.016 (0.005)	(0.006, 0.026)	*0.003*
Perceived built environment (NEWS-A)						
A: Residential density	-	-	-	-	-	-
B: Land-use mix (diversity)	-	-	-	-	-	-
C: Land-use (access)	-	-	-	-	-	-
D: Street connectivity	-	-	-	-0.028 (0.010)	(-0.047, -0.008)	*0.006*
F: Aesthetics	-	-	-	-	-	-
G: Traffic hazards [b]	-	-	-	-	-	-
H: Crime [b]	-	-	-	-	-	-
J: Lack of cul-de-sacs	-	-	-	-	-	-
K: Hilliness [b]	-	-	-	-	-	-
L: Physical barriers [b]	-	-	-	-	-	-
Social cohesion						
SC-5PT	-	-	-	0.029 (0.011)	(0.007, 0.052)	*0.011*
Objective built environment						
Street Smart Walk Score	-	-	-	-	-	-

β=beta coefficient, SE=standard error, CI=confidence interval, *p*=two-sided p-value, NEWS-A= Neighbourhood Environment Walkability Scale – Abbreviated, SC-5PT= Sampson's 5-item measure of collective efficacy.

[a] Backward stepwise elimination was used to remove variables relating to the built and social environments; p > 0.1 was the criterion for removal. Covariates were not eligible for removal.

[b] Reverse coding, where higher score indicate 'less pedestrian friendly'.

Table 4.4. Tobit regression analysis exploring each environmental feature separately, after adjusting for covariates

	EQ-5D-5L (n=160)			ICECAP-O (n=160)		
	β (SE)	95% CI	*p*	*β* (SE)	95% CI	*p*
Perceived built environment (NEWS-A) adjusted for covariates (examined separately)						
A: Residential density	<0.001 (<0.001)	(-<0.001, <0.001)	*0.464*	<0.001 (<0.001)	(-<0.001, <0.001)	*0.158*
B: Land-use mix (diversity)	0.004 (0.013)	(-0.023, 0.032)	*0.749*	-0.011 (0.009)	(-0.028, 0.006)	*0.207*
C: Land-use (access)	0.018 (0.015)	(-0.011, 0.048)	*0.235*	-0.014 (0.010)	(-0.033, 0.004)	*0.123*
D: Street connectivity	-0.002 (0.002)	(-0.032, 0.028)	*0.886*	-0.024 (0.010)	(-0.043, -0.004)	*0.019*
F: Aesthetics	-0.023 (0.017)	(-0.056, 0.010)	*0.171*	-0.018 (0.011)	(-0.039, 0.003)	*0.089*
G: Traffic hazards [b]	-0.022 (0.020)	(-0.061, 0.017)	*0.276*	0.008 (0.012)	(-0.017, 0.033)	*0.560*
H: Crime [b]	-0.020 (0.017)	(-0.054, 0.013)	*0.245*	0.011 (0.011)	(-0.010, 0.032)	*0.299*
J: Lack of cul-de-sacs	-0.016 (0.001)	(-0.039, 0.006)	*0.158*	-0.004 (0.007)	(-0.018, 0.010)	*0.591*
K: Hilliness [b]	-0.006 (0.012)	(-0.028, 0.017)	*0.632*	0.010 (0.007)	(-0.004, 0.024)	*0.157*
L: Physical barriers [b]	<-0.001 (0.015)	(-0.029, 0.028)	*0.976*	0.010 (0.009)	(-0.008, 0.028)	*0.261*
Social cohesion adjusted for covariates						
SC-5PT	-0.011 (0.018)	(-0.046, 0.023)	*0.520*	0.024 (0.011)	(0.001, 0.047)	*0.040*
Objective built environment adjusted for covariates						
Street Smart Walk Score	<0.001 (<0.001)	(-0.001, 0.001)	*0.853*	-<0.001 (<0.001)	(-<0.001, <0.001)	*0.227*

β=beta coefficient, SE=standard error, CI=confidence interval, *p*=two-sided p-value, NEWS-A= Neighbourhood Environment Walkability Scale – Abbreviated, SC-5PT= Sampson's 5-item measure of collective efficacy.

[a] Covariates included in this model: sex; age; living arrangement; functional comorbidity index; loneliness; gait speed; and ambulatory self-efficacy.

[b] Reverse coding, where higher scores indicates 'less pedestrian friendly'.

4.4. Discussion

This chapter contributes to a limited body of literature that examines the association between environmental features and older adults' QoL. The findings suggest that street connectivity and social cohesion were associated with older adults' capability wellbeing. When considering health-related aspects of QoL only, using the EQ-5D-5L, the associations between the built environment and social cohesion with older adults' HRQoL were small and not statistically significant. Such findings suggest that the EQ-5D-5L appears to be too narrow to capture benefits of the environment and could potentially underestimate the impact of environmental interventions. The ICECAP-O appears to be more influenced by characteristics of the environment, meaning that it is more likely to pick up benefits of environmental changes, particularly regarding the social environment.

In the analyses performed in this chapter, only street connectivity emerged as an important built environment feature where greater street connectivity, defined as shorter distances between intersections and many alternative routes, was associated with lower capability wellbeing. Previous literature has shown that better street connectivity facilitates walking in the neighbourhood (Frank *et al.*, 2005; Handy *et al.*, 2002; Rosso *et al.*, 2011; Saelens *et al.*, 2003) and includes other benefits, such as shorter travel trips, increased travel by public transport, and more interactions among neighbourhood residents (Turrell, 2010). Therefore, the finding of a negative association between street connectivity and ICECAP-O scores was unexpected. These findings can possibly be explained within the context of older adults. Although a more walkable neighbourhood supports walking, it may be also associated with more traffic (Villanueva *et al.*, 2013). This exposure to traffic might be negatively associated with older adults' QoL simply because potential reductions in vision, hearing and/or physical robustness could increase the risk of injury (Villanueva *et al.*, 2013). The proposition that lower QoL is attributable to higher exposure to traffic was not supported by other results in this work, where the 'traffic hazard' subscale of the NEWS-A (which contains three items concerning the

amount and speed of traffic in the neighbourhood) was not significant. Other literature links better street connectivity with traffic calming since a large number of intersecting streets tends to slow traffic (Turrell, 2010). More important, however, is the perceived safety by older adults in relation to street connectivity (Yen *et al.*, 2014). While subscale G of the NEWS-A 'traffic hazard', does not refer to safety, subscale E, 'infrastructure and safety for walking', would have provided more insights. Since this subscale could not be generated in the dataset, additional associations between QoL and the presence of sidewalks and pedestrian signals that help walkers to cross busy streets could not be explored in this work.

Interestingly, although eleven perceived and objective built environment features were considered in this chapter, only one association was apparent between the ICECAP-O and the built environment. Other built environment characteristics did not seem to be associated with older adults' QoL and the role of social cohesion appeared to be stronger. Similar results were found in the study by Friedman and colleagues, which examined the impact of three neighbourhood-level factors on QoL among older adults attending New York senior centers. The authors found that social cohesion was a significant correlate of QoL (as was safety), while walkability was not (Friedman *et al.*, 2012). The qualitative component of the WTT study previously published also provided evidence that social needs can often shape individuals' behaviour regardless of whether they live in a highly walkable neighbourhood with amenities in close proximity (Franke *et al.*, 2013). For example, in the qualitative WTT study a participant reported that he prefers to use small alleyways rather than main streets because of the greater likelihood of meeting others (Franke *et al.*, 2013). Further to this, others have shown that, for older adults, features of the built environment might be secondary to attributes of the social environment in promoting physical activity (King, 2008). Although the study by King was concerned with physical activity and not QoL, such findings may explain why an association was found for one built environment feature only. On the contrary, while the social environment, such as social cohesion, is particularly important within the context of older adults, it is important to understand that the social environment often involves interactions that take place within the built environment. It is within the neighbourhood

space where older adults have opportunities to interact with others, access opportunities for socialization, and perform their activities of daily living. The built environment can, therefore, be seen as a facilitator or barrier for the social environment.

4.4.1. Implications for the public health agenda

The potential of the ICECAP-O to be used in public health interventions can be supported by findings of this work. As the results suggest, individuals who live in a less-favourable environment are more likely to experience a negative impact on their capability wellbeing. Such observations open new opportunities for future policy interventions to reconsider the approaches for the evaluation of public health care interventions, i.e., potentially moving away from functionings towards capabilities. It was previously recognized in the literature that healthy aging should not only focus on individual achievement but also consider broader circumstances (Stephens *et al.*, 2015). Stephens and colleagues further highlighted that many public health intervention models focus too much on the delay of illness, disease, disability and mortality. They suggest that rather than placing responsibility on individuals for engaging in exercise, diet, and social engagement prescriptions to produce good health, successful aging should focus on what people can achieve, with support, despite the limitations of physical health changes. As the authors were able to demonstrate in their qualitative study, older people were not simply excluded from achieving valued functionings because of physical limitations but by having to make choices or simply by a lack of money or appropriate transport (Stephens *et al.*, 2015).

Environmental factors play a particularly important role in the capability approach, as can be illustrated using the example of a mobility-impaired person in a wheelchair. In order to convert ‘commodity characteristics’ into capability (e.g., from the mobility and transportation properties of the wheelchair into the ability to move around, or the ability to transport oneself), one needs to consider personal factors (e.g., severity of an individual’s impairment affecting his/her manoeuvre the wheelchair) as well as environmental factors (e.g., type of terrain or street conditions to facilitate or prevent wheelchair movement such as flat, paved roads or dirt roads with bumps) (Welch

Saleeby, 2006). In this context, the consideration of the individual's environment is important since it provides a more realistic assessment of what an individual can really do or his/her real potential to achieve certain functionings within the context of his/her real-life settings factoring in environmental barriers and/or facilitators. In the capability literature, there are three groups of 'conversion factors' that refer to the factors that determine how resources are converted to functionings. While disability is a *personal* conversion factor in the sense that it is harder for a disabled person to convert the characteristics of the commodity into a functioning, *social* factors (e.g., social norms), and *environmental* conversion factors (e.g., the built environment) need also to be considered (Robeyns, 2005). For example, even if a mobility-impaired person has the same income (i.e., personal conversion factor) as an able-bodied person, she/he will not be able to travel on public transport if there are no ramps on the buses (i.e., environmental conversion factor) (Trani *et al.*, 2011). While the application of the capability approach to health is still at an early stage (Kinghorn, 2015; Mitchell *et al.*, 2017a), it is gaining more attention, especially within the context of public health (Lorgelly *et al.*, 2015), and has the potential to play an important role in future studies.

4.4.2. Strengths, limitations and directions for further research

A major strength of the current work relates to the opportunity to explore social cohesion and built environment factors together when assessing the association with older adults' QoL. The application of a broad concept of QoL (i.e., HRQoL *and* capability wellbeing) provided a better understanding of the associations with the built environment and social cohesion. Another strength of this work can be seen in the reduced likelihood of the occurrence of self-selection, a concept which can be understood as the tendency of people to choose locations based on their travel activities, needs and preferences (Mokhtarian & Cao, 2008). There is evidence that individuals whose QoL is higher are likely to be more active and, therefore, these individuals may have chosen to live in a more activity-friendly neighbourhood (Sugiyama & Ward Thompson, 2007). Consequently, all features of the neighbourhood that are potentially related to health and wellbeing may be determined by the characteristics of individuals who reside there. In

this work, self-selection was less of an issue because of individuals' enrolment in the SAFER rental subsidy program. Compared with other individuals, this study population may have fewer opportunities to move to a different residence that better suits their current needs and preferences as a result of their low income. This also implies that the generalizability of the findings to older adults of higher socioeconomic status should to be made with care. Older adults on low income may rely more on their local neighbourhood and amenities that are reachable by walking, as owning a car or taking public transportation may be unaffordable for them.

Another note on the generalizability of this work is that the characteristics of the study population are different compared with other older adults of low income. Typically, individuals of low income attain lower levels of education (i.e., both are characteristics of a low socioeconomic position (McNeill *et al.*, 2006a)); the majority (87.5%) of participants in this work obtained a high school diploma or higher degrees. Previous literature has highlighted that older adults living on a low income may be at increased risk of morbidity and poor physical function (Nilsson *et al.*, 2010). The number of self-reported comorbidities in this population is similar to Canadian population norms of older adults (Wister *et al.*, 2015). With respect to the QoL, the index scores and distribution of responses across the dimensions of the EQ-5D-5L and ICECAP-O are similar to population norms available for older adults aged 65 and older in Canada, UK, and Australia (Couzner *et al.*, 2013a; Flynn *et al.*, 2011; Health Quality Council of Alberta, 2014). Collectively, results suggest that the study population is different compared with other older adults of low income but similar to older adults in general. Since no information is available in the dataset about a person's socioeconomic history, low income could represent either a recent or a lifelong circumstance. Additionally, the low response rate of the WTT study is another consideration for the generalizability of this work.

The primary limitations of this work are the small sample size, which may have resulted in the failure to detect associations, and the cross-sectional design, from which the establishment of causal links cannot be made. The lack of *a priori* information about

which features of the environment may affect QoL, especially for the NEWS-A subscales, is a further limitation. The advantage of using a backward elimination procedure is to reduce the number of variables in the final model to yield a parsimonious model. However, such automated selection procedures are problematic as they may lead to bias in parameter estimates, inconsistencies among model selection algorithms, the inappropriate focus or reliance on a single best model, and an inherent problem of multiple hypothesis testing (Whittingham *et al.*, 2006). Multiple hypothesis testing could lead to a Type 1 error, where certain associations may be found when in reality such associations do not exist. In particular, results from the complete case analysis need to be interpreted with caution because significant predictors may have resulted by chance alone. However, the application of different models in this work has provided more confidence in the results. Particularly, the comparison between the Tobit model and a simple OLS in this Chapter, where similar results were obtained when running a backward stepwise regression, provided further confidence in the results.

Another limitation of the analyses is that neighbourhood was not taken into account. Individuals living in the same neighbourhood could have correlated observations, thus reducing the effective sample size (Maas & Hox, 2005). Multilevel models that account for the fact that individuals are nested within neighbourhoods are needed in future research. It should also be acknowledged that there is a potential risk that some variables that were considered as confounders in the analyses have, in fact, mediated the relation between the built environment and social cohesion and QoL. For example, a previous study provided evidence that self-efficacy mediated the relation between social and physical environments and physical activity (McNeill *et al.*, 2006b). More longitudinal studies are necessary to understand the dynamic changes in QoL among older adults living across diverse built and social environments. Particular focus is required to explore how built and social environments impact a person's functionings and capabilities. Given that this work looked at associations only, the ability of the EQ-5D-5L and the ICECAP-O to capture the benefits of environmental interventions in a longitudinal setting remains unknown. Generally, evidence on the responsiveness (i.e., the ability of an instrument to measure change) of these measures in an older adult

population remains scare. Recent evidence has indicated that the EQ-5D-3L is more responsive than the ICECAP-O in individuals at risk of mobility impairment (Davis *et al.*, 2016). On a final note, future studies should also explore these relationships in samples of low HRQoL and capability levels. Given the distribution of the ICECAP-O and the EQ-5D-5L in this work, none of the participants reported the lowest level ('extreme problems') on the EQ-5D-5L and only a few reported 'no capability' on the ICECAP-O, which could have influenced the findings.

4.5. Conclusion

In conclusion, this chapter has found that street connectivity and social cohesion might be important for older adults' capability wellbeing. The association between capability wellbeing and other NEWS-A subscales, as well as HRQoL and environmental factors could not be confirmed and need to be examined further. The findings should be interpreted with caution and be considered as hypothesis-generating. Larger studies are required to explain the effect of the various features of the built and social environments on older adult's QoL.

This chapter has also shown that there are differences between the EQ-5D-5L and the ICECAP-O in the ability to identity environmental benefits. At this stage, it remains unknown why differences were found between these two measures. Further analyses of the complementarity of the EQ-5D-5L and the ICECAP-O are warranted to explore similarities and differences between these measures. Given the availability of numerous preference-based HRQoL measures, such analyses should be extended beyond the EQ-5D-5L to fully explore the additional information provided by the ICECAP measures. This type of analysis will be the focus of the next chapter.

Chapter 5. Overlap between the ICECAP-A and five preference-based health-related quality of life measures – explanatory factor analyses

5.1. Introduction

Chapter 3 identified capability wellbeing as one form of wellbeing that is currently not captured by the QALY framework, while the previous chapter has shown that the use of the capabilities approach has the potential to provide additional information to health care decision-making. Generally, there is a growing interest in Sen's capability approach within health economics, and for outcome measurement in economic evaluation in particular (Lorgelly, 2015). With the development of new capability measures, including the ICECAP family (i.e., ICECAP-A, ICECAP-O etc.) changes in guidelines for HTA have recognized the potential importance of broader benefits in economic evaluation and have made provision for the measurement of capability wellbeing. For example, NICE in the UK has recommended the use of capability measures in social care economic evaluations for interventions that are associated with non-health benefits (NICE, 2014). Yet, little guidance has been provided in terms of what constitutes a health benefit or a non-health benefit, and which decision rules should be applied if using ICECAP instruments alongside other preference-based HRQoL instruments. Dutch guidelines also advocate for the use of ICECAP instruments in long-term care, where the focus of interventions might be more on improving a person's wellbeing rather than their health (Zorginstituut Nederland, 2016). Since ICECAP instruments do not have 'QALY properties' (i.e., current values are not anchored onto the 'full health' to 'dead' scale but on a 'full capability' to 'no capability' scale (Flynn *et al.*, 2015)), the reference cases described in the UK and Dutch guidelines recommend supplementing CUA (using the EQ-5D (-3L or -5L) (Brooks, 1996; Herdman *et al.*, 2011)) with cost-consequences analysis or cost-effectiveness analysis using an

ICECAP instrument. The underlying intention is to capture explicitly broader aspects of capability wellbeing alongside health benefits.

In practice, decision makers may find it difficult to interpret and reconcile findings from such primary and supplementary analyses without further information describing the extent of overlap between measures of HRQoL and capability wellbeing or any additional non-HRQoL measure. The extent of overlap between the ICECAP instruments and the EQ-5D-3L has been examined in two previous studies. Davis and colleagues performed an exploratory factor analysis (EFA) comparing the ICECAP-O with the EQ-5D-3L in seniors enrolled in a falls prevention clinic (Davis *et al.*, 2013), showing that the two instruments tapped into distinct and complementary factors. These results were confirmed by a second EFA, which compared the ICECAP-A with the EQ-5D-3L in an adult population of patients with knee pain (Keeley *et al.*, 2016).

Further research is needed to explore whether similar relationships hold in other clinical and non-clinical settings, as well as for other preference-based HRQoL instruments. It is well established that preference-based HRQoL instruments differ greatly in their coverage of physical, mental and social health domains (Karimi & Brazier, 2016; Richardson *et al.*, 2015b; Richardson *et al.*, 2014a; Whitehurst *et al.*, 2014a). These issues raise the potential for different degrees of overlap between preference-based HRQoL instruments (i.e., preference-based instruments that define health states) *and* the potential for different degrees of overlap when comparing HRQoL measures with measures of capability wellbeing. Such investigations are particularly important in order to avoid double counting when using HRQoL and capability wellbeing instruments simultaneously in health economic evaluation. In this context, double counting – where the same underlying concept of benefit is measured twice – could occur explicitly (i.e., summing health and non-health benefits into a single metric) or implicitly (e.g., misguided interpretation of outcomes data from a cost-consequences analysis). The objectives of this chapter are to investigate the extent to which five preference-based HRQoL instruments capture aspects of capability wellbeing, as measured by the

ICECAP-A, and to consider the implications of the findings within the context of other literature regarding capability wellbeing and economic evaluation.

5.2. Methods

5.2.1. Data Source

Data were obtained from the Multi Instrument Comparison (MIC) project, a multinational survey funded by Australia's National Health and Medical Research Council. Comprehensive details regarding the background, rationale and administration of the MIC survey have been reported elsewhere (Richardson *et al.*, 2012). Briefly, the aim of the MIC project was to compare several QoL and wellbeing instruments across seven disease areas (in addition to a 'disease free' population defined as reporting 70 on a 0-100 VAS) in six countries: Australia, Canada, Germany, Norway, UK, and USA. The MIC survey was administered online between February 2012 and May 2012 by a global survey company, CINT Pty Ltd, which sent the link to the survey to people on their database until predetermined quotas of patients and demographically representative public respondents were achieved. Data collection was approved by the Monash University Human Research Ethics Committee (CF11/3192-2011001748).

5.2.2. Instruments

The MIC survey contained a comprehensive set of questions and standardized instruments (Richardson *et al.*, 2012). In addition to questions about demographics, self-reported illnesses and subjective wellbeing, all participants were asked to complete the ICECAP-A (with the exception of participants in Norway) and seven preference-based HRQoL instruments: 15D (Sintonen, 2001), Assessment of Quality of Life 4-dimension (AQoL-4D) (Hawthorne *et al.*, 1999), Assessment of Quality of Life 8-dimension (AQoL-8D) (Richardson *et al.*, 2009), EQ-5D-5L (Herdman *et al.*, 2011), Health Utilities Index Mark 3 (HUI-3) (Feeny *et al.*, 2002), Quality of Well-Being Scale Self-Administered (QWB-SA) (Seiber *et al.*, 2008), and SF-6D (Brazier *et al.*, 2002) (based

on the 36-item Short Form health survey version 2 (SF-36v2) (Ware *et al.*, 2008)). Instruments were administered in a randomized order to account for order-effect bias (Perreault, 1976).

For the analyses reported in this chapter, a decision was made to focus on the 35-item AQoL-8D (rather than the 12-item AQoL-4D) because of the more comprehensive descriptive system and the greater potential for overlap with the ICECAP-A. The QWB-SA was also excluded because the measurement scale used for many items provides nominal data. These data would require transformation in order to meet the requirements for the statistical analysis performed, and such transformations effectively render the analysis meaningless because the descriptive system has been modified. An overview of the dimensions and items contained within the preference-based HRQoL instruments included in this analysis (15D, AQoL-8D, EQ-5D-5L, HUI-3, and SF-6D) is provided in Appendix 5.1, with more comprehensive details available elsewhere (Richardson *et al.*, 2014a). The ICECAP-A comprises five dimensions (lay descriptions used by the instrument developers are included in brackets): stability (an ability to feel settled and secure), attachment (an ability to have love, friendship and support), autonomy (an ability to be independent), achievement (an ability to achieve and progress in life), and enjoyment (an ability to experience enjoyment and pleasure). Each dimension comprises one question with four levels of response, ranging from full capability to no capability (Al-Janabi *et al.*, 2012).

5.2.3. Statistical analysis

Exploratory factor analyses (EFA) were conducted in Mplus 7.4 (Muthén & Muthén, 1998-2015). The purpose of EFA is to explore the underlying structure for a set of measures and to ascertain the number of distinct constructs that account for the pattern among the measures. These constructs are called *factors* or *common factors*, reflecting the fact that they are common to more than one measured variable (Fabrigar & Wegener, 2012). Since these constructs are unobserved, they represent *latent variables*. The observed scores in the dataset are called *measured variables* and refer to a set of variables that can be directly measured. Figure 5.1 is a graphical illustration of a hypothetical

common factor model, consisting of six measured variables (MV) (squares) that are influenced by a single unique factor (U). Latent variables (common factors and unique factors) are presented as circles. The model shows the existence of two common factors, with the first common factor strongly influenced by the first three measured variables and the second by the last three measured variables. The strength and direction of association between each item and each of the common factors is reflected by factor loadings. Higher factor loadings indicate that more of the variance in the observed variables is attributable to the latent variable (i.e., the common factor) (Fabrigar & Wegener, 2012). Single-headed arrows depict the causal effects and indicate the direction of the influence. Double-headed arrows represent a non-directional association (correlation). A non-directional association of a variable with itself represents the variance in the variable, which is usually assumed to be 1.0 with a mean of zero.

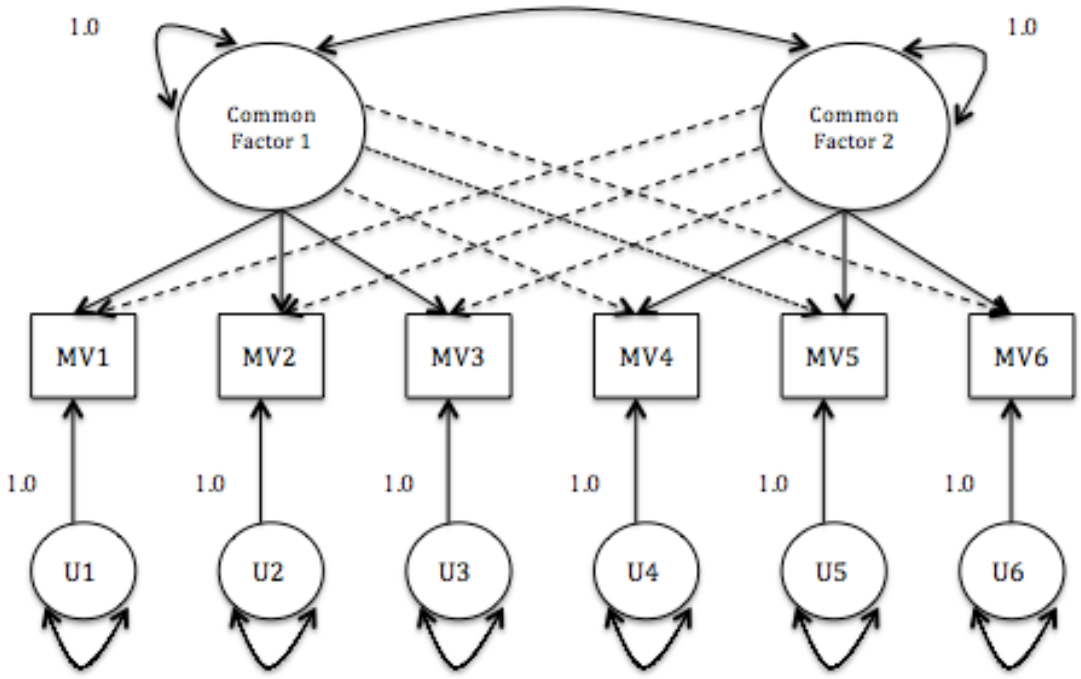


Figure 5.1. Graphical illustration of a hypothetical common factor model

Figure 5.1 also shows that each measured variable is influenced by a single unique factor, with the value of 1 indicating that an increase of one unit in the unique factor corresponds to an increase of one unit in the measured variable. These unique factors represent that portion of the score on a measured variable that is not explained by the common factor. Therefore, the variance of each measured variable can be decomposed into a common variance (*communality*) and unique variance (*uniqueness*). In this figure, the unique factors are independent of each another (i.e., no paths among the unique factors). It is also assumed that the two common factors are related (i.e., oblique design), which is illustrated by the double-headed arrow between the common factors. Orthogonal designs, on the other hand, would assume independency between common factors. Compared with a confirmatory factor analysis (CFA), where explicit assumptions are made regarding the number of factors based on priori theory, an EFA does not have expectations about the underlying structure. Therefore, while CFA is considered as an approach to test a proposed theory, a EFA is used to generate a theory (Williams *et al.*, 2010).

Applied to this work, EFA was used to explore the underlying structure for a set of measures and to determine whether or not the ICECAP-A instrument measures something unique, i.e., a construct or constructs not captured by current preference-based HRQoL instruments. In all pairwise comparisons (i.e., item-level responses for the ICECAP-A compared with item-level responses for each of the other instruments, namely 15D, AQoL-8D, EQ-5D-5L, HUI-3, and SF-6D), the numbers of unique underlying *latent factors* were ascertained that were associated with the items covered by the respective preference-based HRQoL instrument and the ICECAP-A. The axes of the initial factor analysis were rotated using the *geomin* oblique rotation. As explained above, oblique rotation permits correlations between common factors, which is to be expected when all items measure aspects of a person's quality of life. Pearson correlation coefficients were used to examine the extent of the relationship between factors (factors are considered as continuous variables); correlations were interpreted as weak (0.10 to 0.30), moderate (0.30 to 0.50), or strong (greater than 0.50) (Cohen, 1988). Weighted

least square means and variance adjusted (WLSMV) model estimation was applied to account for the ordinal nature of the item-level data.

The factor model and the number of common factors for each pairwise analysis were selected using the following procedure. The first step comprised an examination of eigenvalues. Eigenvalues are numerical values that correspond to the variance in the items accounted for by each of the common factors (Fabrigar & Wegener, 2012). More specifically, an eigenvalue is the sum of the squared factor loadings for a given factor. Model selection based on eigenvalues typically entails comparison of eigenvalues against the Kaiser criterion, where the number of factors with eigenvalues greater than one gives the number of common factors to be specified in the model (Fabrigar & Wegener, 2012). Evaluation against the Kaiser criterion was supplemented with inspection of scree plots, which are graphical representations of the eigenvalues plotted in a descending order. Model selection based on scree plots typically involves identification of the last substantial drop in the magnitude of the eigenvalues and retention of common factors prior to this drop (Fabrigar *et al.*, 1999).

Scree plots also guided the identification of increases (decreases) in the number of factors suggested by the Kaiser criterion that return large gains (small losses) in the variance explained. Three model fit indices were used to further quantify such gains (losses). The root mean square error of approximation (RMSEA) estimates goodness of fit as the discrepancy between the model and the data per degree of freedom for the model (Fabrigar *et al.*, 1999); RMSEA values were interpreted as indicating a good (less than 0.05), acceptable (0.05 to 0.08), marginal (0.081 to 0.1), or poor (greater than 0.1) fit (Fabrigar & Wegener, 2012). The Tucker-Lewis Index (TLI) and Comparative Fit Index (CFI) were also used. These goodness of fit estimates indicate how much better a model fits the data compared with a baseline model that assumes no relationship exists between any of the variables (Geiser, 2013). For both the TLI and CFI, values greater than 0.9 indicate a ‘good’ model fit (Hu & Bentler, 1999).

Using more than one criterion to guide the selection of the number of factors raises the possibility of seemingly conflicting results (e.g., a situation where the Kaiser

criterion suggests a two-factor model, whereas model fit statistics suggest a three-factor model). Within EFA, it is important to recognize that the objective is not to arrive at the 'true' or 'correct' number of factors but to estimate the patterns of correlations among observed variables and to simplify the data so that these patterns of correlations can be more easily interpreted (Fabrigar & Wegener, 2012). Where selection based on the Kaiser criterion, scree plot and model fit did not yield a 'clean' factor structure, models with an increased number of factors were explored to see whether this improved the interpretation of the model (i.e., the interpretability of each set of items in the respective factors). A clean factor structure is given when item loadings are all greater than 0.3 on at least one factor, and there are no or few cross-factor loadings (i.e., items that load greater than 0.3 on more than one factor) (Costello & Osborne, 2005). Where expansion of the number of factors failed to remove cross-loadings, the parsimonious model with fewer factors suggested by the Kaiser criterion, scree plot, and model fit statistics was selected as the preferred model.

Once a preferred factor model was identified for each pairwise comparison, using the procedure described above, overlap between the ICECAP-A and the respective HRQoL instrument was examined using the following criteria: (i) the number of common factors shared by both instruments, and (ii) the extent to which items from each instrument correlate with each shared common factor based on factor loadings. While the former refers to items of the ICECAP-A and the respective HRQoL instrument that contribute to the same underlying latent factor, the latter describes the strength of this contribution. The correlation among common factors was also examined to explore the extent to which the instruments in each pairwise comparison measure separate but correlated factors. The robustness of results was examined by comparing the overlap in the preferred factor model against the overlap in alternative factor models for each pairwise comparison. In addition, supplementary analyses were carried out within each of the seven disease groups, as a previous study has reported that major diseases are associated with relatively different impacts on health and capability (Mitchell *et al.*, 2015a).

5.3. Results

Data from 6,756 individuals were used in the analyses. Edit procedures were performed by the principal investigators of the MIC study, based upon a comparison of duplicated or similar questions, as well as completion time and can be found elsewhere (Richardson *et al.,* 2012). Such editing procedures resulted in the removal of 17% of the total sample. Given that the online program did not permit respondents to proceed until questions were completed, there were no missing data and only 14 individuals did not complete the final question (Richardson *et al.*, 2015a). Table 5.1 provides the characteristics of the study population for the combined sample and by country. Quota sampling was used in the MIC study and, therefore, the distributions of age, gender and education level were similar across the countries. The presence of a chronic disease was self-reported by the majority (78%) of the study population.

5.3.1. 'Preferred' factor models

Scree plots and the Kaiser criterion suggested a two-factor model for the EQ-5D-5L; a three-factor model for the 15D, HUI-3 and SF-6D; and a five-factor model for the AQoL-8D (see Appendices 5.2 – 5.11). In an attempt to improve model fit and interpretability, expansion of the number of factors was explored for all models. For the EQ-5D-5L and 15D this resulted in an improvement in the model fit and factor structure with fewer cross-factor loadings, supporting the superiority of a three- and four-factor model, respectively. For the HUI-3, moving to a four-factor model improved model fit but resulted in a poorer factor structure and the three-factor model was retained as the preferred model. With regard to the SF-6D, a four-factor model was preferred because of a better model fit and a cleaner factor structure. A six-factor model was explored for the AQoL-8D but this did not improve interpretability of the factor structure and the five-factor model was retained. Results for each pairwise EFA are provided in Tables 5.2 – 5.6.

Table 5.1. Characteristics of the study population (values are numbers (percentages) unless stated otherwise) [a]

	Australia (N=1,341)	Canada (N=1,330)	Germany (N=1,269)	UK (N=1,356)	USA (N=1,460)	Total (N=6,756)
Age (mean (SD))	53.9 (14.9)	48.7 (15.4)	49.2 (13.7)	51.6 (16.2)	52.0 (14.9)	51.1 (15.2)
Gender (Female)	686 (48.8)	826 (62.1)	591 (46.6)	670 (49.4)	900 (61.6)	3,707 (54.2)
Education [b]						
High school	472 (35.2)	388 (29.2)	249 (19.6)	517 (38.1)	527 (36.1)	2,193 (32.0)
Diploma or certificate or trade	469 (35.0)	633 (47.6)	698 (55.0)	409 (30.2)	428 (29.3)	2,670 (39.0)
University	400 (29.8)	309 (23.2)	322 (25.2)	430 (31.7)	505 (34.6)	1,982 (29.0)
Self-reported health condition						
Disease-free	265 (19.8)	328 (24.7)	260 (20.5)	298 (22.0)	321 (22.0)	1,472 (21.5)
Asthma	141 (10.5)	138 (10.4)	147 (11.6)	150 (11.1)	150 (10.3)	726 (10.6)
Cancer	154 (11.5)	138 (10.4)	115 (9.1)	137 (10.1)	148 (10.1)	692 (10.1)
Depression	146 (10.9)	145 (10.9)	160 (12.6)	158 (11.7)	168 (11.5)	777 (11.4)
Diabetes	168 (12.5)	144 (10.8)	140 (11.0)	161 (11.9)	168 (11.5)	781 (11.4)
Hearing problems	155 (11.6)	144 (10.8)	136 (10.7)	126 (9.3)	156 (10.7)	717 (10.5)
Arthritis	163 (12.2)	139 (10.5)	159 (12.5)	159 (11.7)	179 (12.3)	799 (11.7)
Heart	149 (11.1)	154 (11.6)	152 (12.0)	167 (12.3)	170 (11.6)	792 (11.6)

[a] Data from Norway is not used in this analysis because the ICECAP-A was not administered. SD=standard deviation.

[b] Categories refer to the 'highest level' of education.

5.3.2. Overlap with the ICECAP-A

Results suggest some degree of overlap between the ICECAP-A and the HRQoL instruments, although the extent of overlap varied across instruments. For the 15D EFA, two common factors were shared (Factors 2 and 4) (see Table 5.2). In each case, ICECAP-A dimensions did not load strongly onto the respective shared factor (autonomy (0.337) on Factor 2 and stability (0.307) on Factor 4) and the shared factor mostly explained variance in the 15D items. All five ICECAP-A dimensions loaded strongly onto Factor 1 – a factor that was not shared by any 15D items. However, Factor 1 was strongly correlated (r = 0.714) with Factor 4, which included the 15D items depression (0.841), distress (0.870), vitality (0.491), mental function (0.309), and sleeping (0.439).

Table 5.2. EFA comparing the ICECAP-A with the 15D (4-factor model) [a]

	Rotated item loadings			
	Factor 1	Factor 2	Factor 3	Factor 4
ICECAP-A				
Stability	**0.601**	0.029	-0.021	**0.307**
Attachment	**0.742**	-0.165	0.061	0.090
Autonomy	**0.497**	**0.337**	0.064	-0.078
Achievement	**0.762**	0.224	-0.034	0.006
Enjoyment	**0.802**	0.019	0.010	0.109
15D				
Mobility	0.060	**0.902**	0.039	-0.167
Vision	0.042	0.147	**0.468**	-0.006
Hearing	-0.028	0.012	**0.654**	-0.121
Breathing	-0.030	**0.563**	0.064	0.145
Sleeping	-0.010	**0.374**	0.010	**0.438**
Eating	-0.003	0.280	**0.648**	-0.024
Speech	0.024	-0.091	**0.815**	0.094
Elimination	-0.112	**0.371**	0.277	0.172
Usual activities	0.152	**0.812**	-0.001	0.029
Mental function	0.085	0.085	**0.474**	**0.309**

Discomfort and symptoms		-0.088	**0.763**	-0.034	0.197
Depression		0.126	0.014	0.009	**0.841**
Distress		0.028	-0.016	0.056	**0.870**
Vitality		0.086	**0.472**	-0.002	**0.491**
Sexual activity		0.096	**0.467**	0.057	0.240
			Correlations among factors		
	Factor 1	1.000			
	Factor 2	0.463*	1.000		
	Factor 3	0.422*	0.562*	1.000	
	Factor 4	0.714*	0.428*	0.396*	1.000
RMSEA	0.042 [90% CI: 0.040 to 0.044]				
CFI	0.991				
TLI	0.985				

RMSEA=root mean square error of approximation, CI=confidence interval, CFI=Comparative Fit Index, TLI=Tucker-Lewis Index.

[a] Loadings greater than 0.30 are presented in bold face. The highest loading for each item is underlined.

* Significant at 5% level.

The degree of overlap was much larger when comparing the ICECAP-A with the AQoL-8D. Three common factors (Factors 1-3) were shared by ICECAP-A and AQoL-8D items (see Table 5.3). Four ICECAP-A dimensions (stability (0.782), autonomy (0.345), achievement (0.634), and enjoyment (0.553)) and eighteen AQoL-8D items loaded onto Factor 1. Factor 2 was shared by ICECAP-A autonomy (0.415) and fourteen AQoL-8D items. Factor 3 included ICECAP-A attachment (0.682) as well as enjoyment (0.338) and six items of the AQoL-8D (social exclusion (0.307), close relationships (0.782), enjoying close relationships (0.842), pleasure (0.365), social isolation (0.338), and intimacy (0.581)). Strong correlations were observed between Factors 1 and 3 (r = 0.643), and Factors 1 and 4 (r = 0.641). Despite the strong correlation with Factor 1, Factor 4 was not a shared factor. Factor 4 comprised AQoL-8D items only, with the largest factor loadings being social exclusion (0.679) and social isolation (0.653).

Table 5.3. EFA comparing the ICECAP-A with the AQoL-8D (5-factor model) [a]

	Rotated item loadings				
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
ICECAP-A					
Stability	**0.782**	0.033	0.084	-0.045	-0.156
Attachment	0.161	-0.045	**0.682**	0.107	-0.070
Autonomy	**0.345**	**0.415**	-0.098	0.080	-0.231
Achievement	**0.634**	0.245	0.081	-0.080	-0.264
Enjoyment	**0.553**	0.117	**0.338**	-0.044	-0.140
AQoL-8D					
Energy level	**0.544**	**0.390**	0.020	-0.049	0.034
Social exclusion	0.060	0.043	**0.307**	**0.679**	-0.004
Getting around	0.073	**0.781**	0.033	0.107	-0.037
Community role	-0.035	**0.701**	0.046	**0.348**	0.005
Sadness	**0.742**	-0.041	0.012	0.234	0.176
Frequency of pain	0.017	**0.883**	-0.001	-0.020	**0.523**
Confidence	**0.721**	-0.054	0.021	0.132	-0.132
Calm	**0.765**	0.010	-0.044	0.026	0.130
Family role	0.042	**0.587**	0.110	**0.346**	0.013
Close relationships	-0.002	0.002	**0.782**	0.228	0.054
Communication	0.002	0.201	0.218	**0.301**	-0.097
Sleep	**0.392**	**0.348**	0.007	0.015	0.196
Feeling worthless	**0.653**	-0.003	0.055	0.275	0.024
Anger	**0.600**	-0.074	0.038	0.072	0.185
Mobility	-0.079	**0.937**	-0.005	-0.005	0.004
Self-harm	**0.488**	-0.004	0.071	0.268	0.039
Enthusiasm	**0.700**	0.068	0.177	-0.056	-0.055
Worry	**0.824**	-0.036	-0.133	0.127	0.102
Self-care	0.018	**0.796**	0.018	0.124	-0.062
Happiness	**0.735**	-0.027	0.287	-0.032	0.039
Coping	**0.775**	0.072	-0.019	0.070	-0.121
Degree of pain	-0.009	**0.920**	-0.001	-0.065	**0.555**
Enjoy close relationships	0.018	-0.053	**0.842**	0.101	0.021
Pain interference	0.026	**0.894**	0.008	0.020	**0.395**
Pleasure	**0.550**	0.034	**0.365**	-0.035	0.011

Feeling burden		**0.367**	**0.333**	-0.085	**0.339**	-0.075
Contentment		**0.723**	0.043	0.235	-0.021	-0.045
Vision		0.004	**0.389**	0.170	-0.046	0.013
Control		**0.809**	0.013	0.019	0.045	-0.144
Household tasks		0.026	**0.868**	-0.068	0.056	-0.036
Social isolation		0.092	0.045	**0.338**	**0.653**	-0.031
Hearing		-0.133	**0.359**	0.210	0.014	-0.006
Depression		**0.781**	-0.019	0.011	0.189	0.174
Intimacy		0.199	0.085	**0.581**	0.034	0.043
Despair		**0.733**	0.019	-0.035	0.214	0.124
			Correlations among factors			
	Factor 1	1.000				
	Factor 2	0.494*	1.000			
	Factor 3	0.643*	0.354*	1.000		
	Factor 4	0.641*	0.301*	0.336*	1.000	
	Factor 5	0.000	-0.222*	-0.081*	-0.048*	1.000
RMSEA	0.061 [90% CI: 0.061 to 0.062]					
CFI	0.974					
TLI	0.965					

RMSEA=root mean square error of approximation, CI=confidence interval, CFI=Comparative Fit Index, TLI=Tucker-Lewis Index.

[a] Loadings greater than 0.30 are presented in bold face. The highest loading for each item is underlined.

* Significant at 5% level.

The EQ-5D-5L shared two common factors with the ICECAP-A (Factors 1 and 3) (see Table 5.4). Four ICECAP-A dimensions (stability (0.803), attachment (0.798), achievement (0.658) and enjoyment (0.826)) and EQ-5D-5L anxiety/depression (0.703) loaded onto Factor 1. Factor 3 was primarily represented by the ICECAP-A autonomy (0.657) and achievement (0.426), as well as EQ-5D-5L self-care (0.301). A moderate correlation was found between Factor 1 and Factor 3 (r = 0.323), whereas a strong correlation (r = 0.685) was observed between Factor 3 and Factor 2, despite Factor 2 comprising EQ-5D-5L items only.

Table 5.4. EFA comparing the ICECAP-A with the EQ-5D-5L (3-factor model) [a]

	Rotated item loadings		
	Factor 1	Factor 2	Factor 3
ICECAP-A			
Stability	**0.803**	0.001	0.147
Attachment	**0.798**	-0.054	-0.003
Autonomy	0.279	0.020	**0.657**
Achievement	**0.658**	-0.015	**0.426**
Enjoyment	**0.826**	0.028	0.132
EQ-5D-5L			
Mobility	-0.137	**0.981**	0.003
Self-care	-0.008	**0.649**	**0.301**
Usual activities	0.041	**0.824**	0.112
Pain/Discomfort	0.002	**0.981**	-0.234
Anxiety/Depression	**0.703**	0.169	-0.020
	Correlation among factors		
Factor 1	1.000		
Factor 2	0.461*	1.000	
Factor 3	0.323*	0.685*	1.000
RMSEA	0.074 [90% CI: 0.069 to 0.078]		
CFI	0.993		
TLI	0.983		

RMSEA=root mean square error of approximation, CI=confidence interval, CFI=Comparative Fit Index, TLI=Tucker-Lewis Index.

[a] Loadings greater than 0.30 are presented in bold face. The highest loading for each item is underlined.

* Significant at 5% level.

The HUI-3 (Table 5.5) and SF-6D (Table 5.6) also shared two common factors with the ICECAP-A in the respective pairwise comparisons. All five ICECAP-A dimensions loaded onto the same factor as a single SF-6D item (energy (0.391)), and two HUI-3 items (emotion (0.895) and cognition (0.455)). In both models, ICECAP-A autonomy cross-loaded onto a second factor that was shared by ambulation (0.883), dexterity (0.576), and pain (0.719) in the HUI-3 EFA, and five items from the physical functioning and role limitation dimensions in the SF-6D EFA. Moderate correlations

were observed between the shared common factors for the respective pairwise comparisons. When comparing the ICECAP-A with SF-6D, a strong correlation (*r = 0.627)* was found between factor one (primarily represented by ICECAP-A dimensions) and factor three, onto which four SF-6D items loaded onto.

Table 5.5. EFA comparing the ICECAP-A with the HUI-3 (3-factor model) [a]

	Rotated item loadings		
	Factor 1	Factor 2	Factor 3
ICECAP-A			
Stability	**0.860**	0.020	-0.096
Attachment	**0.849**	-0.217	0.045
Autonomy	**0.448**	**0.368**	-0.013
Achievement	**0.762**	0.233	-0.106
Enjoyment	**0.912**	-0.023	-0.016
HUI-3			
Vision	-0.016	0.233	**0.324**
Hearing	-0.014	0.064	**0.617**
Speech	0.299	-0.007	**0.704**
Ambulation	-0.023	**0.883**	0.090
Dexterity	0.058	**0.576**	0.249
Emotion	**0.895**	-0.078	0.027
Cognition	**0.455**	0.121	**0.343**
Pain	0.107	**0.719**	0.022
	Correlations among factors		
Factor 1	1.000		
Factor 2	0.460*	1.000	
Factor 3	0.223*	0.282*	1.000
RMSEA	0.054 [90% CI: 0.051 to 0.058]		
CFI	0.990		
TLI	0.981		

RMSEA=root mean square error of approximation, CI=confidence interval, CFI=Comparative Fit Index, TLI=Tucker-Lewis Index.

[a] Loadings greater than 0.30 are presented in bold face. The highest loading for each item is underlined.

* Significant at 5% level.

Table 5.6. EFA comparing the ICECAP-A with the SF-6D (4-factor model) [a]

	Rotated item loadings			
	Factor 1	Factor 2	Factor 3	Factor 4
ICECAP-A				
Stability	**0.726**	-0.034	-0.150	0.017
Attachment	**0.780**	0.130	-0.004	0.066
Autonomy	**0.395**	**-0.423**	-0.102	-0.077
Achievement	**0.777**	-0.293	0.023	-0.068
Enjoyment	**0.859**	0.009	-0.010	0.094
SF-6D				
Vigorous activities	-0.088	**0.738**	-0.202	-0.171
Moderate activities	-0.014	**0.959**	-0.044	-0.030
Bathing or dressing oneself	0.036	**0.735**	0.188	-0.002
Limited in kind of work or other activities	0.098	**0.586**	0.288	-0.225
Accomplished less than you would like	-0.025	**0.311**	**0.656**	0.021
Frequency health problems interfered with social activities	-0.082	0.279	**0.512**	-0.187
Intensity of bodily pain	0.033	-0.075	0.030	**0.809**
Extent pain interfered with normal work	0.001	-0.015	-0.057	**0.968**
Been very nervous	-0.058	-0.027	**0.743**	-0.004
Felt downhearted and depressed	-0.230	-0.025	**0.729**	-0.015
Have a lot of energy	**0.391**	-0.250	-0.097	0.166

	Correlations among factors			
Factor 1	1.000			
Factor 2	-0.372*	1.000		
Factor 3	-0.627*	0.312*	1.000	
Factor 4	0.374*	-0.734*	-0.387*	1.000

RMSEA	0.075 [90% CI: 0.073 to 0.078]
CFI	0.985
TLI	0.971

RMSEA=root mean square error of approximation, CI=confidence interval, CFI=Comparative Fit Index, TLI=Tucker-Lewis Index.

[a] Loadings greater than 0.30 are presented in bold face. The highest loading for each item is underlined.

* Significant at 5% level.

5.3.3. Robustness of the preferred factor model

Comparing the extent of overlap in the preferred factor models against alternative (larger or smaller) factor models identified differences in overlap for the pairwise analyses comprising the 15D, EQ-5D-5L and HUI-3 (see Appendices 5.12 – 5.14, respectively). For the 15D, a three-factor model suggested a higher degree of overlap with the ICECAP-A than the preferred four-factor model. For the three-factor 15D model, four 15D items (depression (0.691), distress (0.619), vitality (0.439), and sleeping (0.314)) and all five ICECAP-A dimensions were explained by Factor 1. For the EQ-5D-5L, a two-factor model confirmed the strong loading of anxiety/depression onto Factor 1, but the remaining four EQ-5D-5L dimensions now shared a common factor with ICECAP-A autonomy, which loaded onto both common factors. Differences with regard to autonomy were also observed for the HUI-3. Unlike the preferred three-factor model, a four-factor model showed that autonomy loaded strongly on a factor that was not shared by any HUI-3 items.

5.3.4. Comparison across seven disease groups

Supplementary analyses, comparing the results of the EFA using pooled data with EFA conducted for each disease group, are reported in Table 5.7. Findings from the pooled sample were compared with the findings of seven disease areas in terms of the number of factors suggested by different selection criteria (i.e., Kaiser criterion, RMSEA, CFI/TLI, and a clean factor structure), the number of shared factors between the HRQoL measures and the ICECAP-A, and the number of items with loadings greater than 0.3 on the respective shared factors. Overall, results from the disease-specific analyses were consistent with the pooled data analysis. Notable deviations were observed in the depression group. Only one shared factor was identified for the 15D, EQ-5D-5L, and SF-6D in the respective pairwise comparisons with the ICECAP-A, indicating a low overlap that was also supported by factor loadings. Whereas the extent of overlap remained high for the AQoL-8D in the depression subgroup (i.e., three shared factors as in the pooled analysis), a fourth shared factor was identified for the HUI-3 EFA.

Table 5.7. Comparison of EFA results for the pooled data (n=6,756) and the seven disease areas

	Pooled (n=6,756)			Asthma (n=726)			Cancer (n=692)			Depression (n=777)		
15D												
# factors [a]	4			4			4			4		
# shared FL > 0.3 [b]	2 (F2 & F4)			2 (F2 & F4)			2 (F2 & F4)			1 (F3)		
ICECAP-A	F2= 1	F4=1		F2=1	F4=1		F2=2	F4=1		F3=1		
15D	F2=8	F4=5		F2=7	F4=4		F2=7	F4=4		F3=5		
AQoL-8D												
# factors [a]	5			5			7			7		
# shared FL > 0.3 [b]	3 (F1, F2, F3)			3 (F1, F2, F3)			3 (F1, F2, F3)			3 (F1, F2, F4)		
ICECAP-A	F1=4	F2=1	F3=2	F1=2	F2=2	F3=2	F1= 3	F2=2	F3= 2	F1= 3	F3=2	F4=2
AQoL-8D	F1= 18	F2= 14	F3=6	F1= 13	F2= 20	F3=7	F1= 20	F2=7	F3= 11	F1= 18	F3=5	F4=9
EQ-5D-5L												
# factors [a]	3			2			2			2		
# shared FL > 0.3 [b]	2			2			2			1		
ICECAP-A	F1=4	F3=2		F1=4	F2=1		F1=5	F2=1		F1=5		
EQ-5D-5L	F1=1	F3=1		F1=1	F2=4		F1=1	F2=4		F2=1		
HUI-3												
# factors [a]	3			4			4			4		
# shared FL > 0.3 [b]	2 (F1 & F2)			2 (F2 & F4)			3 (F1, F2, F3)			3 (F1, F2, F3)		
ICECAP-A	F1=5	F2=1		F2=4	F4=2		F1=4	F2=2	F3=2	F1=4	F2=1	F3=2
HUI-3	F1=2	F2=3		F2=2	F4=2		F1=1	F2=1	F3=3	F1=1	F2=5	F3=2
SF-6D												
# factors [a]	4			4			4			4		
# shared FL > 0.3 [b]	2 (F1, F2)			2 (F1, F2)			1 (F2)			1 (F1)		
ICECAP-A	F1=5	F2=1		F1=5	F2=1		F2=1			F1=5		
SF-6D	F1=1	F2=5		F1=1	F2=4		F2=6			F1=1		

[a] # factors refers to the number of factors suggested by the Kaiser criterion, RMSEA, CFI/TLI, and a clean factor structure.

[b] # shared FL > 0.3 refers to the number of common factors shared by both instruments with factor loadings (FL) greater than 0.3.

F1, F2, F3, and F4 refer to the shared factors in the respective pairwise comparisons, indicating the number of items (of the ICECAP-A and the HRQoL measure) loading onto the shared factors.

Table 5.7 Continued

	Diabetes (n=781)			Hearing problems (n=717)				Arthritis (n=799)			Heart disease (n=792)		
15D													
# factors [a]	4			3				4			4		
# shared FL > 0.3 [b]	2 (F2 & F4)			1 (F2)				3 (F1, F2, F4)			1 (F2)		
ICECAP-A	F2=1	F4=1		F2=1				F1=5	F2= 1	F4=1	F2=1		
15D	F2=4	F4=4		F2=11				F1=1	F2= 8	F4=5	F2=7		
AQoL-8D													
# factors [a]	6			7				5			5		
# shared FL > 0.3 [b]	3 (F1, F2, F3)			4 (F1, F2, F3, F4)				3 (F1, F2, F3)			3 (F1, F2, F3)		
ICECAP-A	F1=4	F2=2	F3=2	F1=4	F2=1	F3=1	F4=1	F1=4	F2=2	F3=2	F1=3	F2=2	F3=2
AQoL-8D	F1=16	F2=11	F3=4	F1=5	F2=11	F3=12	F4=5	F1=19	F2=13	F3=6	F1=17	F2=14	F3=7
EQ-5D-5L													
# factors [a]	2			2				2			2		
# shared FL > 0.3 [b]	2 (F1 & F2)			2 (F1 & F2)				2 (F1 & F2)			2 (F1 & F2)		
ICECAP-A	F1=5	F2=1		F1=5	F2=1			F1=5	F2=1		F1=4	F1=1	
EQ-5D-5L	F1=1	F2=4		F1=1	F2=4			F1=1	F2=4		F1=1	F2=4	
HUI-3													
# factors [a]	4			3				3			3		
# shared FL > 0.3 [b]	1 (F1)			1 (F1)				2 (F1 & F3)			2 (F1 & F3)		
ICECAP-A	F1=4			F1=5				F1=5	F3=1		F1=5	F3=1	
HUI-3	F1=1			F2=1				F1=2	F3=3		F1=1	F3=3	
SF-6D													
# factors [a]	4			4				2			2		
# shared FL > 0.3 [b]	2 (F1 & F2)			2 (F1 & F2)				2 (F1 & F2)			2 (F1 & F2)		
ICECAP-A	F1=5	F2=2		F1=4	F2=1			F1=2	F2=5		F1=5	F2=1	
SF-6D	F1=1	F2=6		F1=1	F2=6			F1=9	F2=5		F1=5	F2=8	

[a] # factors refers to the number of factors suggested by the Kaiser criterion, RMSEA, CFI/TLI, and a clean factor structure.

[b] # shared FL > 0.3 refers to the number of common factors shared by both instruments with factor loadings (FL) greater than 0.3.

F1, F2, F3, and F4 refer to the shared factors in the respective pairwise comparisons, indicating the number of items (of the ICECAP-A and the HRQoL measure) loading onto the shared factors.

A greater overlap between the HUI-3 and the ICECAP-A was also observed for the cancer group, where three shared factors similarly influenced HUI-3 items and ICECAP-A dimensions. On the contrary, for these two disease groups (depression and cancer), the SF-6D showed the lowest overlap with the ICECAP-A. For individuals with hearing problems, the 15D and HUI-3 shared only one common factor with the ICECAP-A, while a fourth shared factor was observed for the AQoL-8D EFA.

5.4. Discussion

The ICECAP-A was developed to overcome perceived limitations associated with existing preference-based instruments that focus primarily (but not only) on health-related aspects of QoL. Analyses in this chapter have shown that the ICECAP-A provides information over-and-above that garnered from several commonly used preference-based HRQoL instruments. However, the level of overlap with the ICECAP-A varied across instruments. Compared with other preference-based HRQoL instruments, more common factors were identified between the ICECAP-A and the AQoL-8D. Based on item loadings, these three common factors can be described as reflecting aspects of wellbeing (Factor 1), physical health (Factor 2), and relationships (Factor 3). Some but not all of these common factors emerged from other pairwise comparisons. The third factor, relationships, was not identified when comparing the ICECAP-A with the SF-6D, EQ-5D-5L, or HUI-3. Only one factor explained overlap with the 15D, which was related to aspects of physical health.

Compared with other studies, similar results were identified by recent studies that conducted an EFA with the ICECAP-A and the EQ-5D-3L (Keeley *et al.*, 2016), as well as with the ICECAP-O and EQ-5D-3L (Davis *et al.*, 2013). In these studies, the respective ICECAP instrument and the EQ-5D-3L measured two separate but correlated factors, with the majority of the EQ-5D-3L items loading onto one factor and the majority of the respective ICECAP items loading onto the second. Only EQ-5D-3L anxiety/depression loaded strongly onto the same factor as four dimensions of the

ICECAP-A (stability, attachment, achievement, and enjoyment) and ICECAP-O (attachment, security, role, and enjoyment), while ICECAP-A autonomy and ICECAP-O control loaded moderately onto both factors. The authors of the two previous EFA studies conclude that the EQ-5D-3L and ICECAP instruments provide complementary information and, therefore, should not be treated as substitute outcome measures. Specific to the EQ-5D-5L, these findings are confirmed by the current study due to the relatively minimal overlap observed with the ICECAP-A. Similar conclusions can be drawn about the 15D, HUI-3 and SF-6D, where relatively few items loaded onto the same common factor(s) as the ICECAP-A items. In contrast, the AQoL-8D provided good coverage of the three factors it shared with the ICECAP-A, with 18 AQoL-8D items loading on Factor 1 (wellbeing), 14 AQoL-8D items loading on Factor 2 (physical health), and six AQoL-8D items loading on Factor 3 (relationships).

The observed differences in overlap across instruments may be due, *inter alia*, to differences in the framing of items (e.g., question formats, response options, or recall time), based on evidence of previous comparative studies of preference-based HRQoL instruments (Richardson *et al.*, 2015b; Whitehurst & Bryan, 2011; Whitehurst *et al.*, 2014a). The combination of different health issues within a single item (e.g., anxiety and depression (EQ-5D-5L); downhearted and depressed (SF-6D); and sad, melancholic, or depressed (15D)) may also contribute to the differences observed between instruments. More generally, the fact that the instruments included in this study differ in the way they conceptualize HRQoL (Karimi & Brazier, 2016), and in their coverage of domains to define health states, is likely to be a primary reason for the variation in study findings (Richardson *et al.*, 2015b; Richardson *et al.*, 2014a). To illustrate, compared with other instruments, the AQoL-8D has a strong focus on the psycho-social domain (25 of 35 items) and contains questions in its descriptive system that have the greatest ability to capture the concept of capability wellbeing, or wellbeing in general. As has been shown in a previous publication using data from the MIC study, which compared three subjective wellbeing instruments (Satisfaction with Life Scale, Personal Wellbeing Index, and the Integrated Household Survey of the Office for National Statistics) with preference-based HRQoL instruments, the AQoL-8D accounted for variation in

subjective wellbeing to a greater extent than the other preference-based HRQoL instruments (Richardson *et al.*, 2015a), particularly in individuals with depression. Also the comparison analyses across seven disease groups in this chapter have shown that the AQoL-8D had the greatest overlap with the ICECAP-A in the depression group, whereas all other measures (except for the HUI-3) seemed to tap into different constructs than the ICECAP-A. It has been previously reported that the AQoL-8D correlates highly with depression-specific instruments (Mihalopoulos *et al.*, 2014), indicating that both the AQoL-8D and the ICECAP-A are sensitive to aspects that are predominantly important to individuals with depression.

5.4.1. Implications and directions for further research

The analyses performed in this chapter have shown that the ICECAP-A, when compared directly to the 15D, EQ-5D-5L, HUI-3 and SF-6D, provides additional, complementary information about an individual's capability wellbeing. Recent studies have demonstrated that the choice of outcome measure for economic evaluation – i.e., selecting a capability measure *or* a HRQoL measure – is not a trivial issue (Goranitis *et al.*, 2017; Makai *et al.*, 2015). In an economic evaluation of an integrated care model for frail seniors, Makai and colleagues found the intervention had a higher probability of being cost-effective when using the ICECAP-O when compared with use of the EQ-5D-3L. This direct comparison of cost-effectiveness findings was made possible because (i) ICECAP-O responses were used to define 'capability QALYs' (which is in tension with the intent of the ICECAP measures) and (ii) the same range of willingness to pay (WTP) values was applied in the analysis of capability QALYs and QALYs derived from EQ-5D-3L responses. Despite the use of identical economic evaluation approaches, Makai and colleagues go on to highlight that there are no estimates of WTP for a capability QALY, and state that it is unlikely that valid comparisons can be made between the ICECAP-O and EQ-5D-3L at a given level of WTP. A second example examined the cost-effectiveness of psychological interventions for drug addiction (Goranitis *et al.*, 2017), concluding that under the health maximization principle (using EQ-5D-5L), the results yielded different treatment recommendations when compared with the application

of the ‘sufficient capability’ approach developed by Mitchell and colleagues (using ICECAP-A) (Mitchell *et al.*, 2015b).

Although methodologies to operationalize the use of ICECAP instruments in economic evaluation are still in their infancy (Mitchell *et al.*, 2015b), findings such as those in the above examples support the use of ICECAP instruments alongside preference-based HRQoL instruments to triangulate results and evaluate the robustness of conclusions regarding cost-effectiveness. However, the use of different metrics to value different health care interventions raises questions about the objective for resource allocation decisions in health care (Mitchell *et al.*, 2015a). For example, does health or wellbeing (or both) enter the objective function, and is the form of this function consistent with the current emphasis on maximization (rather than sufficiency)? To answer this question, further research is needed to determine whether a society is willing to sacrifice health outcomes for improvements in dimensions of wellbeing.

The use of ICECAP instruments within the current QALY-based paradigm for economic evaluation also requires further attention in health economics research. As mentioned above, the ICECAP-A is anchored on a ‘full capability’ and ‘no capability’ scale and the instrument was not intended to be used within the QALY framework (Flynn *et al.*, 2015). Recent advances in this area have proposed to adjust the ICECAP-A for time, to enable the assessment of gains in terms of ‘years of full capability equivalence’ (Flynn *et al.*, 2015), and an approach that focuses on the objective of achieving ‘sufficient capability’ (Mitchell *et al.*, 2015b). Outside the ICECAP instruments, Cookson suggested an application of the capability approach to economic evaluation by re-interpreting the QALY, referred to the ‘capability QALY’ (Cookson, 2005b). Cookson argued that, in practice, HRQoL instruments incorporate some elements of capability because health affects an individual’s freedom to choose non-health activities. Compared with the ‘health QALY’, this operationalization of the ‘capability QALY’ represents individuals’ entire wellbeing (not just the health component) and, therefore, reflects the value of the capability set. Concerns over using preference-based HRQoL instruments as the base of a capability QALY because they may neglect non-health dimensions of

wellbeing led Cookson to conclude that, "*…the QALY approach is compatible with the capability approach only insofar as the health state descriptive systems used for generating QALYs pay close attention to proxy capability variables that cover a wide range of health and non-health dimensions of wellbeing*" (Cookson, 2005a, p.1288). Results from the current chapter suggest the AQoL-8D could be a measure that best fits Cookson's notion of a capability QALY because of the overlap with the ICECAP-A and the presence of non-health items in the AQoL-8D descriptive system.

In a recent review, Karimi and colleagues conclude that existing capability measures (including ICECAP instruments) have important limitations since they do not elicit capability as originally proposed by Sen (Karimi *et al.*, 2016). Accordingly, if the added value of capability instruments in health economics is based solely on broadening the evaluative space to extend beyond a narrow focus on health, the findings provide evidence that such benefits can be potentially captured by the AQoL-8D (i.e., not only through the aggregation of outcomes collected by 'complementary' health-related and capability measures). However, the findings do not imply the AQoL-8D and ICECAP-A are interchangeable instruments. Further work is needed to build on these findings and explore unanswered question, such as whether individuals are able to distinguish between their capabilities and functionings, and the comparative performance of the ICECAP-A and AQoL-8D with regard to capturing the wellbeing impacts of interventions in different clinical contexts. It is also important to note that ICECAP instruments are not the only capability measures that could be combined with QALYs derived from HRQoL instruments to provide a broader assessment of the benefit of interventions. For example, the Adult Social Care Outcomes Toolkit (ASCOT) (Netten *et al.*, 2012) is designed to capture information about an individual's social care-related quality of life and further research is needed to explore the relationships (including overlap) between the ASCOT, ICECAP instruments and preference-based HRQoL instruments.

5.4.2. Strengths & limitations

A major strength of this work is the inclusion of multiple preference-based instruments. While previous studies explored overlap between ICECAP instruments and

the EQ-5D-3L (using much smaller samples (Davis *et al.*, 2013; Keeley *et al.*, 2016)), the analyses reported in this chapter provide EFA results comparing the ICECAP-A with five preference-based HRQoL instruments. Conducting an EFA that uses data from the descriptive systems only (i.e., item-level response data) is a further strength because there is no reliance on country-specific index scores, where variations across national valuation studies could influence the results (Engel *et al.*, 2016; Xie *et al.*, 2013). Given that 'overlap' between instruments can be explored within the descriptive systems or health state valuations, this item-level analysis complements previous work that used correlation analyses and regression-based techniques to assess index scores from the MIC study (Mitchell *et al.*, 2017b). The analysis also addressed the potential problem of factor under- or over-extraction (Prieto *et al.*, 2003) by investigating alternative factor models to examine the robustness of the 'preferred' factor models (see section 5.3.3). Potential limitations associated with using data from a multinational survey include issues with the validity of instrument translations and the representation of the respective populations (for example, participants were required to have Internet access). Survey bias resulting from the repetition of similar items should also be acknowledged due to the administration of seven preference-based HRQoL instruments.

5.5. Conclusion

The ICECAP-A has the potential to capture benefits of interventions and treatments that go beyond those measured by many of the traditional health-focused preference-based instruments, such as the 15D, EQ-5D-5L, HUI-3 and SF-6D. Substantial overlap was observed between the ICECAP-A and AQoL-8D. Researchers and decision makers should be aware that there is a risk of double counting when using the ICECAP-A as a complementary measure, but the level of such a risk varies depending on the choice of HRQoL measure. Further investigations are needed to explore the extent and implications of double counting, particularly when applying the ICECAP-A alongside the AQoL-8D, which will be the focus of the next chapter.

Chapter 6. The relationship between health-related quality of life, capability wellbeing, and subjective wellbeing in the context of spinal cord injury – a path analysis

6.1. Introduction

The analyses conducted in Chapter 5 examined the relationship between preference-based HRQoL measures that are used within the existing QALY framework and the ICECAP-A, which is rooted in the capability theory. The findings have shown that the ICECAP-A contains domains in its descriptive system that are not captured by preference-based HRQoL measures. The exception was the AQoL-8D, where a substantial degree of overlap was observed with the ICECAP-A. While the previous chapter examined the underlying content of alternative measures, further work is needed to analyze whether illness, symptoms and functional limitations have a different impact on HRQoL and capabilities. Given that existing preference-based HRQoL measures aim to assess a person's functionings, while the ICECAP-A focuses on a person's capability wellbeing, the ability of alternative instruments to capture the full impact of an intervention can differ despite conceptual overlap between measures. A previous study compared the impact of seven major health conditions on health status (using the EQ-5D-5L) and capability (using the ICECAP-A) and found that diseases were associated with relatively different impacts on *health* and *capability* (Mitchell *et al.*, 2015a). For example, the study found that the impact of depression on capability is much greater than on health when compared across other conditions. The authors of the study concluded that the relative importance of preventing and treating different conditions would differ depending on whether the focus is on capability or health status.

Although there is an increasing interest for capturing broader benefits of health care interventions, the conceptualization of such a broader approach does not need to be necessarily in terms of capabilities. As was discussed in Chapter 3, another alternative is the measurement of subjective wellbeing (SWB) or happiness for informing policy decisions (Dolan, 2011; van de Wetering *et al.*, 2016). It has been argued that SWB better reflects the 'utility experienced' from changes in health status (van de Wetering *et al.*, 2016), in which context the term utility can be equated with hedonic experience (Kahneman & Sugden, 2005). Compared with preference-based HRQoL measures, SWB uses individuals' own values and not those of the general population, as is predominantly the case within the current QALY framework (Versteegh & Brouwer, 2016). Rather than placing participants in hypothetical scenarios to value different states of health, known as 'decision utility', the focus is entirely on the measurement of an individual's real happiness, life satisfaction, or SWB (Coast *et al.*, 2008d). Hence, decision utility is an *ex ante* concept (Brazier *et al.*, 2005), whereas experienced utility is an *ex post* concept because it reflects the hedonic experiences that result from acts of choices (Kahneman & Sugden, 2005). Dolan *et al.* have criticized the measurement and valuation of health within the current QALY approach based on two reasons: (i) describing health in terms of a fixed and simplified descriptive system may fail to capture what is important to people in terms of their health and important benefits of health care, and (ii) valuing health in terms of hypothetical preferences may fail adequately to anticipate the real impact that different health states have on our lives (Dolan *et al.*, 2012). Since all descriptive systems have a fixed, limited number of dimensions and response levels for practical reasons, the advantage in SWB is that individuals can decide which factors are important for them and the magnitude of these factors. Therefore, SWB accounts for the domains of health that matter most in the experience of people's lives (Dolan *et al.*, 2012).

With this current spread of different measurement approaches for the evaluation of health care interventions, it remains unclear which 'objects of value' should be considered in an economic evaluation. Originally introduced by Sen (Sen, 1993), objects of value relate to *what* should be valued (the domains) and *why* it should be valued (its consequences) that define ultimately the evaluative space of the economic evaluation

(Sampson, 2016). Although it is known that measures of HRQoL, capability and SWB differ in terms of the domains in their descriptive systems and their consequences (functionings, capabilities, or satisfaction), it is not fully understood how diseases influence different objects of value. These three measures under consideration were previously discussed in different contexts. Greco and colleagues outlined methodological challenges for the identification and measurement of broader outcomes of public health interventions in economic evaluations in low-income and middle-income countries (Greco *et al.*, 2016). Also within the context of patients without capacity, which includes dementia, Round and colleagues discussed these three outcome measures and highlighted existing limitations around these measures when applying to individuals with severely restricted capacity (Round *et al.*, 2014). Generally, little is known about the comparative performance of these measurement approaches. Therefore, the aim of this chapter is to provide further evidence on the relationships between HRQoL, capability wellbeing, and SWB. These relationships will be explored within a particular clinical context – here the context of spinal cord injury (SCI), which can be considered as a case study.

Given the improvement in survival rates following injury, measuring QoL (however defined) in individuals with SCI has become an important outcome measure (Strauss *et al.*). As perceived QoL changes, the measurement of QoL requires appropriate measures and a recent panel of experts have recommended the use of the International SCI QoL Basic Data Set, consisting of three questions concerned with the (i) satisfaction with general QoL, (ii) satisfaction with physical health, and (iii) satisfaction with psychological health (Charlifue *et al.*, 2012). While this measure closely aligns with the concept of SWB that reflects the hedonic experiences of SCI, previous research emphasized the use of conventional preference-based HRQoL measures, in particular, when evaluating the cost-effectiveness of interventions in SCI (Whitehurst & Mittmann, 2013; Whitehurst *et al.*, 2012). Additionally, Sen's capability approach is gaining increasing attention in the disability literature, which emphasizes the need to assess what individuals are able to do in their real-life environment (capabilities) rather than capacity or functional status (Mitra, 2006; Welch Saleeby, 2006). Given this interest in different

outcome measures, the clinical context of SCI provides a great opportunity for further exploration of the relationships between such measures.

QoL in individuals with SCI is influenced to a great extent by the presence of secondary health conditions (SHCs) (Craven *et al.*, 2012). SHCs are defined as physical or psychological health conditions that are influenced directly or indirectly by the presence of a disability or underlying physical impairment (Jensen *et al.*, 2012). These include, for example, pressure sores, spasticity, and pain. Prevention, early diagnosis and treatment of SHCs are seen as critical for limiting these complications, improving survival, and enhancing QoL (Sezer *et al.*, 2015). In order to further explore the relationship between different measurement approaches in SCI, a better understanding of the impact of SHCs is needed. Therefore, the objectives of this work are twofold: (i) to examine the effect of SHCs simultaneously on HRQoL, capability wellbeing, and SWB through direct and mediated pathways within the same model; and (ii) to consider the implications for outcome measurement in health economic evaluation.

6.2. Methods

6.2.1. Data source

Data were used from a study that explored the validity of alternative preference-based HRQoL measures in the context of SCI (Whitehurst *et al.*, 2016). Study participants were identified from a list of individuals who participated previously in a Canada-wide SCI research project, known as the Spinal Cord Injury Community Survey (SCICS) (Noreau *et al.*, 2014). A market research company, in collaboration with the research team, invited by email those individuals who indicated a willingness to be contacted for further related research studies during the consent procedure in the SCICS study. Individuals were included if they were 19 years or older, were able to understand the English language, and had an SCI (regardless of the type and cause of injury). Data were collected online through the administration of a survey and participants were reimbursed upon completion of the survey in the form of a $25 gift card. The study was

approved by the University of British Columbia Behavioural Research Ethics Board (H12-01138) and Vancouver Coastal Health Authority (Research Study #V12-01138).

6.2.2. Instruments

The survey consisted of five sections: (i) demographics, (ii) SCI classifications and characteristics, (iii) secondary health complications and conditions, (iv) quality of life, and (v) SCI-specific functioning in activities of daily living. Study participants were expected to respond to all questions (in all five sections), in the order they were presented, with the option to return to previous questions for amendments. To prevent order-effect bias, the appearance of standardized outcome measures included in section four were randomized (Perreault, 1976). Although four preference-based HRQoL instruments were included in the survey (Assessment of Quality of Life 8-dimension (AQoL-8D) (Richardson *et al.*, 2009), EQ-5D-5L (Herdman *et al.*, 2011), Health Utilities Index Mark (HUI) (Feeny *et al.*, 2002), and SF-6D (based on the 36-item Short Form health survey version 2 (SF-36v2) (Ware *et al.*, 2008)) (Brazier *et al.*, 2002)), the analysis reported in this chapter focuses on two instruments only; the EQ-5D-5L and AQoL-8D. The EQ-5D-5L was selected because the previous three-level version, the EQ-5D-3L, is the most often used instrument (Richardson *et al.*, 2015c; Wisløff *et al.*, 2014). The AQoL-8D was chosen as a comparator measure of HRQoL because qualitative evidence indicates that individuals with SCI perceived it to be the most relevant instruments to be administered within the SCI population (Michel *et al.*, 2016; Whitehurst *et al.*, 2014b). Previous studies have also shown that the AQoL-8D generates different results compared with the EQ-5D-5L, mainly due to variation in the descriptive systems (Richardson *et al.*, 2015b; Whitehurst *et al.*, 2016). This was also observed in the previous chapter, where a different degree of overlap was found between the EQ-5D-5L and ICECAP-A when compared with the overlap between the AQoL-8D and the ICECAP-A. While results of the HUI-3 and SF-6D were similar to those of the EQ-5D-5L, the EQ-5D-5L and AQoL-8D seem to deviate most on their descriptive systems, which enables further exploration within this chapter. The Canadian tariff was used for the EQ-5D-5L (Xie *et al.*, 2016)

while for the AQoL-8D only an Australian tariff exist, which was applied in this work (Richardson *et al.*, 2014b).

Capability wellbeing was assessed using the ICECAP-A (Al-Janabi *et al.*, 2012), based on UK societal preferences (Flynn *et al.*, 2015). Further details about the ICECAP-A were provided in the previous Chapter 5. SWB was assessed by a single item, which asks: "*Using a scale of 0 to 10, where 0 means "very dissatisfied" and 10 means "very satisfied", how do you feel about your life as a whole right now?*" Respondents were asked to tick one of eleven boxes (0 to 10), which were presented vertically on screen (see Appendix 6.1). The same question format has been used in waves of the Canadian Community Health Survey (CCHS) (Statistics Canada, 2010). For the assessment of SHCs, a modified version of the Spinal Cord Injury Secondary Conditions Scale (SCI-SCS) was used (Kalpakjian *et al.*, 2007), based on a combination of items from the SCI-SCS and additional conditions that were selected by members of the Research Team from the Rick Hansen Institute (see Appendix 6.2). Although some conditions (e.g., arthritis) are not of 'secondary' nature, the majority of the health conditions included in Appendix 6.2 are directly or indirectly influenced by the presence of SCI and, as such, summarized under the umbrella term SHCs. In total, sixteen SHCs were assessed, each on a four-point ordinal scale, with scores ranging from 0 (not experienced/insignificant problem) to 3 (significant/chronic problem). A sum score was created with total scores ranging from 0 (no SHCs) to 48 (significant/chronic problem in all sixteen SHCs).

6.2.3. Statistical analyses

Descriptive analyses were performed in STATA 14.1 (StataCorp, 2015) for frequency, means, standard deviations, median, and minimum and maximum values of all outcome measures (AQoL-8D, EQ-5D-5L, ICECAP-A, SHCs, and SWB). The primary analyses consisted of two path analysis models that were both conducted in Mplus 7 (Muthén & Muthén, 1998-2015). A path analysis is a multivariate regression model, which is a regression analysis that simultaneously considers multiple dependent (and often multiple independent) variables (Geiser, 2013). In other words, equations are solved simultaneously to determine parameter estimates. It is an extension of multiple regression

analysis in the sense that it is not restricted to a single dependent variable. In a path model, a variable can be a dependent variable in one relationship and an independent variable in another. Given the example in Figure 6.1, if one wants to analyze the effect of X on Z through Y, Y is an independent variable to Z but a dependent variable to X.

Different terminology is used in path analysis compared with a multiple regression analysis, where a distinction is made between *exogenous variables* and *endogenous variables*. Exogenous variables are referred to as independent variables because variability in these variables is not explained by other variables in the model. Variables that are regressed on one or more other variables are called endogenous variables (i.e., dependent variables). Endogenous variables, that serve as independent and dependent variables at the same time, are called *intermittent* or *mediator variables* (Geiser, 2013). In a path analysis, hypotheses can be tested with regard to the direct and indirect effects of variables on each other. While direct effects are effects of one variable on another, indirect effects are effects that are mediated through other variables and are therefore also referred to as mediated effects. Mediated effects are present in a path model when the model contains one or more variables that are dependent and independent variables at the same time.

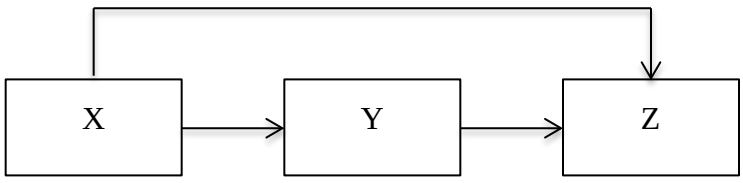


Figure 6.1. Example of a path analysis model

Two separate models were conducted in this chapter; one that measured HRQoL using the EQ-5D-5L (Model 1) and a second model that measured HRQoL using the AQoL-8D (Model 2). For each model the relationship between SHCs, HRQoL, capability wellbeing, and SWB was examined through direct and mediated pathways (further model details are provided in the next section). All paths in the model were adjusted for sex,

age, education, type of injury (paraplegia or tetraplegia), and living arrangements (living alone or living with someone) based on previous evidence (Noonan *et al.*, 2008; Post *et al.*, 1998; Putzke *et al.*, 2002; van Leeuwen *et al.*, 2012). It is generally recommended that the final model in a path analysis should contain only parameters that are statistically significant, i.e., a parsimonious model. Therefore, after running the initial hypothesized model in Mplus, the model was further refined by removing non-significant parameters to yield a parsimonious final model (Raykov & Marcoulides, 1999).

For a path analysis that contains one or more indirect effects, there is often an interest in effect decomposition, where the total effect is split into the sum of all indirect effects plus the direct effect. For the analyses conducted in this chapter, the MODEL INDIRECT command was used in Mplus to estimate the total, direct, and indirect effects. The default maximum likelihood (ML) estimation was applied. Running a path analysis generates path coefficients that characterize the strength of the influence of one variable on another. These coefficients are equivalent to slope coefficients in linear regression analysis (Geiser, 2013). Both unstandardized and standardized beta coefficients are reported (the standardized coefficients enable comparisons to be made across independent variables due to being scaled on the same standardized metric). While a direct effect is expressed by one coefficient, an indirect effect is a product of two or more regression coefficients. Therefore, if one wants, for example, analyze the effect of X on Z through Y, the indirect effect of X on Z is the increase we would see in Z while holding X constant and increasing Y to whatever Y would attain under a unit increase of X (Pearl, 2012). Given that the indirect effect is the product of two or more regression coefficients, the product is often not normally distributed, meaning the application of conventional tests of significance can be problematic (Geiser, 2013). Accordingly, the bias-corrected bootstrap method (1,000 samples) was applied to verify the significance of indirect effects (MacKinnon, 2008). Same model fit indices were used as in the previous chapter. Model fit was deemed to be good given a non-significant chi-square test, an RMSEA $\leq$ 0.5, and a CFI/TLI > 0.95 (Geiser, 2013). Significance level was set at $p < 0.05$.

6.2.4. Conceptual framework

Path analysis is, generally, considered as a model-testing approach and not a model-building approach (Streiner, 2005). This implies that analysts should not include variables just because they are available; instead the selection of variables should be based on theory or empirical evidence. For this reason, having a model that fits the data does not prove that the model is correct. The major criterion for accepting or rejecting the model is the underlying theory. The conceptual model used for this analysis is presented in Figure 6.2. The model considers SWB as an endogenous variable, SHCs as exogenous variables, and HRQoL (i.e., AQoL-8D or EQ-5D-5L) and capability (ICECAP-A) as mediating variables. Each path (indicated by capital letters) in the conceptual model was based on a series of theory-driven hypotheses. The construction of *a priori* hypotheses required careful considerations of key differences between the outcome measures and their theoretical foundation. Table 6.1 provides an overview of the main differences between these three measurement approaches under consideration (i.e., HRQoL, capability wellbeing and SWB).

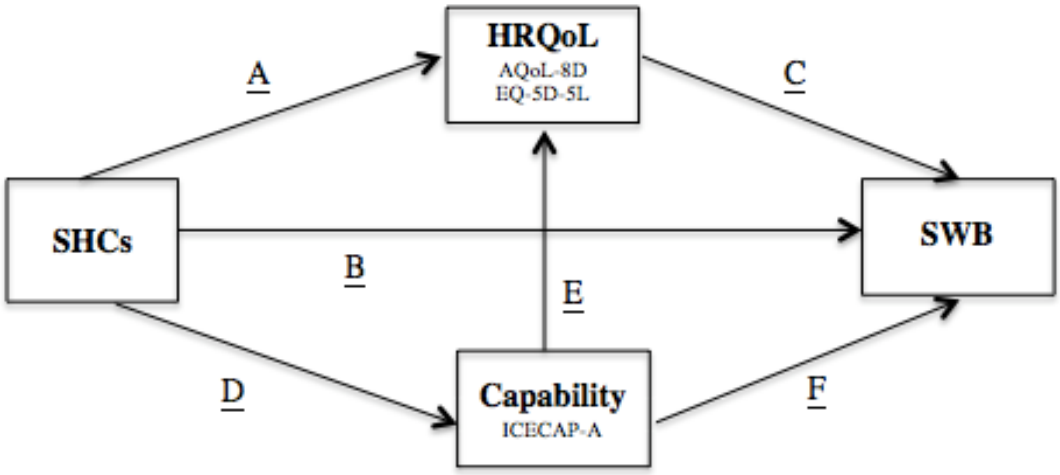


Figure 6.2. Hypothesized conceptual model of the effects of SHCs on HRQoL, capability, and SWB

Table 6.1. Key characteristics of the three measurement approaches under consideration [a]

	Question to be addressed	Decision metric	Objects of value (*what* is valued?)	Consequences (*why* it is valued?)	Source of values
HRQoL	Which health states are more preferred?	Preferences	Fixed descriptive system EQ-5D-5L (Mobility, Self-care, Usual activities, Pain/discomfort, Anxiety/depression) AQoL-8D (Independent living, Senses, Pain, Mental health, Happiness, Self Worth, Coping, Relationships)	Functionings	Society
Capability	What is the effect of these health states on individual's capability to function?	Perceived capabilities	Fixed descriptive system ICECAP-A (Attachment, Stability, Achievement, Enjoyment, Autonomy)	Capabilities	Society
SWB	What is the actual experience of the various health states?	Experiences	No fixed descriptive system	Satisfaction	Affected individual

HRQoL=health-related quality of life, SWB=subjective wellbeing.

[a] The description of the characteristics is specific to the outcome measures applied in this chapter and may deviated from other measures.

Driven by these underlying differences, a-priori hypotheses were formulated, which are presented in Table 6.2. Hypothesis #1 proposed that SHCs would impact HRQoL negatively, as highlighted by a previous model of HRQoL (Wilson & Cleary, 1995) and empirical evidence in SCI (Craven *et al.*, 2012; Rivers *et al.*, 2017). Hypothesis # 2 also suspected a negative impact of SHCs on SWB (Rivers *et al.*, 2017). However, previous evidence has shown that individuals with SCI use coping strategies that are associated with greater life satisfaction (Anderson *et al.*, 2008). As such, compared with the effect of SHCs on HRQoL, the effect on SWB is hypothesized to be of a smaller magnitude (Hypothesis #3). This hypothesis is based on societal preferences reflected in the preference-based HRQoL measures, where the general public might not adequately forecast the coping strategies and adaptation one might experience in such a health state (Versteegh & Brouwer, 2016).

Table 6.2. List of hypotheses tested in the path analysis

#	Path	Hypothesis
1	A	SHCs impact HRQoL negatively
2	B	SHCs impact SWB negatively
3	A vs. B	SHCs have a less negative effect on SWB than on HRQoL
4	A C	There is an indirect effect of SHCs on SWB mediated through HRQoL
5	D	SHCs impact capability wellbeing negatively
6	A vs. D	SHCs have a less negative impact on capability wellbeing than on HRQoL
7	B vs. D	SHCs have a less negative effect on SWB than on capability wellbeing
8	D E	There is an indirect effect of SHCs on HRQoL mediated through capability wellbeing
9	D E C	The indirect effect of SHCs on SWB is mediated through capability wellbeing and health functionings
10	D F	There is an indirect effect of SHCs on SWB mediated through capability wellbeing

[a] HRQoL=health-related quality of life, SHCs=secondary health conditions, SWB=subjective wellbeing.

In addition to the direct effect of SHCs on SWB, an indirect effect mediated through HRQoL was also hypothesized (Hypothesis #4). Broome previously discussed that the harm done by disease should not be understood in terms of the consequences for health but the contribution health makes to wellbeing (Broome, 2002). The rationale behind this notion is that the extent to which a person's wellbeing is affected by the various elements of their health depends on other non-health factors that are unlikely to be included in measures of HRQoL (such as the EQ-5D-5L or AQoL-8D).

Additional evidence-based hypotheses were formulated around the impact of SHCs on capability wellbeing. Hypothesis #5 proposed a negative relationship between SHCs and capability wellbeing (Keeley *et al.*, 2015; Mitchell *et al.*, 2015a). As none of the ICECAP-A dimensions (stability, attachment, autonomy, achievement, and enjoyment) are directly related to health (Kinghorn, 2015), the effect of SHCs on capability wellbeing focuses on the extent to which SHCs will affect a person's freedom to choose non-health activities and states (i.e., non-health capabilities). As previous evidence has indicated that certain conditions have a greater effect on health than the impact on capability wellbeing (Mitchell *et al.*, 2015a), the model explored in this analysis also hypothesized (Hypothesis #6) that the impact of SHCs on HRQoL will be greater than on capability wellbeing. Another hypothesis was made with regard to the impact on capability wellbeing when compared with the impact on SWB (Hypothesis #7). Similar to Hypothesis #3, a greater negative impact of secondary health conditions on capability wellbeing was hypothesized than on SWB because of the adaptation process reflected in SWB measures.

Hypothesis #8 assumed that there would be an indirect effect of SHCs on HRQoL mediated through capability wellbeing. Kinghorn previously discussed different interpretations of the capability approach in a health care context (Kinghorn, 2015). One important consideration is whether health should be treated as the 'end goal' or as a 'means to an end'. If health is to be treated as the end goal, the assessment of capability will be based upon an assessment of the capability to exercise good health. If health is treated as a means, the assessment of capability will be based on a broader capability set,

which is influenced by health and the social care system (Kinghorn, 2015). Given the fact that the ICECAP-A does not include any dimension directly related to health, this model will consider health as an end goal. As such, the model assumed that health achievements observed would depend on the capability to achieve good health. In addition, it was assumed that this indirect effect would also influence SWB (Hypothesis #9). It was assumed that if SHCs lead to a decrease in capability wellbeing, it would also decrease the levels of health functionings, which will decrease a person's SWB. Finally, Hypothsis #10 proposed an indirect effect of SHCs on SWB, mediated through capability wellbeing only. The rationale behind this hypothesis is that secondary health conditions could impact a person's capability wellbeing that influence their SWB, irrespective of the achieved levels of health functionings. For example, secondary health conditions could impact SWB through non-health capabilities (e.g., achievement) independently of HRQoL.

6.3. Results

The total sample size comprised 364 individuals. Characteristics of the study population are provided in Table 6.3. The majority of the study population was male (63%), with 90% self-identifying as Caucasian. With regard to levels of education, 33% of the study population had a Bachelor degree and above. The population contained slightly more individuals with paraplegia (52%) than tetraplegia. The majority of the study population (70%) had lived with their injury for at least 10 years.

Table 6.3. Characteristics of the study population – values are numbers (percentages) unless states otherwise (n=364)

Gender - Male	229 (62.9)
Age (mean (SD))	50.4 (13.2)
Ethnicity	
Caucasian	328 (90.1)
Other	36 (9.9)
Living with someone - Yes	269 (73.9)
Education level	
High school/GED and below	102 (28.0)
Diploma/Certificate	140 (38.5)
Bachelor degree and above	119 (32.7)
Type of injury	
Tetraplegia	175 (48.1)
Paraplegia	189 (51.9)
Time since injury	
1-10 years	109 (29.9)
10+ years	225 (70.1)

GED=General Educational Development, SD=standard deviation.

Table 6.4 reports the results of the descriptive statistics for the instruments under consideration. Total scores of SHCs, indicating frequency and magnitude of conditions, ranged from 2 to 40 with a mean of 18. Mean HRQoL based on the AQoL-8D was higher (0.573) compared with the EQ-5D-5L (0.492). Mean scores for the ICECAP-A and SWB were 0.761 and 6.319 respectively.

Table 6.4. Descriptive statistics for instruments under consideration

Instrument	N	Mean	SD	Min	Max	Range of possible scores
SHCs	364	18.363	7.731	2	40	[0; 48]
ICECAP-A	364	0.761	0.182	0.149	1.000	[0; 1]
EQ-5D-5L	364	0.492	0.196	-0.066	0.911	[-0.148; 0.974]
AQoL-8D	363	0.573	0.197	0.134	0.995	[0.09; 1]
SWB	364	6.319	2.248	0	10	[0; 10]

SD=standard deviation, Min=minimum, Max=maximum, SHCs=secondary health conditions, SWB=subjective wellbeing.

The initial model structures (considering all paths and parameters) for Model 1 and Model 2 are provided in Appendix 6.3 and Appendix 6.4 respectively, showing all parameter estimates (A) and only significant parameter estimates (B). After removing non-significant parameters, the final models for Model 1 and Model 2 are presented in Figure 6.3 and Figure 6.4, respectively. A solid line in Figure 6.3 and 6.4 indicates a direct effect (β coefficients are underlined) and a dashed line indicates an indirect effect (β coefficients are italicized). Abbreviations of the parameters included in the model are provided in Appendix 6.5. For presentation purposes, covariates in Model 1 (type of injury and living arrangements) and Model 2 (type of injury, living arrangements, and age) are not displayed. Model 1 showed an 'acceptable' fit to the data (x^2 = 6.365, df(3), p = 0.0951; RMSEA=0.056; CFI=0.995; TLI=0.979), while Model 2 yielded a 'good' model fit (x^2 = 8.002, df(7), p = 0.3324; RMSEA=0.020; CFI=0.999; TLI=0.998).

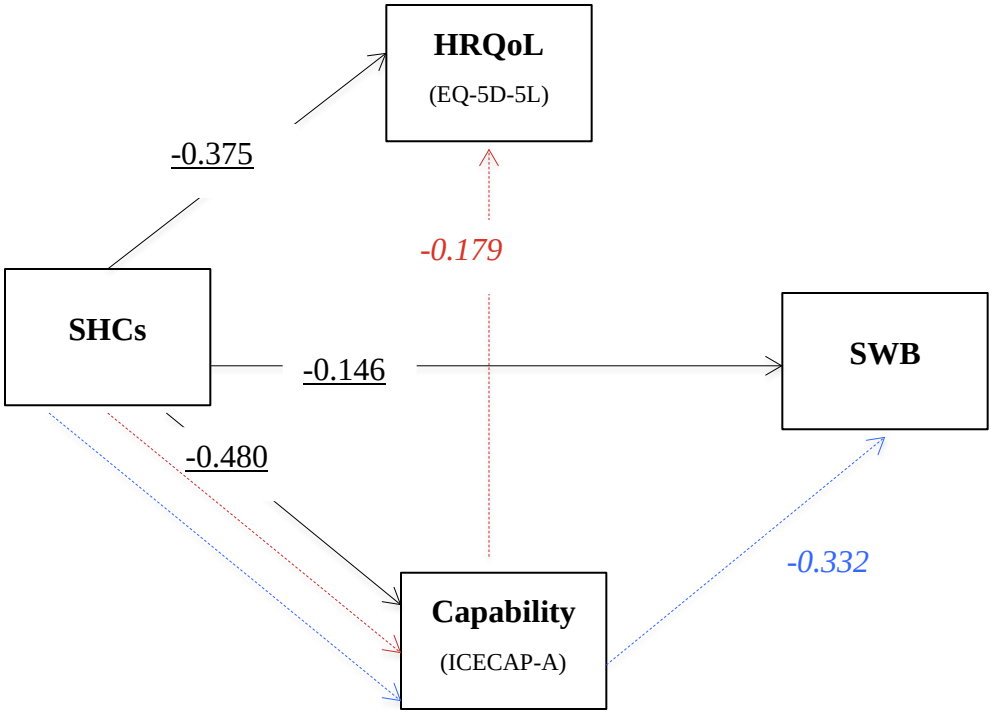


Figure 6.3. Model 1: Path model of the effect of SHCs on HRQoL, capability wellbeing and SWB

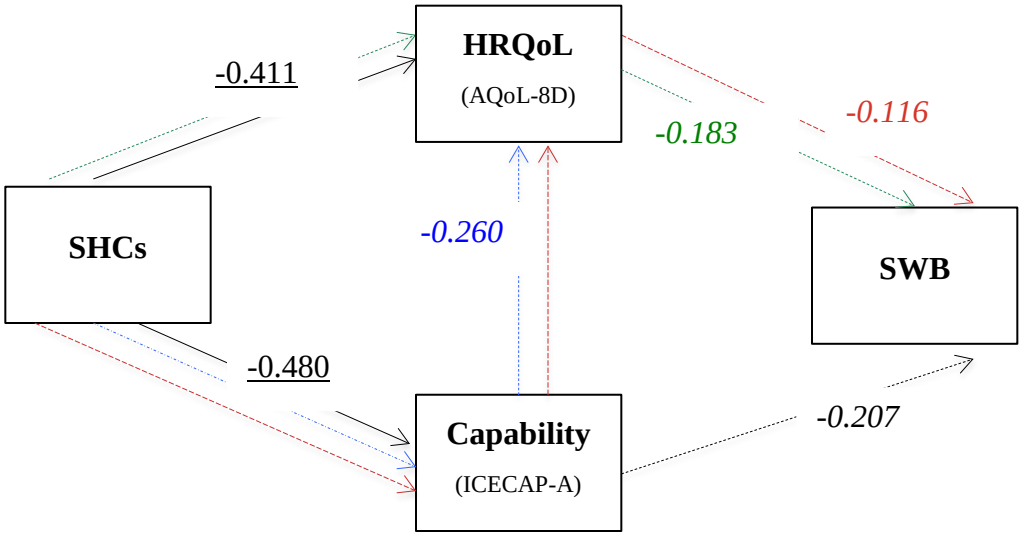


Figure 6.4. Model 2: Path model of the effect of SHCs on HRQoL, capability wellbeing and SWB

Standardized (β) and unstandardized (B) parameter estimates for Model 1 and Model 2 are presented in Table 6.5 and Table 6.6, respectively. An overview of all hypotheses that were (not) confirmed is provided in Table 6.7. With regard to Model 1, seven hypotheses were confirmed, whereas Hypotheses #4, #6, and #9 were not confirmed. SHCs impacted all outcome measures (confirming Hypotheses #1, #2, and #5) but had the greatest negative impact on individuals' capability wellbeing, as indicated by the β-coefficients ($\beta_{ICECAP\text{-}A}$= -0.480, $\beta_{EQ\text{-}5D\text{-}5L}$= -0.375, β_{SWB}= -0.146). While the impact on SWB was lower than on HRQoL and capability wellbeing (confirming hypotheses #3 and #7), the impact was higher on capability wellbeing than HRQoL (hypothesis #6 not confirmed). Capability wellbeing 'partly' mediated the effect of SHCs on HRQoL ($\beta_{indirect}$= -0.179), meaning that the direct effect was greater than the indirect effect. However, when looking at the indirect effect of SHCs on SWB, capability wellbeing 'fully' mediated the effect on SWB ($\beta_{indirect}$= -0.332), which means that the indirect effect exceeded the direct effect. Both hypotheses # 8 and #10 were confirmed. The indirect effect of SHCs on SWB through HRQoL was not statistically significant and as such, hypotheses #4 and #9 could not be confirmed.

Compared with Model 1, all except two hypotheses (#2 and #6) were confirmed in Model 2. The effect of SHCs on capability wellbeing in Model 2 was similar to the effect on HRQoL ($\beta_{ICECAP\text{-}A}$= -0.480, $\beta_{AQoL\text{-}8D}$= -0.411) yet still greater on capability wellbeing, which did not confirm hypothesis #6. Capability wellbeing again mediated the effect of SHCs on HRQoL ($\beta_{indirect}$= -0.260). While no direct effect of SHCs on SWB was observed in Model 2 (i.e., hypothesis #2 could not be confirmed), three indirect effects of SHCs on SWB were found mediated through capability ($\beta_{indirect}$= -0.207), HRQoL ($\beta_{indirect}$= -0.183), and both capability and HRQoL ($\beta_{indirect}$= -0.116).

Table 6.5. Parameter estimates for Model 1 [a]

Path	β (SE)	B (SE)	*p*	95%-CI
SHC → SWB				
Direct	-0.146 (0.038)	-0.043 (0.011)	<0.001	[-0.063; -0.025]
Indirect				
SHC → Capability → SWB	-0.332 (0.032)	-0.097 (0.011)	<0.001	[-0.115; -0.077]
SHC → HRQoL → SWB	- [b]	- [b]	- [b]	- [b]
SHC → Capability → HRQoL → SWB	- [b]	- [b]	- [b]	- [b]
SHC → HRQoL				
Direct	-0.375 (0.042)	-0.009 (0.001)	<0.001	[-0.011; -0.007]
Indirect				
SHC → Capability → HRQoL	-0.179 (0.026)	-0.005 (0.001)	<0.001	[-0.006; -0.003]
SHC → Capability				
Direct	-0.480 (0.040)	-0.011 (0.001)	<0.001	[-0.013; -0.009]

β=standardized coefficient; SE = standard error; B=unstandardized coefficient, CI=confidence interval, SHC=secondary health conditions, SWB=subjective wellbeing.
[a] Model fit: x^2 (df=3)=6.365 (*p*=0.0951); RMSEA=0.056 (*p*=0.375); CFI=0.995; TLI=0.979. Covariates included in the model: type of injury and living arrangements. R^2 SWB=0.601; R^2 HRQoL=0.456; R^2 Capability=0.235.
[b] No results are presented for this path because it was not significant in the initial 'full' model and as such removed from the final model to yield a parsimonious model.

Table 6.6. Parameter estimates for Model 2 [a]

Paths	β (SE)	B (SE)	*p*	95%-CI
SHC → SWB				
Direct:	- [b]	- [b]	- [b]	- [b]
Indirect:				
SHC → Capability → SWB	-0.207 (0.027)	-0.060 (0.009)	<0.001	[-0.080; -0.046]
SHC → HRQoL → SWB	-0.183 (0.024)	-0.053 (0.007)	<0.001	[-0.064; -0.042]
SHC → Capability → HRQoL → SWB	-0.116 (0.016)	-0.034 (0.005)	<0.001	[-0.042; -0.026]
SHC → HRQoL				
Direct:	-0.411 (0.033)	-0.010 (0.001)	<0.001	[-0.012; -0.009]
Indirect:				
SHC → Capability → HRQoL	-0.260 (0.026)	-0.007 (0.001)	<0.001	[-0.008; -0.006]
SHC → Capability				
Direct	-0.480 (0.040)	-0.011 (0.001)	<0.001	[-0.013; -0.010]

β=standardized coefficient; SE = standard error; B=unstandardized coefficient, CI=confidence interval, SHC=secondary health conditions, SWB=subjective wellbeing.
[a] Model fit: x^2 (df=7)=8.002 (p=0.3324); RMSEA=0.020 (p=0.797); CFI=0.999; TLI=0.998. Covariates included in the model: age, type of injury and living arrangements. R^2 SWB=0.673; R^2 HRQoL=0.676; R^2 Capability=0.235.
[b] No results are presented for this path because it was not significant in the initial 'full' model and as such removed from the final model to yield a parsimonious model.

Table 6.7. Overview of hypotheses that were (not) confirmed by Model 1 and Model 2

#	Path	Hypothesis	Model 1	Model 2
1	A	SHCs impact HRQoL negatively	✓	✓
2	B	SHCs have a negative impact on SWB	✓	—
3	A vs. B	SHCs have a less negative effect on SWB than on HRQoL	✓	✓
4	A C	There is an indirect effect of SHCs on SWB mediated through HRQoL	—	✓
5	D	SHCs have a negative impact on capability wellbeing	✓	✓
6	A vs. D	SHCs have a greater negative impact on HRQoL than on capability wellbeing	—	—
7	B vs. D	SHCs have a less negative effect on SWB than on capability wellbeing	✓	✓
8	D E	There is an indirect effect of SHCs on HRQoL mediated through capability wellbeing	✓	✓
9	D E C	The indirect effect of SHCs on SWB is mediated through capability wellbeing and health functionings	—	✓
10	DF	There is an indirect effect of SHCs on SWB mediated through capability wellbeing	✓	✓

HRQoL=health-related quality of life, SHCs=secondary health conditions, SWB=subjective wellbeing, — hypothesis not confirmed, ✓hypothesis confirmed.

6.4. Discussion

This chapter has highlighted the different effects of SHCs on distinct measurement approaches. The results suggest that individuals living with SCI experienced the greatest negative impact of SHCs on their capability wellbeing, which means that SHCs influenced what individuals were able to do and be (what was possible for them). Capability wellbeing also meditated the relationship between SHCs and HRQoL as well SWB, which implies that whether or not individuals with SHCs can achieve good levels of HRQoL and be satisfied with their life depends on their capabilities – in this study non-health capabilities, such as stability, autonomy, and achievement.

In contrast to the strong influence of SHCs on individuals' capability wellbeing, both models showed that SHCs have only a small (Model 1) or no direct effect (Model 2) on SWB. Compared with other types of outcome measures, SWB is different in two respects: (i) SWB uses individuals' values and not those of the general population, and (ii) it is not based on a fixed and simplified descriptive system. With regard to the first point, when studying disabilities, such as SCI, researchers have often found that individuals report high levels of SWB despite living with severe conditions (Schulz & Decker, 1985; van Leeuwen *et al.*, 2012). This is based on individuals' adaptation process (i.e., the process of adjustment to new or changed circumstances) (Kahneman & Sugden, 2005). A previous study by Peeters *et al.* demonstrated that the effect of adaptation seems to be mediated by mental health domains of QoL (Peeters *et al.*, 2010). Given that the ICECAP-A, EQ-5D-5L, and AQoL-8D are all based on societal values, the general public might not adequately forecast such coping strategies and adaptation one might experiences in a health state. As such, this has resulted in lower SWB scores or no effects at all.

Secondly, SWB was measured by a single life satisfaction item, which allowed individuals to focus on any aspects of their lives and captures what actually matters to them with respect to their health (Dolan *et al.*, 2012). Whereas fixed descriptive systems can be criticized in that they restrict the individual's ability to provide an accurate description of their health state, the findings showed that some descriptive systems included in this study can explain variability in SWB that cannot be explained by SHCs directly. This suggests that when individuals are directly asked to report their level of satisfaction with life, individuals do not necessarily think of their conditions. However, SHCs still matter for their SWB because of the decreased level of health functionings and capabilities. As such, descriptive systems appear to be helpful to remind respondents of various aspects of their lives that matter for their level of wellbeing. Nevertheless, when comparing the direct effect of SHCs on SWB and the indirect effect mediated through HRQoL, differences between the two models were observed. Model 1 resulted in a significant direct effect but no indirect effect through the EQ-5D-5L. This suggests that SHCs lead to a loss in SWB not because it leads to poor health states but because of other

reasons not captured by the EQ-5D-5L descriptive system. A recent study, examining the extent to which the EQ-5D-3L is able to capture the effect of mental and physical health of life satisfaction in older adults, showed that the EQ-5D-3L does not adequately capture mental health status on life satisfaction (Sexton *et al.*, 2016). Another study confirming that mental health dimensions have a large association with SWB (Mukuria & Brazier, 2013), highlights that the EQ-5D-3L measure gives more weight to physical functionings that are however less associated with SWB. Such information could also explain the findings in this chapter, where SHCs had no indirect effect on SWB through the EQ-5D-5L. This was the opposite in Model 2, where the loss in SWB was completely mediated through the AQoL-8D and implies that the descriptive system contains dimensions that reflect what is important for an individual's wellbeing. This finding is consistent with a previous study that showed that compared with the EQ-5D-5L, the AQoL-8D has a greater ability to predict SWB (Richardson *et al.*, 2015a).

Differences were also observed between the direct effects of SHCs on the two HRQoL measures and the ICECAP-A. Whereas the direct effect of SHCs on EQ-5D-5L was smaller than on the ICECAP-A, in Model 2 the direct effects on the AQoL-8D and the ICECAP-A were similar. It appears that the AQoL-8D and the ICECAP-A measure similar constructs that are affected in a similar way by SHCs, which would support the conclusion in the previous Chapter 5. The similar mediating effects of the ICECAP-A and AQoL-8D on SWB can further support this statement. In addition to this direct effect, differences were also found for the mediating effect of capability wellbeing on HRQoL. The hypothesized model treated HRQoL as the end goal, mediated through individual's non-health capabilities. It can be argued that HRQoL measures employ broad descriptions of health dimensions that either explicitly or implicitly relate to a wide range of non-health capabilities. While, for example, 'usual activities' or 'self-care' of the EQ-5D-5L cannot be answered without giving thought to non-health capabilities, such as autonomy (an ability to be independent), the AQoL-8D has even more such dimensions in its descriptive system. This could explain the strong mediating effect of capability wellbeing on the AQoL-8D, which is also in line with Cookson's idea of a 'capability QALY' (Cookson, 2005b), that was suggested in the previous chapter. Given the strong

mediating effect of capability on AQoL-8D, it can be argued that the descriptive system of the AQoL-8D *"pays close attention to proxy capability variables that cover a wide range of health and non-health dimensions of wellbeing*" (Cookson, 2005a, p.1288).

In view of the fact that the total effect of SHCs on HRQoL could not be explained by the direct effect only, suggests that the ICECAP-A mediated some effects of SHCs on the HRQoL that could not be captured by HRQoL directly. This could be explained by the different focus on capabilities vs. functionings. It is often criticized that compared with capabilities, functionings would 'hide' information because it is not clear if an individual, for example, does not enjoy 'close relationships' because he/she has not been given the opportunity for social interaction or because of personal choice (influenced by external factors such as social forces like stigma or attitudes). The ICECAP-A, that explicitly tries to capture what a person is able to do/be given his/her current circumstances and innate abilities, pays close attention to such conversion factors (i.e., the factors that determine how resources are converted to functionings). By rephrasing the questions in 'are you able to' or 'can you', the intention of the ICECAP-A measure is to consider such external characteristics, which may explain the mediating effect found in this work. In other words, in addition to the direct effect of SHCs on HRQoL, the ICECAP-A explained another portion of the variability in the HRQoL scores by paying close attention to external characteristics that were not captured by measures of HRQoL.

6.4.1. Implications and directions for further research

Low impact of SHCs on SWB is a reflection of individual's ability to adapt to their conditions. This has implications for resource allocation decisions. The adaptation process is often considered as one of the criticisms in applying an experience-based method, such as SWB, for resource allocation decisions because the more individuals adapt to their conditions the less priority will be given to them (Dolan & Kaheman, 2008). However, while it is generally known that people with long-term health conditions or who are physically disabled are more likely to adapt, mental health conditions, on the other hand, are the hardest to adapt to (Dolan *et al.*, 2012). As a consequence, mental health patients value their own health state lower than the general public (Papageorgiou

et al., 2015; Pyne *et al.*, 2009), which would disadvantage them unfairly compared with physical health conditions. Therefore, particularly in relation to mental health it has been suggested to replace or supplement QALYs with SWB in the assessment of health services because SWB would place a greater emphasis on mental health conditions (Dolan *et al.*, 2012). A contrary view has been provided by a later study that showed that the impact of mental health conditions on a person's life as a whole and upon a person's wellbeing can be better assessed by the general public if the descriptive systems of HRQoL measures contain more concepts that capture the domains that matter most in the experiences of people's lives (Richardson *et al.*, 2015a). As was shown in the study by Richardson *et al.* the AQoL-8D had the greatest ability to account for variation in SWB. Future studies are warranted to provide further evidence on the direct implications for resource allocation decisions when applying a SWB or a QALY approach for both physical and mental conditions.

The findings in this chapter highlight that the capability approach requires more attention in health care decision-making. Applied to the health care context, the evaluation of health care interventions should distinguish between observable health achievements and the capability to achieve good health. Particularly in the disability literature the capability approach is gaining increasingly attention. The disability research has shifted away from focusing merely on the individual at the body level (impairment) towards understanding the contextual situation of the individual in his/her environment (e.g., social, economic, political) (Mitra, 2006; Welch Saleeby, 2006). In other words, the capability approach emphasizes the need to move beyond actual functionings (outcomes or achievements) to promoting capabilities (opportunities or potential) (Welch Saleeby, 2006). Opportunity can be understood as the range of valuable activities and experiences available to individuals, such as opportunity for social interaction, recreation or enjoyment. Therefore, what is important for persons with disabilities is not what they have already achieved, but what they potentially could achieve with the removal of barriers and creation of opportunities (Trani *et al.*, 2011). For example, a physically impaired person requires more resources to achieve the functioning of mobility than a person who has no physical impairment. However, factors influencing the resources

available and the possibilities to use them may not only originate from the health sector but non-health sectors as well, such as the sector for infrastructure and transport or housing. As such, applying the capability approach in health care would not only require a shift from functionings to capabilities but also moving beyond the health care sector.

Although NICE has recently included the use of capability instruments into its guidelines for evaluating social care, the operationalization of the capability approach to guide resource allocation decisions is still in their infancy. Mitchell *et al.* introduced the latest decision-rule for the capability approach based on 'shortfall sufficiency' (Mitchell *et al.*, 2015b). In doing so, the anchoring of the existing scale for the ICECAP instruments is adjusted, such that zero is equivalent to no capability but one is equal to 'sufficient capability', where individuals have at least 'a lot' of capability on all attributes of the ICECAP instrument (i.e., scoring of 33333). In order to estimate sufficient capability over time, years of sufficient capability are then generated, where a year of life in sufficient capability has the value of one (Mitchell *et al.*, 2015b). Applying the sufficient capability approach in practice implies that the evaluation of health care intervention would shift away from efficiency toward prioritizing patient groups who are below the poverty level (i.e., moving from maximization to sufficiency). Such approach would require valid and sensitive measures of a person's capability. However, one of the shortcomings of the ICECAP instruments is that they provide information about perceived rather than actual capabilities. There is a potential risk that individuals may not recognize their own lack of capability because they have adapted to their situation, which may result in lower priority in resources allocation decisions.

The future use of the capability approach in resource allocation decisions remains unclear. Cookson emphasized, the best start in applying the capability approach to health economics, is to maintain a QALY type framework but to select a QALY instrument that pays closer attention to non-health dimensions of wellbeing (Cookson, 2005a). In this context, the QALY will be re-interpreted as a capability QALY. Others have also argued that if the intention is to look for a health capability account that can be used to inform resource allocation decisions, instruments like the EQ-5D is '*not a bad starting point*',

even if it is not a capability measure *per se* (Wolff *et al.*, 2012). Wolff and colleagues even proposed that the EQ-5D measure could be used to measure experienced utility, given that the five dimensions of the EQ-5D are all either aspects of a patient's experience or will affect their experience. However, this work showed that there are differences in effects of SHCs on these three different outcome measures and use of HRQoL instruments as a proxy for capability or experienced utility is not straightforward. As such, the findings suggest that these three measurement approaches cannot be used as substitutes but rather as complements. Given that many jurisdictions focus on achieved health outcomes based on QALYs, the complementary use of the capability measure would extend the QALY framework by considering the ability to pursue health as well as assessing health outcomes (Kinghorn, 2015).

6.4.2. Strengths & limitations

The choice of an outcome measure for use in economic evaluation was discussed a number of times in the literature (Coast *et al.*, 2008d; Dolan *et al.*, 2012; Lorgelly, 2015). Compared with these previous studies, which were mainly concerned with normative issues (i.e., identifying an appropriate maximand for resource allocation decisions), this chapter provides empirical evidence about the impact of SHCs on these different measurement approaches. The context of SCI in this work plays particularly an important role, as the ability of adaptation is an important consideration in outcome measurement. The conduct of a path analysis, which can be seen as an extension of a regression analysis, is another strength of this study that allowed not only the analyses of direct effects but also indirect effects, which have offered further insights into the relationship between the measures. Given that all paths were analyzed simultaneously, issues related to multiple hypotheses testing were avoided. Finally, the application of two models that used different measures of HRQoL makes the analysis stronger and increases the confidence in the findings. On the contrary, differences were found between Model 1 and Model 2, which suggest that further studies are needed to replicate the analysis using other measures of HRQoL, capability and SWB. There is an increasing interest in the capability approach to outcome measurement and other instruments of capability outside

the ICECAP family include the OCAP-18, OxCAP-MH for use in mental health, and the ASCOT for use in social care (Lorgelly, 2015). Also, different experience-based measures exist that do not necessarily consist of one question, which was the case in this work, but measure the level of satisfaction for different domains of life.

The biggest limitation of this study is the cross-sectional design of the study, which is why longitudinal studies are needed to confirm the effects found in this study. Given that the sensitivity of instruments that have a broader evaluative space is not perfectly established, it could be that, compared with a measure of health, broader measures may be insensitive to changes in health alone. A recent study found that the EQ-5D-5L is more responsive than the ICECAP-O among individuals at risk of mobility impairment among adults aged ≤ 70 (Davis *et al.*, 2016) but further studies are needed. Finally, it needs also to be acknowledged that path analysis cannot be used to establish causality. Causality is generally established through the design of the study and not its analysis (Streiner, 2005).

6.5. Conclusion

The choice of an outcome measure to guide resource allocation decisions in health care requires careful consideration. This study has shown that SHCs have different impacts on HRQoL, capability and SWB. The results suggest that the ability of individuals with SHCs to achieve good health functionings depends on their level of capability. Complementing QALY results with individuals' capability wellbeing can better inform economic evaluations, focusing not only on 'observed health' but the 'capability to achieve good health', which takes into account personal factors (e.g., impairment) and circumstantial characteristics (e.g., environment). Further studies are needed to replicate the analysis using other measures of HRQoL, capability and SWB.

Chapter 7. Thesis summary and conclusions

The aim of this thesis was to examine methodological considerations regarding the use of outcome measures that extend beyond HRQoL in the context of QALY-based economic evaluation. In addition, this thesis examined implications for policy and practice when broadening the evaluative space of the QALY. Through a series of studies, this thesis explored the following research questions:

- What are the potential challenges in broadening the evaluative space of the QALY?
- What is understood by 'benefits beyond health'?
- What information is captured by ICECAP measures over and above the information garnered by existing preference-based HRQoL measures?

Chapter 7 provides a summary of the key findings and contributions to the existing literature. Strengths and limitations of the analytic approaches are discussed, followed by a reflection of the extent to which the challenges identified in Chapter 2 have been addressed by this thesis, and which challenges require further research. The thesis ends with some concluding remarks.

7.1. Key findings and contributions to existing knowledge

This section summarizes the key findings of this thesis and the contributions to existing knowledge concerned with the evaluative space of the QALY. It will outline (i) if it is possible to broaden the evaluative space of the QALY, (ii) provide a better understanding of benefits beyond the health-related QALY, and (iii) if there is a need to

broaden the evaluative space, given that the QALY may not be as health-focused as claimed.

7.1.1. Can we broaden the evaluative space of the QALY?

Chapter 2 explored whether the evaluative space of the QALY can be broadened, and set out associated challenges. The breakdown of challenges into theoretical, normative, conceptual, methodological, and practical issues allowed for the identification of areas that require specific attention. The chapter provided clarity about the theoretical concept of extra-welfarism and has highlighted that it is possible to include benefits into the QALY framework that extend beyond health. Advantages of adopting an extra-welfarism approach include the use of outcomes other than utility and the use of sources of valuation other than the affected individuals. There is no theoretical barrier to extending the evaluative space of the QALY beyond health outcomes. Instead, identifying an appropriate maximand and dealing with normative issues around the scope of the evaluative space of the QALY are the main challenges. The chapter reflected on previous literature and outlined two normative considerations for broadening the evaluative space of the QALY: (i) whether individuals value non-health benefits and (ii) whether or not considering such broader outcomes can potentially impact resource allocation decisions. An important observation in this chapter was the ambiguity of the term ‘non-health benefit’. The chapter concluded that conceptual issues need to be solved first (i.e., what do we mean by benefits beyond health?). In addition, this chapter highlighted the need for further examination of the extent to which existing preference-based HRQoL instruments already capture benefits beyond health. These two considerations shaped the rest of the thesis. The extent to which other challenges identified in this chapter were addressed by analyses conducted in this thesis are discussed in Section 7.3.

7.1.2. What is understood by benefits beyond the health-related QALY?

Chapter 3 provided important insights through the conceptualization of benefits beyond the health-related QALY. Although non-health benefits have been extensively

discussed in the literature, no previous studies have tried to bring all evidence together and to conceptualize the relationships between such broader benefits. Considering the current debates to move beyond the health-related QALY (Brazier & Tsuchiya, 2015), a better understanding of such broader benefits was needed and the chapter identified four types of non-health benefits: (i) benefits affecting wellbeing (psychological wellbeing, subjective wellbeing, capability wellbeing, and empowerment); (ii) benefits derived from the process of health care delivery (process utility); (iii) benefits beyond the treated individual (e.g., option value, externalities, spillover effects, and equity); and (iv) benefits beyond the health care sector. Chapter 3 concluded that the exclusion of benefits beyond health does not appear to be problematic for the QALY metric *per se* and, instead, embracing such an approach raises a number of issues for health care decision-making (e.g., identifying an appropriate maximand for resource allocation decision-making). Chapter 3 also highlighted that QALYs are defined by what is measured (e.g., HRQoL, using the EQ-5D-5L) rather than the conceptual origin (Goebbels *et al.*, 2012; Whitehurst & Engel, 2017). Given that preference-based HRQoL measures differ greatly in their coverage of HRQoL dimensions, Chapter 3 cautioned that the extent to which existing preference-based HRQoL measures are able to 'pick up' non-health benefits is currently unknown and any attempt to measure health and non-health benefits simultaneously could potentially result in double counting. This notion was explored further in the subsequent chapters.

7.1.3. Is the evaluative space of the QALY limited to health?

Whether or not the evaluative space of the QALY is limited to health alone was explored by three empirical analyses in this thesis. The approach taken to address this question was through direct comparisons of preference-based HRQoL measures that are used to derive QALYs with outcome measures that adopt a broader evaluative space.

Chapter 4 was an application of the EQ-5D-5L outside the conventional clinical context. Exploring the ability of the EQ-5D-5L to capture broader environmental benefits was the first attempt to investigate how sensitive the EQ-5D-5L is to aspects that are not directly related to health. Given that the analyses in Chapter 4 were also applied to a

measure of capability wellbeing, the ICECAP-O, it was possible to examine whether the ICECAP-O measures something different compared with the EQ-5D-5L. Results from the Walk-The-Talk study showed that older adults' capability wellbeing (based on the ICECAP-O) was associated with street connectivity and social cohesion, while no statistically significant associations were found between environmental factors and HRQoL (based on the EQ-5D-5L). While such findings demonstrate differences between the ICECAP-O and the EQ-5D-5L in terms of the extent to which the respective index scores were explained by environmental features, exactly *what* was driving the differences could not be answered by this chapter and was subsequently explored in Chapters 5 and 6.

Chapter 5 investigated the underlying descriptive content of the EQ-5D-5L and the adult version of the ICECAP through the conduct of an exploratory factor analysis. Results indicated that the EQ-5D-5L and the ICECAP-A measured tapped into distinct factors and provide complementary information, which confirmed previous studies that explored the overlap between the ICECAP measures (ICECAP-A and ICECAP-O) and the EQ-5D-3L (Davis *et al.*, 2013; Keeley *et al.*, 2016). Findings of the exploratory factor analysis also helped to explain differences found between the EQ-5D-5L and the ICECAP-O in the previous chapter. An important contribution of this chapter was the extension of the analyses beyond the EQ-5D-5L – including other preference-based HRQoL measures available in the MIC dataset (15D, AQoL-8D, HUI-3, and SF-6D) – where results indicated a substantial degree of overlap between the AQoL-8D and the ICECAP-A. Although the ICECAP-A was developed to overcome perceived limitations with the evaluative space of the QALY, such findings suggest that broader benefits, as defined by the ICECAP-A, could potentially be captured by the AQoL-8D. In this context, the potential risk of double counting was highlighted when applying ICECAP measures alongside HRQoL measures. Despite the conceptual overlap between the AQoL-8D and the ICEACP-A, the comparative performance of these two measures was unknown and was explored further in the final empirical analysis of the thesis.

Chapter 6 introduced another outcome measurement approach that was compared with preference-based HRQoL measures – the concept of subjective wellbeing (SWB).

SWB accounts for the domains of health that matter most in the experience of people's lives by using individuals' own values and not describing health in terms of a fixed and simplified descriptive system. Chapter 6 was the first empirical analysis in the field of outcome measurement in economic evaluation that looked at all three outcome measurement approaches (HRQoL, capability wellbeing, and SWB) within the same dataset. Findings from this analysis showed that secondary health conditions (here, in the context of spinal cord injury) have different impacts on HRQoL, capability and SWB. However, when measuring HRQoL using the AQoL-8D, a similar impact of secondary health conditions was observed on the ICECAP-A and the AQoL-8D. Furthermore, the AQoL-8D fully mediated the impact of secondary health conditions on SWB, which implies that the descriptive system contains dimensions that reflect what is important for an individual's wellbeing. Such findings confirm that the evaluative space of the QALY is not limited to health, given that the AQoL-8D seems to measure aspects of wellbeing. However, in view of the fact that the effect of secondary health conditions on the AQoL-8D was partly mediated by the ICECAP-A suggests that observed functionings in the AQoL-8D depend on a person's capability wellbeing. In terms of the evaluative space of the QALY, this finding implies that although the 'objects of value' are similar between the AQoL-8D and ICECAP-A, differences exist in what respect they are valued (i.e., functionings versus capabilities). It was proposed that the complementary use of a capability measure would extend the QALY framework, focusing not only on 'observed health' but the 'capability to achieve good health'.

While the thesis was largely focused on methodological considerations when using outcome measures in health economic evaluation that extend beyond HRQoL for the estimation of QALYs, each chapter also discussed the implications for policy and practice. The findings of this thesis have particular relevance for national HTA bodies that set guidelines for the conduct of economic evaluation – some of which have recently recommended the use of broader outcome measures. Key points for decision makers are summarized in Table 7.1.

Table 7.1. Key points for decision makers

#1	The adoption of a broader definition of the extra-welfarism framework is recommended, which would extend the maximization of outcomes beyond health
#2	The evaluative space of the QALY is not necessarily limited to health and is largely driven by dimensions included in preference-based HRQoL measures
#3	Non-health benefits can be incorporated into the QALY framework, although further research is needed to address normative and methodological concerns
#4	Using the ICECAP measures as complementary measures in resource allocation decisions may result in double-counting, depending on the choice of HRQoL measure
#5	If the allocation of health care resources should be informed by the impact of interventions on broader benefits, the selection of a preference-based HRQoL measure should not be limited to the EQ-5D-5L but left open to measures that are more sensitive to such broader benefits (e.g., the AQoL-8D)
#6	Compared with preference-based HRQoL, the use of SWB to guide health care allocation decisions requires careful considerations and further research
#7	The use of the capability approach has the potential to provide additional information to support health care decision-making (i.e., reporting individuals' true opportunities to achieve good health)
#8	Arguments for the use of capability measures in health economic evaluation should not be in terms of their extension of the evaluative space 'beyond health' but rather in moving away from functionings 'towards capabilities'

7.2. Strengths & limitations

While strengths and limitations were reported in the individual chapters, there are some key aspects that apply to the thesis overall. The main objective of the three empirical analyses (Chapters 4, 5, and 6) was to gain a better understanding of the relationships between preference-based HRQoL and measures that extend beyond health. On that note, a key strength of this thesis was the application of different methods that addressed different components of the overall research question: (i) regression analysis compared the sensitivity of a HRQoL measure and a capability measure towards

environmental features; (ii) exploratory factor analysis looked explicitly at the descriptive systems of preference-based HRQoL measures and a capability measure by exploring the underlying constructs; and (iii) path analysis allowed for the exploration of direct and indirect effects between three outcome measurement approaches. While preference-based measures (regardless of whether related to HRQoL or capability wellbeing) consist of a descriptive system and a valuation system, the methods outlined above looked at *both* systems, which allowed for a deeper understanding of the relationships between HRQoL and broader preference-based measures.

Previous literature has repeatedly stated that public health interventions require broader measures of outcome (Lorgelly *et al.*, 2010). This thesis was able to inform these discussions further through the examination of a capability measure in the context of public health. Furthermore, the opportunity for exploration of the capability approach in older adults and individuals with spinal cord injury can be seen as another strength, given that the distinction between capabilities and functionings is likely to be more evident in these two groups compared with other population groups (Mitra, 2006). Another strength of this thesis is the consideration of several preference-based HRQoL measures rather than just the EuroQol instruments. Previous research has highlighted numerous differences between preference-based HRQoL measures (Richardson *et al.*, 2015b) and, as such, it was essential to address the research questions taking not only the frequently used EQ-5D-5L into account.

The selection of preference-based HRQoL measures in this thesis requires a further note. While the availability of datasets that collect multiple preference-based HRQoL measures and/or multiple measures of wellbeing is limited, a strength of this thesis is the use of two datasets (Chapter 5 and Chapter 6) that comprise such outcome measures. Yet, despite the existence of a number of measures, not all were included in the analyses. In Chapter 5, no exploratory factor analysis was conducted comparing the ICECAP-A with the QWB-SA. This was due to the scaling properties of the QWB-SA, which were not suitable for an exploratory factor analysis (see page 100). In Chapter 6, besides the availability of the AQoL-8D and the EQ-5D-5L, this dataset also contained

the SF-36v2 and the HUI. The motivation to focus on the AQoL-8D and EQ-5D-5L only was driven by previous research that supported the use of the AQoL-8D in the context of spinal cord injury (Michel *et al.*, 2016; Whitehurst *et al.*, 2016) and the findings of the exploratory factor analysis, where a strong overlap was observed between the ICECAP-A and the AQoL-8D. As such, further exploration of the AQoL-8D was a natural course of action. The focus on the EQ-5D in all three empirical analyses in this thesis was driven by its widespread use, the preference of some decision makers for this measure (NICE, 2013; Richardson *et al.*, 2015c), and its availability in all three datasets.

With regard to capabilities, this thesis focused exclusively on the ICECAP-A and ICECAP-O measures. While this allowed for a better exploration of the ICECAP instruments using different methodological approaches, this thesis is not able to draw any conclusions for other measures that offer a broader evaluative space for outcome measurement in economic evaluation. An instrument that should be explored further is the ASCOT (Netten *et al.*, 2012). Although the ASCOT is a measure of social care-related QoL, it was developed based on the same motivational grounds as the ICECAP instruments, namely the incorporation of broader benefits within economic evaluation. Furthermore, throughout the thesis, there is often reference made to the 'ICECAP instruments'. This reflects the fact that the ICECAP instruments considered in this thesis (ICECAP-A and ICECAP-O) were developed by the same research team at the University of Birmingham (UK), comprise similar domains, and are characterized by the same language. However, it needs to be acknowledged that the ICECAP-A and the ICECAP-O are two distinct measures. Thus, findings reported about the ICECAP-O do not necessarily apply to the ICECAP-A and *vice versa*.

A further limitation of this thesis is in regard to the data used in the three empirical analyses. All three datasets consisted of cross-sectional data, meaning that the investigation of the research questions in a longitudinal context was not possible. While this may be less of an issue for Chapter 5 (i.e., the exploratory factor analysis), Chapters 4 and 6 would have benefited from further longitudinal explorations. In view of the fact that the ICECAP-O and the ICECAP-A are relatively new instruments, their

responsiveness to change has not fully been explored and contradictive findings have been reported across studies that have looked at responsiveness to change (Davis *et al.*, 2016; Keeley *et al.*, 2015). These issues need to be considered when judging the generalizability of the study findings and drawing conclusions for current practice.

7.3. Challenges addressed and areas for further research

Chapter 2 discussed numerous challenges that require further attention in broadening the evaluative space of the QALY. This section will reflect on these challenges and outline to what extent these were addressed by this thesis and which challenges require further research. To begin with, while an important contribution of this thesis was the exploration of the relationship between preference-based HRQoL measures and measures of capability wellbeing, existing problems associated with the *operationalization* of the capability approach for outcome measurement outlined in Chapter 2 (i.e., identification, measurement and valuation of capabilities) still require further research. Other considerations when using the capability approach to inform resource allocation decisions are discussed below.

7.3.1. Using the capability approach to inform resource allocation decisions

The use of the capability approach to inform resource allocation decisions remains an issue of debate. While it is gaining attention within the context of public health, social care and end-of-life care, Lorgelly has suggested that the capability approach could also be of benefit to evaluations of pharmacotherapies and other technologies (Lorgelly, 2015). In contrast, it has been argued that the capability approach seems unclear and insufficiently justified to be used to inform resource allocation decisions (Sampson, 2016). The ways in which researchers have operationalized the capability approach suggests that its primary goal in health economics is the extension of the evaluative space of the health-related QALY. Sampson has argued that moving 'beyond health' rather than 'towards capabilities' has received less discussion, suggesting there have been fewer arguments provided in favour of such an extension to the

evaluative space (Sampson, 2016). It is a fair question to consider what the benefit of using the capability approach is compared with using preference-based HRQoL measures. For this reason, the underlying factor structure of the capability approach was examined in Chapter 5, where the exploratory factor analysis indicated conceptual overlap between the AQoL-8D and the ICECAP-A. Yet, Chapter 5 was only concerned with the value of objects (*what* is valued (the domains)) and less with the consequences (*why* is it valued (in terms of capabilities or functionings)).

When moving the evaluative space from functionings to capabilities, another way to explore the role of the capability approach is in regard to the evaluation of interventions to inform resource allocation decisions. It can be argued that the focus on capabilities is more appropriate than on functionings because the provision of health care interventions can generally only create opportunities for individuals to realize certain functionings, such as the ability to enjoy close relationships (Al-Janabi, 2014). Health care policy cannot force individuals to enjoy close relationships, as it is the individuals' decision to choose the life they want to live. This becomes particularly important for public health interventions, such as environmental interventions, where places are created to provide opportunity for people for social interaction regardless of whether or nor they would like to be involved in social activities. What matters is that environmental interventions contribute to a person's wellbeing because individuals know they *have an opportunity* for social interaction. However, it is not self-evident how interventions should be evaluated based on the changes in possible (not actual) enjoyment of close relationships (Karimi *et al.*, 2013). While the ICECAP instruments attempt to measure capabilities through the use of particular wording to measure a person's capabilities (such as what a person 'can do' or 'is able to do'), whether individuals are able to truly recognize their opportunities remains an area for further research. Also, considering the concept of opportunity cost, whether resources *should* be taken away to help someone to increase their capability for social interaction, even though they do not like to be socially engaged, is debatable. For decision makers, a focus on measuring functionings may appear more appropriate.

Another notion that warrants further investigation is whether the capability

approach should replace the current health-related QALY approach, or if there are ways to combine these two measurement approaches (i.e., treat them as complements). The findings from Chapter 6 show that the ability of individuals with secondary health conditions to achieve good health functionings was influenced by their level of capability. Therefore, complementing QALY results with individuals' capability wellbeing can better inform economic evaluations, focusing not only on 'observed health' but the 'capability to achieve good health', which takes into account personal factors (e.g., impairment) and circumstantial characteristics (e.g., environment). This becomes particularly important for the population groups explored in this thesis (older adults and individuals with spinal cord injury), where individuals may have mobility issues but policy can change the environment to accommodate such problems.

In order to use the capability approach for resource allocation decisions, there are other points that require attention. Firstly, it is generally known in the capability literature that the selection of capabilities needs to be context specific, and derived from a participatory approach (Kinghorn, 2015). Therefore, the application of the UK-developed ICECAP measures globally can be challenged. In this context, the emphasis of future studies could be on the justification of selected functionings or capabilities. Secondly, in order to determine the value of a capability improvement, further research needs to explore the cost-effectiveness threshold level to value the improvements in capability gains or for a sufficient capability level (Mitchell *et al.*, 2015b). Finally, there remains scope for qualitative research in this area in terms of the exploration of individuals' abilities to distinguish between capabilities and functionings. The think-aloud study by Al-Janabi *et al.* showed that individuals were able to self-report their capabilities and also identified how their capabilities diverged from their level of functionings (Al-Janabi *et al.*, 2013a). However, the study also showed that some participants still understood the ICECAP-A questions in terms of their functionings and had difficulties with the capability concept.

7.3.2. Is it time to redefine health?

This thesis explored the methodological and practical implications in going beyond *health-related* QoL for outcome measurement in economic evaluation. While the main focus was on non-health rather than health benefits, the conceptualization of non-health benefits in Chapter 3 was ultimately driven by the definition of health. Following the WHO typology, which states that health problems result in impairment, disability and participation, existing preference-based HRQoL measures have based their descriptive systems on these concepts, although differences between measures exist (Richardson *et al.*, 2014a). Nevertheless, some researchers have criticized existing preference-based HRQoL for not measuring patients' health as defined by the WHO definition (de Vries *et al.*, 2016). Others have challenged the WHO definition itself and proposed to define health as 'the ability to adapt and to self-manage', which is a better reflection of the human capacity to cope autonomously with life's changing physical, emotional, and social challenges, and to function with fulfilment and a feeling of wellbeing (Huber *et al.*, 2011). An important argument that Huber and colleagues provide is that this definition is more meaningful for health policy, where health gain in terms of survival may be less relevant than societal participation, and an increase in coping capacity may be more relevant and realistic than complete recovery. It is likely that adopting such a broad definition of health would close the gap between HRQoL and QoL (Wichmann *et al.*, 2017). Removing the *health-related* aspect of HRQoL would also no longer disadvantage interventions that do not primarily focus on improving HRQoL (Wichmann *et al.*, 2017). Sampson, furthermore, argued that Huber and colleagues' definition implies capabilities as a consequence (Sampson, 2016). Therefore, the focus on people's *ability* to adapt and to self-manager provides a concept of health that would take capabilities into account.

One of the reasons why Huber and colleagues find the WHO definition counterproductive is that it classifies people with chronic diseases and disabilities as ill (Huber *et al.*, 2011). The WHO definition does not reflect people's ability to develop strategies for coping, and evidence has shown that impaired functioning does not necessarily change people's QoL (von Faber *et al.*, 2001). This was also observed in Chapter 6, where secondary health conditions had no impact (or only a small negative

impact) on individuals' SWB. This finding highlights the fact that individuals with spinal cord injury can be satisfied with their lives despite being impaired and having secondary health conditions. In order to apply Huber and colleagues' definition of health in health economic evaluations, a preference-based HRQoL measure would be required that reflects individuals' ability to adapt and to self-manage. While a previous study showed that adaptation is mediated by mental health domains of QoL (Peeters *et al.*, 2010), findings in Chapter 6 also indicated that effects of secondary health conditions on SWB were mediated by the AQoL-8D, indicating that the AQoL-8D accounts for adaptation. Richardson and colleagues have previously stated that the AQoL-8D is sensitive to changes in the dimensions that drive adaptation, given the psychosocial component of the descriptive system (Richardson *et al.*, 2014b). Further evidence to support the ability of the AQoL-8D to account for adaptation can be drawn from a previous publication, which showed that the AQoL-8D allowed individuals with spinal cord injury to report health state values at the upper end of the 0-1 scale, despite their significant mobility impairments (Whitehurst *et al.*, 2016). As the authors of this study stated, this is most likely because of the definition of mobility employed by the AQoL-8D, which is not limited to individuals' capacity to walk.

7.3.3. Consistency and pragmatism in health care decision making

Making resource allocation decisions in health care is a challenging task and the need to maintain methodological consistency is important (NICE, 2013). In the past decades, decision makers have based their decisions on cost per QALY gains, where QALY weights were often derived from the EQ-5D (-3L or -5L). Moving away from the current definition of health and the preference-based instruments that are widely used is associated with challenges. The empirical analyses conducted in Chapters 5 and 6 showed that the AQoL-8D is a promising measure. The AQoL-8D extends beyond the narrow focus on HRQoL and combines health and non-health outcomes. This combination has been previously suggested to be an important characteristic of measures, from which one could derive capability-QALYs (Cookson, 2005b). The AQoL-8D also seems to better reflect Huber and colleagues' definition of health in terms of the ability to

adapt and to self-manage (Huber *et al.*, 2011). Yet, as was shown in a recent review, the AQoL-8D has received less attention in the health economics literature outside of its country of origin (Australia) (Richardson *et al.*, 2015c). Compared with other preference-based HRQoL measures, a unique feature of the AQoL-8D is that each dimension has more than one item. While this feature draws upon psychometric theory, which suggests a minimum of three items per dimension (Richardson *et al.*, 2014a), the number of items included in the AQoL-8D is likely to be associated with challenges in terms of practicality, due to the length of the questionnaire and the valuation of health states. A previous focus group study found that the length of a questionnaire is not a negative feature *per se*, when compared with perceived relevance (Whitehurst *et al.*, 2014b). However, the importance of the length of questionnaires for use in clinical practice and research will differ depending on the context, although some measures that are commonly used have more items than the AQoL-8D (e.g., the SF-36v2 (Ware *et al.*, 2008)). Another challenge when using a measure with a more comprehensive descriptive system relates to the valuation exercise. While it is usually recommended that health states contain no more than nine attributes (Brazier *et al.*, 2017), the valuation of the AQoL-8D involved the valuation of health states described using all 35 items (Richardson *et al.*, 2014b). Although a number of approaches were employed to make the AQoL-8D valuation task more palatable, the task does appear more complex compared with those completed for other preference-based instruments (Brazier *et al.*, 2017).

At this stage it remains unclear if any changes to the AQoL-8D would result in improved uptake. The fact that UK and Dutch guidelines prefer explicitly the EQ-5D (NICE, 2013; Zorginstituut Nederland, 2016) indicates that at least in these jurisdictions the use of another measure is more challenging. Although the latest Canadian guidelines do not indicate such a preference for the EQ-5D (CADTH, 2017), the fact that the EQ-5D (3L or 5L) is the most often used measure (Richardson *et al.*, 2015c) suggests that analysts and decision makers may still be reluctant to embrace a new measure. The primary reason for this is the need for some degree of pragmatism in health care decision-making. Decision makers may be aware of limitations of the EQ-5D-5L but, without strong evidence to the contrary, the continued use of ‘conventional’ methods will prevail.

The dominant position of the EQ-5D-5L may also be the reason why there is interest in addressing the problem of non-health benefits in terms of developing bolt-ons to the EQ-5D (3L or 5L) rather than new measures. While the AQoL-8D contains health and non-health outcomes, it is currently unknown if one comprehensive measure is preferred in decision-making over a *combination* of HRQoL and wellbeing measures. Employing multiple measures would allow for the separation of HRQoL and non-health benefits, where decisions could be explored separately and where, for example, different thresholds for HRQoL and non-health benefits could be applied. On the other hand, decision makers may find it more challenging to judge the relative value of HRQoL and non-health outcomes. Ultimately, to better inform decision makers, it is necessary to formally identify what patients and citizens fundamentally want from their health service, and whether society is willing to give up health gains for improvements in non-health benefits.

7.4. Concluding remarks

This thesis has focused on the evaluative space of the QALY and examined the relationships between preference-based HRQoL measures and broader measures of wellbeing. It has shown that the application of broader wellbeing measures in economic evaluation requires careful considerations due to the risk of double counting. The conceptual overlap between the ICECAP-A and the AQoL-8D has demonstrated that broader non-health outcomes are not necessarily excluded from the QALY metric *per se*, and that descriptive systems of preference-based HRQoL measures determine the evaluative space of the QALY. The focus on capability measures in this thesis provided support for the use of the capability approach to inform resource allocation decisions in health care. However, the operationalization of this approach and other non-health benefits identified in this thesis, within or outside the QALY framework, requires further research.

References

Al-Janabi H. (2014). Do capability and functioning differ? An analysis of survey responses. *3rd ICECAP Workshop*. Birmingham, UK.

Al-Janabi H, Flynn T, & Coast J. (2012). Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Quality of Life Research,* 21(1): 167-176.

Al-Janabi H, Flynn TN, & Coast J. (2011). QALYs and carers. *Pharmacoeconomics,* 29(12): 1015-1023.

Al-Janabi H, Keeley T, Mitchell P, & Coast J. (2013a). Can capabilities be self-reported? A think aloud study. *Social Science & Medicine,* 87: 116-122.

Al-Janabi H, Peters TJ, Brazier J, Bryan S, Flynn TN, Clemens S, *et al.* (2013b). An investigation of the construct validity of the ICECAP-A capability measure. *Quality of Life Research,* 22(7): 1831-1840.

Al-Janabi H, Van Exel J, Brouwer W, & Coast J. (2015). A framework for including family health spillovers in economic evaluation. *Medical Decision Making,* 36(2): 176-186.

Al-Janabi H, Van Exel J, Brouwer W, Trotter C, Glennie L, Hannigan L, *et al.* (2016). Measuring health spillovers for economic evaluation: a case study in meningitis. *Health Economics,* 25(12): 1529-1544.

Alayli-Goebbels AFG, Dellaert BGC, Knox SA, Ament A, Lakerveld J, Bot SDM, *et al.* (2013). Consumer preferences for health and honhealth outcomes of health promotion: results from a discrete choice experiment. *Value in Health,* 16(1): 114-123.

Ali S, & Ronaldson S. (2012). Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *British Medical Bulletin,* 103(1): 21-44.

Alvarado-Bolanos A, Cervantes-Arriaga A, Rodriguez-Violante M, Llorens-Arenas R, Calderon-Fajardo H, Millan-Cepeda R, *et al.* (2015). Convergent validation of

EQ-5D-5L in patients with Parkinson's disease. *Journal of the Neurological Sciences,* 358(1-2): 53-57.

Anderson CJ, Vogel LC, Chlan KM, & Betz RR. (2008). Coping with spinal cord injury: strategies used by adults who sustained their injuries as children or adolescents. *The Journal of Spinal Cord Medicine,* 31(3): 290-296.

Annemans L, Redekop K, & Payne K. (2013). Current methodological issues in the economic assessment of personalized medicine. *Value in Health,* 16(6): S20-S26.

Arnold D, Girling A, Stevens A, & Lilford R. (2009). Comparison of direct and indirect methods of estimating health state utilities for resource allocation: review and empirical analysis. *British Medical Journal,* 339: b2688.

Asano M, Miller WC, & Eng JJ. (2007). Development and psychometric properties of the ambulatory self-confidence questionnaire. *Gerontology,* 53(6): 373-381.

Attema AE, Edelaar-Peeters Y, Versteegh MM, & Stolk EA. (2013). Time trade-off: one methodology, different methods. *The European Journal of Health Economics,* 14 Suppl 1: S53-64.

Austin PC, Escobar M, & Kopec JA. (2000). The use of the Tobit model for analyzing measures of health status. *Quality of Life Research,* 9(8): 901-910.

Bajaj PS, & Veenstra DL. (2013). A risk-benefit analysis of factor V Leiden testing to improve pregnancy outcomes: a case study of the capabilities of decision modeling in genomics. *Genetics in Medicine,* 15(5): 374-381.

Baker R, Bateman I, Donaldson C, Jones-Lee M, Lancsar E, Loomes G, *et al.* (2010). Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project. *Health Technology Assessment,* 14(27): 1-162.

Banke-Thomas AO, Madaj B, Charles A, & van den Broek N. (2015). Social Return on Investment (SROI) methodology to account for value for money of public health interventions: a systematic review. *BMC Public Health,* 15: 582.

Bansback N, Brazier J, Tsuchiya A, & Anis A. (2012). Using a discrete choice experiment to estimate health state utility values. *Journal of Health Economics,* 31(1): 306-318.

Bansback N, Pollard S, & Bryan S. (2014). Eliciting public values to support health care priority setting: a novel approach. *CADTH Symposium, 6-8 April, 2014*. Gatineau, Quebec, Canada.

Barnett E, & Casper M. (2001). A definition of "social environment". *American Journal of Public Health,* 91(3): 465.

Bayoumi AM. (2004). The measurement of contingent valuation for health economics. *Pharmacoeconomics,* 22(11): 691-700.

BC Housing. (n.d.). Shelter Aid for Elderly Renters (SAFER) rental subsidy: eligibility. Retrieved August 10, 2017, from https://http://www.bchousing.org/housing-assistance/rental-assistance-financial-aid-for-home-modifications/shelter-aid-for-elderly-renters.

Beale SJ, Bending MW, Trueman P, & Naidoo B. (2012). Should we invest in environmental interventions to encourage physical activity in England? An economic appraisal. *The European Journal of Health Economics,* 22(6): 869-873.

Benning TM, Alayli-Goebbels AFG, Aarts MJ, Stolk E, de Wit GA, Prenger R, *et al.* (2015). Exploring outcomes to consider in economic evaluations of health promotion programs: what broader non-health outcomes matter most? *Bmc Health Services Research,* 15: 266.

Birch S, & Donaldson C. (2003). Valuing the benefits and costs of health care programmes: where's the 'extra' in extra-welfarism? *Social Science & Medicine,* 56(5): 1121-1133.

Birch S, Melnikow J, & Kuppermann M. (2003). Conservative versus aggressive follow up of mildly abnormal Pap smears: testing for process utility. *Health Economics,* 12(10): 879-884.

Bobinac A, van Exel NJA, Rutten FFH, & Brouwer WBF. (2010). Caring for and caring about: disentangling the caregiver effect and the family effect. *Journal of Health Economics,* 29(4): 549-556.

Bobinac A, van Exel NJA, Rutten FFH, & Brouwer WBF. (2011). Health effects in significant others: separating family and care-giving effects. *Medical Decision Making,* 31(2): 292-298.

Borghi J, & Jan S. (2008). Measuring the benefits of health promotion programmes: application of the contingent valuation method. *Health Policy,* 87(2): 235-248.

Brazier J, Akehurst R, Brennan A, Dolan P, Claxton K, McCabe C, *et al.* (2005). Should patients have a greater role in valuing health states? *Applied Health Economics and Health Policy,* 4(4): 201-208.

Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T, *et al.* (2014). A systematic review, psychometric analysis and qualitative assessment of generic

preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technology Assessment,* 18(34): 1-188.

Brazier J, Deverill M, Green C, Harper R, & Booth A. (1999). A review of the use of health status measures in economic evaluation. *Health Technology Assessment,* 3(9): i-iv, 1-164.

Brazier J, Ratcliffe J, Salomon J, & Tsuchiya A. (2017). *Measuring and valuing health benefits for economic evaluation. Second Edition*. Oxford, United Kingdom: Oxford University Press.

Brazier J, Ratcliffe J, Tsuchiya A, & Salomon J. (2007). *Measuring and valuing health benefits for economic evaluation* New York: Oxford University Press.

Brazier J, Roberts J, & Deverill M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics,* 21(2): 271-292.

Brazier J, Roberts J, Tsuchiya A, & Busschbach J. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics,* 13(9): 873-884.

Brazier J, Rowen D, Yang Y, & Tsuchiya A. (2012). Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. *The European Journal of Health Economics,* 13(5): 575-587.

Brazier J, & Tsuchiya A. (2010). Preference-based condition-specific measures of health: what happens to cross programme comparability? *Health Economics,* 19(2): 125-129.

Brazier J, & Tsuchiya A. (2015). Improving cross-sector comparisons: going beyond the health-related QALY. *Applied Health Economics and Health Policy,* 13(6): 557-565.

Brazier JE, Dixon S, & Ratcliffe J. (2009). The role of patient preferences in cost-effectiveness analysis a conflict of values? *Pharmacoeconomics,* 27(9): 705-712.

Brazier JE, & Roberts J. (2004). The estimation of a preference-based measure of health from the SF-12. *Medical Care,* 42(9): 851-859.

Brennan VK, & Dixon S. (2013). Incorporating process utility into quality adjusted life years: a systematic review of empirical studies. *Pharmacoeconomics,* 31(8): 677-691.

Briggs AH, & O'Brien BJ. (2001). The death of cost-minimization analysis? *Health Economics,* 10(2): 179-184.

Brooks R. (1996). EuroQol: the current state of play. *Health Policy,* 37(1): 53-72.

Broome J. (2002). Measuring the burden of disease by aggregating well-being. In Murray C J L, Salomon J A, Mathers C D & Lopez A D (Eds.), *Summary measures of population health: concepts, ethics, measurement and applications* (pp. 91-113). Geneva: World Health Organization.

Brouwer WB, & Koopmanschap MA. (2000). On the economic foundations of CEA. Ladies and gentlemen, take your positions! *Journal of Health Economics,* 19(4): 439-459.

Brouwer WBF. (2008). Perspective, costs, outcomes and discounting in pharmacoeconomic evaluations. *European Journal of Hospital Pharmacy,* 14(3): 20-22.

Brouwer WBF, Culyer AJ, van Exel NJA, & Rutten FFH. (2008). Welfarism vs. extra-welfarism. *Journal of Health Economics,* 27(2): 325-338.

Brouwer WBF, Koopmanschap MA, & Rutten FFH. (1997). Productivity costs measurement through quality of life? A response to the recommendation of the Washington Panel. *Health Economics,* 6(3): 253-259.

Brouwer WBF, van Exel NJA, Baltussen R, & Rutten FFH. (2006). A dollar is a dollar is a dollar - or is it? *Value in Health,* 9(5): 341-347.

Brouwer WBF, van Exel NJA, van den Berg B, van den Bos GAM, & Koopmanschap MA. (2005). Process utility from providing informal care: the benefit of caring. *Health Policy,* 74(1): 85-99.

Bryan S, & Dolan P. (2004). Discrete choice experiments in health economics. For better or for worse? *The European Journal of Health Economics,* 5(3): 199-202.

Buchanan J, & Wordsworth S. (2015). Welfarism versus extra-welfarism: can the choice of economic evaluation approach impact on the adoption decisions recommended by economic evaluation studies? *Pharmacoeconomics,* 33(6): 571-579.

Buchanan J, Wordsworth S, & Schuh A. (2013). Issues surrounding the health economic evaluation of genomic technologies. *Pharmacogenomics,* 14(15): 1833-1847.

Buxton M. (2008). Discussion of Paul Dolan and Aki Tsuchiya's paper: 'The measurement and valuation of public safety'. In Mason A & Towse A (Eds.), *The*

ideas and influence of Alan Williams: be reasonable, do it my way! (pp. 147-151). Oxon, UK: Radcliffe Publishing Ltd.

Byford S, & Sefton T. (2003). Economic evaluation of complex health and social care interventions. *National Institute Economic Review,* 186(1): 98-108.

CADTH. (2017). Guidelines for the economic evaluation of health technologies: Canada. 4th Edition. Ottawa, Canada: CADTH.

Campbell R, Pound P, Morgan M, Daker-White G, Britten N, Pill R, *et al.* (2011). Evaluating meta-ethnography: systematic analysis and synthesis of qualitative research. *Health Technology Assessment,* 15(43): 1-164.

Canaway A, Al-Janabi H, Kinghorn P, Bailey C, & Coast J. (2016). Development of a measure (ICECAP-Close Person Measure) through qualitative methods to capture the benefits of end-of-life care to those close to the dying for use in economic evaluation. *Palliative Medicine,* 31(1): 53-62.

Cerin E, Saelens BE, Sallis JF, & Frank LD. (2006). Neighborhood environment walkability scale: validity and development of a short form. *Medicine and Science in Sports and Exercise,* 38(9): 1682-1691.

Chalkidou K, Culyer A, Naidoo B, & Littlejohns P. (2008). Cost-effective public health guidance: asking questions from the decision-maker's viewpoint. *Health Economics,* 17(3): 441-448.

Charlifue S, Post MW, Biering-Sorensen F, Catz A, Dijkers M, Geyh S, *et al.* (2012). International spinal cord injury quality of life basic data set. *Spinal Cord,* 50(9): 672-675.

Chisholm D, Healey A, & Knapp M. (1997). QALYs and mental health care. *Social Psychiatry and Psychiatric Epidemiology,* 32(2): 68-75.

Chuang L-H, & Whitehead SJ. (2012). Mapping for economic evaluation. *British Medical Bulletin,* 101: 1-15.

Chudyk AM, Winters M, Moniruzzaman M, Ashe MC, Sims-Gould J, & McKay H. (2014). Destinations matter: the association between where older adults live and their travel behavior. *Journal of Transport and Land Use,* 2(1): 50-57.

Clark D, & Olsen JA. (1994). Agency in health care with an endogenous budget constraint. *Journal of Health Economics,* 13(2): 231-251.

Claxton K, Sculpher M, & Culyer AJ. (2007). Mark versus Luke? Appropriate methods for the evaluation of public health interventions. CHE Research Paper 31. York: Centre for Health Economics, The University of York.

Claxton K, Walker S, Palmer S, & Sculpher M. (2010). Appropriate perspectives for health care decisions. CHE Research Paper 54. York: Centre for Health Economics, The University of York.

Cleemput I, Neyt M, Thiry N, De Laet C, & Leys M. (2011). Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. *International Journal of Technology Assessment in Health Care,* 27(1): 71-76.

Coast J. (2004). Is economic evaluation in touch with society's health values? *British Medical Journal,* 329(7476): 1233-1236.

Coast J. (2009). Maximisation in extra-welfarism: A critique of the current position in health economics. *Social Science & Medicine,* 69(5): 786-792.

Coast J. (2014). Strategies for the economic evaluation of end-of-life care: making a case for the capability approach. *Expert Review of Pharmacoeconomics & Outcomes Research,* 14(4): 473-482.

Coast J, & Donovan J. (1996). Conflict, complexity and confusion: the context for priority setting. In Coast J, Donovan J & Frankel S (Eds.), *Priority setting: the health care debate*. Chichester, England: John Wiley & Sons Ltd.

Coast J, Flynn T, Sutton E, Al-Janabi H, Vosper J, Lavender S, *et al.* (2008a). Investigating Choice Experiments for Preferences of Older People (ICEPOP): evaluative spaces in health economics. *Journal of Health Services Research & Policy,* 13 Suppl 3: 31-37.

Coast J, Flynn TN, Natarajan L, Sproston K, Lewis J, Louviere JJ, *et al.* (2008b). Valuing the ICECAP capability index for older people. *Social Science & Medicine,* 67(5): 874-882.

Coast J, Peters TJ, Natarajan L, Sproston K, & Flynn T. (2008c). An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people. *Quality of Life Research,* 17(7): 967-976.

Coast J, Smith RD, & Lorgelly P. (2008d). Welfarism, extra-welfarism and capability: the spread of ideas in health economics. *Social Science & Medicine,* 67(7): 1190-1198.

Cohen J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Connell J, O'Cathain A, & Brazier J. (2014). Measuring quality of life in mental health: are we asking the right questions? *Social Science & Medicine,* 120: 12-20.

Cookson R. (2005a). QALYs and capabilities: a response to Anand. *Health Economics,* 14(12): 1287-1289.

Cookson R. (2005b). QALYs, and the capability approach. *Health Economics,* 14(8): 817-829.

Costello AB, & Osborne JW. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation,* 10(7): 1-9.

Coulter ID, Herman PM, & Nataraj S. (2013). Economic analysis of complementary, alternative, and integrative medicine: considerations raised by an expert panel. *BMC Complementary and Alternative Medicine,* 13: 191.

Couzner L, Crotty M, Norman R, & Ratcliffe J. (2013a). A comparison of the EQ-5D-3L and ICECAP-O in an older post-acute patient population relative to the general population. *Applied Health Economics and Health Policy,* 11(4): 415-425.

Couzner L, Ratcliffe J, Lester L, Flynn T, & Crotty M. (2013b). Measuring and valuing quality of life for public health research: application of the ICECAP-O capability index in the Australian general population. *International Journal of Public Health,* 58(3): 367-376.

Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, *et al.* (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *British Medical Journal,* 337: a1655.

Craven C, Hitzig SL, & Mittmann N. (2012). Impact of impairment and secondary health conditions on health preference among Canadians with chronic spinal cord injury. *The Journal of Spinal Cord Medicine,* 35(5): 361-370.

Crosignani PG, Baird DT, Barri PN, Bhattacharya S, Devroey P, Evers JLH, *et al.* (2015). Economic aspects of infertility care: a challenge for researchers and clinicians. *Human Reproduction,* 30(10): 2243-2248.

Culyer AJ. (1989). The normative economics of health care finance and provision. *Oxford Review of Economic Policy,* 5(1): 34-58.

Culyer AJ, & Evans RG. (1996). Mark Pauly on welfare economics: normative rabbits from positive hats. *Journal of Health Economics,* 15(2): 243-251.

Dakin H, Devlin N, Feng Y, Rice N, O'Neill P, & Parkin D. (2013). The influence of cost-effectiveness and other factors on NICE decisions. Research Paper 05/14. Oxford: Health Economics Research Centre (HERC), University of Oxford.

Davidson T, & Levin LA. (2010). Is the societal approach wide enough to include relatives? Incorporating relatives' costs and effects in a cost-effectiveness analysis. *Applied Health Economics and Health Policy,* 8(1): 25-35.

Davis JC, Best JR, Bryan S, Li LDC, Hsu CL, Gomez C, *et al.* (2015). Mobility is a key predictor of change in well-being among older adults who experience falls: evidence from the Vancouver falls prevention clinic cohort. *Archives of Physical Medicine and Rehabilitation,* 96(9): 1634-1640.

Davis JC, Best JR, Dian L, Khan KM, Hsu CL, Chan W, *et al.* (2016). Are the EQ-5D-3L and the ICECAP-O responsive among older adults with impaired mobility? Evidence from the Vancouver Falls Prevention Cohort Study. *Quality of Life Research,* 26(3): 737-747.

Davis JC, Liu-Ambrose T, Richardson CG, & Bryan S. (2013). A comparison of the ICECAP-O with EQ-5D in a falls prevention clinical setting: are they complements or substitutes? *Quality of Life Research,* 22(5): 969-977.

Day R. (2008). Local environments and older people's health: dimensions from a comparative qualitative study in Scotland. *Health & Place,* 14(2): 299-312.

de Vries M, Emons WHM, Plantinga A, Pietersma S, van den Haut WB, Stiggelbout AM, *et al.* (2016). Comprehensively measuring health-related subjective well-being: dimensionality analysis for improved outcome assessment in health economics. *Value in Health,* 19(2): 167-175.

Devlin N, & Sussex J. (2011). Incorporating multiple criteria in HTA: methods and processes. London: Office of Health Economics.

Ding A, Eisenberg JD, & Pandharipande PV. (2011). The economic burden of incidentally detected findings. *Radiologic Clinics of North America,* 49(2): 257-265.

Dirksen CD. (2014). The use of research evidence on patient preferences in health care decision-making: issues, controversies and moving forward. *Expert Review of Pharmacoeconomics & Outcomes Research,* 14(6): 785-794.

Dixon S, Shackley S, & Brazier J. (2011). What is dignity? A literature review and conceptual mapping. HEDS Discussion Paper 11/13. Sheffield: The University of Sheffield.

Dixon-Woods M, Cavers D, Agarwal S, Annandale E, Arthur A, Harvey J, *et al.* (2006). Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Med Res Methodol,* 6: 35.

Dolan P. (1999). Whose preferences count? *Medical Decision Making,* 19(4): 482-486.

Dolan P. (2011). Using happiness to value health. London, UK: Office of Health Economics.

Dolan P, & Kaheman D. (2008). Interpretations of utility and their implications for the valuation of health. *The Economic Journal,* 118(525): 215-234.

Dolan P, Kavetsos G, & Tsuchiya A. (2013). Sick but satisfied: the impact of life and health satisfaction on choice between health scenarios. *Journal of Health Economics,* 32(4): 708-714.

Dolan P, Lee H, & Peasgood T. (2012). Losing sight of the wood for the trees: some issues in describing and valuing health, and another possible approach. *Pharmacoeconomics,* 30(11): 1035-1049.

Dolan P, & Olsen JA. (2002). *Distributing health care: economic and ethical issues*. Oxford: Oxford University Press.

Dolan P, Shaw R, Tsuchiya A, & Williams A. (2005). QALY maximisation and people's preferences: a methodological review of the literature. *Health Economics,* 14(2): 197-208.

Donaldson C, & Shackley P. (1997). Does "process utility" exist? A case study of willingness to pay for laparoscopic cholecystectomy. *Social Science & Medicine,* 44(5): 699-707.

Doward LC, & McKenna SP. (2004). Defining patient-reported outcomes. *Value in Health,* 7: 4-8.

Dowie J. (2001). Analysing health outcomes. *Journal of Medical Ethics,* 27(4): 245-250.

Drummond M, Brixner D, Gold M, Kind P, McGuire A, & Nord E. (2009). Toward a consensus on the QALY. *Value in Health,* 12 Suppl 1: S31-35.

Drummond M, Sculpher M, Claxton K, Stoddart G, & Torrance G. (2015). *Methods for the economic evaluation of health care programmes* (Fourth ed.). Oxford: Oxford University Press.

Drummond M, Weatherly H, Claxton K, Cookson R, Ferguson B, Godfrey C, *et al.* (2007). Assessing the challenges of applying standard methods of economic

evaluation to public health interventions. York: Public Health Research Consortium.

Eden M, Payne K, Combs RM, Hall G, McAllister M, & Black GCM. (2013). Valuing the benefits of genetic testing for retinitis pigmentosa: a pilot application of the contingent valuation method. *British Journal of Ophthalmology,* 97(8): 1051-1056.

Edwards RT, Bryning L, & Crane R. (2015). Design of economic evaluations of mindfulness-based interventions: ten methodological questions of which to be mindful. *Mindfulness,* 6(3): 490-500.

Edwards RT, Charles JM, & Lloyd-Williams H. (2013). Public health economics: a systematic review of guidance for the economic evaluation of public health interventions and discussion of key methodological issues. *BMC Public Health,* 13: 1001.

Elliott J, Gale CR, Parsons S, Kuh D, & Team HAS. (2014). Neighbourhood cohesion and mental wellbeing among older adults: a mixed methods approach. *Social Science & Medicine,* 107: 44-51.

Engel L, Bansback N, Bryan S, Doyle-Waters MM, & Whitehurst DG. (2016). Exclusion criteria in national health state valuation studies: s systematic review. *Medical Decision Making,* 36(7): 798-810.

Entwistle V, Firnigl D, Ryan M, Francis J, & Kinghorn P. (2012). Which experiences of health care delivery matter to service users and why? A critical interpretive synthesis and conceptual map. *Journal of Health Services Research & Policy,* 17(2): 70-78.

Essink-Bot ML, Korfage IJ, & De Koning HJ. (2003). Including the quality-of-life effects in the evaluation of prostate cancer screening: expert opinions revisited? *Bju International,* 92: 101-105.

Fabrigar LR, & Wegener DT. (2012). *Understanding statistics: exploratory factor analysis*. New York: Oxford University Press.

Fabrigar LR, Wegener DT, MacCallum RC, & Strahan EJ. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods,* 4(3): 272-299.

Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, *et al.* (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical Care,* 40(2): 113-128.

Feng Y, Devlin N, & Herdman M. (2015). Assessing the health of the general population in England: how do the three- and five-level versions of EQ-5D compare? *Health and Quality of Life Outcomes,* 13: 171.

Ferrans CE, Zerwic JJ, Wilbur JE, & Larson JL. (2005). Conceptual model of health-related quality of life. *Journal of Nursing Scholarship,* 37(4): 336-342.

Flood EM, De Cock E, Mork AC, & Revicki DA. (2006). Evaluating preference weights for the Asthma Symptom Utility Index (ASUI) across countries. *Health & Quality of Life Outcomes,* 4: 51.

Flynn TN. (2010a). Using conjoint analysis and choice experiments to estimate QALY values: issues to consider. *Pharmacoeconomics,* 28(9): 711-722.

Flynn TN. (2010b). Valuing citizen and patient preferences in health: recent developments in three types of best–worst scaling. *Expert Review of Pharmacoeconomics & Outcomes Research,* 10(3): 259-267.

Flynn TN, Chan P, Coast J, & Peters TJ. (2011). Assessing quality of life among British older people using the ICEPOP CAPability (ICECAP-O) measure. *Applied Health Economics and Health Policy,* 9(5): 317-329.

Flynn TN, Huynh E, Peters TJ, Al-Janabi H, Clemens S, Moody A, *et al.* (2015). Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. *Health Economics,* 24(3): 258-269.

Flynn TN, Louviere JJ, Peters TJ, & Coast J. (2007). Best--worst scaling: what it can do for health care research and how to do it. *Journal of Health Economics,* 26(1): 171-189.

Frank L. (2013). Enhancing walk score's ability to predict physical activity and active transportation. Retrieved August 10, 2017, from http://activelivingresearch.org/files/2013_Bike-WalkScore_Frank.pdf.

Frank LD, Schmid TL, Sallis JF, Chapman J, & Saelens BE. (2005). Linking objectively measured physical activity with objectively measured urban form: findings from SMARTRAQ. *American Journal of Preventive Medicine,* 28(2 Suppl 2): 117-125.

Franke T, Tong C, Ashe MC, McKay H, Sims-Gould J, & Walk The Talk T. (2013). The secrets of highly active older adults. *Journal of Aging Studies,* 27(4): 398-409.

Friedman D, Parikh NS, Giunta N, Fahs MC, & Gallo WT. (2012). The influence of neighborhood factors on the quality of life of older adults attending New York City senior centers: results from the Health Indicators Project. *Quality of Life Research,* 21(1): 123-131.

Furlong WJ, Feeny DH, Torrance GW, & Barr RD. (2001). The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Annals of Medicine,* 33(5): 375-384.

Gafni A. (1994). The standard gamble method: what is being measured and how it is interpreted. *Health Services Research,* 29(2): 207-224.

Gafni A. (2006). Economic evaluation of health-care programmes: Is CEA better than CBA? *Environmental & Resource Economics,* 34(3): 407-418.

Gafni A, & Birch S. (1995). Preferences for outcomes in economic evaluation: an economic approach to addressing economic problems. *Social Science & Medicine,* 40(6): 767-776.

Gale CR, Dennison EM, Cooper C, & Sayer AA. (2011). Neighbourhood environment and positive mental health in older people: the Hertfordshire Cohort Study. *Health & Place,* 17(4): 867-874.

Gandjour A. (2001). Is subjective well-being a useful parameter for allocating resources among public interventions? *Health Care Analysis,* 9(4): 437-447.

Garin N, Olaya B, Miret M, Ayuso-Mateos JL, Power M, Bucciarelli P, *et al.* (2014). Built environment and elderly population health: a comprehensive literature review. *Clinical Practice and Epidemiology in Mental Health,* 10: 103-115.

Gebel K, Bauman AE, Sugiyama T, & Owen N. (2011). Mismatch between perceived and objectively assessed neighborhood walkability attributes: prospective relationships with walking and weight gain. *Health & Place,* 17(2): 519-524.

Geiser C. (2013). *Data analysis with Mplus*. New York: The Guilford Press.

Goebbels AFG, Lakerveld J, Ament AJHA, Bot SDM, & Severens JL. (2012). Exploring non-health outcomes of health promotion: the perspective of participants in a lifestyle behaviour change intervention. *Health Policy,* 106(2): 177-186.

Gold MR, Siegel JE, Russel LB, & Weinstein MC. (1996). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.

Golicki D, Niewada M, Buczek J, Karlinska A, Kobayashi A, Janssen MF, *et al.* (2015). Validity of EQ-5D-5L in stroke. *Quality of Life Research,* 24(4): 845-850.

Goranitis I, Coast J, Day E, Copello A, Freemantle N, & Frew E. (2017). Maximizing health or sufficient capability in economic evaluation? A methodological experiment of treatment for drug addiction. *Medical Decision Making,* 37(5): 498-511.

Greco G, Lorgelly P, & Yamabhai I. (2016). Outcomes in Economic Evaluations of Public Health Interventions in Low- and Middle-Income Countries: Health, Capabilities and Subjective Wellbeing. *Health Economics,* 25 Suppl 1: 83-94.

Grewal I, Lewis J, Flynn T, Brown J, Bond J, & Coast J. (2006). Developing attributes for a generic quality of life measure for older people: preferences or capabilities? *Social Science & Medicine,* 62(8): 1891-1901.

Groll DL, To T, Bombardier C, & Wright JG. (2005). The development of a comorbidity index with physical function as the outcome. *Journal of Clinical Epidemiology,* 58(6): 595-602.

Grosse SD, McBride CM, Evans JP, & Khoury MJ. (2009). Personal utility and genomic information: look before you leap. *Genetics in Medicine,* 11(8): 575-576.

Grosse SD, Wordsworth S, & Payne K. (2008). Economic methods for valuing the outcomes of genetic testing: beyond cost-effectiveness analysis. *Genetics in Medicine,* 10(9): 648-654.

Grossman M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy,* 80(2): 223-255.

Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, *et al.* (1994). A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *Journal of Gerontology,* 49(2): M85-94.

Gyrd-Hansen D. (2005). Willingness to pay for a QALY: theoretical and methodological issues. *Pharmacoeconomics,* 23(5): 423-432.

Handy SL, Boarnet MG, Ewing R, & Killingsworth RE. (2002). How the built environment affects physical activity: views from urban planning. *American Journal of Preventive Medicine,* 23(2 Suppl): 64-73.

Hansen P. (2012). Health sector decision-making: more than just cost per QALY calculations. *Journal of Health Services Research & Policy,* 17(3): 129-130.

Hanson HM, Schiller C, Winters M, Sims-Gould J, Clarke P, Curran E, *et al.* (2013). Concept mapping applied to the intersection between older adults' outdoor walking and the built and social environments. *Preventive Medicine,* 57(6): 785-791.

Harris J. (1987). QALYfying the value of life. *Journal of Medical Ethics,* 13(3): 117-123.

Hartley RJ, Keen EM, Large JA, & Tedd LA. (1990). Search strategies. In Hartley R J (Ed.), *Online searching: Principles and practice* (pp. 153-173). London: Bowker-Saur.

Hausman DM. (2012). Health, well-being, and measuring the burden of disease. *Population Health Metrics,* 10(1).

Hawthorne G, Richardson J, & Osborne R. (1999). The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health-related quality of life. *Quality of Life Research,* 8(3): 209-224.

Health Quality Council of Alberta. (2014). 2014 Alberta Population Norms for EQ-5D-5L. Retrieved August 10, 2017, from https://d10k7k7mywg42z.cloudfront.net/assets/542f01f2edb2f37083002e54/2014_EQ_5D_5L_report_FINALFINAL.pdf.

Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, *et al.* (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research,* 20(10): 1727-1736.

Herlitz A, & Horan D. (2016). Measuring needs for priority setting in healthcare planning and policy. *Social Science & Medicine,* 157: 96-102.

Higgins A, Barnett J, Meads C, Singh J, & Longworth L. (2014). Does convenience matter in health care delivery? A systematic review of convenience-based aspects of process utility. *Value in Health,* 17(8): 877-887.

Hoefman RJ, van Exel J, & Brouwer W. (2013). How to include informal care in economic evaluations. *Pharmacoeconomics,* 31(12): 1105-1119.

Howard K, Salkeld G, McCaffery K, & Irwig L. (2008). HPV triage testing or repeat Pap smear for the management of atypical squamous cells (ASCUS) on Pap smear: Is there evidence of process utility? *Health Economics,* 17(5): 593-605.

Hu L, & Bentler PM. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling,* 6(1): 1-55.

Huber M, Knottnerus JA, Green L, van der Horst H, Jadad AR, Kromhout D, *et al.* (2011). How should we define health? *British Medical Journal,* 343: d4163.

Husereau D, Marshall DA, Levy AR, Peacock S, & Hoch JS. (2014). Health technology assessment and personalized medicien: are economic evaluation guidelines sufficient to support decision making? *International Journal of Technology Assessment in Health Care,* 30(2): 179-187.

Huter K, Kocot E, Kissimova-Skarbek K, Dubas-Jakobczyk K, & Rothgang H. (2016). Economic evaluation of health promotion for older people-methodological problems and challenges. *BMC Health Services Research,* 16 Suppl 5: 328.

Jacobsson F, Carstensen J, & Borgquist L. (2005). Caring externalities in health economic evaluation: how are they related to severity of illness? *Health Policy,* 73(2): 172-182.

Jan S. (1998). A holistic approach to the economic evaluation of health programs using institutionalist methodology. *Social Science & Medicine,* 47(10): 1565-1572.

Jensen MP, Molton IR, Groah SL, Campbell ML, Charlifue S, Chiodo A, *et al.* (2012). Secondary health conditions in individuals aging with SCI: terminology, concepts and analytic approaches. *Spinal Cord,* 50(5): 373-378.

Johnson FR. (2009). Editorial: Moving the QALY forward or just stuck in traffic? *Value in Health,* 12 Suppl 1: S38-39.

Kahneman D, & Sugden R. (2005). Experienced utility as a standard of policy evaluation. *Environmental & Resource Economics,* 32(1): 161-181.

Kalpakjian CZ, Scelza WM, Forchheimer MB, & Toussaint LL. (2007). Preliminary reliability and validity of a Spinal Cord Injury Secondary Conditions Scale. *The Journal of Spinal Cord Medicine,* 30(2): 131-139.

Karimi M, & Brazier J. (2016). Health, health-related quality of life, and quality of life: what is the difference? *Pharmacoeconomics,* 34(7): 645-649.

Karimi M, Brazier J, & Basarir H. (2016). The capability approach: a critical review of its application in health economics. *Value in Health,* 19(6): 795-799.

Karimi M, Brazier J, & Paisley S. (2013). The capability approach in health economics: are we applying it correctly? *Health Economists' Study Group winter 2013 meeting*. Exeter, UK.

Kawachi I, & Berkman L. (2000). Social cohesion, social capital, and health. In Berkman L & Kawachi I (Eds.), *Social Epidemiology*. New York: Oxford University Press.

Keeley T, Al-Janabi H, Nicholls E, Foster NE, Jowett S, & Coast J. (2015). A longitudinal assessment of the responsiveness of the ICECAP-A in a randomised controlled trial of a knee pain intervention. *Quality of Life Research,* 24(10): 2319-2331.

Keeley T, Coast J, Nicholls E, Foster NE, Jowett S, & Al-Janabi H. (2016). An analysis of the complementarity of ICECAP-A and EQ-5D-3 L in an adult population of patients with knee pain. *Health and Quality of Life Outcomes,* 14(1): 36.

King D. (2008). Neighborhood and individual factors in activity in older adults: results from the neighborhood and senior health study. *Journal of Aging and Physical Activity,* 16(2): 144-170.

Kinghorn P. (2015). Exploring different interpretations of the capability approach in a health care context: where next? *Journal of Human Development and Capabilities,* 16(4): 600-616.

Kinghorn P, Robinson A, & Smith RD. (2015). Developing a capability-based questionnaire for assessing well-being in patients with chronic pain. *Social Indicators Research,* 120: 897-916.

Kotzian P. (2009). Value for money: health system efficiency and preferences for health care. *Canadian Journal of Political Science,* 42(3): 729-748.

Krabbe PFM, Stouthard MEA, Essink-Bot M-L, & Bonsel GJ. The effect of adding a cognitive dimension to the EuroQol multiattribute health-status classification system. *Journal of Clinical Epidemiology,* 52(4): 293-301.

Labelle RJ, & Hurley JE. (1992). Implications of basing health-care resource allocations on cost-utility analysis in the presence of externalities. *Journal of Health Economics,* 11(3): 259-277.

Lancaster K. (1966). A new approach to consumer theory. *Journal of Political Economy,* 74(April): 132-157.

Lancsar E, & Louviere J. (2008). Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics,* 26(8): 661-677.

Lawton MP. (1977). An ecological theory of aging applied to elderly housing. *Environments and Aging,* 31(1): 8-10.

Laxminarayan R, Jamison DT, Krupnick AJ, & Norheim OF. (2014). Valuing vaccines using value of statistical life measures. *Vaccine,* 32(39): 5065-5070.

Leslie E, Saelens B, Frank L, Owen N, Bauman A, Coffee N, *et al.* (2005). Residents' perceptions of walkability attributes in objectively different neighbourhoods: a pilot study. *Health & Place,* 11(3): 227-236.

Liljas B. (2011). Welfare, QALYs, and costs - a comment. *Health Economics,* 20(1): 68-72.

Lochner K, Kawachi I, & Kennedy BP. (1999). Social capital: a guide to its measurement. *Health & Place,* 5(4): 259-270.

Longworth L, Yang Y, Young T, Mulhern B, Hernández Alava M, Mukuria C, *et al.* (2014). Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technology Assessment,* 18(9): 1-224.

Lorgelly P, Lorimer K, Fenwick EA, & Briggs AH. (2008). The capability approach: developing an instrument for evaluating public health interventions: Final Report. Glasgow: Glasgow Centre for Population Health.

Lorgelly PK. (2015). Choice of outcome measure in an economic evaluation: a potential role for the capability approach. *Pharmacoeconomics,* 33(8): 849-855.

Lorgelly PK, Lawson KD, Fenwick EA, & Briggs AH. (2010). Outcome measurement in economic evaluations of public health interventions: a role for the capability approach? *International Journal of Environmental Research and Public Health,* 7(5): 2274-2289.

Lorgelly PK, Lorimer K, Fenwick EAL, Briggs AH, & Anand P. (2015). Operationalising the capability approach as an outcome measure in public health: The development of the OCAP-18. *Social Science & Medicine,* 142: 68-81.

Lu CY, & Cohen JP. (2015). Can genomic medicine improve financial sustainability of health systems? *Molecular Diagnosis & Therapy,* 19(2): 71-77.

Maas CJM, & Hox JJ. (2005). Sufficient sample sizes for multilevel modeling. *Methodology,* 1(3): 86-92.

MacKinnon DP. (2008). *Introduction to statistical mediation analysis*. New York: Erlbaum.

Makai P, Beckebans F, van Exel J, & Brouwer WB. (2014a). Quality of life of nursing home residents with dementia: validation of the German version of the ICECAP-O. *PLOS ONE,* 9(3): e92016.

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

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

Makai P, Looman W, Adang E, Melis R, Stolk E, & Fabbricotti I. (2015). Cost-effectiveness of integrated care in frail elderly using the ICECAP-O and EQ-5D: does choice of instrument matter? *The European Journal of Health Economics,* 16(4): 437-450.

Marsh K, Phillips CJ, Fordham R, Bertranou E, & Hale J. (2012). Estimating cost-effectiveness in public health: a summary of modelling and valuation methods. *Health Economics Review,* 2(1): 17.

Mason A, Goddard M, Weatherly H, & Chalkley M. (2015). Integrating funds for health and social care: an evidence review. *Journal of Health Services Research & Policy,* 20(3): 177-188.

Mathes T, Jacobs E, Morfeld J-C, & Pieper D. (2013). Methods of international health technology assessment agencies for economic evaluations- a comparative analysis. *BMC Health Services Research,* 13(1): 371.

McAllister M, Dunn G, Payne K, Davies L, & Todd C. (2012). Patient empowerment: The need to consider it as a measurable patient-reported outcome for chronic conditions. *Bmc Health Services Research,* 12: 157.

McDonough CM, & Tosteson ANA. (2007). Measuring preferences for cost-utility analysis: how choice of method may influence decision-making. *Pharmacoeconomics,* 25(2): 93-106.

McGinn AP, Evenson KR, Herring AH, Huston SL, & Rodriguez DA. (2007). Exploring associations between physical activity and perceived and objective measures of the built environment. *Journal of Urban Health,* 84(2): 162-184.

McNamee P, & Seymour J. (2008). Incorporation of process preferences within the QALY framework: a study of alternative methods. *Medical Decision Making,* 28(3): 443-452.

McNeill LH, Kreuter MW, & Subramanian SV. (2006a). Social environment and physical activity: a review of concepts and evidence. *Social Science & Medicine,* 63(4): 1011-1022.

McNeill LH, Wyrwich KW, Brownson RC, Clark EM, & Kreuter MW. (2006b). Individual, social environmental, and physical environmental influences on physical activity among black and white adults: a structural equation analysis. *Annals of Behavioral Medicine,* 31(1): 36-44.

Michel YA, Engel L, Rand-Hendriksen K, Augestad LA, & Whitehurst DG. (2016). "When I saw walking I just kind of took it as wheeling": interpretations of

mobility-related items in generic, preference-based health state instruments in the context of spinal cord injury. *Health and Quality of Life Outcomes,* 14(1): 164.

Mihalopoulos C, Chen G, Iezzi A, Khan MA, & Richardson J. (2014). Assessing outcomes for cost-utility analysis in depression: comparison of five multi-attribute utility instruments with two depression-specific outcome measures. *The British Journal of Psychiatry*, 205(5): 390-397.

Mitchell PM, Al-Janabi H, Richardson J, Iezzi A, & Coast J. (2015a). The relative impacts of disease on health status and capability wellbeing: a multi-country study. *PLOS One,* 10(12): e0143590.

Mitchell PM, Roberts TE, Barton PM, & Coast J. (2015b). Assessing sufficient capability: A new approach to economic evaluation. *Social Science & Medicine,* 139: 71-79.

Mitchell PM, Roberts TE, Barton PM, & Coast J. (2017a). Applications of the Capability Approach in the Health Field: A Literature Review. *Social Indicators Research,* 133(1): 345-371.

Mitchell PM, Venkatapuram S, Richardson J, Iezzi A, & Coast J. (2017b). Are quality-adjusted life years a good proxy measure of individual capabilities? *Pharmacoeconomics,* 35(6): 637-646.

Mitra S. (2006). The Capability Approach and Disability. *Journal of Disability Policy Studies,* 16: 236-247.

Mokhtarian PL, & Cao X. (2008). Examining the impacts of residential self-selection on travel behavior: a focus on methodologies. *Transportation Research. Part B: Methodological,* 42(3): 204-228.

Mooney G. (1989). QALYs: are they enough? A health economist's perspective. *Journal of Medical Ethics,* 15(3): 148-152.

Mooney G. (1994). What else do we want from our health services? *Social science & medicine,* 39(2): 151-154.

Mooney G. (1998). Beyond health outcomes: the benefits of health care. *Health Care Analysis,* 6(2): 99-105.

Mooney G. (2009). *Challanging health economics*. New York: Oxford University Press.

Mooney GH, & Drummond MF. (1982). Essentials of health economics: Part I - what is economics? *British Medical Journal (Clinical Research Edition),* 285(6346): 949-950.

Morgan SG, McMahon M, Mitton C, Roughead E, Kirk R, Kanavos P, *et al.* (2006). Centralized drug review processes in Australia, Canada, New Zealand, and the United kingdom. *Health Affairs,* 25(2): 337-347.

Morris KS, McAuley E, & Motl RW. (2008). Self-efficacy and environmental correlates of physical activity among older women and women with multiple sclerosis. *Health Education Research,* 23(4): 744-752.

Mukuria C, & Brazier J. (2013). Valuing the EQ-5D and the SF-6D health states using subjective well-being: a secondary analysis of patient data. *Social Science & Medicine,* 77: 97-105.

Mulhern B, Rowen D, Jacoby A, Marson T, Snape D, Hughes D, *et al.* (2012). The development of a QALY measure for epilepsy: NEWQOL-6D. *Epilepsy & Behavior,* 24(1): 36-43.

Muthén LK, & Muthén BO. (1998-2015). *Mplus User's Guide. Seventh Edition.* Los Angeles: Muthén & Muthén.

Netten A, Burge P, Malley J, Potoglou D, Towers AM, Brazier J, *et al.* (2012). Outcomes of social care for adults: developing a preference-weighted measure. *Health Technology Assessment,* 16(16): 1-166.

Neumann PJ, Cohen JT, Hammitt JK, Concannon TW, Auerbach HR, Fang CH, *et al.* (2012). Willingness-to-pay for predictive tests with no immediate treatment implications: a survey of US residents. *Health Economics,* 21(3): 238-251.

Neumann PJ, Goldie SJ, & Weinstein MC. (2000). Preference-based measures in economic evaluation in health care. *Annual Review of Public Health,* 21: 587-611.

NICE. (2008). Social Value Judgements: Principles for the development of NICE guidance. Second Edition. London: NICE.

NICE. (2012). Methods for the development of NICE public health guidance (third edition) London: NICE.

NICE. (2013). Guide to the methods of technology appraisal 2013. London: NICE.

NICE. (2014). Developing NICE guidelines: the manual. London: NICE.

NICE. (2016). The social care guidance manual. London: NICE.

Nicholas K. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *The Economic Journal,* 49(195): 549-552.

Nilsson CJ, Avlund K, & Lund R. (2010). Social inequality in onset of mobility disability among older Danes: the mediation effect of social relations. *Journal of Aging and Health,* 22(4): 522-541.

Noonan VK, Kopec JA, Zhang H, & Dvorak MF. (2008). Impact of associated conditions resulting from spinal cord injury on health status and quality of life in people with traumatic central cord syndrome. *Archives of Physical Medicine and Rehabilitation,* 89(6): 1074-1082.

Nord E, Daniels N, & Kamlet M. (2009). QALYs: some challenges. *Value in Health,* 12 Suppl 1: S10-15.

Noreau L, & Boschen K. (2010). Intersection of participation and environmental factors: a complex interactive process. *Archives of Physical Medicine and Rehabilitation,* 91(9 Suppl): S44-53.

Noreau L, Noonan VK, Cobb J, Leblond J, & Dumont FS. (2014). Spinal cord injury community survey: a national, comprehensive study to portray the lives of canadians with spinal cord injury. *Topics in Spinal Cord Injury Rehabilitation,* 20(4): 249-264.

Nussbaum M. (2003). Capabilities as fundamental entitlements: Sen and social justice. *Feminist Economics,* 9(2-3): 33-59.

Nyman JA. (2011). Measurement of QALYs and the welfare implications of survivor consumption and leisure forgone. *Health Economics,* 20(1): 56-67.

Olsen JA. (2017). *Principles in Health Economics and Policy. Second Edition*. Oxford, UK: Oxford.

Olsen JA, & Smith RD. (2001). Theory versus practice: a review of 'willingness-to-pay' in health and health care. *Health Economics,* 10(1): 39-52.

Ong KS, Kelaher M, Anderson I, & Carter R. (2009). A cost-based equity weight for use in the economic evaluation of primary health care interventions: case study of the Australian Indigenous population. *International Journal for Equity in Health,* 8: 34.

Opmeer BC, de Borgie CA, Mol BW, & Bossuyt PM. (2010). Assessing preferences regarding healthcare interventions that involve non-health outcomes: an overview of clinical studies. *The Patient,* 3(1): 1-10.

Papageorgiou K, Vermeulen KM, Schroevers MJ, Stiggelbout AM, Buskens E, Krabbe PFM, *et al.* (2015). Do individuals with and without depression value depression differently? And if so, why? *Quality of Life Research,* 24(11): 2565-2575.

Papaioannou D, Sutton A, Carroll C, Booth A, & Wong R. (2010). Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Information & Libraries Journal,* 27(2): 114-122.

Pawson R, Greenhalgh T, Harvey G, & Walshe K. (2005). Realist review--a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy,* 10 Suppl 1: 21-34.

Payne K, McAllister M, & Davies LM. (2013). Valuing the economic benefits of complex interventions: when maximising health is not sufficient. *Health Economics,* 22(3): 258-271.

Payne K, & Thompson AJ. (2013). Economics of pharmacogenomics: rethinking beyond QALYs? *Current Pharmacogenomics & Personalized Medicine,* 11(3): 187-195.

Pearl J. (2012). The causal mediation formula--a guide to the assessment of pathways and mechanisms. *Prevention Science,* 13(4): 426-436.

Peeters Y, Ranchor AV, Vliet Vlieland TP, & Stiggelbout AM. (2010). Effect of adaptive abilities on utilities, direct or mediated by mental health? *Health and Quality of Life Outcomes,* 8: 130.

Peeters Y, & Stiggelbout AM. (2010). Health state valuations of patients and the general public analytically compared: a meta-analytical comparison of patient and population health state utilities. *Value in Health,* 13(2): 306-309.

Perreault WD. (1976). Controlling order-effect bias. *Public Opinion Quarterly,* 39(4): 544-551.

Petrou S, & Wolstenholme J. (2000). A review of alternative approaches to healthcare resource allocation. *Pharmacoeconomics,* 18(1): 33-43.

Porter ME. (2010). What is value in health care? *The New England Journal of Medicine,* 363(26): 2477-2481.

Post MW, de Witte LP, van Asbeck FW, van Dijk AJ, & Schrijvers AJ. (1998). Predictors of health status and life satisfaction in spinal cord injury. *Archives of Physical Medicine and Rehabilitation,* 79(4): 395-401.

Prieto L, Alonso J, & Lamarca R. (2003). Classical test theory versus rasch analysis for quality of life questionnaire reduction. *Health and Quality of Life Outcomes,* 1: 27.

Pullenayegum EM, Tarride JE, Xie F, Goeree R, Gerstein HC, & O'Reilly D. (2010). Analysis of health utility data when some subjects attain the upper bound of 1: are Tobit and CLAD models appropriate? *Value Health,* 13(4): 487-494.

Putzke JD, Richards JS, Hicken BL, & DeVivo MJ. (2002). Predictors of life satisfaction: a spinal cord injury cohort study. *Archives of Physical Medicine and Rehabilitation,* 83(4): 555-561.

Pyne JM, Fortney JC, Tripathi S, Feeny D, Ubel P, & Brazier J. (2009). How bad is depression? Preference score estimates from depressed patients and the general population. *Health Services Research,* 44(4): 1406-1423.

Rabin R, & de Charro F. (2001). EQ-5D: a measure of health status from the EuroQol Group. *Annals of Medicine,* 33(5): 337-343.

Rantakokko M, Iwarsson S, Kauppinen M, Leinonen R, Heikkinen E, & Rantanen T. (2010). Quality of life and barriers in the urban outdoor environment in old age. *Journal of the American Geriatrics Society,* 58(11): 2154-2159.

Ratcliffe J, Brazier J, Palfreyman S, & Michaels J. (2007). A comparison of patient and population values for health states in varicose veins patients. *Health Economics,* 16(4): 395-405.

Ratcliffe J, Couzner L, Flynn T, Sawyer M, Stevens K, Brazier J, *et al.* (2011). Valuing child health utility 9D health states with a young adolescent sample. *Applied Health Economics and Health Policy,* 9(1): 15-27.

Rawles J. (1989). Castigating QALYs. *Journal of Medical Ethics,* 15(3): 143-147.

Raykov T, & Marcoulides GA. (1999). On desirability of parsimony in structural equation model selection. *Structural Equation Modeling,* 6(3): 292-300.

Richardson J, Chen G, Khan MA, & Iezzi A. (2015a). Can multi-attribute utility instruments adequately account for subjective well-being? *Medical Decision Making,* 35(3): 292-304.

Richardson J, Iezzi A, & Khan MA. (2015b). Why do multi-attribute utility instruments produce different utilities: the relative importance of the descriptive systems, scale and 'micro-utility' effects. *Quality of Life Research,* 24(8): 2045-2053.

Richardson J, Iezzi A, Khan MA, Sinha K, Mihalopoulos C, Herrman H, *et al.* (2009). Data used in the development of the AQoL-8D (PsyQoL) Quality of Life Instrument. Melbourne: Centre for Health Economics, Monash University.

Richardson J, Iezzi A, & Maxwell A. (2012). Cross-national comparison of twelve quality of life instruments. MIC Paper 1 - Background, questions, instruments. Research Paper 76. Melbourne: Centre for Health Economics, Monash University.

Richardson J, Khan MA, Iezzi A, & Maxwell A. (2015c). Comparing and explaining differences in the magnitude, content, and sensitivity of utilities predicted by the EQ-5D, SF-6D, HUI 3, 15D, QWB, and AQoL-8D multiattribute utility instruments. *Medical Decision Making,* 35(3): 276-291.

Richardson J, McKie J, & Bariola E. (2011). Review and critique of health related multi attribute utility instruments. Research Paper 64. Melbourne: Centre for Health Economics, Monash University.

Richardson J, McKie J, & Bariola E. (2014a). Multiattribute utility instruments and their use. In Culyer A J (Ed.), *Encyclopedia of Health Economics* (Vol. 2, pp. 341-357). San Diego: Elsevier.

Richardson J, Sinha K, Iezzi A, & Khan MA. (2014b). Modelling utility weights for the Assessment of Quality of Life (AQoL)-8D. *Quality of Life Research,* 23(8): 2395-2404.

Rivers CS, Fallah N, Noonan VK, Whitehurst DGT, Schwartz C, Finkelstein J, *et al.* (2017). Health conditions: impact on function, health-related quality of life, and life satisfaction following traumatic spinal cord injury. A prospective observational registry cohort study. *Archives of Physical Medicine and Rehabilitation.*

Robeyns I. (2005). The capability approach: a theoretical survey. *Journal of Human Development and Capabilities,* 6(1): 93-117.

Rogowski WH, Grosse SD, Schmidtke J, & Marckmann G. (2014). Criteria for fairly allocating scarce health-care resources to genetic tests: which matter most? *European Journal of Human Genetics,* 22(1): 25-31.

Rogowski WH, & Schleidgen S. (2015). Using needs-based frameworks for evaluating new technologies: An application to genetic tests. *Health Policy,* 119(2): 147-155.

Rosso AL, Auchincloss AH, & Michael YL. (2011). The urban built environment and mobility in older adults: a comprehensive review. *Journal of Aging Research,* 2011: 816106.

Round J. (2012). Is a QALY still a QALY at the end of life? *Journal of Health Economics,* 31(3): 521-527.

Round J, Sampson EL, & Jones L. (2014). A framework for understanding quality of life in individuals without capacity. *Quality of Life Research,* 23(2): 477-484.

Rowen D, Brazier J, & Van Hout B. (2015). A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Medical Decision Making,* 35(3): 328-340.

Rowles GD. (1983). Place and personal identity in old age: observations from Appalachia. *Journal of Environmental Psychology,* 3(4): 299-313.

Rubin DB. (2004). *Multiple imputation for nonresponse in surveys*. New York: John Wiley and Sons.

Russell D, Peplau LA, & Cutrona CE. (1980). The revised UCLA loneliness scale: concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology,* 39(3): 472-480.

Ryan M. (1999). A role for conjoint analysis in technology assessment in health care? *International Journal of Technology Assessment in Health Care,* 15(3): 443-457.

Ryan M, Kinghorn P, Entwistle VA, & Francis JJ. (2014). Valuing patients' experiences of healthcare processes: towards broader applications of existing methods. *Social Science & Medicine,* 106: 194-203.

Ryan M, & Shackley P. (1995). Assessing the benefits of health care: how far should we go? *Quality in Health Care,* 4(3): 207-213.

Sach TH, Smith RD, & Whynes DK. (2007). A 'league table' of contingent valuation results for pharmaceutical interventions a hard pill to swallow? *Pharmacoeconomics,* 25(2): 107-127.

Saelens BE, Sallis JF, & Frank LD. (2003). Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine,* 25(2): 80-91.

Salkeld G. (1998). What are the benefits of preventive health care? *Health Care Analysis,* 6(2): 106-112.

Salkeld G, Quine S, & Cameron ID. (2004). What constitutes success in preventive health care? A case study in assessing the benefits of hip protectors. *Social Science & Medicine,* 59(8): 1593-1601.

Sallis JF. (2009). Measuring physical activity environments: a brief history. *American Journal of Preventive Medicine,* 36(4 Suppl): S86-92.

Sampson CJ. (2016). Identifying objects of value at the end of life In Round J (Ed.), *Care at the End of Life - An Economic Perspective* (pp. 103-122). Switzerland: Springer International Publishing

Sampson RJ, Raudenbush SW, & Earls F. (1997). Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science,* 277(5328): 918-924.

Sanders GD, Neumann PJ, Basu A, & et al. (2016). Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second panel on cost-effectiveness in health and medicine. *JAMA,* 316(10): 1093-1103.

Sassi F. (2006). Calculating QALYs, comparing QALY and DALY calculations. *Health Policy Plan,* 21(5): 402-408.

Schlander M, Garattini S, Holm S, Kolominsky-Rabas P, Nord E, Persson U, *et al.* (2014). Incremental cost per quality-adjusted life year gained? The need for alternative methods to evaluate medical interventions for ultra-rare disorders. *Journal of Comparative Effectiveness Research,* 3(4): 399-422.

Schulz DN, Smit ES, Stanczyk NE, Kremers SPJ, de Vries H, & Evers S. (2014). Economic evaluation of a web-based tailored lifestyle intervention for adults: findings regarding cost-effectiveness and cost-utility from a randomized controlled trial. *Journal of Medical Internet Research,* 16(3): 383-400.

Schulz R, & Decker S. (1985). Long-term adjustment to physical disability: the role of social support, perceived control, and self-blame. *Journal of Personality and Social Psychology,* 48(5): 1162-1172.

Seiber WJ, Groessl EJ, David KM, Ganiats TG, & Kaplan RM. (2008). Quality of Well Being Self Administered (QWB-SA) Scale. User's Manual. San Diego: Health Services Research Center, University of California.

Sen A. (1993). Capability and well-being. In Nussbaum M & Sen A (Eds.), *Quality of Life Research*. Oxford: Oxford University Press.

Sexton E, Bennett K, Fahey T, & Cahir C. (2016). Does the EQ-5D capture the effects of physical and mental health status on life satisfaction among older people? A path analysis approach. *Quality of Life Research,* 26(5): 1-10.

Sezer N, Akkus S, & Ugurlu FG. (2015). Chronic complications of spinal cord injury. *World Journal of Orthopedics,* 6(1): 24-33.

Shackley P, Slack R, & Michaels J. (2001). Vascular patients' preferences for local treatment: an application of conjoint analysis. *Journal of Health Services Research & Policy,* 6(3): 151-157.

Shah K, Praet C, Devlin N, Sussex J, Appleby J, & Parkin D. (2012). Is the aim of the English health care system to maximize QALYs? *Journal of Health Services Research & Policy,* 17(3): 157-163.

Simon J, Anand P, Gray A, Rugkasa J, Yeeles K, & Burns T. (2013). Operationalising the capability approach for outcome measurement in mental health research. *Social Science & Medicine,* 98: 187-196.

Singer P, McKie J, Kuhse H, & Richardson J. (1995). Double jeopardy and the use of QALYs in health-care allocation. *Journal of Medical Ethics,* 21(3): 144-150.

Sintonen H. (2001). The 15D instrument of health-related quality of life: properties and applications. *Annals of Medicine,* 33(5): 328-336.

Stanczyk NE, Smit ES, Schulz DN, de Vries H, Bolman C, Muris JWM, *et al.* (2014). An economic evaluation of a video- and text-based computer-tailored intervention for smoking cessation: a cost-effectiveness and cost-utility analysis of a randomized controlled trial. *PLOS ONE,* 9(10): e110117.

Starnes HA, McDonough MH, Tamura K, James P, Laden F, & Troped PJ. (2014). Factorial validity of an abbreviated neighborhood environment walkability scale for seniors in the Nurses' Health Study. *The International Journal of Behavioral Nutrition and Physical Activity,* 11: 126.

StataCorp. (2007). Stata Statistical Software: Release 10. College Station, TX: StataCorp LP.

StataCorp. (2015). Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.

Statistics Canada. (2010). Canadian Community Health Survey (CCHS). Annual Component - 2010 Questionnaire.

Stephens C, Breheny M, & Mansvelt J. (2015). Healthy ageing from the perspective of older people: A capability approach to resilience. *Psychology & Health,* 30(6): 715-731.

Stevens K, & Palfreyman S. (2012). The use of qualitative methods in developing the descriptive systems of preference-based measures of health-related quality of life for use in economic evaluation. *Value in Health,* 15(8): 991-998.

Stevens KJ. (2016). How well do the generic multi-attribute utility instruments incorporate patient and public views into their descriptive dystems? *The Patient,* 9(1): 5-13.

Stolk EA, van Donselaar G, Brouwer WBF, & Busschbach JJV. (2004). Reconciliation of economic concerns and health policy. *Pharmacoeconomics,* 22(17): 1097-1107.

Strauss DJ, DeVivo MJ, Paculdo DR, & Shavelle RM. Trends in life expectancy after spinal cord injury. *Archives of Physical Medicine and Rehabilitation,* 87(8): 1079-1085.

Streiner DL. (2005). Finding our way: an introduction to path analysis. *The Canadian Journal of Psychiatry,* 50(2): 115-122.

Sugiyama T, & Ward Thompson C. (2007). Outdoor environments, activity and the well-being of older people: conceptualising environmental support. *Environment and Planning,* 39: 1943-1960.

Sullivan PW. (2011). Are utilities bounded at 1.0? Implications for statistical analysis and scale development. *Medical Decision Making,* 31(6): 787-789.

Sundaram M, Smith MJ, Revicki DA, Miller LA, Madhavan S, & Hobbs G. (2010). Estimation of a valuation function for a diabetes mellitus-specific preference-based measure of health: the Diabetes Utility Index. *Pharmacoeconomics,* 28(3): 201-216.

Sutton EJ, & Coast J. (2014). Development of a supportive care measure for economic evaluation of end-of-life care using qualitative methods. *Palliative Medicine,* 28(2): 151-157.

Swan JS, Pandharipande PV, & Salazar GM. (2016). Developing a patient-centered radiology process model. *Journal of the American College of Radiology,* 13(5): 510-516.

Swan JS, Sainfort F, Lawrence WF, Kuruchittham V, Kongnakorn T, & Heisey DM. (2003). Process utility for imaging in cerebrovascular disease. *Academic Radiology,* 10(3): 266-274.

Swan JS, Ying J, Stahl J, Kong CY, Moy B, Roy J, *et al.* (2010). Initial development of the Temporary Utilities Index: a multiattribute system for classifying the functional health impact of diagnostic testing. *Quality of Life Research,* 19(3): 401-412.

Swinburn P, Lloyd A, Boye KS, Edson-Heredia E, Bowman L, & Janssen B. (2013). Development of a disease-specific version of the EQ-5D-5L for use in patients suffering from psoriasis: lessons learned from a feasibility study in the UK. *Value in Health,* 16(8): 1156-1162.

Thorn JC, Noble SM, & Hollingworth W. (2014). Methodological developments in randomized controlled trial-based economic evaluations. *Expert Review of Pharmacoeconomics & Outcomes Research,* 14(6): 843-856.

Tinelli M, Ryan M, Bond C, & Scott A. (2013). Valuing benefits to inform a clinical trial in pharmacy do differences in utility measures at baseline affect the effectiveness of the intervention? *Pharmacoeconomics,* 31(2): 163-171.

Torgerson D, & Raftery J. (1999). Economics notes: measuring outcomes in economic evaluations. *British Medical Journal,* 318(7195): 1413.

Torrance GW. (1986). Measurement of health state utilities for economic appraisal. *Journal of Health Economics,* 5(1): 1-30.

Towse A, & Garrison LP. (2013). Economic incentives for evidence generation: promoting an efficient path to personalized medicine. *Value in Health,* 16(6): S39-S43.

Trani J-F, Bakhshi P, Bellanca N, Biggeri M, & Marchetta F. (2011). Disabilities through the Capability Approach lens: Implications for public policies. *ALTER - European Journal of Disability Research / Revue Européenne de Recherche sur le Handicap,* 5(3): 143-157.

Trueman P, & Anokye NK. (2013). Applying economic evaluation to public health interventions: the case of interventions to promote physical activity. *Journal of Public Health,* 35(1): 32-39.

Turrell G. (2010). Understanding the local physical activity environment and obesity. In Pearce J & Witten K (Eds.), *Geographies of obesity - environmental understandings of the obesity epidemic* (pp. 151-174). Aldershot, UK: Ashgate Publishing Group.

van de Wetering EJ, Stolk EA, van Exel NJ, & Brouwer WB. (2013). Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall. *The European Journal of Health Economics,* 14(1): 107-115.

van de Wetering EJ, van Exel J, & Brouwer W. (2016). Health or happiness? A note on trading off health and happiness in rationing decisions. *Value in Health,* 19(5): 552-557.

van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, *et al.* (2012). Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value in Health,* 15(5): 708-715.

van Leeuwen CM, Post MW, van Asbeck FW, Bongers-Janssen HM, van der Woude LH, de Groot S, *et al.* (2012). Life satisfaction in people with spinal cord injury during the first five years after discharge from inpatient rehabilitation. *Disability and Rehabilitation,* 34(1): 76-83.

van Leeuwen KM, Bosmans JE, Jansen AP, Hoogendijk EO, van Tulder MW, van der Horst HE, *et al.* (2015a). Comparing measurement properties of the EQ-5D-3L, ICECAP-O, and ASCOT in frail older adults. *Value in Health,* 18(1): 35-43.

van Leeuwen KM, Jansen AP, Muntinga ME, Bosmans JE, Westerman MJ, van Tulder MW, *et al.* (2015b). Exploration of the content validity and feasibility of the EQ-5D-3L, ICECAP-O and ASCOT in older adults. *BMC Health Services Research,* 15: 201.

van Mastrigt GA, Paulus AT, Aarts MJ, Evers SM, & Alayli-Goebbels AF. (2015). A qualitative study on the views of experts regarding the incorporation of non-health outcomes into the economic evaluations of public health interventions. *BMC Public Health,* 15(1): 954.

Veenstra DL, Roth JA, Garrison LP, Ramsey SD, & Burke W. (2010). A formal risk-benefit framework for genomic tests: facilitating the appropriate translation of genomics into clinical practice. *Genetics in Medicine,* 12(11): 686-693.

Versteegh MM, & Brouwer WBF. (2016). Patient and general public preferences for health states: A call to reconsider current guidelines. *Social Science & Medicine,* 165: 66-74.

Versteegh MM, Leunis A, Uyl-de Groot CA, & Stolk EA. (2012). Condition-specific preference-based measures: benefit or burden? *Value in Health,* 15(3): 504-513.

Villanueva K, Pereira G, Knuiman M, Bull F, Wood L, Christian H, *et al.* (2013). The impact of the built environment on health across the life course: design of a cross-sectional data linkage study. *British Medical Journal Open,* 3(1).

Viney R, Norman R, Brazier J, Cronin P, King MT, Ratcliffe J, *et al.* (2014). An Australian discrete choice experiment to value eq-5d health States. *Health Economics,* 23(6): 729-742.

von Faber M, Bootsma–van der Wiel A, van Exel E, & et al. (2001). Successful aging in the oldest old: Who can be characterized as successfully aged? *Archives of Internal Medicine,* 161(22): 2694-2700.

von Neumann J, & Morgenstern O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University.

Wagstaff A. (1991). QALYs and the equity-efficiency trade-off. *Journal of Health Economics,* 10(1): 21-41.

Wailoo A, Tsuchiya A, & McCabe C. (2009). Weighting must wait: incorporating equity concerns into cost-effectiveness analysis may take longer than expected. *Pharmacoeconomics,* 27(12): 983-989.

Ware JE, Jr., Kosinski M, Bjorner JB, Turner-Bowker DM, Gandek B, & Maruish ME. (2008). *SF-36v2® Health Survey: Administration guide for clinical trial investigators* (Vol. Quality Metric Incorporated). Lincoln, RI.

Weernink MGM, Groothuis-Oudshoorn CGM, Ijzerman MJ, & van Til JA. (2016). Valuing Treatments for parkinson disease incorporating process utility: performance of best-worst scaling, time trade-off, and visual analogue scales. *Value in Health,* 19(2): 226-232.

Weinstein MC, Torrance G, & McGuire A. (2009). QALYs: the basics. *Value in Health,* 12 Suppl 1: S5-9.

Welch Saleeby P. (2006). Applications of a Capability approach to disability and the International Classification of Functioning, Disability and Health (ICF) in social work practice. *Journal of Social Work in Disability & Rehabilitation,* 6(1-2): 217-232.

Whitehead SJ, & Ali S. (2010). Health outcomes in economic evaluation: the QALY and utilities. *British Medical bulletin,* 96: 5-21.

Whitehurst DG, & Bryan S. (2011). Another study showing that two preference-based measures of health-related quality of life (EQ-5D and SF-6D) are not interchangeable. But why should we expect them to be? *Value in Health,* 14(4): 531-538.

Whitehurst DG, & Engel L. (2017). Disability discrimination and the quality-adjusted life year framework: misplaced criticism that sheds light on the importance of 'inclusive' descriptive systems. *2nd EuroQol Academy Meeting, 7 and 8 March 2017*. Noordwijk aan Zee, the Netherlands.

Whitehurst DG, & Mittmann N. (2013). The value of health economics research in spinal cord injury. *Spinal Cord,* 51(8): 586-587.

Whitehurst DG, Mittmann N, Noonan VK, Dvorak MF, & Bryan S. (2016). Health state descriptions, valuations and individuals' capacity to walk: a comparative evaluation of preference-based instruments in the context of spinal cord injury. *Quality of Life Research,* 25(10): 2481-2496.

Whitehurst DG, Noonan VK, Dvorak MF, & Bryan S. (2012). A review of preference-based health-related quality of life questionnaires in spinal cord injury research. *Spinal Cord,* 50(9): 646-654.

Whitehurst DG, Norman R, Brazier JE, & Viney R. (2014a). Comparison of contemporaneous EQ-5D and SF-6D responses using scoring algorithms derived from similar valuation exercises. *Value in Health,* 17(5): 570-577.

Whitehurst DG, Suryaprakash N, Engel L, Mittmann N, Noonan VK, Dvorak MF, *et al.* (2014b). Perceptions of individuals living with spinal cord injury toward preference-based quality of life instruments: a qualitative exploration. *Health and Quality of Life Outcomes,* 12: 50.

Whittingham MJ, Stephens PA, Bradbury RB, & Freckleton RP. (2006). Why do we still use stepwise modelling in ecology and behaviour? *The Journal of Animal Ecology,* 75(5): 1182-1189.

WHO. (1948). Constitution of the World Health Organization. Basic documents. Geneva: WHO.

WHO. (2001). International classification of functioning, disability and health. Geneva: WHO.

WHO. (2002a). Definition of an older or elderly person. Retrieved August 10, 2017, from http://www.who.int/healthinfo/survey/ageingdefnolder/en/.

WHO. (2002b). Towards a common language for functioning, disability and health: ICF. Geneva.

Wichmann AB, Adang EM, Stalmeier PF, Kristanti S, Van den Block L, Vernooij-Dassen MJ, *et al.* (2017). The use of quality-adjusted life years in cost-effectiveness analyses in palliative care: mapping the debate through an integrative review. *Palliative Medicine,* 31(4): 306-322.

Williams A. (1996). QALYS and ethics: a health economist's perspective. *Social Science & Medicine,* 43(12): 1795-1804.

Williams A. (1997). Intergenerational equity: an exploration of the 'fair innings' argument. *Health Economics,* 6(2): 117-132.

Williams B, Brown T, & Onsman A. (2010). Exploratory factor analysis: a five-step guide for novices. *Australasian Journal of Paramedicine,* 8(3).

Wilson IB, & Cleary PD. (1995). Linking clinical variables with health-related quality of life. a conceptual model of patient outcomes. *JAMA,* 273(1): 59-65.

Wilson MG, Ellen ME, Lavis JN, Grimshaw JM, Moat KA, Shemer J, *et al.* (2014). Processes, contexts, and rationale for disinvestment: a protocol for a critical interpretive synthesis. *Systematic Reviews,* 3: 143.

Wisløff T, Hagen G, Hamidi V, Movik E, Klemp M, & Olsen JA. (2014). Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. *Pharmacoeconomics*, 32(4): 367-375.

Wister AV, Levasseur M, Griffith LE, & Fyffe I. (2015). Estimating multiple morbidity disease burden among older persons: a convergent construct validity study to discriminate among six chronic illness measures, CCHS 2008/09. *BMC Geriatrics,* 15: 12.

Wolff J, Edwards S, Richmond S, Orr S, & Rees G. (2012). Evaluating interventions in health: a reconciliatory approach. *Bioethics,* 26(9): 455-463.

Wright DR, Wittenberg E, Swan JS, Miksad RA, & Prosser LA. (2009). Methods for measuring temporary health states for cost-utility analyses. *Pharmacoeconomics*, 27(9): 713-723.

Xie F, Gaebel K, Perampaladas K, Doble B, & Pullenayegum E. (2013). Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. *Medical Decision Making,* 34(1): 8-20.

Xie F, Pullenayegum E, Gaebel K, Bansback N, Bryan S, Ohinmaa A, *et al.* (2016). A time trade-off-derived value Set of the EQ-5D-5L for Canada. *Medical Care,* 54(1): 98-105.

Yang Y, Brazier J, & Tsuchiya A. (2014). Effect of adding a sleep dimension to the EQ-5D descriptive system: a "bolt-on" experiment. *Medical Decision Making,* 34(1): 42-53.

Yen IH, Fandel Flood J, Thompson H, Anderson LA, & Wong G. (2014). How design of places promotes or inhibits mobility of older adults: realist synthesis of 20 years of research. *Journal of Aging and Health,* 26(8): 1340-1372.

Zorginstituut Nederland. (2016). Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. Diemen: Zorginstituut Nederland.

Appendices to Chapter 3

Appendix 3.1. Characteristics of the included papers

Author(s), year	Type of work	Objective	Concept of interest
Core papers (n=21)			
Al-Janabi *et al*., 2011	Discussion paper	To discuss the evaluative scope of QALYs – going beyond the patient	Care-related quality of life
Al-Janabi *et al*., 2012	Empirical analysis	To develop the ICECAP-A	Capability wellbeing
Birch & Donaldson, 2003	Discussion paper	To discuss the welfarism and the extra-welfarism approach	Welfarism/ extra-welfarism
Brennan & Dixon, 2013	Review	To summarize studies that provide an empirical measure of process utility	Process utility
Brouwer *et al*., 2008	Discussion paper	To compare welfarism and extra-welfarism	Welfarism/ extra-welfarism
Chisholm *et al*., 1997	Discussion paper	To discuss strengths and shortcomings of QALYs in the context of mental health	QALYs in mental health
Coast *et al*., 2008a	Discussion paper	To discuss the evaluative framework for the ICEPOP	Capability and the quality of life of older people; unpaid care for older people; and end-of-life care for older people
Coast *et al*., 2008d	Discussion paper	To discuss welfarism, extra-welfarism, and the capability approach for use in health economics	Welfarism/ extra-welfarism/ capability approach
Cookson, 2005b	Discussion paper	To discuss the application of the capability approach in QALYs	Capability QALY

Donaldson & Shackley, 1997	Empirical analysis	To proof the existence of process utility in health care	Process utility
Dowie, 2001	Discussion paper	To analyze health outcomes	Health outcomes
Gandjour, 2001	Discussion paper	To discuss the role of subjective wellbeing for resource allocation decisions	Subjective wellbeing
Goebbels *et al*., 2012	Qualitative research	To explore non-health outcomes in health promotion	Non-health outcomes
Lorgelly *et al*., 2010	Review	To review outcome measure used in public health and to discuss the role of the capability approach	Capability approach
Mooney, 1998	Discussion paper	To discuss outcome beyond health	Outcomes beyond health
Opmeer *et al*., 2010	Review	To review the literature to assess the extent to which valuations of process and non-health outcomes have been elicited in patients	Non-health outcomes
Payne *et al*., 2013	Discussion paper	To discuss outcomes of clinical genetics services and the evaluation of such services	Complex interventions
Round *et al*., 2014	Discussion paper	To develop a framework to better understand the quality of life in individuals living without capacity	Individuals without capacity
Ryan *et al*., 2014	Review & qualitative research	To analyze patients' benefits of health care processes	Process utility
Ryan & Shackley, 1995	Discussion paper	To assess benefits of health care	Benefits of health care
Simon *et al*., 2013	Empirical analysis	To operationalize the capability approach in mental health	Capability approach
1st wave of search (n=47)			
Al-Janabi *et al*., 2013b	Empirical analysis	To investigate the construct validity of the ICECAP-A	Capability wellbeing

Alayli-Goebbels *et al*., 2013	Empirical analysis	To explore consumer preferences for health and non-health outcomes of health promotion	Non-health outcomes in health promotion
Ali & Ronaldson, 2012	Discussion paper	To discuss ordinal preference elicitation methods in health economics	Elicitation techniques
Annemans *et al*., 2013	Discussion paper	To discuss methodological issues in economic evaluations of personalized medicine	Personalized medicine
Birch *et al*., 2003	Empirical analysis	To test for process utility in the context of pap smears	Process utility
Brouwer, 2008	Discussion paper	To discuss perspectives, costs, outcomes and discounting in pharmacoeconomic evaluations	Pharmacoeconomic evaluations
Brouwer *et al*., 2006	Discussion paper	To discuss whether costs should be equally considered in economic evaluation	Cost perspective
Coast, 2014	Discussion paper	To examine different theoretical perspectives from which economic evaluation of end of life care could be conducted	End of life care
Connell *et al*., 2014	Qualitative research	To identify the domains of quality of life that are important to people with mental health problems	Mental health
Coulter *et al*., 2013	Discussion paper	Expert panel views on Economic analysis of complementary, alternative, and integrative medicine	Complementary, alternative, and integrative medicine
Crosignani *et al*., 2015	Discussion paper	To discuss the economic aspects of infertility care	Infertility care
Davidson & Levin, 2010	Discussion paper	To analyze how relatives' costs and effects could be measured, valued and incorporated into a cost-effectiveness analysis	Relatives' QALY (R-QALY)
Davis *et al*., 2013	Empirical analysis	To compare the ICECAP-O with the EQ-5D	EQ-5D vs. ICECAP-O

Dirksen, 2014	Discussion paper	To provide an overview of the issues and controversies related to the use of research evidence on patient preferences	Patient preferences
Edwards *et al*., 2013	Review	To examine what guidance currently exists in the field of economic evaluations of public health economics	Economic evaluation in public health
Essink-Bot *et al*., 2003	Discussion paper	To apply the CUA framework in prostate cancer screening	Prostate cancer screening
Gandjour, 2014	Letter to the Editor	To discuss disutility from waiting times	Waiting times
Grosse *et al*., 2008	Review	To discuss the relative merits of different economic measures and methods to inform recommendations relative to genetic testing for risk of disease	Genetic testing
Hausman, 2012	Discussion paper	To discuss whether the global burden of diseases, injuries, and risk factors should be measured in terms of their consequences for health or in terms of their consequences for wellbeing	Measuring burden of disease
Herlitz & Horan, 2016	Discussion paper	To measure needs for priority setting in healthcare planning and policy	Needs
Hoefman *et al*., 2013	Discussion paper	To discuss how to Include Informal Care in Economic Evaluations	Informal care
Howard *et al*., 2008	Empirical analysis	To elicit the utilities of women of screening age for different management approaches for the evaluation of an Pap smear result	Process utility
Kotzian, 2009	Discussion paper	To discuss why existing health care system efficiency evaluations fail to take fully into account the effect of preferences and their variation	Health System Efficiency

Laxminarayan *et al*., 2014	Discussion paper	To describe the pros and cons of using benefit-cost analysis and cost-effectiveness analysis in the context of vaccine priorities	Valuing vaccines
Makai *et al*., 2012	Empirical analysis	To validate the ICECAP-O in psycho-geriatric elderly in nursing homes	Capabilities and quality of life in nursing homes
Makai *et al*., 2014b	Review	To assess the usefulness of HRQoL and wellbeing instruments for economic evaluations in older people	Economic evaluation in older adults
Makai *et al*., 2013a	Empirical analysis	To validate the ICECAP-O in post-hospitalized older people	Capability wellbeing
McAllister *et al*., 2012	Discussion paper	To discuss why patient empowerment itself is a directly measureable patient reported outcome for chronic conditions.	Patient empowerment
Netten *et al*., 2012	Empirical analysis	To develop a preference-based measure of social care outcome	Social care-related quality of life
Nyman, 2011	Discussion paper	To discuss whether to include survivor consumption and leisure forgone should be included in a cost-utility analysis	Survivor consumption and leisure forgone
Olsen & Smith, 2001	Review	To outline the arguments advanced for the superiority of WTP over QALYs	WTP versus QALYs
Ong *et al*., 2009	Discussion paper	To propose an alternative to cost-based equity weight for use in economic evaluation	Equity weights
Petrou & Wolstenholme, 2000	Review	To review alternative approaches to healthcare resource allocation	Healthcare resource allocation
Rogowski *et al*., 2014	Discussion paper	To assess substantive ethical and economic criteria to prioritize genetic services	Genetic services
Round, 2012	Discussion paper	To discuss the QALY at the end of life	End of life

Ryan, 1999	Empirical analysis	To demonstrate the use of conjoint analysis in health services research	Conjoint analysis
Salkeld, 1998	Discussion paper	To explore the concept of process utility in the context of preventive goods	Process utility and preventive goods
Salkeld *et al*., 2004	Empirical analysis	To outline the nature of the ex ante and ex post perspective in valuing benefits and the presence of process utility and the utility of gambling in individual's utility function for preventive health care	Process utility and preventive goods
Schlander *et al*., 2014	Discussion paper	To discuss issues in using the QALY approach to evaluate interventions for ultra-rare disorders	Ultra-rare disorders
Schulz *et al*., 2014	Empirical analysis	To assess the cost-effectiveness and cost-utility of a sequential and a simultaneous Web-based computer-tailored lifestyle intervention for adults	Web-Based Tailored Lifestyle Intervention
Shackley *et al*., 2001	Empirical analysis	To investigate whether and to what extent vascular patients are willing to trade expected health outcomes for improvements in non-health benefits	Non-health outcomes
Stanczyk *et al*., 2014	Empirical analysis	To investigate the cost-effectiveness and cost-utility of two web-based computer-tailored smoking cessation interventions	Web-based computer-tailored smoking cessation intervention
Swan *et al*., 2016	Discussion paper	To develop a patient-centered model for the experience of radiologic care	Patient experiences
Swan *et al*., 2003	Empirical analysis	To use WTT method to evaluate preferences of patients for magnetic resonance angiography and conventional x-ray angiography	Waiting trade-off (WTO)

Swan *et al*., 2010	Empirical analysis	To develop a temporary utility index and to test it within the context of diagnostic testing	Temporary Utility Index (TUI)
Thorn *et al*., (2014)	Discussion paper	To provide an overview of advances in methodology that guides health economists working on applied economic evaluations	Trial-based economic evaluations
Torgerson & Raftery, 1999	Discussion paper	To provide an overview of outcomes for different types of economic evaluation	Outcomes in economic evaluations
2nd wave of search (n=31)			
Bajaj & Veenstra, 2013	Empirical analysis	To assess the benefits, risks, and personal utility of factor V Leiden mutation testing	Genomics
Bayoumi, 2004	Review	To review the theory and conduct of contingent valuation studies	Contingent Valuation
Benning *et al*., 2015	Empirical analysis	To investigating the relative importance of non-health outcomes in a health promotion context	Non-health outcomes in health promotion
Bobinac *et al*., 2011	Empirical analysis	To investigate the occurrence of the family and care-giving effect	Family and care-giving effect
Borghi & Jan, 2008	Empirical analysis	To explore the use of the contingent valuation (CV) method to value the broader benefits in health promotion programs	CV in health promotion
Brouwer et al., 2005	Empirical analysis	To test the hypothesis that informal caregivers derive utility not only from the outcome of informal care, but also from the process of providing informal care	Process utility from providing informal care
Buchanan & Wordsworth, 2015	Review	To evaluate the potential for the choice of economic evaluation approach to impact on the adoption decisions	Welfarism vs Extra-welfarism

Buchanan *et al*., 2013	Review	To summarize the methodological issues associated with conducting economic evaluations of genomic interventions	Economic evaluations of genomic interventions
Coast *et al*., 2015	Discussion paper	To discuss the development of capability measures in health economics	Capability measures
Davis *et al*., 2015	Empirical analysis	To determine the factors that predict change in well-being over time	ICECAP-O
Ding *et al*., 2011	Discussion paper	To examine current related work on imaging expenditures for incidental findings	Incidental findings
Eden *et al*., 2013	Empirical analysis	To explore if contingent valuation (CV) method can be used to value the benefits of genetic testing	Genetic testing
Edwards *et al*., 2015	Discussion paper	Design of Economic Evaluations of Mindfulness-Based Interventions	Mindfulness-Based Interventions
Grosse *et al*., 2010	Review	To address critical evidentiary, economic, and ethical issues that arise in the appraisal of screening tests	Screening tests
Grosse *et al*., 2009	Discussion paper	Personal utility and genomic information	Personal utility
Higgins *et al*., 2014	Review	To systematically review the existing literature on the value associated with convenience in health care delivery,	Process utility
Liljas, 2011	Discussion paper	To discuss survivor consumption and leisure forgone and whether or not future non-medical cost should be included	Survivor consumption and leisure forgone
Lippert-Rasmussen & Lauridsen, 2010	Discussion paper	Discussion about the distribution of indirect, non-health benefits in allocating healthcare resources	Indirect non-health effects

Lorgelly, 2015	Discussion paper	To provide an introduction to the capability approach and to review the measures that are available for use in an economic evaluation	Capability approach
Lorgelly *et al*., 2015	Empirical analysis	To operationalize the capability approach as an outcome measure in public health and the development of the OCAP-18	OCAP-18
Lu & Cohen, 2015	Discussion paper	To discuss the potential economic impact of genomic medicine and the challenges	Genomic medicine
Makai *et al*., 2014a	Empirical analysis	To validate the ICECAP-O capability wellbeing measure's German translation in older people with dementia living in a nursing home	Application of the ICECAP-O in residents with dementia
McNamee & Seymour, 2008	Empirical analysis	To explore the implications of incorporating process preferences within the QALY framework	Process preferences
Neumann *et al*., 2012	Empirical analysis	To assess how much people would pay for a laboratory test that predicted their future disease status	Value of predictive tests
Rogowski & Schleidgen, 2015	Discussion paper	To develop a prioritization score for genetic tests to facilitate equitable allocation based on need-based claims	Needs-based framework
Sach *et al*., 2007	Review	To develop a League Table of Contingent Valuation Results	Contingent Valuation
Sutton & Coast, 2014	Empirical analysis	To develop of a supportive care measure for economic evaluation of end-of-life care	End-of-life
Tinelli *et al*., 2013	Empirical analysis	To test the sensitivity of the EQ-5D and SF-6D within pharmacy	Benefits in Pharmacy
Towse & Garrison, 2013	Discussion paper	To identify implications for the economic incentives for evidence generation	Personalized medicine

van Mastrigt et al., 2015	Empirical analysis	To investigate expert views regarding the incorporation of NHOs into the economic evaluations of public health interventions	Non-health outcomes
Veenstra *et al*., 2010	Discussion paper	To present a risk-benefit framework for assessing the health-related utility of genomic tests	Genomic tests
Reference list search (n=10)			
Baker *et al*., 2010	HTA report	To identify characteristics of beneficiaries of health care over which relative weights should be derived and to estimate relative weights to be attached to health gains according to characteristics of recipients of these gains	Social value of a QALY
Beale *et al*., 2012	Empirical analysis	To provide a summary of economic appraisals for investing in environmental interventions to encourage physical activity	Environmental interventions
Bobinac *et al*., 2010	Empirical analysis	To establish the existence of the caregiving and family effect	Spillover effects
Brazier *et al*., 2009	Discussion paper	To review the role of patient preferences within the framework of cost-effectiveness analysis	Patient preferences
Brouwer *et al*., 1997	Response letter	To comment on the published guidelines of the Washington Panel on incorporation of indirect non-medical costs	Productivity costs
Jacobsson *et al*., 2005	Empirical analysis	To compare caring externalities and internal preferences of health states with different severity levels	Caring externalities
Kinghorn *et al*., 2015	Qualitative research	To identify important capabilities in patients with chronic pain	Capability approach in chronic pain

Labelle & Hurley, 1992	Discussion paper	To explore the implications of omitting external effects from cost-utility studies	Externalities
Singer *et al*., 1995	Discussion paper	To discuss the issue of double jeopardy in using QALYs	Double jeopardy
Williams, 1997	Discussion paper	To discuss the concept of fair innings	Fair innings

CUA=cost-utility analysis; CV=contingent valuation; NHO=Non-health outcomes; OCAP-18=18-item capability measure; QALY=Quality-adjusted life year; R-QALY=Relatives' quality-adjusted life year; SF-6D=Short Form six Dimensions; TUI=Temporary Utility Index; WTO=waiting trade-off; WTP=willingness to pay.

Appendix 3.2. Thematic framework of benefits beyond the health-related QALY

Theme	Construct	Attribute/domain/item	Source (First author (Year))
Wellbeing	Subjective wellbeing	Activity, Leisure	Al-Janabi (2011), Connell (2014), Cookson (2005b), Davis (2013)
		Body satisfaction	Alayli-Goebbels (2013), Goebbels (2012)
		Burden	Coast (2008a), Coast (2014)
		Creativity and play	Benning (2015), Davis (2013), Makai (2014b),
		Cleanliness (& comfort)	Goebbels (2012)
		Disappointment	Salkeld (2004)
		Emotions	van Mastrigt (2015)
		Endurance	Alayli-Goebbels (2013), Goebbels (2012)
		Environment	Makai (2014b), Ong (2009), van Mastrigt (2015)
		Family, family life	Cookson (2005b), van Mastrigt (2015)
		Financial risk protection/ financial wellbeing, wealth	Davis (2013), Hoefmann (2013), Laxminarayana (2014), Makai (2014b)
		Happiness, optimism	Benning (2015), Higgings (2014), Hoefmann (2013), Thorn (2014)
		Motivation	Goebbels (2012)
		Religion, spirituality	Coast (2008a), Coast (2014), Cookson

			(2005b), Ong (2009), Round (2012)
		Relationships, social interaction/maintaining friendships, companionship; social participation; social live, one's place in society	Chisholm (1997), Connell (2014), Coast (2008a), Davis (2013), Goebbels (2012), Hausman (2012), Lorgelly (2010), Makai (2014b), van Mastrigt (2015)
		Regret	Salkeld (1998), Salkeld (2004)
		Relaxation, stress reduction	Alayli-Goebbels (2013), Benning (2015), Goebbels (2012)
		Satisfaction	Schulz (2014), Stanczyk (2014)
		Security and safety	Ali (2012), Makai (2014b)
		Sexual performance	Benning (2015)
		Subjective wellbeing, general wellbeing, quality of life, view of life	Benning (2015), Davis (2015), Gandjour (2001), Hoefman (2013), Makai (2014a), Thorn (2014)
		Support	Benning (2015)
	Psychological wellbeing	Autonomy	Mooney (1998), Ryan (1995)
		Cope	Payne (2013)
		Control	Benning (2015), van Mastrigt (2015)
		Freedom	Makai (2014b)
		Goals	Goebbels (2012)
		Hope and hopelessness	Connell (2014), Salkeld (1998)
		Independence	Chisholm (1997), Makai (2014b), Makai (2013a)
		Purpose in life and achievement	Makai (2014b)

		Self-confidence	Benning (2015), Borghi (2008), van Mastrigt (2015)
		Self-esteem	Borghi (2008), Cookson (2005b), Hausman (2012)
		Self-efficacy	Borghi (2008)
		Self-respect	Coast (2008d)
	Capabilities	Access	Simon (2013)
		Accommodation cleanliness & comfort	Lorgelly *et al*. (2015), Netten (2012), Simon (2013)
		Achievement, accomplishment	Al-Janabi (2012), Al-Janabi (2013b), Cookson (2005b)
		Assaults	Lorgelly *et al*. . (2015), Simon (2013)
		Attachment, love & support, love & social exclusion, love & affection	Al-Janabi (2012), Al-Janabi (2013b), Coast (2008a), Kinghorn (2015), Lorgelly *et al*. (2015), Makai (2014b), Simon (2013), Sutton (2014)
		Autonomy (social autonomy) and choice, independence & autonomy	Al-Janabi (2012), Al-Janabi (2013b), Coast (2015), Kinghorn (2015), Netten (2012), Sutton (2014)
		Capabilities	Buchanan (2015), Crosignani (2015), Lorgelly *et al*. (2015), Thorn (2014)
		Choice	Sutton (2014)
		Control, control over daily life	Al-Janabi (2011), Coast (2008a), Netten (2012)
		Cleanliness	Netten (2012)
		Daily activities, activities or employment	Lorgelly *et al*. (2015), Simon (2013)

		Dignity	Coast (2015), Netten (2012), Sutton (2014)
		Discrimination	Lorgelly *et al*. (2015), Simon (2013)
		Employment discrimination	Lorgelly *et al*. (2015)
		Enjoyment	Al-Janabi (2012), Al-Janabi (2013b), Coast (2008a), Kinghorn (2015)
		Food & Drink	Netten (2012)
		Freedom for expression	Lorgelly *et al*. (2015), Simon (2013)
		Imagination and creativity	Lorgelly *et al.* (2015), Simon (2013)
		Influence local decisions	Lorgelly *et al*. (2015), Simon (2013)
		Love	Coast (2015)
		Life expectancy	Lorgelly *et al*. (2015), Simon (2013)
		Nature	Lorgelly *et al*. (2015), Simon (2013)
		Neighbourhood safety	Lorgelly *et al.* (2015), Simon (2013)
		Occupation	Netten (2012)
		Physical and mental health	Kinghorn (2015), Sutton (2014)
		Planning one's life	Lorgelly *et al*. (2015), Simon (2013)
		Preparation (for death)	Coast (2008a), Coast (2014), Coast (2015), Round (2012), Sutton (2014)
		Property ownership	Lorgelly *et al.* (2015), Simon (2013)
		Recreation	Lorgelly *et al*. (2015), Simon (2013)
		Respect and appreciation; respect & identity	Kinghorn (2015), Lorgelly *et al.* (2015), Simon (2013)
		Role, societal and family roles	Coast (2008a), Kinghorn (2015)

		Security (feeling safe), Safety	Coast (2008a), Coast (2008d), Kinghorn (2015), Netten (2012)
		(Social) support, social networks, social participation and involvement	Al-Janabi (2011), Coast (2015), Lorgelly *et al.* (2015), Netten (2012), Simon (2013), Sutton (2014)
		Sleep	Lorgelly *et al.* (2015), Simon (2013)
		Stability	Al-Janabi (2012), Al-Janabi (2013b)
		Suffering (physical, emotional)	Coast (2015)
	Empowerment	Addictions - overcome	Goebbels (2012)
		Awareness	Goebbels (2012)
		Behaviour change	Lorgelly (2015), van Mastrigt (2015)
		Behavioural control	McAllister (2012)
		Control	Goebbels (2012)
		Cognitive control	McAllister (2012)
		Decision-making	Goebbels (2012), Payne (2013), Ryan (2014)
		Decisional control	McAllister (2012)
		Effort	Goebbels (2012)
		Emotional regulation	McAllister (2012)
		Empowerment	Alayli-Goebbels (2013), Coulter (2013), Lorgelly (2010), Lorgelly (2015), McAllister (2012), Payne (2013), Schulz (2014)
		Health literacy	Benning (2015)
		Hope for the future	McAllister (2012)

		Information, knowledge, knowledge sharing	Benning (2015), Borghi (2008), Buchanan (2015), Dowie (2001), Grosse (2008), Mooney (1998), Payne (2013), Ryan (1995), Ryan (1999), Salkeld (2004)
		Participation	Essink-Bot (2003)
		Self-management capacities	Benning (2015)
Process utility	Intervention characteristics	Accuracy	Payne (2013)
		Administration	Ali (2012)
		Frequency	Ali (2012), Brouwer (2008a)
		Complications, morbidity	Swan (2003), Swan (2010), Swan (2016)
		Invasiveness	Alayli-Goebbels (2013), Opmeer (2010)
		Tailored	Alayli-Goebbels (2013), Grosse (2008)
	Provider characteristics	Attitudes, Relationship to the healthcare provider	Alayli-Goebbels (2013), Coulter (2013), Ryan (1999), Payne (2013), Ryan (2014)
		Communication, competence	Annemans (2013), Grosse (2008), Swan (2016)
		Recommendation	Annemans (2013)
		Non-medical reasons for visiting the doctor	Ryan (1995)
	Structural	Access/ Location/ Distance	Ali (2012), Benning (2015), Birch (2003), Dowie (2001), Grosse (2008),

	characteristics		Payne (2013), Petrou (2000), Ryan (1999), Ryan (2014), Shackley (2001), Swan (2010), Swan (2016)
		Accommodation	Kotzian (2009), Birch & Donaldson (2003)
		Coordination with services	Swan (2016)
		Cost, affordability	Ali (2012), Benning (2015), Swan (2016), Swan (2010)
		Cleanliness	Ali (2012), Swan (2016)
		Follow-up care, continuity of care	Ryan (1999), Shackley (2001)
		Services	Ali (2012), Grosse (2008)
		Treatment environment	Shackley (2001)
	Process characteristics	Anxiety, fear before or during process	Borghi (2008), Ding (2011), Salkeld (1998), Swan (2016), Swan (2010)
		Adherence	Towse (2013)
		Burden	Brouwer (2005)
		Comfort, discomfort	Al-janabi (2012), Birch (2003), Cookson (2005b), Kotzian (2009), Makai (2014b), Makai (2013a), Opmeer (2010)
		Convenience/ inconvenience	Ding (2011), Higgings (2014)
		Embarrassment	Borghi (2008), Opmeer (2010), Swan (2016)
		Good death, end-of-life aspects	van Mastrigt (2015)
		Care	Brouwer (2005), Davidson (2010), Mooney (1998)

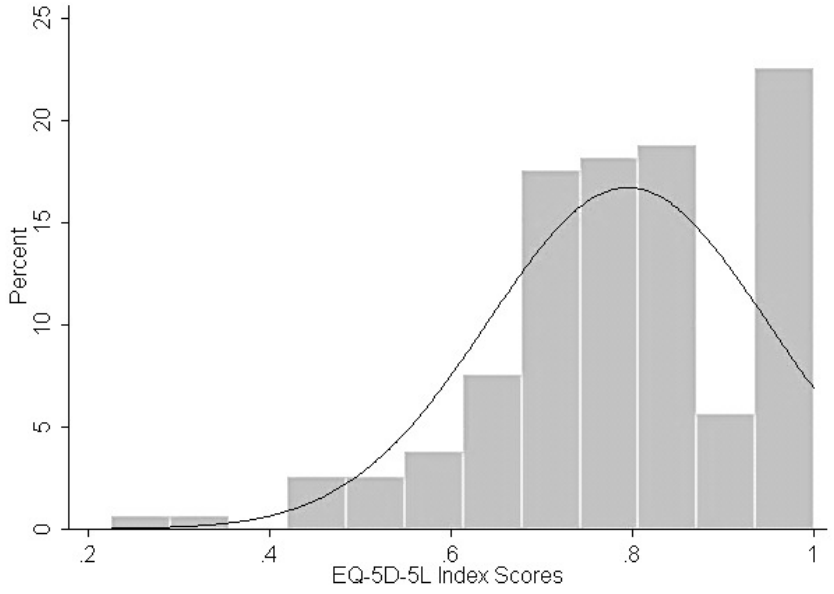
		Choice	Ali (2012), Brouwer (2005), Kotzian (2009), Ryan (1995)
		Control	Petrou (2000)
		Cosmetic	Opmeer (2010)
		Dignity	Ali (2012), Al-Janabi (2012), Coast (2008a), Coast (2008d), Coast (2014), Cookson (2005b), Donaldson (1997), Makai (2014b), Makai (2013a), Mooney (1998), Sutton (2014), Netten (2012), Ryan (1995), Ryan (2014), Ryan (1999), Schlander (2014), Torgerson (1999)
		Duration, time	Alayli-Goebbels (2013), Dowie (2001), Grosse (2008), Opmeer (2010), Payne (2013), Ryan (2014), Ryan (1999), Shackley (2001), Swan (2010), Swan (2016)
		Information, health-related utility	Bajaj (2013), Birch (2003), Buchanan (2013), Donaldson (1997), Eden (2013), Grosse 2009, Grosse (2010), Howard (2008), Lu (2015), Neumann (2012), Ryan (1995), Swan (2003), Towse (2013), Veesntra (2010)
		Physical and mental health after treatment	Swan (2016)
		Pleasure/ pleasantness	Brennan (2013), Brouwer *et al.* (2008), Donaldson (1997), Makai (2014b)

		Process of treatment (process utility; process attributes; process characteristics; process preferences; quality of care)/ non-outcome attributes	Ali (2012), Annemans (2013), Birch (2003), Birch & Donaldson, Brouwer *et al*. (2008), Brouwer (2006), Brouwer (2005), Coast (2008a), Crosignani (2015), Dirksen (2014), Donaldson (1997), Dowie (2001), Essink-Bot (2003), Gandjour (2014), Howard (2008), Higgings (2014), Kotzian (2009), McNamee (2008), Mooney (1998), Opmeer (2010), Olsen (2001), Payne (2013), Round (2014), Petrou (2000), Ryan (1995), Ryan (2014), Salkeld (1998), Salkeld (2004), Swan (2003), Sach (2007)
		Privacy	Ryan (2014)
		Satisfaction/ Experience	Brazier (2009), Buchanan (2015), Dirksen (2014), Ryan (2014), Swan (2016), Tinelli (2013)
		Stress	Ding (2011)
		Reassurance	Cookson (2005b), Donaldson (1997), Grosse (2008), Howard (2008), Ryan (1999), Ryan (2014), Ryan (1995), Salkeld (1998), Swan (2003)
		Uncertainty	Opmeer (2010)
Benefits beyond the affected individual	Spillover effects	Family members/ spillover effects/ caregiving effect/ family effect/ process utility from providing informal care, population level benefits	Ali (2012), Al-Janabi (2011), Bobinac (2010) Bobinac (2011), Brouwer (2005), Brouwer (2006), Brouwer (2008a), Coast (2014), Coulter (2013), Davidson (2010), Eden

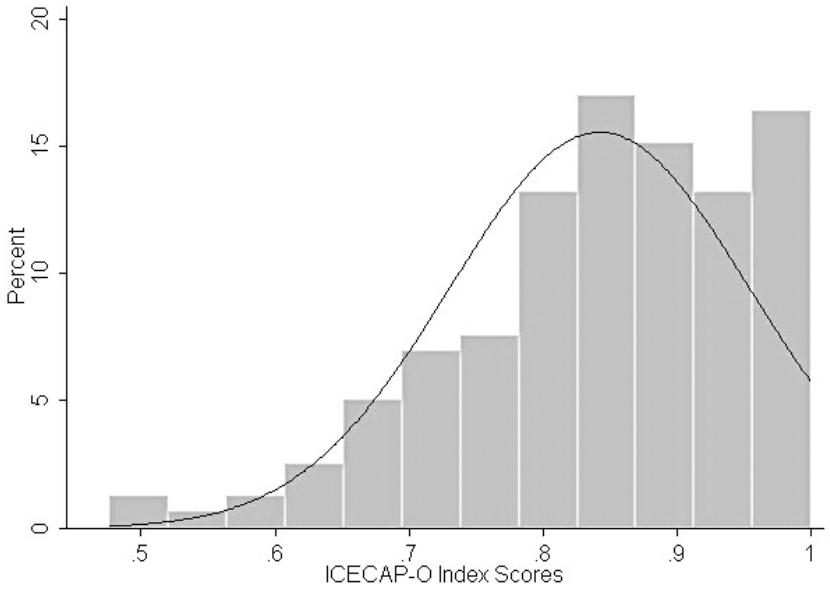
			(2013), Grosse (2008), Hoefmann (2013), Lu (2015), Rogowski (2014), Mooney (2000)
	Externalities	Externalities (existence or non-use value), altruism, passive use value	Borghi (2008), Donaldson (1997), Jacobsson (2005), Labelle (1992), Olsen (2001), Ryan (1995), Sach (2007)
	Option value		Borghi (2008), Donaldson (1997), Eden (2013), Olsen (2001), Ryan (1995), Sach (2007)
	Distributional benefits	Equity, fairness and need	Baker (2010), Hansen (2012), Herlitz (2016), Lippert-Rasmussen (2010), Rogowski (2015), Schlander (2014), Singer (1995), Williams (1997)
Benefits outside the health care sector		Other sectors, Survivor consumption and survivor forgone leisure	Ali (2012), Bayoumi (2004), Benning (2015), Beale (2012), Coast (2008a), Coulter (2013), Edwards (2013), Edwards (2015), Gandjour (2001), Hausman (2012), Laxminarayana (2014), Lorgelly (2010), Lorgelly (2015), Lorgelly *et al*. (2015), Lu (2015), Liljas (2011), Nyman (2009), van Mastrigt (2015)

Appendices to Chapter 4

Appendix 4.1. Histogram of the EQ-5D-5L index scores (n=160)



Appendix 4.2. Histogram of the ICECAP-O index scores (n=158)



Appendix 4.3. Correlation analyses between environmental features

	A: Residential density	B: Land use (diversity)	C: Land use (access)	D: Street connectivity	F: Aesthetics	G: Traffic Hazards[a]	H: Crime[a]	J: Lack of cul-de-sacs	K: Hilliness [a]	L: Physical Barriers[a]	SC-5PT	SSWS
A: Residential density	1											
B: Land use (diversity)	0.666	1										
C: Land use (access)	0.449	0.589	1									
D: Street connectivity	0.266	0.298	0.354	1								
F: Aesthetics	0.284	0.254	0.358	0.286	1							
G: Traffic Hazards [a]	-0.108	0.088	0.031	0.056	0.102	1						
H: Crime [a]	-0.110	-0.046	-0.091	-0.131	-0.106	0.224	1					
J: Lack of cul-de-sacs	0.062	0.043	0.045	0.175	-0.003	-0.026	-0.009	1				
K: Hilliness [a]	0.024	-0.098	-0.124	-0.018	-0.085	0.166*	0.248	0.077	1			
L: Physical Barriers [a]	-0.043	-0.119	-0.025	-0.144	-0.099	0.098	0.287	-0.081	0.408	1		
SC-5PT	0.099	0.034	0.136	0.215	0.163	-0.065	-0.187	-0.017	0.018	0.089	1	
SSWS	0.534	0658	0.612	0.317	0.169	0.128	-0.019	0.143	0.069	-0.003	0.097	1

Letters refer to the NEWS-A subscales; SC-5PT=Sampson's 5-item measure of collective efficacy; SSWS=Street Smart Walk Score.
[a] NEWS-A subscales: Reverse codes for these scales; higher score = less pedestrian friendly.

Appendix 4.4. Correlation analyses between environmental features and the EQ-5D-5L and ICECAP-O

	EQ-5D-5L [a]	ICECAP-O [b]
A: Residential density	0.131	0.011
B: Land use (diversity)	0.214	0.041
C: Land use (access)	0.158	-0.006
D: Street connectivity	0.079	-0.017
F: Aesthetics	-0.062	-0.086
G: Traffic Hazards [c]	-0.152	-0.007
H: Crime [c]	-0.172	-0.027
J: Lack of cul-de-sacs	-0.002	0.000
K: Hilliness [c]	-0.183	0.019
L: Physical Barriers [c]	-0.146	0.058
SC-5PT [d]	0.071	0.293
Street Smart Walk Score [e]	0.104	0.022

Letters refer to the NEWS-A subscales.
[a] EQ-5D-5L index score: the higher the score, the better an individuals‘ HRQoL.
[b]ICECAP-O index score: the higher the score, the better an individual’s' capability wellbeing.
[c] Reverse codes for these scales; higher score = less pedestrian friendly.
[d] SC-5PT: the higher the score, the better the perceived collective efficacy.
[e] Street Smart Walk Score: the higher the score, the greater the walkability.

Appendix 4.5. Missing value analysis of the variables used in the analyses

Variable	# missing	% missing
Outcome variables		
EQ-5D-5L index score	0	0
ICECAP-A index score	2	1.3
Covariates		
Sex	0	0
Age	0	0
Live with someone	0	0
Functional Comorbidity Index	2	1.3
Loneliness	0	0
Gait Speed	0	0
Ambulatory	0	0
Exploratory variables		
NEWS-A (A): Residential density	3	1.9
NEWS-A (B): Land-use mix (diversity)	2	1.3
NEWS-A (C): Land-use (access)	2	1.3
NEWS-A (D): Street connectivity	4	2.5
NEWS-A (F): Aesthetics	0	0
NEWS-A (G): Traffic hazards	5	3.1
NEWS-A (H): Crime	8	5.0
NEWS-A (J): Lack of cul-de-sacs	1	0.6
NEWS-A (K): Hilliness	1	0.6
NEWS-A (L): Physical barriers	1	0.6
SC-5PT	4	2.5
SSWS	0	0

NEWS-A=Neighbourhood Environment Walkability Scale – Abbreviated, SC-5PT= Sampson's 5-item measure of collective efficacy, SSWS=Street Smart Walk Score.

Appendix 4.6. Backward stepwise OLS regression analysis (complete case analysis) [a]

	EQ-5D-5L (n=137)			ICECAP-O (n=135)		
	β (SE)	95% CI	*p*	*β* (SE)	95% CI	*p*
Constant	0.508 (0.201)	(0.111, 0.906)	*0.013*	1.036 (0.140)	(0.759, 1.314)	*<0.001*
Covariates						
Sex (1=female)	-0.011 (0.030)	(-0.056, 0.034)	*0.638*	0.024 (0.016)	(-0.007, 0.056)	*0.133*
Age	0.002 (0.002)	(-0.001, 0.006)	*0.161*	-0.002 (0.001)	(-0.004, 0.001)	*0.216*
Living with someone (1=yes)	0.009 (0.026)	(-0.043, 0.060)	*0.742*	0.019 (0.018)	(-0.017, 0056)	*0.292*
Functional Comorbidity Index	-0.026 (0.005)	(-0.036, -0.015)	*<0.001*	-0.003 (0.004)	(-0.010, 0.004)	*0.386*
Loneliness	-0.041 (0.024)	(-0.087, 0.006)	*0.086*	-0.139 (0.017)	(-0.174, -0.106)	*<0.001*
Gait Speed	-0.04 (0.029)	(-0.043, 0.035)	*0.836*	0.014 (0.014)	(-0.013, 0.041)	*0.322*
Ambulatory	0.043 (0.007)	(0.030, 0.058)	*<0.001*	0.016 (0.005)	(0.005, 0.026)	*0.004*
Perceived built environment (NEWS-A)						
A: Residential density	-	-	-	-	-	-
B: Land-use mix (diversity)	-	-	-	-	-	-
C: Land-use (access)	0.028 (0.014)	(0.000, 0.056)	*0.049*	-	-	-
D: Street connectivity	-	-	-	-	-	-
F: Aesthetics	-0.038 (0.016)	(-0.069, -0.007)	*0.017*	-0.032 (0.011)	(-0.054, -0.011)	*0.003*
G: Traffic hazards [b]	-	-	-	-	-	-
H: Crime [b]	-	-	-	-	-	-
J: Lack of cul-de-sacs	-0.024 (0.010)	(-0.043, -0.005)	*0.016*	-	-	-
K: Hilliness [b]	-	-	-	-	-	-

L: Physical barriers [b]	-	-	-	-	-	-
Social cohesion						
SC-5PT	-	-	-	0.033 (0.011)	(0.011, 0.053)	*0.003*
Objective built environment						
Street Smart Walk Score	-	-	-	-	-	-
Statistics						
BIC	-164.369			-256.707		
AIC	-196.489			-291.570		
LR chi^2 (*p*-value)	109.24 (<0.001)			157.79 (<0.001)		

β=Beta coefficient, SE=standard error, CI=confidence interval, NEWS-A=Neighbourhood Environment Walkability Scale – Abbreviated, SC-5PT=Sampson's 5-item measure of collective efficacy, SSWS=Street Smart Walk Score, AIC=Akaike information criterion, BIC=Bayesian information criterion, LR chi^2= likelihood ratio chi-square test.

[a] Backward stepwise elimination was used to remove variables relating to the built and social environments; $p > 0.1$ was the criterion for removal. Covariates were not eligible for removal.

[b] Reverse coding, where higher scores indicates 'less pedestrian friendly'.

Appendix 4.7. Backward stepwise Tobit regression analysis (complete case analysis) [a]

	EQ-5D-5L (n=137)			ICECAP-O (n=135)		
	β (SE)	95% CI	*p*	*β* (SE)	95% CI	*p*
Constant	0.459 (0.225)	(0.014, 0.905)	*0.043*	1.098 (0.133)	(0.836, 1.361)	*<0.001*
Covariates						
Sex (1=female)	-0.010 (0.028)	(-0.066, 0.046)	*0.720*	0.021 (0.017)	(-0.012, 0.054)	*0.210*
Age	0.003 (0.002)	(-0.001, 0.007)	*0.195*	-0.002 (0.001)	(-0.004, 0.001)	*0.152*
Living with someone (1=yes)	0.016 (0.032)	(-0.047, 0.080)	*0.614*	0.020 (0.019)	(-0.018, 0057)	*0.296*
Functional Comorbidity Index	-0.030 (0.006)	(-0.043, -0.017)	*<0.001*	-0.005 (0.004)	(-0.012, 0.003)	*0.225*
Loneliness	-0.048 (0.029)	(-0.105, 0.008)	*0.094*	-0.139 (0.017)	(-0.174, -0.105)	*<0.001*
Gait Speed	-0.021 (0.056)	(-0.133, 0.090)	*0.705*	0.039 (0.034)	(-0.027, 0.106)	*0.246*
Ambulatory	0.050 (0.009)	(0.032, 0.068)	*<0.001*	0.014 (0.006)	(0.003, 0.025)	*0.011*
Perceived built environment (NEWS-A)						
A: Residential density	-	-	-	-	-	-
B: Land-use mix (diversity)	-	-	-	-	-	-
C: Land-use (access)	0.038 (0.017)	(0.005, 0.071)	*0.026*	-	-	-
D: Street connectivity	-	-	-	-	-	-
F: Aesthetics	-0.043 (0.019)	(-0.081, -0.006)	*0.025*	-0.038 (0.011)	(-0.060, -0.017)	*0.001*
G: Traffic hazards [b]	-	-	-	-	-	-
H: Crime [b]	-	-	-	-	-	-
J: Lack of cul-de-sacs	-0.027 (0.012)	(-0.050, -0.003)	*0.029*	-0.015 (0.007)	(-0.029, -0.001)	*0.031*

K: Hilliness [b]	-	-	-	-	-	-
L: Physical barriers [b]	-	-	-	-	-	-
Social cohesion						
SC-5PT	-	-	-	0.029 (0.011)	(0.008, 0.051)	*0.008*
Objective built environment						
Street Smart Walk Score	-	-	-	-	-	-
Statistics						
BIC	-9.498			-200.191		
AIC	-44.538			-235.054		
LR chi^2 (*p*-value)	88.55 (<0.001)			97.63 (<0.001)		

β=Beta coefficient, SE=standard error, CI=confidence interval, NEWS-A=Neighbourhood Environment Walkability Scale – Abbreviated, SC-5PT=Sampson's 5-item measure of collective efficacy, SSWS=Street Smart Walk Score, AIC=Akaike information criterion, BIC=Bayesian information criterion, LR chi^2= likelihood ratio chi-square test.

[a] Backward stepwise elimination was used to remove variables relating to the built and social environments; p > 0.1 was the criterion for removal. Covariates were not eligible for removal.

[b] Reverse coding, where higher scores indicates 'less pedestrian friendly'.

Appendices to Chapter 5

Appendix 5.1. Overview of dimensions and items of the preference-based HRQoL instruments included in the analysis [a]

Instrument	Dimension	Item
15D	Mobility	-
	Vision	-
	Hearing	-
	Breathing	-
	Sleeping	-
	Eating	-
	Speech	-
	Elimination	-
	Usual activities	-
	Mental function	-
	Discomfort and symptoms	-
	Depression	-
	Distress	-
	Vitality	-
	Sexual activity	-
AQoL-8D	Independent living	3. Getting around
		15. Mobility
		19. Self-care
		30. Household tasks
	Pain	6. Frequency of pain
		22. Degree of pain
		24. Pain interference
	Senses	11. Communication
		28. Vision
		32. Hearing
	Mental health	5. Sadness

		8. Calm
		12. Sleep
		14. Anger
		16. Self-harm
		18. Worry
		33. Depression
		35. Despair
	Happiness	17. Enthusiasm
		20. Happiness
		25. Pleasure
		27. Contentment
	Coping	1. Energy level
		21. Coping
		29. Control
	Relationships	2. Social exclusion
		4. Community role
		9. Family role
		10. Close relationships
		23. Enjoy close relationships
		31. Social isolation
		34. Intimacy
	Self worth	7. Confidence
		13. Feeling worthless
		26. Feeling burden
EQ-5D-5L	Mobility	-
	Self-care	-
	Usual activities	-
	Pain / Discomfort	-
	Anxiety / Depression	-
HUI-3[b]	Vision	-
	Hearing	-
	Speech	-
	Ambulation	-
	Dexterity	-
	Emotion	-
	Cognition	-

	Pain	-
SF-6D (SF-36v2)	Physical functioning	3a.Vigorous activities
		3b. Moderate activities
		3j. Bathing or dressing oneself
	Role limitation	4c. Limited in kind of work or other activities
		5b. Accomplished less than you would like
	Social functioning	10. Frequency health problems interfered with social activities
	Pain	7. Intensity of bodily pain
		8. Extent pain interfered with normal work
	Mental health	9b. Been very nervous
		9f. Felt downhearted and depressed
	Vitality	9e. Have a lot of energy

[a] The 15D and EQ-5D-5L contain only one item per dimension. Numbers in the item column refer to the number of the item in the respective questionnaire.

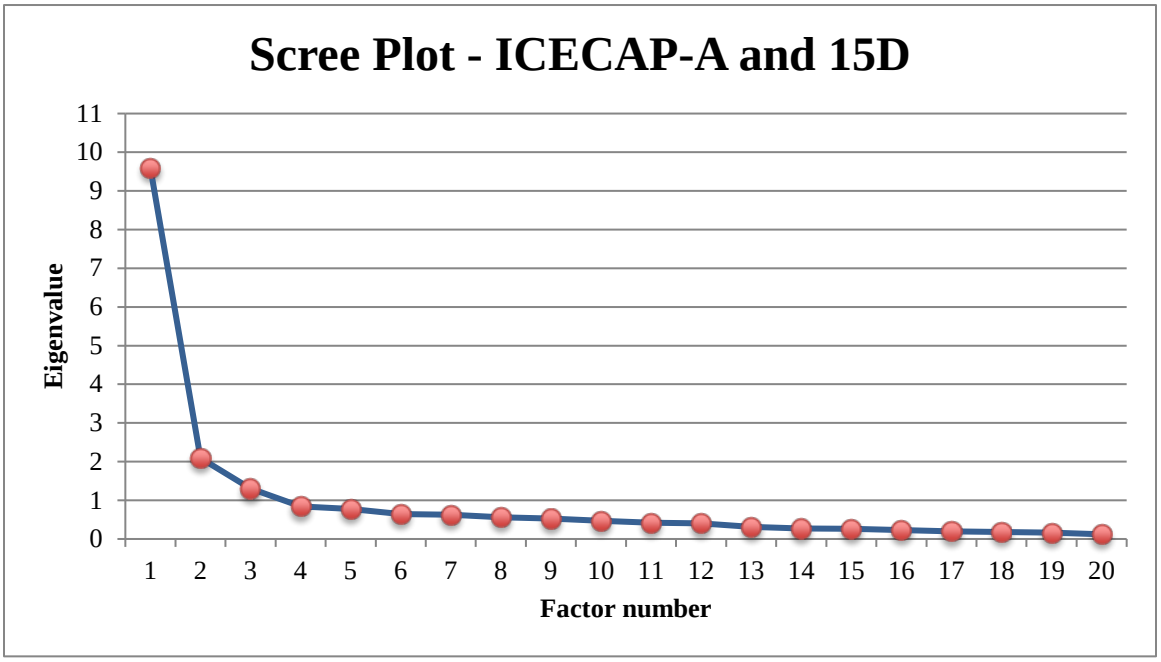
[b] Typically, HUI-3 index scores are derived from 12 of the 15 items comprising the HUI questionnaire. In the MIC study, these 12 items were collapsed to reflect the eight dimensions of the HUI-3 and only those eight questions were part of the survey.

Appendix 5.2. Descriptive indices of model fit comparing the ICECAP-A with the 15D

	Eigenvalues > 1	CFI	TLI	RMSEA	Clean structure	Chosen factor solution
	3					
1 Factor		0.886	0.873	0.124	n/a	
2 Factors		0.959	0.948	0.079	n/a	
3 Factors		0.978	0.969	0.061	No	
4 Factors		0.991	0.985	0.042	Yes	4 Factors
5 Factors		0.995	0.991	0.034	Yes	

RMSEA = root mean square error of approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index.

Appendix 5.3. Scree plot comparing the ICECAP-A with the 15D

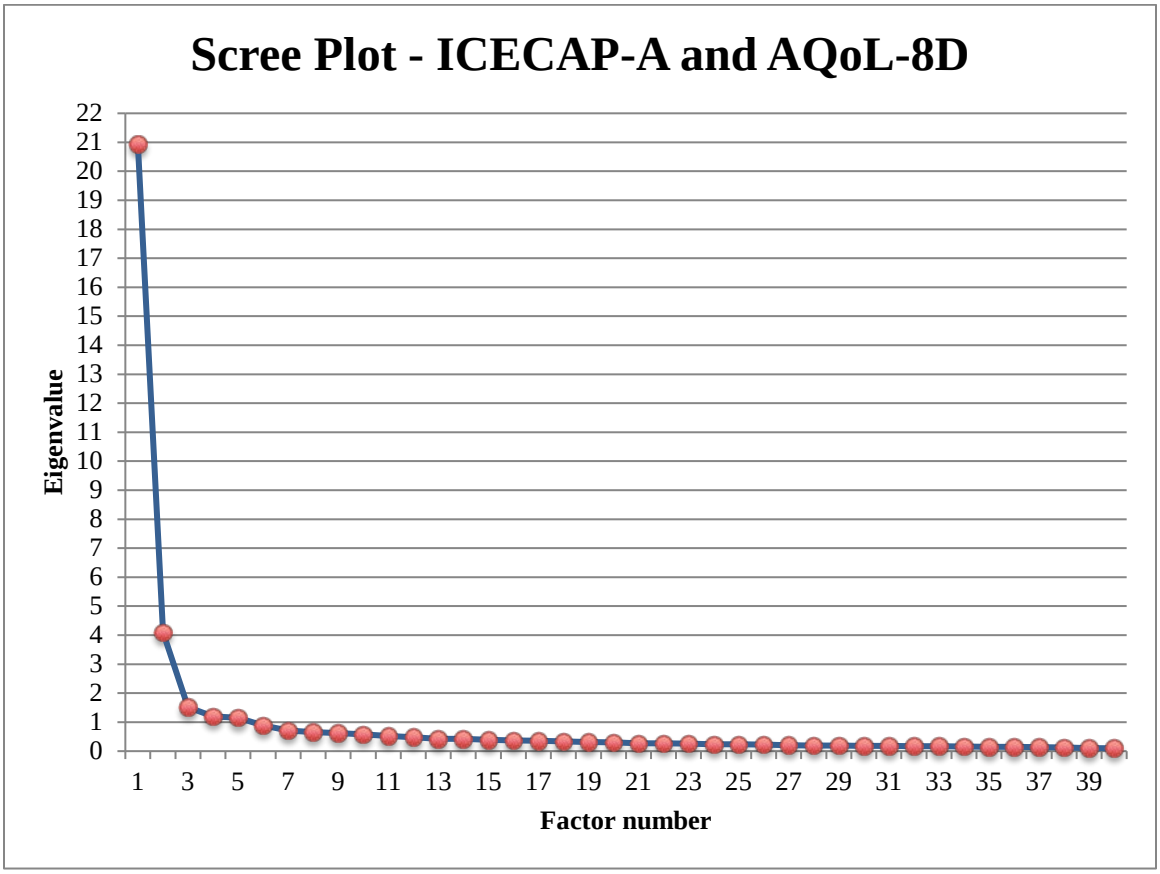


Appendix 5.4. Descriptive indices of model fit comparing the ICECAP-A with the AQoL-8D

	Eigenvalues > 1	CFI	TLI	RMSEA	Clean structure	Chosen factor solution
	5					
1 Factor		0.847	0.839	0.132	n/a	
2 Factors		0.922	0.913	0.097	n/a	
3 Factors		0.948	0.938	0.082	n/a	
4 Factors		0.963	0.954	0.070	n/a	
5 Factors		0.974	0.965	0.061	Yes	5 Factors
6 Factors		0.980	0.972	0.055	No	

RMSEA = root mean square error of approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index.

Appendix 5.5. Scree plot comparing the ICECAP-A with the AQoL-8D

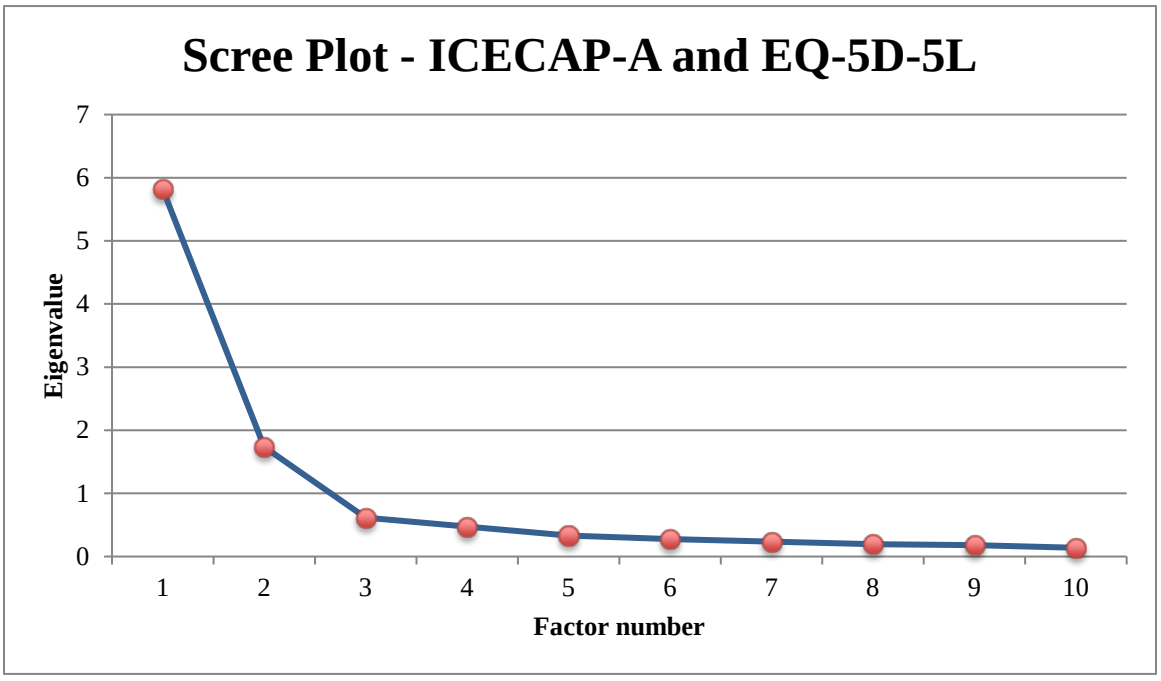


Appendix 5.6. Descriptive indices of model fit comparing the ICECAP-A with the EQ-5D-5L

	Eigenvalues > 1	CFI	TLI	RMSEA	Clean structure	Chosen factor solution
	2					
1 Factor		0.900	0.871	0.210	n/a	
2 Factors		0.987	0.977	0.084	No	
3 Factors		0.993	0.983	0.074	Yes	3 Factors
4 Factors		0.999	0.997	0.032	No	

RMSEA = root mean square error of approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index.

Appendix 5.7. Scree plot comparing the ICECAP-A with the EQ-5D-5L

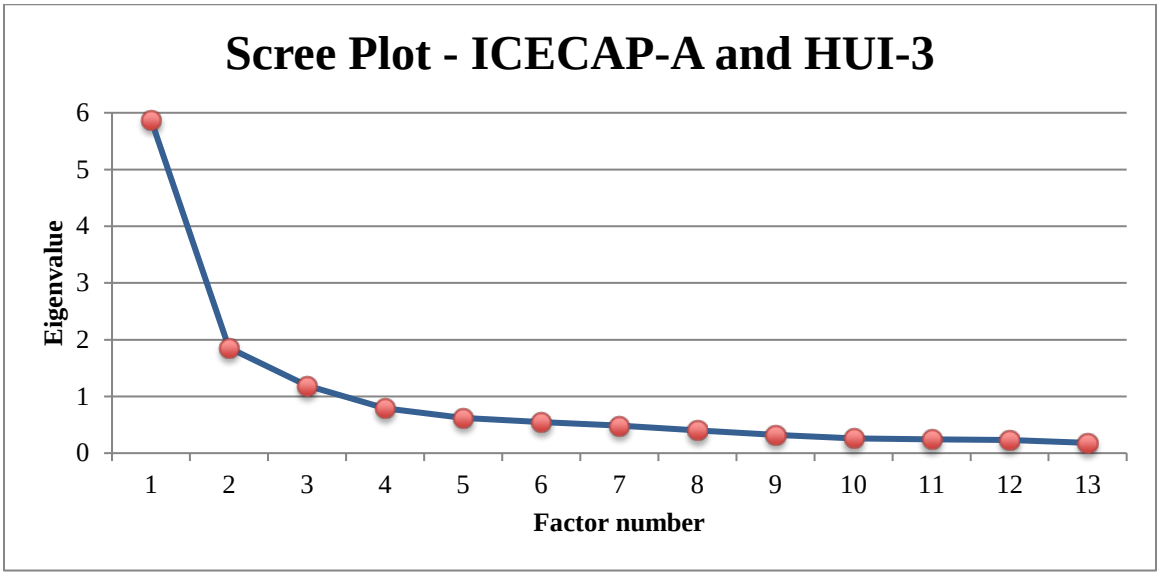


Appendix 5.8. Descriptive indices of model fit comparing the ICECAP-A with the HUI-3

	Eigenvalues > 1	CFI	TLI	RMSEA	Clean structure	Chosen factor solution
	3					
1 Factor		0.897	0.876	0.139	n/a	
2 Factors		0.974	0.961	0.077	n/a	
3 Factors		0.990	0.981	0.054	Yes	3 Factors
4 Factors		0.996	0.990	0.038	No	

RMSEA = root mean square error of approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index.

Appendix 5.9. Scree plot comparing the ICECAP-A with the HUI-3

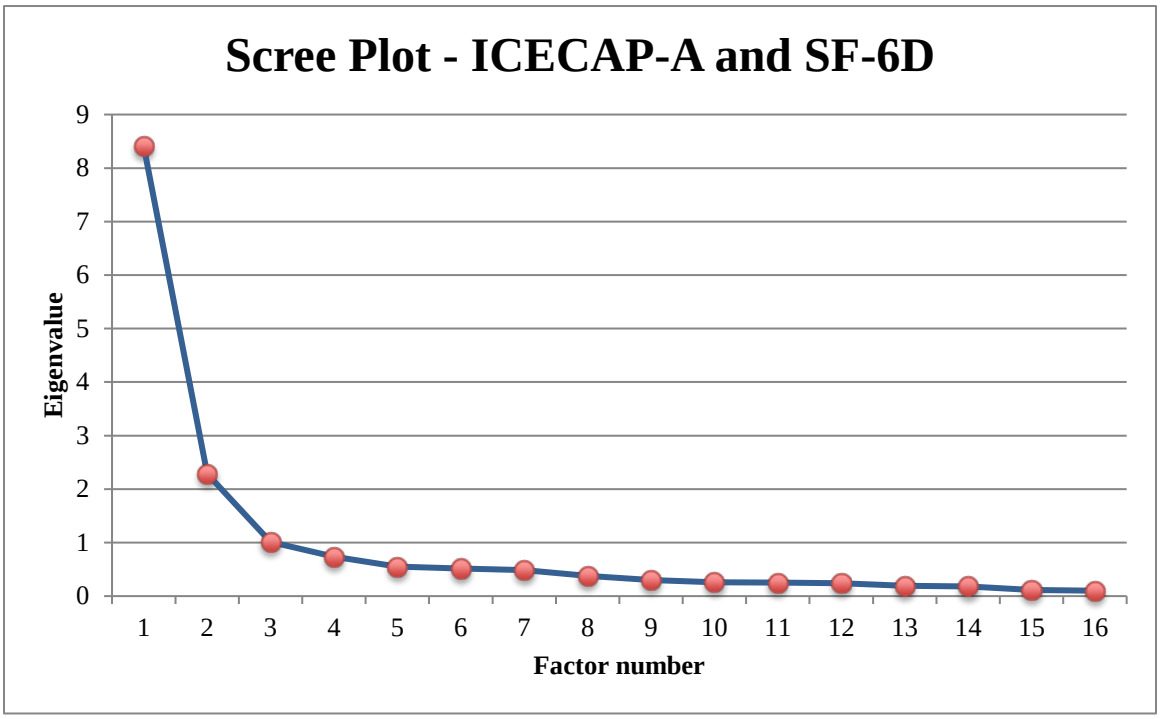


Appendix 5.10. Descriptive indices of model fit comparing the ICECAP-A with the SF-6D

	Eigenvalues >1	CFI	TLI	RMSEA	Clean factor structure	Chosen factor solution
	3					
1 Factor		0.849	0.826	0.186	n/a	
2 Factors		0.937	0.916	0.129	n/a	
3 Factors		0.969	0.950	0.100	No	
4 Factors		0.985	0.971	0.075	Yes	4 Factors
5 Factors		0.991	0.977	0.067	No	

RMSEA = root mean square error of approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index.

Appendix 5.11. Scree plot comparing the ICECAP-A with the SF-6D



Appendix 5.12. EFA comparing the ICECAP-A with the 15D (3-factor model) [a]

	Rotated item loadings		
	Factor 1	Factor 2	Factor 3
ICECAP-A			
Stability	**0.817**	-0.002	0.116
Attachment	**0.800**	-0.134	0.028
Autonomy	**0.454**	**0.362**	-0.075
Achievement	**0.784**	0.184	-0.086
Enjoyment	**0.895**	0.006	-0.014
15D			
Mobility	-0.009	**0.919**	-0.201
Vision	0.013	**0.435**	0.186
Hearing	-0.145	**0.417**	0.221
Breathing	0.103	**0.603**	0.020
Sleeping	**0.314**	**0.369**	0.190
Eating	-0.090	**0.698**	**0.310**
Speech	0.007	**0.448**	**0.431**
Elimination	0.017	**0.535**	0.155
Usual activities	0.213	**0.805**	-0.101
Mental function	0.268	**0.381**	**0.342**
Discomfort and symptoms	0.098	**0.730**	-0.010
Depression	**0.691**	0.002	**0.492**
Distress	**0.619**	-0.003	**0.529**
Vitality	**0.439**	**0.459**	0.221
Sexual activity	0.283	**0.492**	0.087

	Correlations among factors		
Factor 1	1.000		
Factor 2	0.494*	1.000	
Factor 3	0.240*	0.245*	1.000

RMSEA	0.061 [90% CI: 0.060 to 0.063]
CFI	0.978
TLI	0.969

RMSEA=root mean square error of approximation, CI=confidence interval, CFI=Comparative Fit Index, TLI=Tucker-Lewis Index.

[a] Loadings greater than 0.30 are presented in bold face. The highest loading for each item is underlined.

* Significant at 5% level.

Appendix 5.13. EFA comparing the ICECAP-A with the EQ-5D-5L (2-factor model)[a]

	Rotated item loadings	
	Factor 1	Factor 2
ICECAP-A		
Stability	**<u>0.867</u>**	-0.008
Attachment	**<u>0.837</u>**	-0.161
Autonomy	**0.400**	**<u>0.403</u>**
Achievement	**<u>0.740</u>**	0.195
Enjoyment	**<u>0.891</u>**	0.004
EQ-5D-5L		
Mobility	-0.121	**<u>0.991</u>**
Self-care	0.072	**<u>0.821</u>**
Usual activities	0.087	**<u>0.875</u>**
Pain/Discomfort	-0.004	**<u>0.801</u>**
Anxiety/Depression	**<u>0.739</u>**	0.060
	Correlation among factors	
Factor 1	1.000	
Factor 2	0.530*	1.000
RMSEA	0.084 [90% CI: 0.080 to 0.088]	
CFI	0.987	
TLI	0.977	

RMSEA=root mean square error of approximation, CI=confidence interval, CFI=Comparative Fit Index, TLI=Tucker-Lewis Index.

[a] Loadings greater than 0.30 are presented in bold face. The highest loading for each item is underlined.

* Significant at 5% level.

Appendix 5.14. EFA comparing the ICECAP-A with the HUI-3 (4-factor model) [a]

	Rotated item loadings			
	Factor 1	Factor 2	Factor 3	Factor 4
ICECAP-A				
Stability	**0.737**	0.204	-0.034	-0.011
Attachment	**0.867**	-0.133	0.027	-0.062
Autonomy	0.073	**0.635**	0.176	0.045
Achievement	**0.540**	**0.427**	0.017	0.045
Enjoyment	**0.854**	0.058	-0.003	0.046
HUI-3				
Vision	0.021	-0.171	**0.354**	0.234
Hearing	-0.022	-0.278	**0.673**	0.057
Speech	0.024	0.020	**0.948**	-0.236
Ambulation	-0.020	0.156	0.178	**0.684**
Dexterity	-0.001	0.082	**0.332**	**0.451**
Emotion	**0.849**	0.002	0.042	0.006
Cognition	**0.308**	0.062	**0.443**	0.028
Pain	0.221	-0.027	-0.033	**0.798**

	Correlations among factors			
Factor 1	1.000			
Factor 2	0.578*	1.000		
Factor 3	0.451*	0.410*	1.000	
Factor 4	0.268*	0.460*	0.468*	1.000
RMSEA	0.038 [90% CI: 0.035 to 0.042]			
CFI	0.996			
TLI	0990			

RMSEA=root mean square error of approximation, CI=confidence interval, CFI=Comparative Fit Index, TLI=Tucker-Lewis Index.

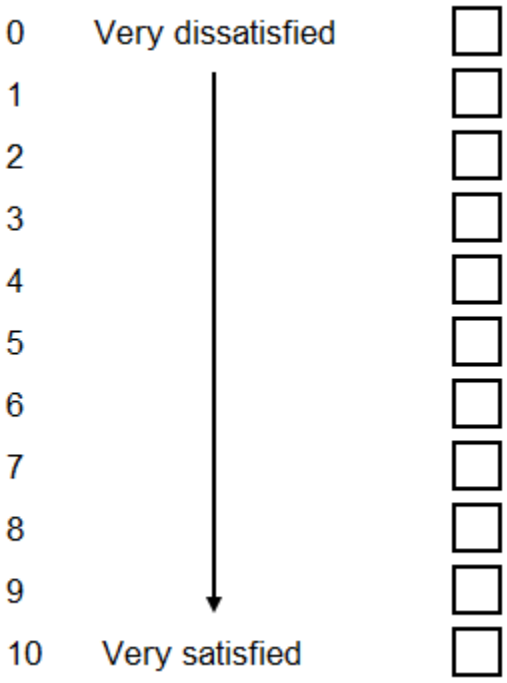
[a] Loadings greater than 0.30 are presented in bold face. The highest loading for each item is underlined.

* Significant at 5% level.

Appendices to Chapter 6

Appendix 6.1. Assessment of subjective wellbeing in the SCI Community Survey

Using a scale of 0 to 10, where 0 means "very dissatisfied" and 10 means "very satisfied", how do you feel about your **life as a whole right now**?



Appendix 6.2. Assessment of secondary health conditions in the SCI Community Survey

Section C: The following are questions that ask about health conditions and complications you may have.

For the following 11 health problems, please rate how much each one affected your activities and independence in the last 3 months. If you have not experienced the health problem in the last 3 months, or if it is an insignificant problem for you, please circle ‘‘0.’’

Use the following scale to rate **each** of the health problems.

0 = NOT experienced in the last 3 months, or it is an insignificant problem
1 = MILD or INFREQUENT problem
2 = MODERATE or OCCASIONAL problem
3 = SIGNIFICANT or CHRONIC problem

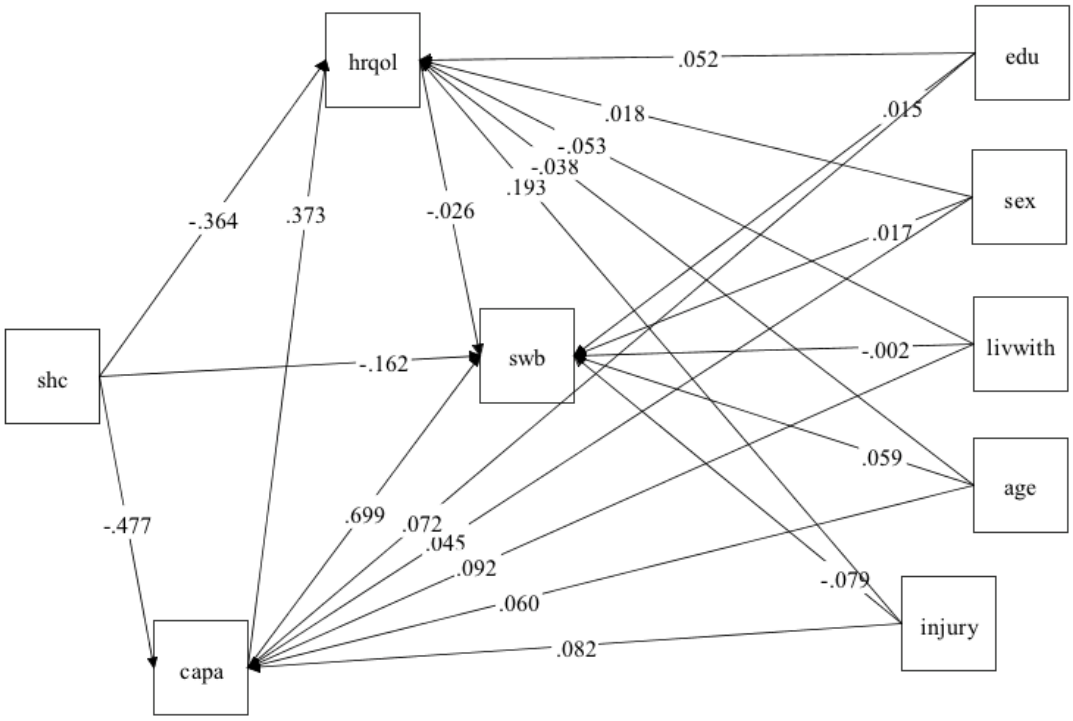
Health Problem	Description	Rating			
		0	1	2	3
1. Muscle Spasms (Spasticity)	Spasticity refers to uncontrolled, jerky muscle movements, such as uncontrolled muscle twitch or spasm. Often spasticity increases with infection or some kind of restriction, like a tight shoe or belt.	☐	☐	☐	☐
2. Depression / Mood Problems	A state of intense sadness that lasts for more than two weeks and has advanced to the point of interfering with daily life – feeling “down”, being tired, or feeling irritable for no apparent reason.	☐	☐	☐	☐
3. Pressure Sore(s)	These develop as a skin rash or redness and progress to an infected sore. Also called skin ulcers, bedsores, and decubitus ulcers.	☐	☐	☐	☐
4. Bladder Dysfunction	Incontinence, bladder or kidney stones, kidney problems, urine leakage and urine back up are all symptoms of bladder dysfunction. NOTE:	☐	☐	☐	☐

	There is a separate item for urinary tract infections.				
5. Trouble Sleeping	Difficulty falling asleep, waking during the night, pauses in breathing during sleep, etc.	☐	☐	☐	☐
6. Joint and Muscle Pain	This includes pain in specific muscle groups or joints. People who must overuse a particular muscle group, such as shoulder muscles, or who put too much strain on their joints are at risk of developing pain.	☐	☐	☐	☐
7. Neuropathic Pain	Pain, that is often ongoing and intense, caused by damage to nerves, that occurs spontaneously or by light touching and is characterized by feelings of burning, shooting, tingling, etc.	☐	☐	☐	☐
8. Sexual Dysfunction	This includes dissatisfaction with sexual functioning. Causes for dissatisfaction can be decreased sensation, changes in body image, difficulty in movement, and problems with bowel or bladder, like infections.	☐	☐	☐	☐
9. Autonomic Dysreflexia	Autonomic dysreflexia, sometimes called hyperreflexia, results from interference in the body's temperature regulating systems. Symptoms of dysreflexia include sudden rises in blood pressure and sweating, skin blotches, goose bumps, pupil dilation and headache. It can also occur as the body's response to pain where an individual doesn't experience sensation.	☐	☐	☐	☐
10. Bowel Dysfunction	Diarrhea, constipation, "accidents," and associated problems are signs of bowel dysfunction.	☐	☐	☐	☐
11. Fatigue	Constantly feeling tired, having low energy, feeling listless, etc.	☐	☐	☐	☐
12. Fractures	A crack or break in the bone.	☐	☐	☐	☐

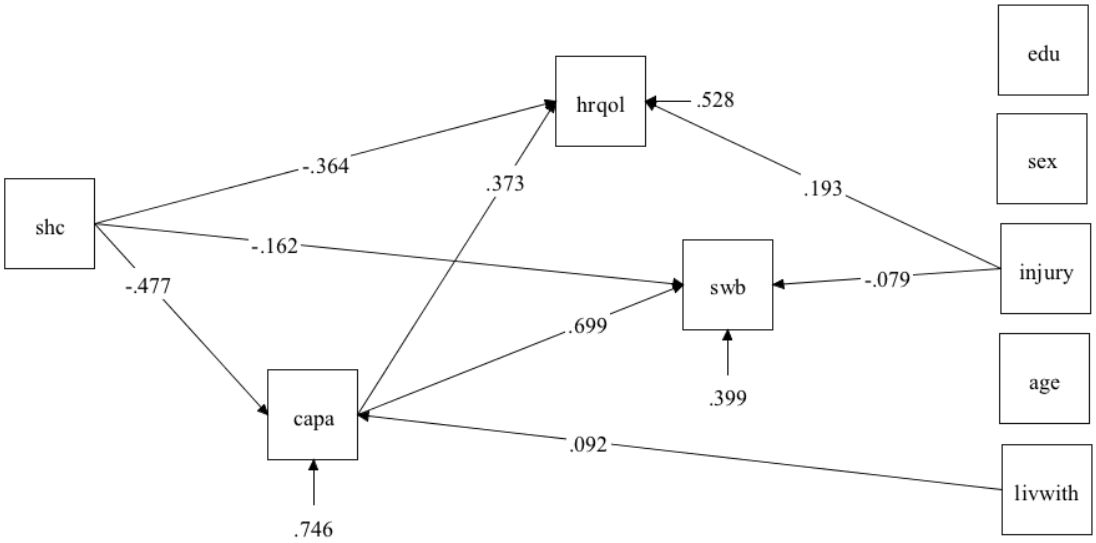
13. Urinary Tract Infection	This includes infections such as cystitis and pseudomonas. Symptoms include pain when urinating, a burning sensation throughout the body, blood in the urine and cloudy urine.	☐ ☐ ☐ ☐
14. Respiratory Problems	Symptoms of respiratory infections or problems include difficulty in breathing and increased secretions.	☐ ☐ ☐ ☐
15. Osteoarthritis / Degenerative Arthritis	"Wear and tear" on joints causing pain, swelling, and reduced movement / function of the joint.	☐ ☐ ☐ ☐
16. Cerebrovascular disease, stroke, trans-ischemia attack (i.e. TIA)	Permanent or temporary loss or reduction of brain function due to an interruption of blood flow to the brain.	☐ ☐ ☐ ☐

Appendix 6.3. Initial model structure for Model 1

A: Initial model structure with standardized parameter estimates

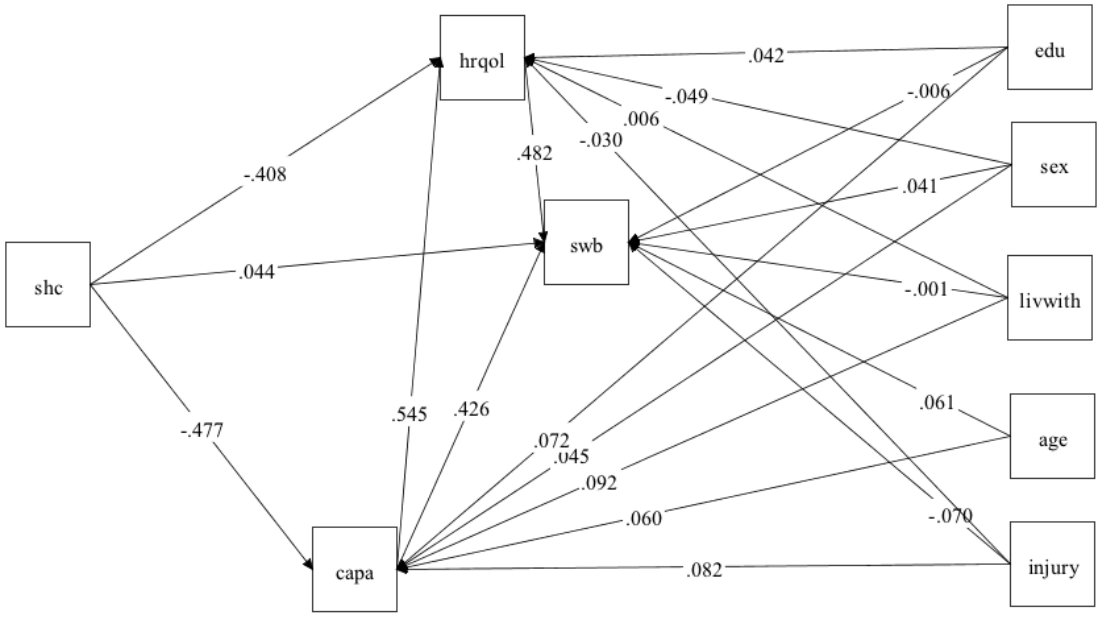


B: Initial model structure showing only significant standardized parameter estimates

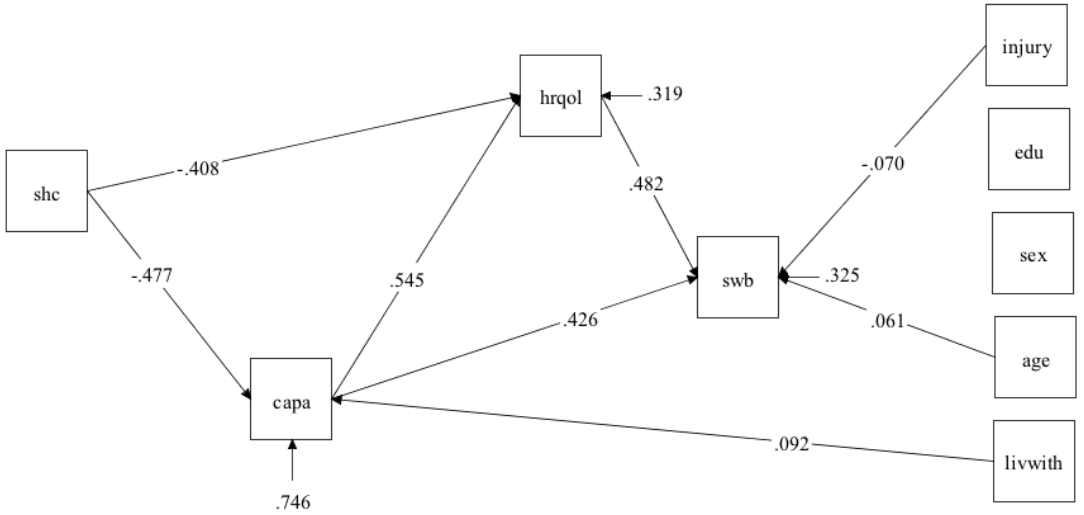


Appendix 6.4. Initial model structure for Model 2

A: Initial model structure with standardized parameter estimates



B: Initial model structure showing only significant standardized parameter estimates



Appendix 6.5. Abbreviations of the parameters included in the model

Abbreviation	Description
SHC	Secondary health condition
HRQoL	Health-related quality of life
Capa	Capability wellbeing
SWB	Subjective wellbeing
Edu	Education
Sex	Sex
Livwith	Living with somebody
Age	Age
Injury	Level of injury (tetraplegia & paraplegia)