Different Aspects of municipal Rehabilitation and the Use of a Capability Measure

the importance of outcome measures

Holst-Kristensen, Annette Willemoes

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

DIFFERENT ASPECTS OF MUNICIPAL REHABILITATION AND THE USE OF A CAPABILITY MEASURE

THE IMPORTANCE OF OUTCOME MEASURES

BY
ANNETTE WILLEMOES HOLST-KRISTENSEN

DISSERTATION SUBMITTED 2020

AALBORG UNIVERSITY
DENMARK

DIFFERENT ASPECTS OF MUNICIPAL REHABILITATION AND THE USE OF A CAPABILITY MEASURE

THE IMPORTANCE OF OUTCOME MEASURES

by

Annette Willemoes Holst-Kristensen

AALBORG UNIVERSITY
DENMARK

Dissertation submitted 2020

.

Dissertation submitted: March 10th, 2020

PhD supervisor: Professor Kjeld Møller Pedersen,
Aalborg University

Assistant PhD supervisor: Clinical Professor Kirsten Fonager,
Aalborg University

PhD committee: Professor Lars Holger Ehlers (chairman)
Aalborg University

Associate Professor Line Bjørnskov Pedersen
Southern Denmark University

DR. PHD Hareth Al-Janabi
University of Birmingham

PhD Series: Faculty of Social Sciences, Aalborg University

ISSN (online): 2246-1256
ISBN (online): 978-87-7210-616-8

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2020

Standard pages: 50.5 pages (2,400 characters incl. spaces).

CV

Annette Willemoes Holst-Kristensen obtained a bachelor's degree in Medicine with Industrial Specialisation in 2012 and a master's degree in Medical Market Access in 2014, both from Aalborg University, Denmark. After graduation, Annette worked as a research assistant at the Danish Center for Healthcare Improvements at Aalborg University. A year later, collaboration with the municipal healthcare centre and the Centre for Applied municipal Health Research in the municipality of Aalborg was established, and Annette enrolled as a PhD student in the autumn of 2015.

Research during Annette's PhD period has focused on rehabilitation and the development and validation work of the Danish ICECAP-A questionnaire. This has resulted in presentations at the annual meetings of the Danish Society for Health Economics and the International Health Economics Association (iHEA), and at the iHEA pre-conference 'Capability workshop' in Basel, 2019, and the Nordic Health Economic Study Group in Reykjavik, 2019.

In the autumn of 2018, Annette visited Bristol for her research stay. She spent a month with Professor Joanna Coast and the Health Economics Bristol (HEB) group at Bristol University, which resulted in scientific guidance and collaboration, network building and a presentation for the group.

During her enrolment as a PhD student, Annette has given lectures and supervised master's students of Medical Market Access. Furthermore, she has been a member of the PhD Study Board at the Faculty of Social Science, and of the Study Board for Medicine.

ACKNOWLEDGEMENTS

During the past four years and the making of this thesis, I have been lucky to work with a lot of wonderful people who have inspired and contributed to my research. Thank you to all!

I would especially like to thank Kjeld Møller Pedersen, my main supervisor. Thank you for sharing all your expertise and for all the educative and challenging conversations and discussions. It has been very rewarding. Also, a thank you to Kirsten Fonager, my co-supervisor, for all the right questions and challenging discussions. I have learned a great deal from you both.

I would like to thank Joanna Coast and the Health Economics Bristol group at Bristol University for hosting my research stay in November 2018, and for giving me a unique insight into the work you do. Also, a big thanks to Joanna for giving me feedback on the thesis.

To all my colleagues at the Danish Center for Healthcare Improvements, thank you for all the tremendous and educative hours and all the academic discussions during the early morning coffees. None mentioned, none forgotten – thank you all.

Thank you to my family and friends, particularly my husband, Tim, for all the moral support, understanding, and love. A thank you to Nicoline Sjøberg, my close friend, for visiting me in Bristol, and for knowing that I was missing a piece of home. That, I will never forget.

Last but not least, I would like to thank those who have supported my research financially. Without support from the municipality of Aalborg – a front runner when it comes to co-financing and co-operation with research – and the Health Foundation (Helse Fonden), this thesis would not have been possible.

An adventure is not knowing what you set out on. This, indeed, has been an adventure.

Annette Willemoes Holst-Kristensen

ENGLISH SUMMARY

The burden of chronic illness is increasing in Denmark and across the globe, and we live longer. As populations age, the amount of time that people spend living with a disability and with the consequences of chronic illness increases. This increase affects the healthcare and social care systems. Therefore, a societal focus on the, health, quality of life and well-being are essential, and as the Danish population ages and seriously ill individuals survive with disabilities and lower health status levels, the field of rehabilitation becomes increasingly important. Rehabilitation programmes for people with a chronic illness will become increasingly important over the upcoming decades. Therefore it is crucial to continually assess, evaluate, adjust and develop the programmes offered today so that people with a chronic illness can master their disease and live a meaningful life with good quality of life. This thesis is an attempt to highlight the importance of broad measurement in public health interventions such as rehabilitation programmes. Furthermore, to highlight some of the less explored aspects of municipal rehabilitation. Hence the research questions of interest are: is the capability approach and the ICECAP-A a potential outcome measure in rehabilitation, and is ICECAP-A a substitute or supplement to the health-related quality of life outcome measures. Lastly, what are the different non-quality of life aspects of municipal rehabilitation in terms of participation and effects of the programme? Based on the results of four scientific papers, this thesis finds the capability approach and the ICECAP-A to be potential outcome measures for use in public health intervention. A reliable, valid and responsive Danish version of ICECAP-A is now available for use in chronically ill populations. The thesis also demonstrated the different reasons for attendance, non-attendance and dropping out of municipal rehabilitation programmes, resulting in different 'profiles' and focus areas for use in everyday work at the healthcare centre. Furthermore, the patients change in health-related quality of life, calculated as quality-adjusted life-years was investigated before and after rehabilitation, and found small positive net effects, with differences across socioeconomic status. In conclusion, the results from this thesis highlight the necessity for a broader view of outcome measurements of public health interventions, where the aim is broader than health. Based on the findings of this thesis, the recommendation is to draw up standardised guidelines for measuring effects and broaden the view on effects. Furthermore, the healthcare centre should continue to focus on the individual characteristics, possible comorbidities and personal objectives with the presented 'profiles' in mind.

DANSK RESUME

Antallet af kronisk syge er voksende i hele verden, herunder også i Danmark samtidig med vi lever længere. Efterhånden som befolkningen bliver ældre, øges den periode, hvor folk lever med funktionsnedsættelse og konsekvenserne af kronisk sygdom. Dette sætter sundhedsvæsenet under pres. Samfundsmæssigt fokus på sundhed, livskvalitet og velvære er derfor vigtigt. Når befolkningen ældes og alvorligt syge borgere lever længere med funktionsnedsættelse og lavere sundhedsniveau, bliver rehabiliteringsområdet stadig vigtigere. Derfor er løbende vurdering, evaluering, justering og udvikling af de rehabiliteringsprogrammer, der tilbydes i dag nødvendig. Borgere med en kronisk sygdom skal gennem disse programmer lære mestre deres sygdom og leve et meningsfuldt liv med god livskvalitet og velvære. Denne afhandling forsøger at fremhæve betydningen af hvilket redskab der anvendes til at måle effekt af folkesundhedsinterventioner såsom rehabilitering. Endvidere fremhæver afhandlingen nogle af de mindre udforskede perspektiver af kommunal rehabilitering. Følgende forskningsspørgsmål er undersøgt: er *capability approach* og ICECAP-A mulige metoder til måling af effekt på velvære i rehabilitering, og er ICECAP-A en erstatning eller et supplement til de nuværende sundhedsrelaterede livskvalitetsmål. Afslutningsvis, undersøges de sundhedsøkonomiske effekter af det kommunale rehabiliteringsprogram i Aalborg. Resultaterne i denne afhandling baseres på fire videnskabelige artikler, og finder at ICECAP-A er et potentielt måleinstrument for folkesundhedsinterventioner til at måle effekten af velvære. Den danske version af ICECAP-A har påvist at være pålidelig, valid og følsom. Afhandlingen undersøger endvidere forskellige årsager til borgeres deltagelse, udeblivelse og afbrydelse af kommunale rehabiliteringsprogrammer, hvilket resulterede i forskellige 'profiler' og fokusområder til brug i det daglige arbejde på sundhedscentre. Endvidere oplevede de borgere der gennemførte en forøgelse af deres sundhedsrelateret livskvalitet. Denne effekt påvirkes desuden af borgernes socioøkonomiske status. Afslutningsvis fremhæver resultaterne fra denne afhandling nødvendigheden af et konkret og bredt effektmål til folkesundhedsinterventioner, da målet netop er bredere end sundhed. Baseret på resultaterne af denne afhandling er anbefalingen at der udarbejdes standardiserede danske retningslinjer for måling af effekter med et udvidet syn på livskvalitet. Desuden bør sundhedscentrene fortsætte med at fokusere på den enkelt borgeres egenskaber, mulige komorbiditeter og personlige mål, men med et større afsæt i de præsenterede ‘profiler’.

ABBREVIATIONS

ASCOT	Adult social care outcomes toolkit
BWS	Best–worst scaling
CBA	Cost–benefit Analysis
CEA	Cost–effectiveness analysis
CE-plane	Cost–effectiveness plane
CUA	Cost–utility analysis
DCE	Discrete-choice experiment
HrQoL	Health-related quality of life
ICECAP	ICEpop CAPability measure
ICECAP-A	ICEpop CAPability measure for adults
ICER	Incremental cost-effectiveness analysis
OCAP	Oxford capability measure
OCAP-18	18-item capability questionnaire for public health
OCAP-MH	Oxford capability measure for mental health
QoL	Quality of life
QALY	Quality-adjusted life year
SCRQoL	Social care-related quality of life
SC-QALY	Social care quality-adjusted life year
TTO	Time trade-off

LIST OF PUBLICATIONS

PAPER I **TEST–RETEST RELIABILITY OF ICECAP-A IN THE ADULT DANISH POPULATION**

Annette Willemoes Holst-Kristensen, Kirsten Fonager and Kjeld Møller Pedersen.
Quality of Life Research (2020) 29:547–557

PAPER II **AN INVESTIGATION OF CONSTRUCT VALIDITY AND RESPONSIVENESS OF THE DANISH ICECAP-A**

Annette Willemoes Holst-Kristensen, Paul Mark Mitchell, Myles-Jay Linton, Joanna Coast, Kirsten Fonager and Kjeld Møller Pedersen
Submitted

PAPER III **DETERMINANTS OF PARTICIPATION AND DROP-OUT IN MUNICIPAL REHABILITATION**

Annette Willemoes Holst-Kristensen, Kirsten Fonager and Kjeld Møller Pedersen.
Submitted

PAPER IV **THREE PERSPECTIVES ON ECONOMIC ANALYSIS OF HEALTH-RELATED QUALITY OF LIFE AND HEALTHCARE UTILISATION IN MUNICIPAL REHABILITATION PROGRAMMES – A COMPARISON OF SOCIOECONOMIC GROUPS**

Annette Willemoes Holst-Kristensen, Kirsten Fonager and Kjeld Møller Pedersen.
Manuscript

TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER 1. INTRODUCTION

Across the globe, including Denmark, the burden of chronic diseases is increasing, with 16–57% of adults in developed countries suffering from more than one chronic condition [1–3]. Along with this, we are living longer and longer, and in 2017, the number of Danes over 65 years of age was 1,169,563, equivalent to 19% of the Danish population. This percentage is expected to reach 24% by 2050 [4]. As populations age, the amount of time that people spend living with a disability and with the consequences of chronic illness increases. This undoubtedly affects the healthcare and social care systems. Therefore, a societal focus on quality of life (QoL), health and well-being are essential, and as the Danish population ages and seriously ill individuals survive with disabilities and lower health status levels, the field of rehabilitation becomes increasingly important. There is a great deal of knowledge and evidence regarding the effects of various rehabilitation programmes aimed at people with chronic illness [5–7]. Rehabilitation programmes for people with a chronic illness will become increasingly important over the upcoming decades, and therefore it is crucial to continually assess, evaluate, adjust and develop the programmes offered today so that people with a chronic illness can master their disease and live a meaningful life with good QoL [8].

Municipal rehabilitation programmes differ from other aspects of the healthcare system as they focus simultaneously on health outcomes, ranging from clinical to social, while also focusing on a broader aim besides health and curing diseases. In rehabilitation, this broader aim comprises what is essential for the patient in everyday life [9]. The question is whether to measure the effects in terms of narrow health outcomes or if there is a need for broader measurement in order to capture the entire purpose of rehabilitation. In order to evaluate a municipal rehabilitation programmes, it is important to investigate aspects beyond the health and non-health dimensions, such as participation and dropout rates, and the effects on QoL.

The starting point for this thesis was the acknowledged concern that the evaluative space of quality-adjusted life years (QALYs) focuses too narrowly on *health-related* aspects of QoL, and that there is a lack of instruments that can capture the effects of interventions with broader aims beyond health and, thus, those of some interventions in the public health setting – e.g., municipal rehabilitation. The capability approach is the theoretic foundation of new instruments considered as a possible solution for this concern. One such instrument is the ICECAP-A (ICEpop CAPability measure for Adults) questionnaire, which captures QoL in a broader manner and for use in health research and health economic evaluation. A new and broader measurement is required due to the limitations of the current methods of assessment and to the increasing acknowledgment that health interventions often result in – and are intended to result in – outcomes beyond health [10, 11].

CHAPTER 2. RESEARCH QUESTIONS

The thesis presents work carried out in order to operationalise the capability approach in terms of the ICECAP-A questionnaire in the context/setting of a Danish municipal rehabilitation programme. It further investigates some of the different aspects of rehabilitation, such as participation and dropout rates at the municipal rehabilitation programme, along with the importance of the choice of outcome measures in order to cover the aim of such programmes. The thesis includes four scientific papers about issues and perspectives that substantiate the research questions.

The present thesis is based on the following research questions:

Are the capability approach and the ICECAP-A a (future) potential outcome measure in, for instance, rehabilitation?

Is the ICECAP-A a substitute or supplement to the health-related quality of life outcome in health economic evaluation such as QALYs?

What are the different non-QoL aspects of municipal rehabilitation? For instance the rate of participation and dropout and the effect of the programme.

These research questions have not previously been explored in a municipal rehabilitation context. To investigate the research questions, the thesis has two dimensions. Firstly, it examines the psychometric properties of a Danish version of ICECAP-A and its operationalisation in a rehabilitation setting. Operationalisation of the capability approach in terms of ICECAP-A has been investigated in other populations [12, 13], but not in a chronically ill population and not in a municipal rehabilitation setting. Secondly, it investigates different aspects of rehabilitation, with the municipality of Aalborg and the healthcare centre used as a case study. Little is known about participation in and effects of rehabilitation in a municipal setting, which makes this of great interest.

Chapter 3 presents a description of the background in terms of the organisation of the Danish healthcare system, public healthcare sector and prevention programmes, along with a description of rehabilitation programmes. The chapter defines rehabilitation internationally and nationally, along with a description of the general purpose and aim of rehabilitation programmes offered to chronically ill patients in DK.

Chapter 4 presents the theoretical background, the concepts of health economic evaluation, and how and why they are used. Furthermore, it outlines the theoretical rationale behind QALYs and the most relevant instruments for collecting health-related quality of life (HrQoL) data, along with a description of the known limitations of QALYs when used in settings other than those strictly related to health.

Chapters 5, 6 and 7 present the theory of the capability approach and the development of the ICECAP-A questionnaire, along with a description of the psychometric methods used to determine the reliability, validity and responsiveness of the Danish ICECAP-A questionnaire. Furthermore, Chapter 7 summarises and discusses the findings of Papers I and II, where the psychometric properties of the Danish ICECAP-A were investigated. The overall research questions are additionally answered in-depth in Chapter 9.

Chapter 8 presents a case concerning the rehabilitation programme in the municipality of Aalborg, Denmark. Little is known about Danish municipal rehabilitation in terms of participation and HrQoL. Hence, the contribution of this thesis is its elucidation of some of the less explored aspects of municipal rehabilitation. In this context, the aspects are participation, effects, costs and hospital utilisation. Paper III presents the attendance rates and possible reasons for dropping out and non-attendance by combining data from the municipality with data from the Danish registries. Another aspect is the effects of the programmes. In Paper IV, this is investigated as a cost–utility analysis based on somewhat limited cost data along with a difference-in-difference analysis, where the novelty lies in the comparison of the effects across different socioeconomic groups.

Chapter 9 presents a unifying discussion and reflection on the research questions and the individual papers included in the thesis, as well as its contribution and implications. Lastly, it includes suggestions for future work concerning municipal rehabilitation and the possible implementation of ICECAP-A in a Danish setting.

CHAPTER 3. BACKGROUND

This chapter describes the rehabilitation background of the thesis. Rehabilitation will be described as a general phenomenon both internationally and nationally, and more specifically in terms of municipal rehabilitation in Denmark.

3.1. ON HEALTHCARE SYSTEMS

What is the primary purpose of healthcare systems? The most natural answer would be 'to improve health'. The World Health Organization (WHO) has defined health as '*a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity*' [14]. The primary purpose of the Danish healthcare system is stated in the Danish Health Care Act: '*The healthcare system aims to increase the health of the population and to prevent and treat illness, suffering, and functional limitations for the individual*' (author's translation) [15]. Thus, the primary purpose of the healthcare system and, therefore, the interventions it provides is to provide relief from pain and illness and improve the health of individuals and society as a whole. The aim is not, therefore, to increase activities in the healthcare system, such as bed days, medication, surgery, etc., since these measures do not increase patient utility but are a means to an end – the achievement of health [16].

A healthcare system is essential to the maintenance of public health and the services offered in the public health sector. Public health is defined as '*the art and science of preventing disease, prolonging life and promoting health through the organised efforts of society*' (Acheson, 1988)[17]. The overall vision is to encourage better health and well-being sustainably while strengthening integrated public health services and reducing inequalities. In order to achieve this, the public health sector must collaborate with other sectors and health professionals to address the broader determinants of health. Here, primary healthcare professionals play an essential role in preventing illness and promoting health [17]. Public health interventions tend to be complex and context-dependent, and often involve a combination of social, educational and health-promoting elements, with aspects of empowerment, capacity building and knowledge across the different players – e.g., the healthcare system and social system. These services may include drug abuse treatment, mental healthcare, or rehabilitation. The nature of public health interventions means that some effects are likely to fall outside the healthcare sector, which impacts the choice of health or non-health outcomes as measures and value interventions – e.g., in a health economic evaluation. The evidence for their effectiveness must be sufficiently comprehensive to encompass that complexity [18, 19].

3.2. ORGANISATION OF THE DANISH HEALTHCARE SYSTEM

The Ministry of Health and the Elderly is responsible for establishing the general framework for the provision of healthcare and elderly care. This includes regulation of the organisation and the provision of treatment and healthcare in hospitals, home nursing, and all other types of health services such as pharmacies, pharmaceuticals, vaccinations, maternity care and child healthcare. The Danish healthcare system works across three political and administrative levels: the state, the regions and the municipalities. The state has the overall regulatory and supervisory functions in healthcare and elderly care. The regulation covers the tasks of the regions, municipalities and other authorities within the area of health. The primary task for the five regions is hospital care, including emergencies, psychiatry, and the health services provided by general practitioners (GP) and office-based specialists in private practice. The 98 municipalities are local administrative bodies responsible for a range of primary healthcare, elderly care and social services. Healthcare and elderly care services include preventive care and health promotion, rehabilitation outside of the hospital, home nursing, physiotherapy, school health services, child nursing, child dental treatment, alcohol and drug abuse treatment, home care services, nursing homes, and other services. The municipalities co-finance regional services, both hospital treatment and rehabilitation services during hospitalisation. In general, the health and social services are funded by general taxes and supported by a system of central government block grants and reimbursements. Approximately 80% of healthcare expenditure is publicly funded, and 20% is through patient co-payments. Healthcare accounts for 30% of total public expenditure [20, 21].

Preventive care and health promotion are often defined as being aimed at eliminating disease or risk factors. Prevention is most often divided into primary, secondary and tertiary prevention. Primary prevention comprises interventions aimed at preventing the onset of illness; it is also known as '*civil citizens-directed prevention*', since it takes place among healthy people. Secondary prevention is aimed at detecting people with the initial signs of disease to prevent further progression – e.g., by screening. Tertiary prevention comprises interventions aimed at preventing the worsening of illness that has already arisen, and maintaining the functioning of patients; it is also called '*patient-directed*' prevention, and includes rehabilitation of chronically ill people [20, 21]. The Danish Health Act states that regions, together with municipalities, share responsibility for patient-directed prevention efforts [22]. Every four years, each regional council and the associated municipal council sign a political health agreement, which sets out the framework and objectives for collaboration between the parties in four areas of action in the field of health: prevention, treatment and care, rehabilitation, and health IT and digital work procedures [23]. Regarding rehabilitation, the political health agreement for 2015–2018 aimed at increasing the number of referred patients, the rate of patients completing rehabilitation and the rate of patients completing with an effect [24].

3.3. DEFINITION OF REHABILITATION

Municipal rehabilitation programmes have a new set of requirements due to a number of factors: changing demographics (with more elderly people living with the consequences of chronic illness), centralisation of hospitals, shorter lengths of stay, and the Danish Health Act transferring a greater part of the responsibility for follow-up, during and after treatment, to municipalities [7]. Rehabilitation is essential in enabling people with restrictions in functioning to remain in or return to their home, live independently, and participate in education, the job market and everyday life [14, 25].

Rehabilitation is defined by the WHO, in its broadest sense, as:

> *A set of measures that assist individuals, who experience or are likely to experience disability, to achieve and maintain optimum functioning in interaction with their environments (WHO, 2011).* [14]

Rehabilitation is a multidisciplinary activity carried out by health professionals in conjunction with specialists in, e.g., education, employment, social welfare and other fields. Rehabilitation is aimed at achieving the broad outcomes of preventing or slowing function loss, improving or restoring function, compensating for function loss, and maintaining current function [14, 25]. Rehabilitation can be provided in a variety of settings – e.g., acute-care in hospitals, specialised rehabilitation centres, nursing homes, and so forth. Longer-term rehabilitation may be provided within municipal settings and facilities such as primary healthcare centres, rehabilitation centres, schools, workplaces or homes [14, 25].

3.3.1. REHABILITATION IN DENMARK

In practice, rehabilitation, as a general term, is seen as a set of principles and methods that includes medical, psychological, social, educational and occupational elements aimed at helping sick and disabled people to regain and/or maintain the best possible functional level, engage in everyday life activities and achieve a desired QoL [9].

In 2005, The Danish Health Authority [26] defined rehabilitation according to a 'white paper [9] as:

> *A concentrated and time-limited collaboration between a patient, relatives and health professionals. The rehabilitation programmes are for patients who have, or are at risk of having, significant physical, mental and/or social limitations after an acute event. Rehabilitation is based on the patient's entire life situation, and decisions consist of a coordinated, coherent and knowledge-based effort. (author's translation)* [9]

Since rehabilitation is defined as a process centred on an entire life situation, the overall aim of the rehabilitation programmes can be defined, according to the white paper, as helping patients with health-impaired functioning to regain a meaningful and independent life – a process centred on individuals, enabling them to maintain and promote QoL, regain previous functioning levels (or the highest possible functioning level), prevent relapses or further reduction of functioning levels, learn to live with chronic illness and encourage health by mastering their situation. The extent to which the aim is fulfilled is defined by an individual's own values, determination and resources, as well as those of wider society [27].

Along with a municipal reform in 2007, the Danish municipalities were given co-responsibility for rehabilitation services for non-hospitalised individuals with chronic diseases – referred to as 'patient-directed prevention initiatives'. This involves responsibility for home nursing, nursing homes and rehabilitation [7]. The formal definition of patient-directed prevention is:

> *An effort that prevents further development of a disease and seeks to diminish or postpone possible complications. Tertiary and patient-orientated prevention aims to optimise treatment and enable patients to take care of themselves to the best of their ability, as well as to gain competencies, knowledge and skills in order to practice good self-care – e.g., through rehabilitation and patient education. (author's translation)*[28]

Since the reform, the municipalities have had to develop new prevention programmes, including rehabilitation programmes for chronically ill patients – i.e., patients with, e.g., CVD, COPD and diabetes. However, it has been unclear how the 98 Danish

municipalities have implemented programmes to meet the requirements, as there was no overall guideline defining the aim of such programmes or how they are to be implemented.

In 2016, the Danish Health Authority published new official recommendations concerning rehabilitation for chronically ill patients with CVD, COPD and diabetes. Rehabilitation is now defined according to the previously mentioned WHO definition, and the realm of understanding is based on the International Classification of Functioning, Disability and Health (ICF) [29], a framework for measuring health and disability at both individual and population levels. It is the theoretical basis of the definition, measurement and policy formulations for health and disability, and it is used by both professionals and people with disabilities to evaluate healthcare settings that deal with disabilities and chronic illness, such as rehabilitation programmes, nursing homes, psychiatric institutions, and community services. The model is a bio-psycho-social model with function as an essential part of a dynamic interaction with health and context (environmental and personal factors) [30]. Rehabilitation includes, according to patient needs, physical training, disease management, dietary efforts, and supportive and compensatory efforts, as well as social, educational and employment activities [29].

CHAPTER 4. THEORETICAL BACKGROUND

This chapter describes and discusses the theories and concepts underlying this thesis and, therefore, the basis for investigating the research questions.

4.1. THEORY AND CONCEPTS OF HEALTH ECONOMICS

Health economics is founded on microeconomics and welfare economics and is recognised as a sub-discipline within economics, applying economic methods to questions of health and the healthcare sector. Health economics encompasses elements from various scientific fields, including health science and social science [16, 31]. The methods used all adopt the expectation that resources, both monetary and non-monetary, are scarce. Furthermore, it is expected that all decisions are based on rationality, with a desire to maximise benefits [16, 32].

The two essential paradigms of economics for guiding researchers in the healthcare sector are welfarism and extra-welfarism. Welfarism is an element of the welfare economic framework and deals with maximisation of the overall sum of individual utility as a primary outcome, whereas extra-welfarism involves economic analysis based on a broad set of information that goes beyond exclusive consideration of the utility attained by individuals in society [33].

4.1.1. WELFARISM

In standard welfare economics, the overall welfare of society is a function of individuals' utility, which is a function of the goods and services consumed by the individuals themselves [34, 35]. Individuals' utility is used as an outcome as individuals are believed to be the best judge of what provides the most significant personal benefit – i.e., individual consumer sovereignty [36]. Welfare economics draws upon normative theories developed within the parental discipline of economics. In economics, there is a distinction between positive and normative economics. Positive economics is defined as 'what is', and it is concerned exclusively with analysing the consequences of different changes or policies, without making decisions about the desirability of alternative allocations of resources. Normative economics is defined as 'what we ought to do', and it focuses on the economic evaluation of interventions, mainly clinical or organisational, to help decision makers in allocating resources. Normative economics is value and judgement based. Through economic evaluations of costs and benefits, it pursues to inform decision makers about how resources ought to be allocated [31, 36, 37]. It is standard to consider health economics as the application of the discipline of economics to the questions of health.

There are four central tenets on which welfare economics seeks to achieve economic efficiency, and these are the cornerstones of the framework for welfare economic evaluation: *utilitarianism* – assumes that each individual in society rationally orders options based on their expected benefit, maximising their welfare by choosing the optimum or highest possible level of utility; *individual sovereignty* – assumes that individuals are the best judge of how to maximise their welfare and utility; *consequentialism* – the only consideration for assessing individual goodness is the outcome of choices made by the individual, and how the ends or outcomes are achieved is irrelevant; *welfarism* – the goodness of any resource allocation is judged merely on the aggregation of individual utility information [36].

To determine whether an improvement in social welfare has taken place or not, the 'Pareto principle' is introduced. Practically, welfare economics analyses possible welfare improvements by fulfilling the 'Pareto criterion' – i.e., 'Pareto improvements'. A Pareto improvement occurs if an intervention in healthcare makes one or more persons better off without making another person worse off [35, 38]. However, the allocation of resources cannot produce winners without involving losers and is therefore not very useful, since most decisions, including those in healthcare, involve a choice between alternatives where the additional benefits offered will accrue for some, but the additional cost will mean that sources of value must be given up by others. Thus, accomplishing the Pareto criterion is problematic, because when are there no losers? Alternative criteria for evaluating welfare-improving interventions have been developed. Cost–benefit analysis employs the Kaldor–Hicks hypothetical compensation criterion. Under this approach, society as a whole has benefited from a particular allocation decision if the winners could, in theory, compensate the losers and remain better off than they were before the decision [16, 35, 38]. It is important to note that since individuals can derive utility from different sources, such as from the consumption of health services or education, individuals not receiving health services 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 [35].

4.1.2. EXTRA-WELFARISM

Extra-welfarism requires that normative assessments and economic analysis be based on a broad set of information that does not exclusively consider the utility attained by individuals. Extra-welfarism adopts the same objective for action in healthcare as that of welfare economics: increasing a maximand within budgetary constraints. One principal difference between the two is the notion of what should be maximised [16]. In contrast to welfarism, health, as opposed to utility, is most often seen as the primary outcome of interest. Extra-welfarist approaches to health have changed the evaluative space within economic evaluation away from utility and towards a broader space that can comprise capabilities and characteristics, including, health [35, 36]. The formation of extra-welfarism in health has been heavily influenced by the work of

Amartya Sen and Anthony Culyer. Sen developed an alternative framework based on the concepts of human functionings and capabilities – his capability approach. Sen rejected the limited focus on individuals' utility and suggested replacing it with a broader perspective, considering the quality of utility and individuals' capabilities rather than their emotional reaction to the possession of goods and capabilities [36, 39]. Later, Culyer recognised the potential of adapting Sen's ideas to the healthcare sector [40].

The refusal of individual utility as the solitary outcome of interest in an evaluation marks a clear separation between welfarist economics and extra-welfarism. The extra-welfarist approach is believed to differ from welfarism in four general ways. Firstly, it permits the use of non-utility outcomes, since a sole focus on utility is too narrow for healthcare analysis. Secondly, it permits the use of sources of valuation other than directly affected individuals, meaning consideration of relevant population groups. Thirdly, extra-welfarism allows the weighting of outcomes according to factors other than individual preferences. Lastly, it permits interpersonal comparisons of well-being in a variety of dimensions, thus enabling a comparison between the health of different people, which departs from welfare economic principles [36].

Extra-welfarism has received criticism regarding how it has followed the ideas of Sen, with the most predominant criticism being that it focuses exclusively on health and thereby ignores individuals' capabilities and other factors relevant to their welfare and that of society. It also focuses on functionings, rather than people's ability to function. Therefore, it also criticised for not adding anything 'extra', and instead narrowing the evaluation space [37, 40]. Thus, although not a full expression of the capability approach in healthcare, extra-welfarism is very associated with some aspects of this approach, predominantly the concept of the importance of functionings and the belief that the basis of values does not need to be individual [31]. With the increasing interest and need for economic evaluations in public health settings and other complex interventions, the focus on health alone has been questioned as many interventions often provide a broad range of benefits beyond health – e.g., for those living with a chronic illness, health may not be the single, or even the most important, outcome of treatment. If the effects of public health intervention fall outside the range of health, such intervention is believed to be undervalued, with the sole use of, e.g., QALYs [31]. It is still unclear how to evaluate such interventions and how, or if, new frameworks and outcome measures could help. Recent work has been developed on the basis of Sen's capability approach and the concept of well-being. This is described and discussed in detail in Chapter 5.

4.2. ECONOMIC EVALUATION

A primary challenge in healthcare is the allocation of scarce resources among many competing needs. The healthcare policymakers and decision makers responsible for prioritising resource use are faced with ever-increasing numbers of treatments, ever-

increasing costs, and constantly changing rates of illness [16]. There is a widespread assumption that decision makers intend to maximise the population's health, but they must do so within existing budgetary constraints [16]. Accepting the premise that the purpose of the healthcare system is to maximise health and that this must be done with available resources, the average expected additional health benefits and additional costs related to actions within the healthcare system should be a decisions maker's main concern [38]. Economists and health economists have sought to assist policymakers and decision makers by applying health economic evaluations [36].

Economic evaluation is an analytic approach used to weigh up costs and consequences of interventions competing for the same resources. It provides a systematic way of dealing with scarcity [32, 38]. Economic evaluation is defined as '*the comparative analysis of alternative courses of action in terms of both their costs and consequences*', and its purpose is to inform decision makers on whether the amount of extra health benefits that healthcare interventions provide justifies the additional costs related to the interventions [38]. The term '*benefits*' is also referred to as '*consequences*', '*outcomes*', or '*effects*' within the economic evaluation literature. Decisions about what services to provide (or not), to whom, and where and when usually have resource implications. Following one course of action means that other possible actions are not taken, and using resources in one place means there is less to use elsewhere. Thus, there are lost opportunities – referred to as opportunity costs. This means that decisions about the allocation of scarce resources involve inevitable trade-offs, which are captured in the opportunity cost. An opportunity cost is defined as the benefits foregone from those resources not being used in the most highly valued alternative [32, 38].

Health economics has developed several methods for measuring the benefits of health interventions, focusing on their quantification. Evaluations that consider both costs and effects can be considered 'full' economic evaluations, of which there are three main types: cost-effectiveness analysis (CEA), cost–utility analysis (CUA), and cost–benefit analysis (CBA) [38]. The unit for measuring the effects of healthcare is the key feature that distinguishes the different types of economic evaluation [32, 38]. All three methods aim to improve society's value for money when investigating healthcare and welfare by establishing the comparative value of competing alternatives through evaluation of their associated costs and benefits.

4.2.1. COST–EFFECTIVENESS ANALYSIS

Cost-effectiveness analysis is the most frequently used form of economic evaluation in healthcare sectors. Effects are valued in natural health units – e.g., life-years saved, reduction in blood pressure or decreased number of hospital admissions [32, 38]. The analysis usually takes a narrow health sector perspective by only informing decision makers in healthcare, but it could also take a societal perspective. CEA informs healthcare decision makers about the cost of obtaining one extra unit of health

outcome when comparing the cost-effectiveness of two or more interventions. Whether the cost is worthwhile depends on the value that decision makers place on such health effects. The limitation of CEA is that it only focuses on a single outcome common to the alternative being evaluated. Therefore, it cannot compare across sectors, diseases or different outcomes without missing many of the effects. Moreover, trade-offs cannot be made explicitly. Consequently, HrQoL measures have become an increasingly important method of assessing the efficacy of an intervention or treatment. CEA is, therefore, most useful for prioritisation when evaluating new treatments for specific disease groups or similar health interventions where CUA would be more suitable for optimising across the healthcare system [16, 38].

4.2.2. COST–UTILITY ANALYSIS

To enable comparisons across different patient groups, and areas of healthcare, a standard measure is necessary. Cost–utility analysis is the most frequently used form of economic analysis for decisions involving healthcare resource allocation. Here, utility refers to the preferences for a health outcome. This method allows health outcomes to be 'valued according to their desirability' [38]. Therefore, while CUA is often considered a subgroup of CEA, since it compares interventions in terms of their cost per unit of effect, CUA has clear and important distinctions – e.g., the effect measure and the recognition of the importance of considering population preferences [38]. In CUA, effects are measured using QALYs, which combine measures of length of life and HrQoL. When calculating QALYs, the HrQoL associated with a particular health state is multiplied by the length of time spent in that health state, often life-years. Measures of HrQoL often refer to the preferences that individuals have for a health outcome, and they allow health outcomes to be values according to their desirability [38]. With CUA, it is only possible to optimise within the healthcare system, and thus if we are interested in broader allocative efficiency of investments, CBA is the preferred choice of evaluation.

4.2.3. COST–BENEFIT ANALYSIS

In CBA, the benefits of an intervention are expressed in monetary terms. Cost–benefit analysis is the only type of economic evaluation to put costs and benefits in monetary terms, and it is, therefore, able to compare interventions across sectors, as well as aiding decisions regarding how much money to invest in an intervention. CBA implies placing a value on life and health, which is difficult [41].

Table 1 The differences between the three main evaluation types.

	Summary of characteristics associated with economic evaluation		
	CEA	**CUA**	**CBA**
Theoretic	Partly extra-welfarism	Extra-welfarism	Welfarism
Perspective	Healthcare sector or societal	Healthcare sector or societal	Societal
Description	Values all cost in monetary terms. Health effects are measured in natural units (e.g., life-years gained, blood pressure reduction, bed days, etc.).	Values all cost in monetary terms. Health effects are measured in terms of QALYs.	Values all cost in monetary terms. Health effects are measured in terms of monetary valuations.
Decision rule	A new intervention is cost-effective if the ICER falls below the threshold value in the CE plane (Figure 2).	A new intervention is cost-effective if the ICER falls below the threshold value the CE plane (Figure 2).	A new intervention is cost-effective if the net monetary benefit is positive.
Advantages	Easy to understand and relevant to clinicians.	Incorporates QoL by adjusting changes in life-years for differences in health effects. Comparable across disease areas and interventions.	Applicable across different sectors to inform on allocative efficiency.
Disadvantages	No comparison across diseases	Can be challenging to interpret QALYs. Quality of life requires evaluation of preferences.	Difficult to apply within healthcare as patients and decision makers have difficulties valuing effects in monetary terms.

4.3. QUALITY OF LIFE AND HEALTH-RELATED QUALITY OF LIFE

There is ever-increasing interest in including patient outcomes in clinical studies, as well as in a wider range of interventions. These outcome measures often include QoL and HrQoL [38]. The construct of HrQoL refers to the impact of the health aspects of an individual's life on their QoL or overall well-being, including their health state [32].

The interpretation of health and QoL is essential when evaluating the effectiveness of interventions in the healthcare system. Interventions aimed at maximising the health of patients are (often) interpreted as an improvement of QoL. Health is defined by the WHO as: '*a state of complete physical, mental and social wellbeing and not merely the absence of disease and infirmity*' [42]. A key aspect of this definition is the

inclusion of social well-being and the emphasis on health being more than the absence of disease. However, this inclusion of social well-being is not accepted by everyone. Patric et al. defined health as '*an individual's level of function*', where a function is compared with '*society's standards of physical and mental well-being*' [43]. Is improving health an improvement of quality of life? QoL has proven challenging to define, but one definition is: '*a conscious cognitive judgment of satisfaction with one's life*' [44] and '*an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns*' [45]. Though most definitions of QoL focus on subjective judgements, some authors have suggested that objective aspects should be included [46–48]. In that form, QoL has been defined as '*an overall general well-being that comprises objective descriptors and subjective evaluations of physical, material, social, and emotional well-being together with the extent of personal development and purposeful activity, all weighted by a personal set of values*' [47]. It is therefore more than just health.

4.3.1. GENERIC AND PREFERENCE-BASED MEASURES

Health related quality of life can be ascertained by generic and preference-based measures. There is an assortment of generic and preference-based measures that can be used to obtain values of health states; some of the most commonly used measures are mentioned here. The generic outcome measures are essential for comparisons of populations across disease states and interventions. Generic instruments such as the Sickness Impact Profile and SF-36 provide insights into various attributes of health with established population norms. The SF-36 is a questionnaire comprised of eight health attributes: physical functioning, physical role, bodily pain, general health, vitality, social functioning, emotional role and mental health. The outcome is comprised of two summary scores of physical health and mental health [49]. Preference-based measures, including the EQ-5D, Health Utilities Index (HUI), and SF-6D (a derivative of the SF-36 and SF-12), are used to generate utility scores[32]. The EQ-5D is a measure which comprises the quality of life in terms of five attributes: mobility, self-care, usual activities, pain/discomfort and anxiety/depression [50]. Two versions are currently available and differs in number of response categories to each question [51]. From these questions, a single index score can be calculated through a preference elicitation procedure. The HUI orders HrQoL under eight attributes: vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain. The SF-6D, developed in 2002 by Brazier et al., is based on SF-36 measure [52]. The measures all claim to be generic and preference-based, although they have some differences in their content, size, and the way they calculate the preference weights. A methodological similarity is that for all above mentioned measures, the index score is a summary of health state and not just a description of its valuations. This methodological approach enables the calculation of QALYs based on population preferences [53].

4.3.2. QUALITY-ADJUSTED LIFE YEARS

Quality adjusted life years is a generic measure that takes into account both the quantity and quality of life, thus capturing the effect of a treatment on a patient's length of life and the impact on their HrQoL within a single measure. QALYs are widely used in health economics as a summary measure of health outcome, which can be combined with cost data to inform healthcare resource allocation decisions [32, 38]. The conventional application of QALYs in CUA is justified from the extra-welfarist framework developed i.a. by Culyer [40, 54]. When applying QALYs, a year is adjusted for the life quality during that year, and thereafter reduced to a single value. In order to generate that single value of QALYs, health utility is necessary. Utility acts as a preference weight that can be equated with a value or desirability. The concept is that individuals move through different health states over time, and each health state has a value on a scale from 0 to 1. A score of 1 indicates a person living in perfect health for one year. A score lower than 1 suggests that a person is either living in a degree of poor health for that year or they have lived less than a year. An intervention, treatment or programme in the healthcare system can increase QALYs by either increasing the QoL during a time or extending the time a person lives [38, 55]. In order to use QALYs in economic evaluation and to aggregate the changes, value must be measured on an interval scale for the weights. The interval scale is required because intervals of equal length on the scale must have equal interpretation; for example, a gain from 0.3 to 0.5 is equally valuable as a gain from 0.6 to 0.8. States worse than death is possible and would have a negative value [38, 54].

The response to the HrQoL measure can be scored, using weighted responses, to give an output that distinguishes patient or the general population. The methods by which these preferences are elicited vary based on individual preference using one of three main processes: standard gamble (SG), time trade-off (TTO) [56], and the discrete-choice experiment (DCE) [57]. In the SG approach, individuals are presented with two options. In option one, the individual stays in a chronic state for life. Option two has two possible outcomes, with probabilities attached. In outcome one, the person returns to full health and lives for a number of years, and in outcome two, the person dies immediately. The probabilities are varied systematically until the individual is indifferent to the two options [38]. The TTO approach gives the individual two options. Option one is to live in the diseased state for a given time, and option two is to live in a healthy state for a time period smaller than the first option. The time is then varied until the individual is indifferent [38, 56]. The DCE uses questionnaires based on different choices. The respondent is presented with different sets of hypothetical scenarios. In each scenario, the respondent is asked to choose between two or more options that vary in essential characteristics. Resulting choices reveal an underlying (latent) utility function [53, 57, 58].

4.3.3. THE LIMITATION OF QUALITY-ADJUSTED LIFE YEARS

The credibility of QALYs comes from the concept of a person preferring a shorter healthy life than a longer life in a state of severe discomfort. Despite the fact that QALYs are widely applied in health economic evaluations, they have attracted criticism. QALYs are used to determine which treatment to give a patient group or which procedure to use for a specific disease. But they may also be used to decide which groups of patients to treat and what diseases should be prioritised in the allocation of resources. Evaluations are based on population preferences, which raises ethical concerns, since each person is equally important as any other and should be given equal weight [59]. The advantage of QALYs is the possibility of measuring health outcomes from interventions across a range of clinical areas on a standard scale. However, there is increasing concern that the HrQoL measures that underpin QALYs are not sufficiently sensitive in a number of healthcare areas. Specific areas such as social care [60], mental health [61], public health [62], complex interventions [63] and chronic pain [64], as well as certain groups, such as the elderly [65] and carers [66], have been identified as having a broader set of benefits than those currently measured by HrQoL instruments. Healthcare interventions in mentioned areas may be geared towards helping individuals maintain independence, dignity, comfort or social interaction, and these benefits may be neglected if only the measures of health is being used[10, 60, 67]. According to Brazier et al.[32], there is no reason why the concept of QALYs needs to be limited to health alone, neglecting the inclusion of a broader view of effects and benefits such as well-being. A simple generic measure of well-being, which goes beyond health, may be useful for comparing the benefits of a diverse range of healthcare and social care policies. Such a measure could be ICECAP, which is described and discussed in Chapters 6 and 7.

4.4. DECISION RULES IN ECONOMIC EVALUATION

The primary purpose of CEA and CUA is to identify which of two or more alternatives provides the best alternative – the intervention that is value for money. By calculating the incremental cost-effectiveness ratio (ICER), the cost and effects of the alternative (treatment B) can be compared with current practice (treatment A). Here, the ICER is used as a decision rule and is shown in the equation below [38, 68].

$$\text{ICER} = \frac{Cost^B - Cost^A}{Effect^B - Effect^A} \rightarrow \frac{\Delta Costs}{\Delta Effects}$$

To illustrate the ICER and enable a decision, the ICER can be presented in an incremental cost-effectiveness plane (CE plane; Figure 1). The y-axis represents the incremental cost – i.e., the difference between the costs of B and A – and the x-axis represents the incremental effects of B compared to A. The four quadrants represent four scenarios – the north-east quadrant presents a new treatment that is more effective and more costly. In the south-east quadrant, the new treatment dominates the old

treatment by being more effective and less costly. In the south-west quadrant, the new treatment is less effective and less costly. Finally, in the north-west quadrant, the old treatment dominates the new treatment. Although the ICER is a useful summary of the cost-effectiveness of B compared to A, the decision concerning which alternative to choose between the north-east and south-west quadrant remains unclear. An assessment of the opportunity cost must be made in order to inform decision makers, which is often referred to as the threshold. The threshold is introduced to help determine whether a particular ICER indicates that an intervention represents a good use of resources. The decision rule is that the ICER should reflect the size of the budget and the other opportunities available for using these resources, and any ICER below the threshold should be implemented [32, 69].

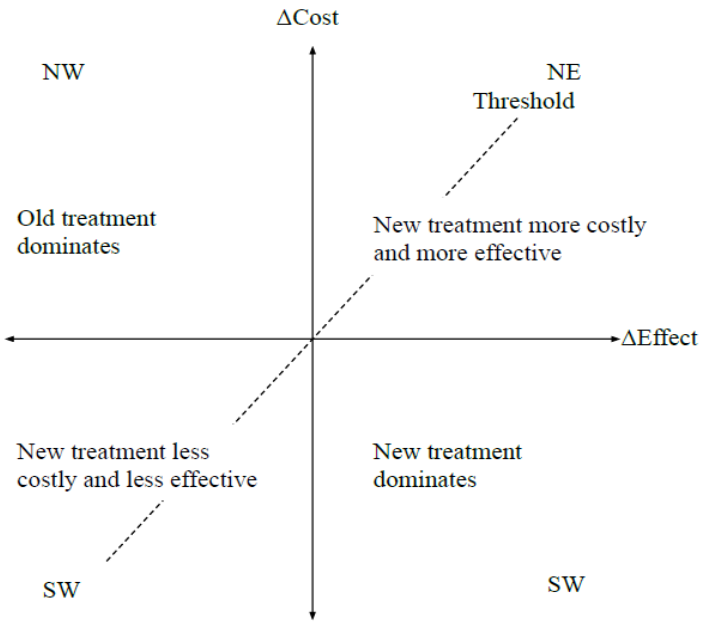


Figure 1 Cost-effectiveness plane. The y-axis represents incremental cost and the x-axis represents incremental effect [70].

CHAPTER 5. THE CAPABILITY APPROACH

This chapter explores the concepts and ideas upon which the capability approach was founded, and how it is operationalised.

5.1. WHAT IS THE CAPABILITY APPROACH?

A broader view on health measurements is becoming especially relevant when moving towards greater integration between health, public health and social care. Economic evaluation for public health interventions as well as other interventions may, therefore, require measures that go beyond health while still being applicable in health economic evaluation [31].

The capability approach is an economic theory formed in the 1980s as a normative and extra-welfarist alternative approach. Amartya Sen pioneered the approach as the theoretical and philosophical inspiration, and it was thereafter further developed by Martha Nussbaum and others [71]. The capability approach is a broad normative framework for the evaluation and assessment of an individual's well-being and social arrangements. The notion of capabilities is derived from Sen's work on functioning and capability [39]. The capability approach was initially developed with the aim of evaluating inequality as an answer to the question 'equality of what?' Sen states that as well as the question 'equality of what?' there is a parallel question of 'efficiency of what?' Thus, Sen's work can be applied to efficiency questions – including questions concerning economic evaluation of healthcare programmes. Sen's work differs from standard welfare economics, were utility is the basis for evaluating programmes or interventions. Instead, he advocates for evaluating programmes based on functioning and, ideally, capabilities. Therefore, an alternative framework for conceptualising well-being for public policy has a core idea of focusing on what people are effectively able to 'do' and 'be' in their life [72].

The capability approach has been defined as follows:

> *'The capability approach is a broad normative framework for the evaluation and assessment of individual well-being and social arrangements, the design of policies, and proposals about social change in society.'* [73]

Sen argues that evaluations and policies should focus on people's quality of life, and on removing obstacles in their lives so that they have more freedom to live the life that, upon reflection, they have reason to value [73]. The capability approach has emerged as a theoretical framework concerning well-being, freedom to achieve well-being, and all the public values in which either of these can play a role – e.g., development and social justice. The capability approach is generally understood as a conceptual framework for the assessment of individual levels of achieved well-being and freedom, the evaluation and assessment of arrangements and institutions and the design of policies and other forms of social change in society [73–75].

5.1.1. FUNCTIONING AND CAPABILITIES

The distinction between functionings and capabilities is essential for understanding how individual welfare through the capability approach is assessed. The capability approach states that freedom to achieve well-being is a matter of what people are able to do and be, and thus the kind of life they are effectively able to lead [76]. Functionings consist of 'beings and doings' and are defined by Sen as representing '*parts of the state of a person – in particular, the various things that he or she manages to do or be in leading a life*' [39]. Examples of functionings can differ from basic things, such as being healthy, having a good career, and feeling secure safe, to more complex things, such as being happy, having self-respect, and being independent. Examples of 'beings' include being well-nourished, being housed in a decent house, being educated and being part of a supportive social network, but also include negative 'beings' such as being depressed [71]. Sen states that functionings are central to a sufficient understanding of the capability approach; capability is conceptualised as a reflection of the freedom to achieve valuable functionings. Examples of 'doings' are travelling, caring for a relative or child, taking part in politics, doing charity etc. [71]. Although functionings are essential, it is the capability, the extent to which a person is able to function in a particular way, whether or not he or she chooses to do so [39], that is the particularly novel and interesting part of Sen's theory and the basis upon which he recommends evaluation [31].

Sen provides this description of capability:

> *The 'capability' of a person represents the freedom to achieve valuable human 'functionings', which can vary from such elementary things as being well nourished and avoiding escapable morbidity and mortality, to such complex achievements as having self-respect, being well-integrated with society, and so on. Capabilities thus reflect the actual freedoms that people respectively enjoy in being able to lead the kind of lives they have reasons to value.* [77]

Capabilities are combinations of the functionings that a person can achieve. According to Sen, a person's well-being, consists not only of current states and activities (functionings), but may also include the activity of choice but also of freedom or actual opportunities to function in ways alternative to current functioning. Sen defines these actual opportunities or freedoms for functioning as '*capabilities*'. Ultimately, capabilities represent a person's opportunity and ability to produce valuable outcomes, taking into account relevant personal characteristics and external factors. Capabilities are individuals' real freedom or opportunities to achieve the functionings [71, 76], and the ends of well-being, justice, and development should be conceptualised in terms of people's '*capabilities to function*' – that is, their opportunities to carry out the actions and activities that they want to do, and be the person they want to be. These doings and beings, and the freedom to engage in them, are the things that make a life valuable [73].

5.2. OPERATIONALISING THE CAPABILITY APPROACH

As the above definition indicates, the capability approach is an open-ended and underspecified framework for use in multiple areas. Open-ended means that, in general, the capability approach can be developed in a range of different directions with different aims, and it is underspecified because further specifications are needed before the approach can become useful for a specific purpose, and therefore there are several ways of closing and specifying the concept [71]. Use of the capability approach has different implications. First, there is the difference in evaluation space. Culyer defined a different evaluative space offered by the capability approach in his description of extra-welfarism, limiting the focus to one dimension (health) and to functioning rather than capability (health status rather than the freedom to pursue health improvement) [40]. The capability approach offers a potentially rich set of dimensions for evaluation. Although Sen's theory of functionings and capabilities does not stipulate any specific functioning and capability lists, he states that different capabilities are likely to be important in different contexts. The lack of a specific list

of capabilities has been seen as a limitation by other researchers in the capabilities field [78]. Nussbaum has, on that note, developed a comprehensive list of ten '*central human capabilities*' – life, bodily health, bodily integrity, senses, emotions, practical reason, affiliation, other species, play and control – for which she believes that all humans are morally entitled [78]. She states that the list is formed at the abstract level and should be translated into implementation at the local level, and that the list is *'humble and open-ended'*. Nussbaum's critics state that the list lacks legitimacy and consensus, since Nussbaum has no right to speak on behalf of the people to whom the list would apply [79]. There is growing interest in operationalising the capability approach for use in health economic evaluations and to measure QoL in a broader manner of well-being. In the following sections, different attempts to operationalise the capability approach will be described and analysed.

5.2.1. CAPABILITY AND THE QALY

In 2005, Cookson made an attempt to incorporate the capability approach within a health economic evaluation format [80]. He recommends continuing use of the QALY but suggests reshaping the current QALY as a capability-set, representing a '*capability QALY*' by proposing a re-interpretation of the QALY as: '*a cardinal and interpersonally comparable index of value of the individual's capability-set*', thus moving beyond the consensus of a '*health QALY*'. He also argues that the QALY outcome is a feasible option for assessing health interventions through the capability approach. He believes that direct estimation and valuation of capability-sets are not feasible, and he also rejects alternative preference-based measures used in public policy evaluation, such as WTP, as being '*inadequate*' for capturing capability as intended by Sen and others. Cookson refers to the use of his capability QALY as a measure of: '*capability efficiency alone (i.e. maximising the aggregate value of individual capability-sets, ignoring equity considerations)*' [80]. He believes that the capability QALY can be used as an alternative to the health QALY for a number of reasons. Firstly, the capability QALY captures health and non-health within the same measure. Secondly, the 'process of care' can be captured by shifting to a focus on non-health functionings within capability-sets. Finally, his approach can also account for the value that different people attach to achieving the same level of functioning. He recognises the need for incorporating broader non-health functionings within the QALY, but he thinks that the EQ-5D-3L dimension of '*usual activities*' fulfils this role [80]. This is not, however, in line with Sen's thoughts on freedom to achieve, but it still refers merely to functionings. Re-interpretation of the QALY as a capability measure is criticised for not attempting to measure capabilities, but instead focusing on the assessment of functionings used to calculate QALYs [31]. This appears to be a misinterpretation of what the capability approach tries to encompass. Adopting the capability approach should provide a more encompassing basis for evaluation beyond that involved within the QALY measure. For example, Coast et al. [75] suggest that a broader evaluative space based on capabilities would capture non-health benefits for interventions like those introduced in public health.

5.2.2. CAPABILITY AS A MEASURING INSTRUMENT

A number of researchers have attempted to develop questionnaires representing the capability approach, generally focusing on the use of questionnaires to assess capability and more closely for use in health economic evaluations, and in decision-making in healthcare resource allocation [79]. Three groups of questionnaires have been developed based on the capability approach: the OxCAP (Oxford Capability Questionnaire Survey) measures, ASCOT (adult social care outcome toolkit) and ICECAP capability questionnaires. Some measures were designed to explicitly measure capabilities, where others 'adopted' the capability approach as a framework.

The OxCAP instruments are a group of measures developed from the theoretical work of Martha Nussbaum and her list of ten central human capabilities (Table 2). The OCAP was the first attempt to measure capabilities directly, and it was developed by Anand and colleagues [81], who proposed that '*capability indicators*' could be created from Nussbaum's list. Capabilities are derived through questions of individuals' achieved functionings. The approach has 64 indicators of capabilities, thus limiting its usability, particularly on a wide scale [82]. Lorgelly et al. refined the OCAP into the OCAP-18 (18-item capability questionnaire for public health), not only for further development but also to validate the questionnaire in order that it could be used for evaluating interventions in public health. Using a mixed-methods approach, they reduced the number of attributes to 18. Additionally, some questions were re-worded so that the capability of an individual, and not their functioning levels, was captured. To develop the measure for use in health economic evaluations, the OCAP is provided with an index score. All of the 18 questions in the OCAP-18 hold equal weight, and the responses are coded on a 0–1 scale [82, 83]. The latest questionnaire, the OxCAP-MH (Oxford Capability Measure for Mental Health) is a self-reported 16-item questionnaire, where attributes are rated on a 1–5 scale, with scores ranging from 16–80. Some of the questions in the OxCap-MH were re-worded, but the intention is to capture the same capability principle as the OCAP-18 [84]. It is a refinement of the OCAP for use in mental health contexts. It covers individual well-being, including overall health, enjoying social and recreational activities, losing sleep over worry, friendship and support, having suitable accommodation, feeling safe, the likelihood of discrimination and assault, freedom of personal and artistic expression, appreciation of nature, self-determination and access to interesting activities or employment.

Table 2 The OxCAP family of questionnaires' attributes based on the 10 central human capabilities developed by Nussbaum

Nussbaum's capability list [85]	**OCAP** [81]	**OCAP-18** [83]	**OxCAP-MH** [84]
Life	Life expectancy	Life expectancy	Life expectancy
Bodily health	Health limits activities Reproductive health Adequately nourished Adequate shelter	Health limits activities Adequate shelter	Adequate shelter
Bodily integrity	Safe during day Safe during night Previous violent assault Future violent assault Past sexual assault Future sexual assault Past domestic violence Future domestic violence Sexual satisfaction Reproduction choice	Safe walking alone near your home Future assault (any)	Safe walking alone near your home Future assault (any)
Senses, imagination, thought	Education Uses imagination Political expression Exercise religion Enjoys activities	Political and religious expression Uses imagination	Political and religious expression Uses imagination Access to interesting activities (or employment)
Emotions	Make friends Family love Expresses feelings Lost Sleep Under Strain	Enjoy love and friendship of family and friends Lost sleep	Enjoy love and friendship of family and friends Lost sleep

Practical reason	Concept of good life Plan of Life Evaluates Life Useful role	Free to decide how to live life	Free to decide how to live life
Affiliation	Respects others Takes holidays Meets friends Thinks of others Feels worthless Past Discrimination Future Discrimination	Respect others Able to meet people socially Likelihood of discrimination outside of work	Respect others Able to meet people socially Likelihood of discrimination
Other species	Appreciates plants, animals, nature	Able to appreciate plants, animals, nature	Able to appreciate plants, animals, nature
Play	Enjoy recreation	Ability to enjoy recreation	Ability to enjoy recreation
Control over one's life	Participate in politics Owns home Past and Future Discrimination (work) Expect stop and search Skills used at work Useful role at work Relate to colleagues Respected by colleagues	Participate in local decisions Owns home Current or future discrimination within work	Participate in local decisions Owns home

The ASCOT measure is another collection of instruments for measuring service user and carer outcomes across social care. The questionnaires are developed to capture capability for use in health economic evaluations and additionally aims to measure social care-related quality of life (SCRQoL), which is believed to be applicable for measuring a social care QALY (SC-QALY) and make comparisons between health-related QALY interventions. The ASCOT questionnaire for service users has evolved through four versions to the present version, which attempts to account for Sen's capability theory within the latest version of the questionnaire's development. The questionnaires attributes are personal cleanliness and comfort, food and drink, safety, clean and comfortable accommodation, social participation and involvement, control over daily life, occupation, and dignity [86, 87]. The ASCOT developed a preference weighting of states, such that '1' represents the ideal state of SCRQOL and '0' represents a state equivalent to being dead. Valuation exercises were conducted through a combination of TTO and best-worst scaling (BWS), a type of DCE. The final value set implemented allowed for the calculation of an SC-QALY, which could range from -0.19 to 1 [86]. Though the ASCOT measure make explicit reference to the capability approach, the measure seems to focus more on assessing achieved functioning rather than capabilities[53].

Lastly, the ICECAP is a family of capability questionnaires, with the ICECAP-A being the type of capability analysis operationalised in this thesis and described and discussed in detail in Chapter 6.

CHAPTER 6. THE ICECAP MEASURES

This chapter will describe and discuss the ICECAP family of questionnaires, with detailed examination of the ICECAP-A questionnaire.

6.1. THE ICECAP QUESTIONNAIRES

The ICECAP questionnaires are all designed to measure a particular set of capabilities related to the ability to achieve valuable functionings in life. They provide a generic measure of capability well-being for use in the economic evaluation of, e.g., public healthcare and social care interventions, where a broader aim is to be explored. The ICECAP family consists of four measures, all developed by qualitative methods: ICECAP-A (Adult), ICECAP-O (Older), ICECAP-SCM (Supportive Care Measure) and ICECAP-CPM (Close Person Measure). There is also a questionnaire for kids, which is under development. Thus, the ICECAP 'family' represents the entire 'life cycle'. The ICECAP-SCM and CPM are developed as tools for use in an 'end of life' setting, from the patients' and relatives' perspective. In Table 3, the attributes in each questionnaire are presented [10, 88–90].

Table 3 The ICECAP family and the individual attributes in each questionnaire.

ICECAP-A [10]	**ICECAP-O** [88]	**ICECAP-SCM**[89]	**ICECAP-CPM** [90]
Stability	Attachment	Choice	Good communication with services
Attachment	Security	Love and affection	Privacy and space to be with loved ones
Autonomy	Role	Physical suffering	Emotional support
Achievement	Enjoyment	Emotional suffering	Practical support
Enjoyment	Control	Dignity	Being able to prepare and cope
		Being supported	Being free from emotional distress related to the condition of the decedent
		Preparation	

6.2. THE ICECAP-A

The ICECAP-A represents a generic capability measure for the general adult population, for use in economic evaluation. [91]. The questionnaire is designed to capture capability across five attributes, each with four response levels ranging from full capability (level 4) to no capability (level 1), as shown in Figure 2. The five attributes were identified using qualitative methods in terms of semi-structured interviews among members of the UK population identifying capabilities that are important to people, and they cover an individual's capability and freedom to have stability, attachment, autonomy, achievement, and enjoyment in their life [10, 92].

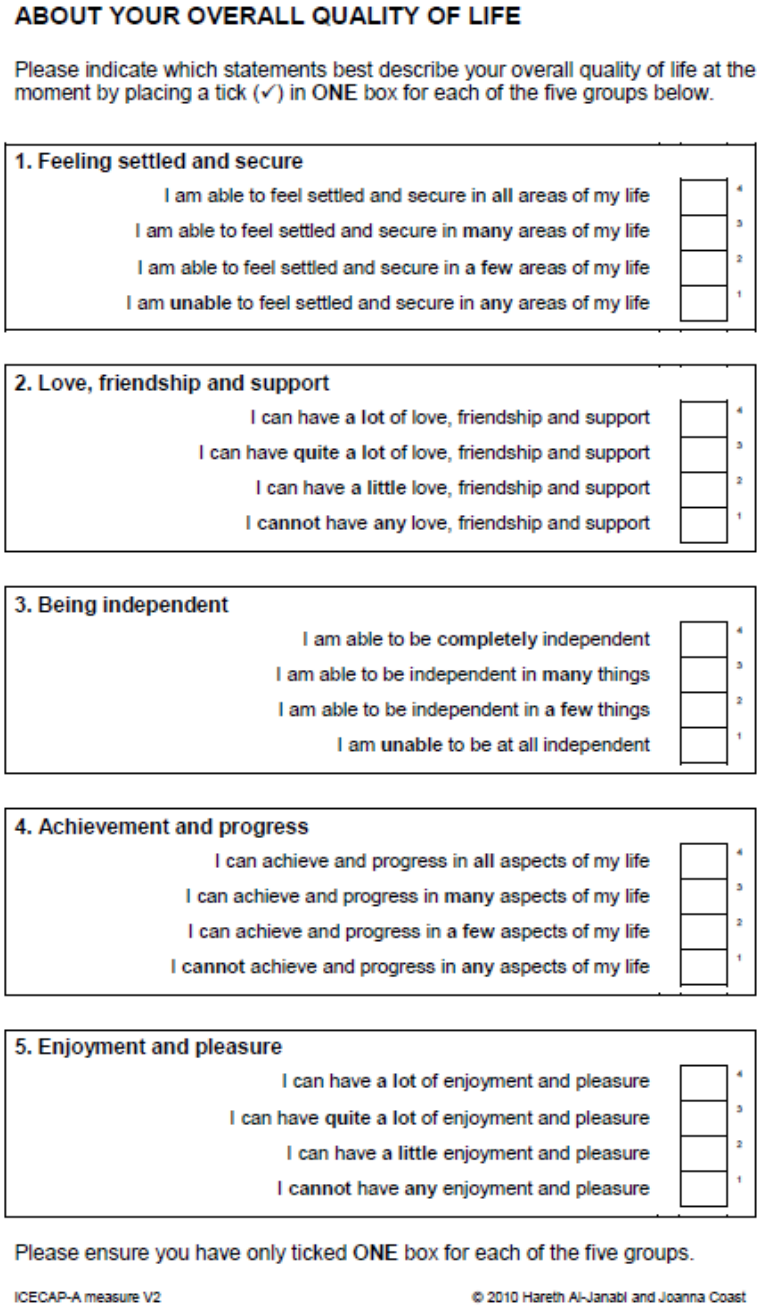
Stability refers to 'the ability to feel settled and secure' and to have stability in one's life regarding work and friends. Stability is affected positively by consistent friendship, work, and secure income, and negatively by factors such as unemployment, crime and reduced health.

Attachment refers to 'the ability to have love, friendship, and support', and social contact, which involves being close to people, feeling affection, and having a sense of belonging. The ability to feel attached is strongly related to the presence of family and friends.

Autonomy refers to 'the ability to be independent', look out for yourself, and have the freedom to be your 'own person'. Homeownership and self-employment are associated with high autonomy, while reduced health limits autonomy.

Achievement refers to 'the ability to achieve and progress', move forward in life, look back with satisfaction and attain goals. Achievement is strongly related to opportunities to be successful at work, to own things and have a family.

Enjoyment refers to 'the ability to have enjoyment and pleasure' and to experience the little joys in life. Enjoyment tends to be generated by the presence of family and friends. Limiting factors for enjoyment are feeling down, being in pain, having financial difficulties and experiencing poor health [10, 93].



ABOUT YOUR OVERALL QUALITY OF LIFE

Please indicate which statements best describe your overall quality of life at the moment by placing a tick (✓) in **ONE** box for each of the five groups below.

1. Feeling settled and secure

I am able to feel settled and secure in **all** areas of my life	4
I am able to feel settled and secure in **many** areas of my life	3
I am able to feel settled and secure in **a few** areas of my life	2
I am **unable** to feel settled and secure in **any** areas of my life	1

2. Love, friendship and support

I can have **a lot** of love, friendship and support	4
I can have **quite a lot** of love, friendship and support	3
I can have **a little** love, friendship and support	2
I **cannot** have **any** love, friendship and support	1

3. Being independent

I am able to be **completely** independent	4
I am able to be independent in **many** things	3
I am able to be independent in **a few** things	2
I am **unable** to be at all independent	1

4. Achievement and progress

I can achieve and progress in **all** aspects of my life	4
I can achieve and progress in **many** aspects of my life	3
I can achieve and progress in **a few** aspects of my life	2
I **cannot** achieve and progress in **any** aspects of my life	1

5. Enjoyment and pleasure

I can have **a lot** of enjoyment and pleasure	4
I can have **quite a lot** of enjoyment and pleasure	3
I can have **a little** enjoyment and pleasure	2
I **cannot** have **any** enjoyment and pleasure	1

Please ensure you have only ticked **ONE** box for each of the five groups.

ICECAP-A measure V2

Figure 2 The original ICECAP-A questionnaire, with five attributes, each with four levels of answers [10].

The ICECAP-A attributes have a set of preference-based tariff scores. The tariff scores are developed by Flynn et al. [94], using BWS approach, a method to measure preferences. This method of valuation is applied because it does not necessarily rely on individual preferences, because individuals are not directly asked to choose between two different scenarios. In the BWS approach, respondents are presented with scenarios and, for the ICECAP-A, asked to state their most and least favoured attribute from the five options presented to them [94–96]. In the case of ICECAP-A respondents were interviewed and presented with a set of hypothetical scenarios with different best–worst options representing all five attributes of the ICECAP-A. Respondents were then asked to choose within each profile which attribute is best and which is worst given the hypothetical scenario. Respondents could be asked to choose the best and worst capability states when the ICECAP-A attributes of stability is at the highest level, attachment and autonomy are at the lowest levels, and achievement and enjoyment are at their second-lowest levels. The estimated capability values are then a function of the choice frequencies, based on scale-adjusted latent class estimates. The study by Flynn et al. showed that all five attributes are of importance, especially the values for stability and attachment, which are somewhat stronger than the remaining three. The individual responses are scored on a 0–1 scale (tariff), and ICECAP-A values are anchored to the 'no capability' state, which is the zero point, with 1 being 'full capability'. A zero tariff score is not anchored as dead, as a QALY score is. However, the possible of interpret it in this way is argued; hence, 'no capabilities' provides a meaningful lower anchor [10, 92, 94].

A weighted tariff score for an overall state can be calculated simply by summing the values from each level across the individual attributes. For example, a tariff score for levels 43221 based on Table 4 would be calculated as follows:

$$0.222 + 0.189 + 0.084 + 0.091 - 0.003 = 0.583 \text{ weighted tariff score}$$

Table 4 The ICECAP-A weighted tariff scores, developed through BWS methods.

	Stability	**Attachment**	**Autonomy**	**Achievement**	**Enjoyment**
Level 4	0.222	0.228	0.188	0.181	0.181
Level 3	0.191	0.189	0.156	0.159	0.154
Level 2	0.101	0.096	0.084	0.091	0.069
Level 1	−0.001	−0.024	0.006	0.021	−0.003

CHAPTER 7. METHODOLOGY OF RELIABILITY AND VALIDITY

This chapter will describe the methods used to investigate the psychometric properties – reliability, validity and responsiveness – of the Danish ICECAP-A. The chapter closes with a description and discussion of the development and testing of the Danish ICECAP-A in terms of Papers I and II.

7.1. PSYCHOMETRIC PROPERTIES OF INSTRUMENTS

The ability of a measure to assist decision-making relies on psychometric properties, such as capturing the burden of disease or effect of a treatment. Reliability, validity, and responsiveness are essential psychometric properties for any measurement. Evidence for the reliability and validity of measurement falls along a range from no evaluation to full evaluation for a study population. The description of both reliability and validity as psychometric indicators are more accurately 'continuous' rather than 'dichotomous'. Therefore, concluding that a measuring instrument is completely 'reliable' or 'valid' is inaccurate. On the contrary, reliability, e.g., refers to the results attained with an evaluation instrument and not to the instrument itself. Reliability is population-specific because of interaction among the measure, the population and the situation. Therefore, the focus should be on the test scores and the reliability of the test with that population. Correspondingly, stating an instrument has been 'validated' conveys no information other than that its performance or psychometric properties have been evaluated. The more evidence there is that the instrument is reliably measuring the specific measure it is supposed to be measuring, the more confidence one has in it [97, 98].

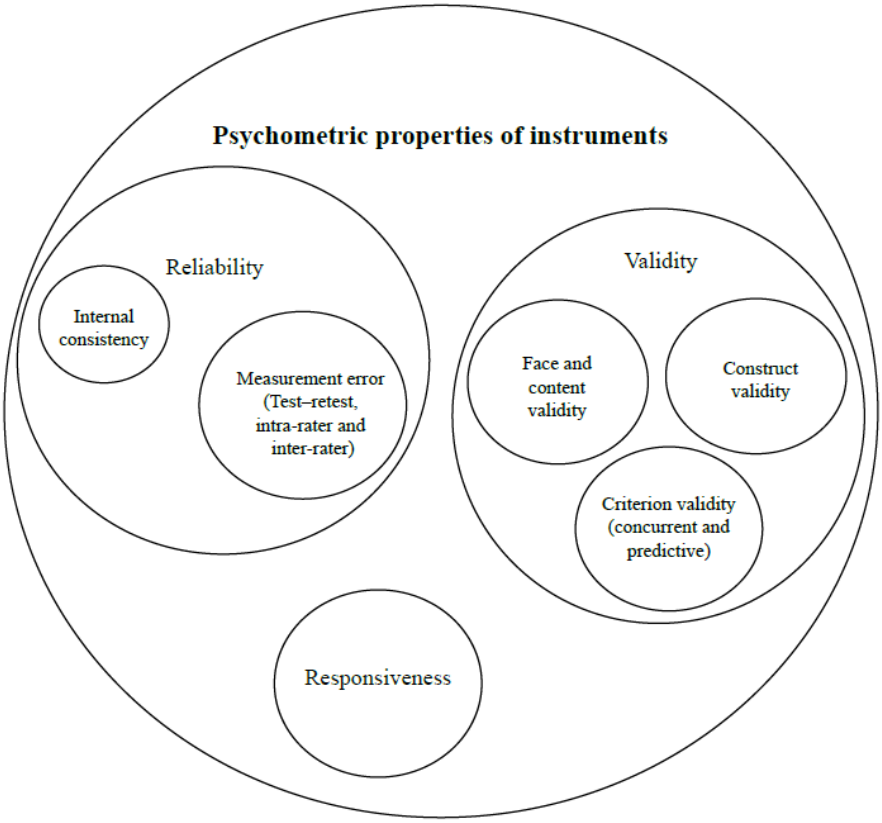


Figure 3 The different types of psychometric properties of an instrument [99]

7.2. RELIABILITY

An essential requirement of all measurements in research is that they are reliable. Reliability is defined as:

> The degree to which measurement is free from measurement error [100].

Mokkink et al. extend the definition of reliability: 'the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g. using different sets of items from the same health-related patient-reported outcomes (internal consistency); over time (test–retest); by different persons on the same occasion (inter-rater); or by the same persons (i.e. raters or responders) on different occasions (intra-rater)' [99]. Before one can obtain evidence that an instrument is measuring what it is intended to measure, it is necessary to gather

evidence that the scale is measuring something in a reproducible manner. The concept of reliability is an essential way of reflecting the number of errors, both systematic and random, innate in any measurement [98]. Reliability is measured as the consistency of measures, defined by the extent to which a measurement yields the same result on repeated measures with a low incidence of inconsistency and error [101]. Inconsistency is present in all observations and measurements. Small, subtle variations in the measure, the appearance of the construct, and the respondents all contribute to inconsistency. Inconsistency or error reduce trust in measurement and its usefulness. In order to have trust in measurement, the degree to which a measure is compromised by error must be assessed [98].

A fundamental assumption of reliability is that every observed score consists of two components: a true score (the one to be measured and that is unknown) and a random measuring error as a possible result of inaccuracies in the instrument. If the error is small, the observations are reliable, and vice versa. What is essential to know is the size of the error in relation to the true value. Reliability will increase as true variation increases and error variance decreases. Based on the assumptions of true scores, random error and observed scores, we can determine the reliability coefficient, which expresses the proportion of total variance due to true difference between subjects [98]. Thus, the formal definition of reliability is:

$$\text{Reliability} = \frac{Subject\ Variability}{Subject\ Variability + Measurement\ Error}$$

Reliability is estimated as a ratio of the variability between individuals to the total variability in the scores – thus, reliability is a measure of the proportion of the variability in scores, which is due to true differences between individuals. Subject variability will always be less than subject variability + measurement error. Therefore, reliability will vary between 0 (no reliability) and 1 (complete reliability) [98].

Reliability is used as a general term for agreement, but also as a more technical and specific term for the correlation and consistency between sets of values – e.g., two or more assessments of a number of observations. Therefore, there is an essential difference between reliable agreement and reliable consistency, whereby an instrument can have high consistency but poor agreement. Such a situation arises, for example, due to systematic biases in one observer's assessment so that it is consistently different from that of another observer. However, the correlation will be perfect (1), but the agreement is not [98].

7.3. METHODS FOR RELIABILITY ASSESSMENT

When investigating reliability, one distinguishes between four concepts: test–retest, internal consistency, and inter-rater and intra-rater reliability. The methods used in Paper I will be described in detail, and the remainder is summarised in Table 6.

7.3.1. TEST–RETEST

Stability refers to the consistency over time and is investigated with the test–retest method. This involves administering a measure to a sample once and then administering it again on another occasion. The correlation is a measure of the strength of the relationship between two variables, and the expectation is for a high correlation between observations one and two, with a low correlation indicating an unreliable and unstable measure. The problem with the test–retest method is that the first observation at time 1 can influence the observation at time 2, suggesting greater consistency than is indeed the case. Secondly, events, such as bad health, may influence consistency. The solution is to balance the time horizon between observations 1 and 2. There are no fixed rules for the period between the two observation points, and therefore this can vary. A period of two weeks is often used, because this is believed to be long enough for the respondent or observer to have forgotten their first answer or observation, and short enough for no changes in their life to have occurred. A low test–retest correlation over a long time is therefore not necessarily an expression of low reliability but an expression of the instrument being sensitive to changes. There are three main indications of low values of a test–retest. First, the test may be reliable but the phenomenon may have changed over time. Second, the scale itself may be unreliable. Third, taking a test on one occasion may affect people's response on the second administration, because they have been introduced to the phenomenon, or may be prompted to think about it more, and therefore the test is 'reactive' [98].

The different methods for analysing correlation and agreement are: correlation coefficients, the intra-class correlation coefficient (ICC), Cohen's kappa coefficients, and Bland–Altman plots. To reflect both the degree of both correlation and agreement between measures ICC can be applied. The ICC can be calculated using a one-way random model, a two-way random model or a two-way mixed model between baseline and the two-week follow-up, depending on the data [102]. Cohen's kappa coefficient is a chance-adjusted agreement coefficient, with a non-weighted and weighted version the weighted version accounting for that inconsistent responses could vary in their level of inconsistency. Cohen's kappa coefficients, can take any value from −1 to +1, negative values indicate that the observed agreement is less than that expected from chance alone, a value of 0 indicates exact chance agreement, and positive values indicate that the observed agreement is higher than that expected from chance. Values ranging from 0.41–0.60 are considered moderate; values from 0.61–0.80 indicate substantial agreement and values from 0.81–1 stand for almost perfect agreement (Table 5) [98, 103].

Table 5 Interpretation of kappa and intra-class correlation values.

Kappa		**Intra-class correlation**	
Value of κ	Interpretation	Value of ICC	Interpretation
<0.20	Poor	<0.50	Poor
0.21–0.40	Fair	0.50–0.75	Moderate
0.41–0.60	Moderate	0.75–0.90	Good
0.61–0.80	Good	<0.90	Excellent
>0.80	Very good		

A Bland–Altman plot and limits of agreement is an analysis of agreement between two measures and the measurement errors. Bland–Altman limits of agreement examine the absolute reliability with bias and upper and lower limits of agreement (LoA), which calculates the difference in a score from baseline to follow-up for each respondent and the 95% limits of agreement of the mean difference for the whole group. The 95% LoA is estimated by the mean difference ±1.96 standard deviations of the differences. Bland and Altman recommend that 95% of the differences between measurement at baseline and follow-up should lie within the limits of agreement. This is visualised in a Bland–Altman plot, where the individual differences are plotted against the mean of the baseline to follow-up [100, 104].

Table 6 Reliability is the extent to which the outcomes are consistent when the experiment is repeated more than once. Investigated with different types and methods.

Reliability type	**Description**	**Assessment**
Test–retest	Measures the stability of a test over time. Includes intra-rater reliability, which reflects the variation of data measured by one rater across two or more trials	Conducts the same test on the same group of raters at two different time points. Then calculates the correlation between the two sets of results using, e.g., Pearson's coefficient, ICC, Cohen's kappa or a Bland–Altman plot
Inter-rater	Assesses the degree of agreement between two or more different raters and their appraisal of the same group of subjects	Observation and calculation of the correlation between their different sets of results.
Internal	Assesses the degree to which a measure is consistent within itself and the extent to which all parts of the test contribute equally to what is being measured.	Tested by the split-half method. Cronbach's alpha calculates the average of all possible split-half reliability coefficients. Alpha varies between 1 (perfect) and 0 (no internal reliability). 0.80 is acceptable.

7.4. VALIDITY

Validity is the accuracy of a measurement, and whether or not an indicator really measures the concept it is intended to measure. There are several types of validity; here, face validity and construct validity will be discussed in detail, with the remaining types briefly described in Table 7.

7.4.1. FACE VALIDITY

Face validity is established by asking respondents whether or not the measurement seems to be getting at the concept that is the focus of attention. It is an essential intuitive process – does it look reasonable? Do the items appear, on the surface, to be measuring what they are actually supposed to measure? Good face validity leads to increased motivation, attracts potential respondents, increases satisfaction among respondents, encourages acceptance of the results, and improves public relations.

7.4.2. CONSTRUCT VALIDITY

Construct validity is concerned with assessing the extent to which the scores of an instrument correlate with other hypothesised measures or indicators of the construct of interest – e.g., health or capabilities. It is tested empirically and can be assessed by considering the degree to which an expected relationship between a measure and other factors is confirmed. Best-practice guidance on psychometric analyses highlights the importance of an a priori statement of hypotheses regarding the anticipated relationship between the constructs explored [105]. There are two main approaches to examine construct validity: one is to examine whether the measure can differentiate between groups believed to differ – e.g., in terms of their health – while the other is referred to as convergent validity, the extent to which the measure correlates with another measure of the construct [98, 106].

Table 7 Validity is the extent to which the instruments that are used in the experiment measure exactly what you expect them to measure. Investigated with different types and methods.

Validity types	Description	Assessment
Face and content	The extent to which the test appears to test what it aims to test. The extent to which the items of an instrument are appropriate for the dimensions being measured, and the degree to which the items cover all important aspects of the whole dimension.	Assessed by observational appraisal and qualitative methods
Construct	The extent to which the test relates to the underlying theoretical concepts	Assessed by correlation coefficients
Criterion (concurrent and predictive)	The relationship to other measures. The extent to which the measure relates to an existing similar measure. The extent to which the test predicts later performance on a related criterion.	Assessed by correlation coefficients

7.4.3. RESPONSIVENESS – THE MEASUREMENT OF CHANGE

For outcome measures to be useful in healthcare and social care interventions, it is essential that they are able to detect meaningful changes. There are two core ideas in the assessment of evaluative instruments: sensitivity to change, and responsiveness. Sensitivity to change refers to the ability of instruments to measure change statistically. Responsiveness addresses detection of the clinically relevant change. Thus, it refers to an instrument's ability to measure meaningful or essential change – for example, anchor-based and distribution-based approaches. The anchor-based method is sample-independent and examines the relationship with an anchor, such as a QoL measure, to explain the meaning of a particular degree of change. The anchors can either be cross-sectional or longitudinal. Anchor-based analysis aims to assess whether scores on the target measure change in an anticipated way, as indicated by changes in the scores on the anchor [107, 108]. The distribution-based method uses the effect size of the difference between groups to measure variability. There are two main effect size statistics: standard effect size and the standardised response mean (SRM). The effect size indicates the relative size of the 'signal' in comparison with the underlying 'noise' in the data. A common assumption is that for a given health change, the measure with the larger effect size is the better measure. However, when the purpose is to compare the size of change between treatments, it is the value of

change that matters. The standard effect size is calculated by dividing the change between baseline and follow-up by the standard deviation of the baseline scores. The SRM is calculated by dividing the change between baseline and follow-up with the standard deviation of this change. Effect size does not indicate the value or importance of a change. For an economic measure, responsiveness is whether or not the descriptive system reflects a change in health in order that it could be valued [98, 108]. Using Cohen's rule, correlations are considered strong when the coefficients are >0.50, moderate when >0.30, and weak when <0.30 [109].

7.5. PSYCHOMETRIC PROPERTIES OF ICECAP-A

There is increasing evidence regarding the psychometric properties of the ICECAP-A[110]. Research is available on the reliability [92, 111], content validity [112], construct validity [12, 13, 93, 113–119], feasibility [120, 121] and responsiveness [12, 93, 112, 117, 122, 123] measured in various populations. So far, most evidence relates to the original UK version. The questionnaire is translated into Chinese, Dutch, German, Italian, Persian, Welsh and Danish [92, 124, 125]. Seven published studies have assessed the construct validity of the ICECAP-A, using Pearson's or Spearman rank correlation coefficients. Furthermore, they compared the ICECAP-A with a variety of HrQoL instruments such as EQ-5D, 15D, AQoL-8D, HUI3, and SF-6D in different populations – e.g., healthy general populations and ill and chronically ill populations. There is variation among the studies regarding the correlation measures used (ranging from values of 0.31 to 0.80), the instruments compared, the characteristics of the population, and the number of respondents. Hence, it is difficult to conclude on the comparison of ICECAP-A with other outcome measures, or to conduct statistical pooling of the results. Most often, a capability instrument's ability to measure changes is reported to be higher than in the case of HrQoL measures. The literature shows that a capability instrument captures changes related to the wider meaning of health more efficiently compared with the EQ-5D instrument. There is strong evidence for all capabilities used in ICECAP-A and general health, with the exception of the attachment attribute [13, 84, 110, 123, 126].

7.6. PSYCHOMETRIC PROPERTIES OF THE DANISH ICECAP-A

To operationalise the ICECAP-A measure in a Danish setting, an investigation of some of the psychometric properties was necessary. Two scientific papers – Papers I and II – performed this investigation. Their overall aim was to translate the original ICECAP-A into Danish and to investigate its reliability, construct validity and responsiveness.

7.6.1. TEST–RETEST OF THE DANISH ICECAP-A

The aim of Paper I was to translate the original ICECAP-A questionnaire into Danish and to investigate the test–retest reliability. The original English ICECAP-A was translated into Danish by forward-backwards translation using the guidelines by Beaton et al. [127]. To establish face validity, a pilot test was undertaken. The purpose was to investigate whether the translated questionnaire appeared, on the surface, to be relevant to the respondents, and to examine their willingness and ability to answer, and possible doubts about the meaning of questions [92].

Data came from a web-based study conducted by the survey agency EPINION in December 2017 with 804 participants at baseline, aged 18 years or older. Respondents were recruited among EPINION online panel members representing the general Danish population. Respondents completed the ICECAP-A questionnaire on two occasions, two weeks apart. A total of 332 respondents participated at both time points. Data concerning agreement was analysed with ICC and Bland–Altman plot with limits of agreement. The overall and item consistency was investigated by weighted kappa statistics from baseline to two-week follow-up. Logistic regression was used to study the effect of the socio-demographic characteristics, with inconsistent responses as the dependent dummy variable [92].

The results of the study show that baseline ICECAP-A preference-based tariff score was 0.84, and at follow-up, was 0.83. The preference weights used came from the UK value set [128]. The ICC was 0.86 (95% CI 0.826–0.884), and limits of agreement were 0.164 and −0.151. The kappa coefficient ranged from 45–65%, between random and perfect agreement. The logistic regression used to analyse inconsistent responses showed no significant association between the overall index score and sociodemographic characteristics, and no clear pattern was found concerning the individual item inconsistency. The test–retest reliability results of the Danish ICECAP-A capability measure suggest had good test–retest reliability in terms of ICC and moderate agreement for each item, using the weighted kappa when tested in the general population. The moderate agreement could be explained by the respondents defined as outliers in the Bland–Altman plot. Outliers were the respondents that changed their answers more than one level – for example, they answered at level 4 at baseline but at level 2 at follow-up. These changes result in lower kappa coefficients because of the use of weighted kappa statistics. The use of ICC, a Bland–Altman plot,

and weighted kappa statistics provided different evidence about the test–retest reliability of ICECAP-A, which gave a better picture of its reliability [92].

7.6.2. CONSTRUCT VALIDITY AND RESPONSIVENESS OF THE DANISH ICECAP-A

Paper II aimed to provide the first assessment of construct validity in patients with CVD, COPD and diabetes, and to assess the responsiveness of the ICECAP-A for this group in a Danish municipal rehabilitation setting. Data were collected from March 2018 to April 2019 on a routine basis from patients attending rehabilitation in the municipality of Aalborg. Sociodemographic characteristics included age, gender, cohabitation, education and socioeconomic status. All attending patients were asked to complete the ICECAP-A questionnaire and a questionnaire developed by the healthcare centre, at baseline and at a 12-week follow-up after completion of the rehabilitation programme. To assess construct validity, a priori hypotheses were developed. Based on these hypotheses, associations between sociodemographic characteristics, 'general health', a freedom dimension, and ICECAP-A were analysed through chi-squared tests and Spearman rank correlations for categorical and ordinal variables, respectively. To investigate responsiveness, the anchor-based method was used. Patients were divided into categories of 'improved', 'worsened' or 'no change' according to changes between baseline and follow-up. To quantify responsiveness, both the weighted and unweighted ICECAP-A scores' effect sizes, standardised response means and t-tests were used. Findings were explored across different age groups. Additionally, to assess the responsiveness of the individual ICECAP-A items, a response profile (frequency of participants answering each level for each item, at baseline and follow-up) was completed for the two anchors. Change in response profiles between baseline and follow-up was analysed for each item to indicate which items were the 'drivers' of change in the overall measure. A total of 155 patients answered the ICECAP-A at baseline and follow-up. Of all the hypothesised associations, 16 of 26 (62%) were in the expected direction. The expected relationships were found between ICECAP-A scores and general health and 'freedom to do things'. ICECAP-A was responsive in terms of capturing the effects on general health and the freedom to do things. Differences were found across age groups, with greater responsiveness to change in those aged under 65 years. The ES and SRM were larger in the <65 groups, and both the improved and worsened mean changes were statistically significantly different between baseline and follow-up. In the ≥65 subgroups, this was only the case with the improved group. Results concerning freedom showed small ES and SRM in both age groups, but the smallest in the ≥65 subgroups. The item-by-item analysis showed that in the group of patients reporting an improvement in general health, the largest increase was in stability, and in the patients reporting worsening of general health, the biggest decrease was in autonomy. In the group of patients reporting an improvement in 'feeling fit to do the things I want to', the increase was comparable across attributes, with increases in attachment being

the lowest, and in the patients reporting worsening in ‘feeling fit to do the things I want to’, the biggest decreases were seen in autonomy [129].

7.6.3. CONCLUDING REMARKS ABOUT PAPERS I AND II

The overall limitation of the psychometric testing of the Danish ICECAP-A lies with the selected populations. If the methods of reliability and validity were followed, the testing should have been performed on the same population. Both studies should have been on either the general population or the patients in the healthcare centre participating in the rehabilitation programme. Paper I is a mix of a pilot test in the healthcare centre and a test–retest on the general population, and Paper II is entirely on the rehabilitation population in the healthcare centre, but only in one municipality, Aalborg. This is a limitation, but this approach was taken due to resource constraints. Preferably, all test should have been performed in the healthcare centre and across more municipalities [92, 129].

Concluding this chapter, the psychometric testing of the Danish ICECAP-A showed reliability for both the index score agreement and the individual item consistency in the general population. Furthermore, the Danish ICECAP-A demonstrated potential for accurately measuring the effect of rehabilitation. The construct validity showed a positive indication and appeared to be responsive in terms of capturing the effects on general health and the ‘freedom to do things’. Hence, according to these results in this population, the Danish ICECAP-A is a reliable, valid and responsive measure for use in a Danish context and future health economic evaluations. The evidence of reliability, validity and responsiveness adds to the psychometric profile of the ICECAP-A measure, and the results provide an initial indication that the ICECAP-A may be responsive in public health research and chronically ill populations [92, 129].

CHAPTER 8. MUNICIPAL REHABILITATION

This chapter describes and discusses the setting of the thesis in terms of the organisation of the municipality of Aalborg, the data foundation used in Papers III and IV in terms of municipal data, and the use of Danish registries. The chapter also describes and discusses the municipal rehabilitation programme, the attendance rates, and the effects and healthcare utilisations presented in Papers III and IV. The municipality of Aalborg was chosen as a case study for the thesis, since it is a municipality that offers rehabilitation for all three of the chosen chronic illnesses, with high participation rates, and it is the only municipality with a centre for research–practice cooperation, and therefore has a special interest in research and a relatively high level of data quality.

8.1. MUNICIPAL REHABILITATION IN AALBORG

Denmark is divided into five regions, and further split into 98 municipalities, with a various number of cities. The municipality of Aalborg is in the North Denmark Region, with 215,312 citizens, making it the third largest municipality in terms of population. The city of Aalborg is the fourth biggest in Denmark. The municipality consists of 39 towns and rural districts, and thus the population is heterogenic [130]. In most municipalities, rehabilitation, along with several preventive offers, is organised by healthcare centres [21]. Rehabilitation programmes in the municipality of Aalborg take place at the healthcare centre, which has its primary location in Aalborg city, with buildings in a number of satellite towns in rural areas.

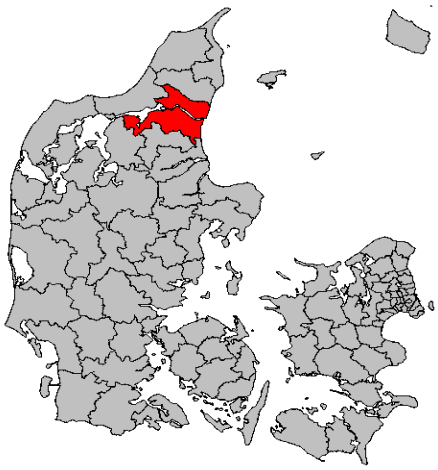


Figure 4 Denmark, with the municipality of Aalborg marked in red.

8.2. THE BASIS OF DATA

All information and data about the patients participating in a rehabilitation programme at the healthcare centre are registered in the electronic care journal KMD CARE, which contains all data at Civil Registration number (CR-number) level. The registration of information about individual patients within KMD CARE varies with the diagnosis and programme in question, and with agreement from the patient [131].

The rehabilitation programmes at the Aalborg healthcare centre included repeated measurement of the SF-36 questionnaire (from 200–-2014). The data was used as a tool to adjust the programmes and as information for management. The patients completed the questionnaire before and after rehabilitation, as well as at follow-up after 12 months. The aim was to assess the effects of rehabilitation for the individual. Due to the extensive time and analytical resource consumption, compared with the relatively low outcome of the questionnaire, the healthcare centre decided, in 2014, to cease using this questionnaire as the primary outcome measure. Instead, from the autumn of 2014 to the autumn of 2015, the municipality tested the MoEva (Monitoring and Evaluation of Patient Education) questionnaire battery (including SF12), developed by the Danish Central Regions Centre for Public Health and Quality Development (now DEFACTUM). However, after a short test period, this was considered too extensive as a primary outcome measure. Therefore, from 2018 onwards, a self-developed evaluation questionnaire – 'The Aalborg Questionnaire' (non-validated or tested for reliability) was used, serving as a follow-up to the programme and recording answers to a few simple questions about patients' experience of improvement in the parameters of quality of life and mastering physical functioning [131].

Table 8 Data collected in the municipality of Aalborg

	Content	Year
KMD CARE	Marital status, education, occupation, dietary information, smoking, alcohol, physical activity level, weight, height, BMI, blood pressure, and several physical tests.	2007–2019
Survey data	SF-36	2007–2014
The Aalborg Questionnaire	Marital status, education, occupation, physical activity level, expectations, satisfaction, and questions concerning freedom and the quality of life.	2018 to present

Registry-based research is an essential source for understanding health and diseases. Denmark has more health registers than most other countries, and it provides exceptional opportunities to perform registry-based research because of the unique CR-number available to all individuals with permanent residence. The Danish Registration System (CRS), established in 1968, comprises all people living in Denmark. This includes, for example, individuals' CR-number, gender, date and place of birth, place of residence, the identity of parents and spouse, and continuous updates on vital status. The CR-number is used as a personal identifier in all Danish national registers, enabling accurate linkage among them [132].

The ability to link data from various registries and databases via CR-numbers provides unique opportunities to find answers to health- and social-related questions. In this thesis, the data from the municipality of Aalborg is combined with the registers listed in Table 9.

Table 9 List of Danish registries used in the four papers.

Name of register	Translated name	Content
Central person registered	Civil Registration System	Information on residence and relationships of all citizens
Landspatientregisteret	National Patient Registry	Information on diagnoses and operations performed at the hospital
Sygesikringsregisteret	The National Health Insurance Service Register	Information on providers, health service, and citizens receiving primary healthcare treatment
Uddannelsesregisteret	Population's Education Register	Information on citizens' approved education
Skatteregisteret	Tax registry	Taxable income – personal and family

8.3. PARTICIPATION AND EFFECT OF REHABILITATION PROGRAMMES

Rehabilitation can be a crucial factor for people with chronic illness when it comes to regaining and maintaining functional levels, and thereby a meaningful and independent life [133], after hospitalisation or referral from general practitioners. Despite evidence of rehabilitation's positive effects, many patients fail to adhere to or complete rehabilitation programmes. Participation among CVD and COPD patients is usually low (20–50%)[134, 135]. Specifically, patients who are smokers, physically inactive, unmarried, unemployed, have a low level of education and are of lower socioeconomic status are less likely to attend [134, 136–138]. Additional predictors for dropping out and non-attendance among patients with CVD are being female, being older than 70 years, having depression, and having a low perception of illness. Similar reasons were found for COPD patients, along with medical reasons and baseline health status. No studies concerning rehabilitation and diabetes were identified. However, studies concerning self-management/self-care education were found. Here, comorbidity, lack of perceived benefit and the content of the course were associated with non-attendance [134–142].

Although men and women achieve the same effects from rehabilitation, studies have shown that dropout and non-attendance rates are higher among women [134, 143], and the reasons for dropping out and non-attendance are different between the genders [134, 138]. However, the degree to which the sociodemographic predictors are the same for men and women is less explored, and no studies concerning rehabilitation for chronic illness in general for one non-disease-specific group were found.

8.4. MUNICIPAL REHABILITATION PROGRAMME IN AALBORG

To participate in a rehabilitation programme, patients need to be referred by the hospital or their GP. In the municipality of Aalborg, the rehabilitation programme consists of a start-up interview in order to assess the need of services and the motivation of the individual, and this is followed by 8–12 weeks of disease self-management courses as well as exercise classes, dietary counselling, and an interview at the end of the programme. The start-up interview includes baseline characteristics, completion of questionnaires containing primary outcome measures, and a variety of physical tests. During 2007–2014, the SF-36 was collected. Furthermore, a clarification of the rehabilitation problem and the motivation for rehabilitation are established. Hereafter, it is determined if the programme is well-matched for the patient and which specific items of the programme are suitable. The education encompasses knowledge of the disease, dietary advice, and the importance of physical activity, smoking cessation, medicine consumption, as well as determining goals and motivation. The programmes have a particular focus on respiration for COPD patients and anxiety for CVD patients. At the end of a programme, the patients complete

outcome questionnaires and perform physical tests if possible. At three and six months after completion, the municipality telephones patients to assess their progress [27].

8.4.1. DETERMINANTS OF PARTICIPATION AND DROP-OUT IN MUNICIPAL REHABILITATION

The average lifespan is increasing, but this comes with an increase in the rate of people living with chronic illness. This calls for efficient resource use and preventive initiatives such as rehabilitation programmes. The purpose of Paper III was to investigate the participation and dropout rates of municipal rehabilitation. In continuation of this, Paper IV aimed to investigate the effects of the rehabilitation programme among the patients who participated. The data for both papers was a combination of data from the healthcare centre in Aalborg (CR-number, SF-36 data and participation status) obtained from 2007–2014 linked with data from the Danish National Registers (civil status, income, socioeconomic status, and healthcare utilisation) [144].

Paper III used multinomial logistic regression to investigate the three attendance groups. The interpretation was based on the relative risk ratio. As a subgroup analysis, gender was investigated as two separate regressions, one for females and one for males. The paper indicated that the risk of dropping out is significantly higher among patients who are employed or unemployed compared to those in retirement if patients have a low household income, and if patients have 1–4 comorbidities compared with none or more than four. The strongest predictors for non-attendance were living without a spouse, having a low level of education and being unemployed. The regression models were tested for association and showed that the three groups were significantly different from each other. Comparing the two regression models for gender for association with a seemingly unrelated estimation test indicated that there were no significant differences between females and males in the dropout group and no significant differences between genders in the non-attendance group [144].

8.4.2. THREE PERSPECTIVES ON ECONOMIC ANALYSIS OF HEALTH-RELATED QUALITY OF LIFE AND HEALTHCARE UTILISATION IN MUNICIPAL REHABILITATION PROGRAMMES – A COMPARISON OF SOCIOECONOMIC GROUPS

Paper IV aimed to evaluate the effects of a municipal rehabilitation programme. This paper reports a cost–utility analysis, with a narrow municipal payer perspective, and QALYs were used as the outcome measure obtained by crosswalking the SF-36-scores to SF-6D scores using a regression algorithm. The analysis was conducted with a 12-week time horizon consistent with the end of the rehabilitation programme. The economic analysis was a comparison between the intervention and control groups, where the intervention group comprises those who completed the rehabilitation programme and the control group consists of those with baseline data but no follow-

up because they have dropped out. Meaning zero cost and only baseline QALY scores, assuming the patients not being rehabilitated would maintain the baseline QALY score. Subgroup analysis investigated if any socioeconomic subgroup was more cost-effective than others. The direct cost of providing the rehabilitation programme consisted of staff costs, based on salaries for the staff related to the rehabilitation teams allocated to each programme (for each disease – CVD, COPD and diabetes) based on hours contributed to the programme. The mean salary represents the time spent by the wage earner, including overtime. A DID analysis was conducted in order to investigate the healthcare utilisation differences between the patients being rehabilitated and those who were referred but never completed. Logistic regression was used to elucidate the DID, and the following outcomes were analysed: outpatient visits, hospital admissions, and GP/specialist visits. The results showed that among the referred patients, 555 patients had completed SF-36 at baseline and follow-up. The baseline characteristics show that 62% of the sample was categorised as retired, 25% as employed, and 13% as unemployed / other cash benefits. The ICER for the base-case analysis suggests that the rehabilitation programme provides a cost of €19,056 per QALY gained when compared to the control group, making rehabilitation more costly and more effective. The subgroup analysis indicated that the employed are gaining the most. The DID analysis found no significant differences before and after rehabilitation [145].

8.4.3. CONCLUDING REMARKS ON PAPERS III AND IV

Overall, Papers III and IV revealed differences between participation and dropping out. Moreover, there were small differences between subgroups in terms of who gains more HrQoL than others. The risk of dropping out was significantly higher if patients were employed or unemployed (compared to those in retirement), had a low household income, and had 1–4 comorbidities compared with none or more than four. The strongest predictors for non-attendance were living without a spouse, having a low level of education and being unemployed. Paper IV showed that the employed gained the most QALYs. Linking the two papers indicate that those in higher risk of dropping out or being non-attenders are those potentially gaining the most quality of life [144, 145].

CHAPTER 9. CONTRIBUTION AND IMPLICATIONS

This thesis explores neglected aspects of municipal rehabilitation in Denmark and presents new findings such as: The measurement of non-health outcome, analysis of drop-out, and economic evaluation of municipal rehabilitation. The findings are new nationally and to a considerable extent also internationally. Part of the thesis addresses the important issue of how to address non-health related outcome measures based on capability theory. Not necessarily as a replacement of QALY but in some cases like rehabilitation as an important adjunct to QALY. There are several learning points for the municipalities in particular in regard to measurement of outcome and better costing data.

Based on the findings of Papers I and II, this thesis finds the capability approach and the ICECAP-A to be potential valid and reliable outcome measures for use in public health intervention. The main contribution is that a reliable, valid and responsive Danish version of ICECAP-A is now available. Paper II is an analysis of construct validity and responsiveness to change for the Danish translation of the ICECAP-A, and the first international investigation of responsiveness to change for any ICECAP measure for CVD, COPD and diabetes.

A unique feature of the thesis is its demonstration of different reasons for attendance, non-attendance and dropping out of municipal rehabilitation programmes. This, results in different 'profiles' and focus areas of relevance for the day-to-day work at the healthcare centre (Paper III). Furthermore, the QALY gain was investigated in Paper IV, and it was found to have small positive net effects and QALY differences across socioeconomic status, and no significant differences in healthcare utilisations.

This thesis has emphasised the need for broader non-health measurement in public health interventions nationally and internationally using municipal rehabilitation programmes as a case study. Choosing an outcome measure based on the aim of an intervention is an essential first step in a health economic evaluation. The starting point for choosing the capability approach and the ICECAP-A measure was the lack of a broad generic non-health measure for use in Denmark. The capability approach and the literature on operationalising the capability approach in public health seem like a possible option for measuring the effects of rehabilitation. In this thesis, the ICECAP-A was chosen because it is a short generic and preference-based questionnaire based on the principles of the capability approach. Based on the results from Paper I, the Danish version is reliable for use in the general population. Paper II found the Danish ICECAP-A to be valid and to have demonstrable potential for accurately measuring the effect of rehabilitation. Furthermore, it appears to be responsive in terms of capturing the effects on general health and the freedom to do

things. Hereby, the thesis has contributed to the health economic research field by making the ICECAP-A available for use in Denmark and adding to the relatively small international literature and opening up a broader view on outcome measurement in, e.g., public health interventions and health economic evaluations.

In this thesis the ICECAP-A was used. However, the ICECAP-O might as well have been chosen. The study population had an average age of 65, which is the intersection between ICECAP-A and ICECAP-O, but ICECAP-A was applied since it was known that patients under the age of 65 would participate in the rehabilitation programme. The subgroup analysis in Paper II indicates better responsiveness for the group <65 years, implying that it might be preferable to have used the ICECAP-O for those >65 years. Future studies should investigate the possibility of using both questionnaires depending on age, thus facilitating further investigation of the intersection for age and whether it is age or the impact of for instance of employment that is the determining factor.

In terms of measuring the effects of rehabilitation as regards to QoL the problem is that no one has defined, in detail, the expected effects. In the municipality of Aalborg, the previous way of measuring the effects was in terms of clinical and physical measures, and HrQoL in terms of SF-36. However, the work of analysing and interpreting the results was not done until now (Paper IV). The municipality of Aalborg no longer uses the SF-36 questionnaire. This leaves strictly health (clinical) and functioning outcomes, for which data quality is often low in terms of completion rate. Most municipalities are in search of finding a replacement – possibly self-developed without looking into reliability and validity of self-developed measures.

In more general terms for the Danish Municipalities this may leave us with 98 potential ways (98 municipalities) of measuring the effects of rehabilitation, or having no effects measures at all. This is of course problematic for quality and resource use in general, and there should be more joint action and work. There is ongoing work between some of the biggest municipalities to develop a questionnaire battery of patient-reported outcomes (PROs) to enable the municipalities to compare results; however, until now, the intended purpose has been to use the forthcoming PRO questionnaire as a dialogue tool and not an outcome measurement. In the long run, the aim for all 98 municipalities should be to collect the same data and agree on joint outcome measures to document effects. In this process, it could be advantageous to make national guidelines more specific in terms of the definition of effects. The latest health agreements (2015–2018) state that more patients should be referred to rehabilitation, and that more should complete, and complete with effects. However, the status report says nothing about the effects or if any municipalities have measured effects successfully [146]. The health agreements are regional and may lead to the undesirable situation from a national point of view that we may end of with five different aims and goals across the country. This could be avoided by introducing national guidelines for public health interventions as is the case in England where

NICE (the National Institute for Health and Care Excellence) improves and secures health and social care through evidence-based guidelines. An institute such as NICE is not necessarily the right model for Denmark but as a minimum national guidelines stating aim, method and data to be collected should be considered. The right decision in a Danish setting may not be one single outcome measure, but rather a combination that covers the broad aim of the interventions in healthcare centres. Whatever outcomes measure is chosen, it needs to be deliberate, valid and reliable and useful, not only in a dialogue with the patients but also in connection with decision-making based on good outcome measurements.

Based on the stated Danish definition and aim of rehabilitation in Chapter 3, it is clear that in order to measure the effect of rehabilitation, it is necessary to consider QoL in a relatively broad sense, i.e. capturing the entire aim of rehabilitation, health, HrQoL and capabilities in terms of well-being. As illustrated in Figure 5, QoL could be interpreted as being a measure consisting of several elements: Health, HrQoL, and well-being. In the setting of rehabilitation where a specific definition of effect(s) is still lacking this means that in order to cover as broad a spectrum of QoL as possible, the healthcare centre could collect data on health in terms of functioning and health status. HrQoL could be measured with one of the generic and preference-based questionnaires – e.g., EQ-5D. Also to cover capabilities, the use of the ICECAP-A is a possibility. In Denmark, the use and recognition of QALYs are relatively new dating back to the turn of the millennium. The only HrQoL questionnaire with a Danish preference-based value-set is the EQ-5D-3L, [147] with 5L on the way. NICE has a guideline for public health interventions where, depending on the intervention, and the anticipated effects of the intervention, it is recommended that the economic analysis considers effects in terms of capabilities and well-being. Furthermore, NICE recommends that if an intervention is associated with both health- and non-health-related outcomes, both elements are presented [148]. As Figure 5 indicates, there is an overlap between health and HrQoL, and between HrQoL and capabilities. Thus, the risk of double-counting when applying well-being measures in economic evaluations requires careful consideration. In a study by Engel et al. [114], the overlap between ICECAP-A and five preference-based HrQoL measures (15D, Assessment of Quality of Life 8-dimension (AQoL-8D)), EQ-5D-5L, Health Utilities Index Mark 3 (HUI-3)), and SF-6D) was investigated. The results showed that the ICECAP-A provided additional complementary information when compared with the 15D, EQ-5D-5L, HUI-3, and SF-6D, while there was substantial overlap between the ICECAP-A and AQoL-8D [114]. Given that the ICECAP-A provides additional information and is now available as a Danish version, municipal healthcare centres should employ HrQoL and well-being measures in combination, thus facilitating a broader QoL perspective (Figure 5). A recommendation in the Danish context would be to include, e.g., the ICECAP-A questionnaire in the forthcoming PRO battery, as a supplement along with a HrQoL instrument, e.g. the EQ-5D-5L, while keeping the risk of overlap in mind.

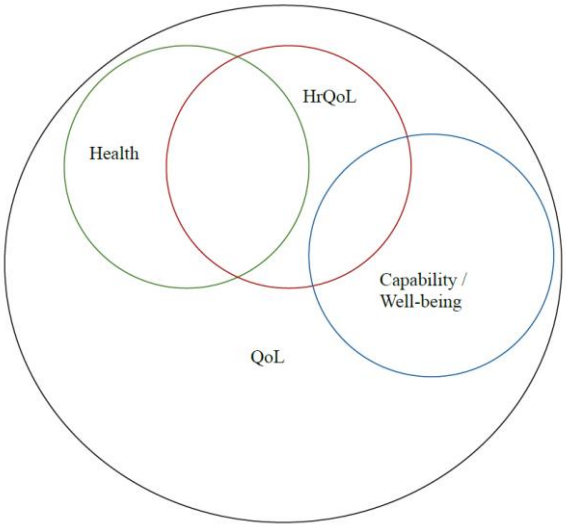


Figure 5 The aspects of quality of life.

This thesis also investigated some of the less explored aspects of municipal rehabilitation in a Danish context: participation and the HrQoL effect. In Paper III, the key finding was that there are significant differences in baseline characteristics in terms of socioeconomic status, marital status and comorbidities among the three participation groups: attenders, non-attenders, and dropouts. This analysis was possible only because of the detailed participation data in the Aalborg rehabilitation programme. Compared to attenders, non-attenders were more likely to be living alone, have a low level of education and be unemployed. Those who dropped out were more likely to be employed or unemployed / other cash benefits, and have more than one comorbidity. This may indicate that the rehabilitation programme is more appealing for those with more resources. The QALY gain was investigated in Paper IV, which found small but positive net effects. Furthermore, the study found QALY differences across socioeconomic status. Linking these results with the findings in Paper III indicates that those gaining most from rehabilitation are those who are likely to drop out – the employed and unemployed – with a QALY gain of 0.016 and 0.013, respectively, compared with the retired group, with a QALY gain of 0.011. The results should, however, be interpreted with caution, because the control group was small and was not generated through randomisation. However, there are trends indicating possible socioeconomic differences in terms of HrQoL effects.

In conclusion, the results from this thesis highlight the necessity for a broader view of outcome measurements of public health interventions, where the aim is broader than health. Based on the findings of this thesis, the recommendation is to draw up standardised guidelines for measuring effects and broaden the view on effects, here the ICECAP-A could be a considerable measure along with a HrQoL measure. Furthermore, the healthcare centre should continue to focus on the individual characteristics, possible comorbidities and personal objectives with the presented 'profiles' in mind.

LITERATURE LIST

1. The Danish Health Authority. (2018). *The Danes health - The national health profile (Danskernes sundhed - den nationale sundhedsprofil)*. Retrieved from www.sst.dk

2. Falk Hvidberg, M., Johnsen, P. S., Davidsen, M., & Ehlers, L. (2019). A Nationwide Study of Prevalence Rates and Characteristics of 199 Chronic Conditions in Denmark. *Pharmaco Economics*. https://doi.org/10.1007/s41669-019-0167-7

3. Hajat, C., & Stein, E. (2018). The global burden of multiple chronic conditions: A narrative review. *Preventive Medicine Reports*, *12*, 284–293. https://doi.org/10.1016/j.pmedr.2018.10.008

4. The Danish Health Authority. (2019). *Ældres Sundhed Og Trivsel*. Retrieved from https://www.sst.dk/da/udgivelser/2019/~/media/388EDD0F865045F1BABA4542FA83FCD2.ashx

5. Anderson, L., Oldridge, N., Thompson, D. R., Zwisler, A.-D., Rees, K., Martin, N., & Taylor, R. S. (2016). Exercise-Based Cardiac Rehabilitation for Coronary Heart Disease. *Journal of the American College of Cardiology*, *67*(1), 1–12.

6. McCarthy, B., Casey, D., Devane, D., Murphy, K., Murphy, E., & Lacasse, Y. (2015). Pulmonary rehabilitation for chronic obstructive pulmonary disease (Review). *Cochrane Database of Systematic Reviews*, (2).

7. Pedersen, B. M., & Petersen, S. R. (Eds.). (2014). *Det kommunale sundhedsvæsen*. Hans Reitzels forlag.

8. Handberg, C., & Nielsen, C. (2012). *Evaluering af rehabiliteringsindsatsen overfor borgere med kronisk sygdom i Brønderslev Kommune 2012*. Retrieved from http://www.marselisborgcentret.dk/siteassets/viden-om-rehabilitering/publikationer/publikationer-om-rehabilitering/evaluering_af_rehabiliteringsindsatsen.pdf

9. Marselisborg Centret. (n.d.). *Rehabilitering i Danmark Hvidbog om rehabiliteringsbegrebet*. Retrieved from http://www.rehabiliteringsforum.dk/contentassets/519c219e42a3454da34db60a8454de58/hvidbog.pdf

10. Al-Janabi, H., Flynn, T. N., & 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.

11. 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*.

12. Goranitis, I., Coast, J., Al-Janabi, H., Latthe, P., & Roberts, T. E. (2016). The validity and responsiveness of the ICECAP-A capability-well-being measure in women with irritative lower urinary tract symptoms. *Quality of Life Research*, *25*, 2063–2075.

13. Mitchell, P. M., Al-Janabi, H., Byford, S., Kuyken, W., Richardson, J., Iezzi, A., & Coast, J. (2017). Assessing the validity of the ICECAP-A capability measure for adults with depression. *BMC Psychiatry, 17:46*.

14. World Health Organization. (2011). *World report on disability*.

15. Ministry of Health. (2005). Forslag til Sundhedsloven. Retrieved September 16, 2019, from https://www.retsinformation.dk/Forms/R0710.aspx?id=89049

16. Pedersen, K. M. (2013). *Sundhedsøkonomi* (1st ed.). Munksgaard.

17. World Health Organization. (2012). *European Action Plan for Strengthening Public Health Capacities and Services*. Retrieved from http://www.euro.who.int/__data/assets/pdf_file/0005/171770/RC62wd12rev1-Eng.pdf?ua=1

18. 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*, 83–94. https://doi.org/10.1002/hec.3302

19. Rychetnik, L., Frommer, M., Hawe, P., Shiell, A., Rychetnik, L., Healthcare Australia, E., & Coppleson Building, V. (2002). Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health*, *56*, 119–127. https://doi.org/10.1136/jech.56.2.119

20. The Danish Ministry of health. (2017). *Healthcare in Denmark - an overview*. Retrieved from https://www.sum.dk/English/~/media/Filer - Publikationer_i_pdf/2016/Healthcare-in-dk-16-dec/Healthcare-english-V16-dec.ashx

21. Vallgårda, S., & Krasnik, A. (Eds.). (2011). *Sundhedsvæsen og sundhedspolitik* (2nd ed.). Munksgaard Danmark.

22. The Danish Ministry of health. Bekendtgørelse af sundhedsloven (2019). Retrieved from https://www.retsinformation.dk/Forms/R0710.aspx?id=210110

23. The Danish Ministry of health. (2018). *Vejledning om sundhedskoordinationsudvalg og sundhedsaftaler*. Retrieved from https://www.retsinformation.dk/Forms/R0710.aspx?id=202287

24. Region Nordjylland. (2015). Den Politiske Sundhedsaftale 2015-2018. Retrieved August 14, 2019, from https://rn.dk/sundhed/til-sundhedsfaglige-og-samarbejdspartnere/sundhedsaftalen-2015-2018/-/media/Rn_dk/Sundhed/Til-sundhedsfaglige-og-samarbejdspartnere/Sundhedsaftaler/Sundhedsaftale-2015/Status-på-sundhedsaftalen-2015-2018.ashx

25. World Health Organization. (n.d.). *World health organization -concept paper WHO Guidelines on Health-Related Rehabilitation*. Retrieved from https://www.who.int/disabilities/care/rehabilitation_guidelines_concept.pdf

26. The Danish Health Authority. (2005). *Terminologi - forebyggelse sundhedsfremme og folkesundhed*. Retrieved from http://sundhedsstyrelsen.dk/publ/publ2005/cff/termpjece/termpjece3jun05.pdf

27. Johansen, J., Rahbek, J., Møller, K., & Jensen, L. (2004). *Hvidbog om rehabiliteringsbegrebet. Rehabilitering i Danmark.* Retrieved from http://www.marselisborgcentret.dk/fileadmin/filer/Publikationer/PDF_er/Hvidbog.pdf

28. The Danish Health Authority. (2008). *Forløbsprogrammer for kronisk sygdom - Generisk model og forløbsprogram for diabetes*. Retrieved from http://www.sst.dk

29. The Danish Health Authority. (2016). *Anbefalinger for forebyggelsestilbud til borgere med kronisk sygdom*. Retrieved from https://www.sst.dk/-/media/Udgivelser/2016/Forebyggelsestilbud-borgere-med-kronisk-sygdom/Anbefalinger-for-forebyggelsestilbud-til-borgere-med-kronisk-sygdom.ashx?la=da&hash=3D90B150C5840C9E9966A3EFC192AA04EC63C1F3

30. World Health Organization. (2002). *Towards a Common Language for Functioning, Disability and Health: ICF The International Classification of Functioning, Disability and Health*.

31. Coast, J., Smith, R. D., & Lorgelly, P. (2008). Welfarism, extra-welfarism and capability: The spread of ideas in health economics. https://doi.org/10.1016/j.socscimed.2008.06.027

32. Brazier, J., Ratcliffe, J., Salomon, J. A., & Tsuchiya, A. (2007). *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford University Press.

33. Hurley, J. (2014). Welfarism and Extra-Welfarism. In *Encyclopedia of Health*

Economics (Volume 3.).

34. 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. https://doi.org/10.1016/0277-9536(94)E0109-6

35. Coast, J. (2009). Maximisation in extra-welfarism: A critique of the current position in health economics. *Social Science & Medicine*, *69*(5), 786–792. https://doi.org/10.1016/j.socscimed.2009.06.026

36. Brouwer, W. B. F., Culyer, A. J., Job, N., Van Exel, A., & Rutten, F. F. H. (2008). Welfarism vs. extra-welfarism. *Journal of Health Economics*, *27*, 325–338.

37. 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. https://doi.org/10.1016/S0277-9536(02)00101-6

38. Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes* (4th ed.). Oxford University Press.

39. Sen, A. (1993). Capability and well-being. In *M. Nussbaum & A. Sen (Eds.), Quality of Life* (pp. 30–53). Oxford: Oxford University Press.

40. Culyer, A. J. (1989). The Normative economics finance and of health provision. *Oxford Review of Economic Policy*, *5*(1), 34–58.

41. Fox-Rushby, J., & Cairns, J. (2005). *Economic Evaluation* (1st ed.). Open University Press.

42. World Health Organization. (2006). *Constitution of the world health organization*.

43. Patrick, L. D., Bush, J. W., & Chen, M. M. (2016). Toward an Operational Definition of Health. *American Sociological Association*, *14*(1), 6–23.

44. Phillips, S. M., Wójcicki, T. R., & McAuley, E. (2013). Physical activity and quality of life in older adults: An 18-month panel analysis. *Quality of Life Research*, *22*(7), 1647–1654. https://doi.org/10.1007/s11136-012-0319-z

45. The World Health Organization quality of life assessment (WHOQOL): Position paper from the World Health Organization. (1995). *Social Science & Medicine*, *41*(10), 1403–1409. https://doi.org/10.1016/0277-9536(95)00112-K

46. Cummins, R. A. (2005). Moving from the quality of life concept to a theory. *Journal*

of Intellectual Disability Research, *49*(10), 699–706. https://doi.org/10.1111/j.1365-2788.2005.00738.x

47. Felce, D., & Perry, J. (1995). Quality of life: Its definition and measurement. *Research in Developmental Disabilities*, *16*(1), 51–74. https://doi.org/10.1016/0891-4222(94)00028-8

48. Meeberg, G. A. (1993). Quality of life: a concept analysis. *Journal of Advanced Nursing*, *18*(1), 32–38. https://doi.org/10.1046/j.1365-2648.1993.18010032.x

49. Ware, J. E., & Sherbourne, C. D. (1992). *I. Conceptual Framework and Item Selection* (Vol. 30). Retrieved from https://about.jstor.org/terms

50. Brooks, R. G., Jendteg, S., Lindgren, B., Persson, U., & Björk, S. (1991). EuroQol©: health-related quality of life measurement. Results of the Swedish questionnaire exercise. *Health policy*. https://doi.org/10.1016/0168-8510(91)90142-K

51. Janssen, M. F., Birnie, E., Haagsma, J. A., & Bonsel, G. J. (2008). Comparing the standard EQ-5D three-level system with a five-level version. *Value in Health*. https://doi.org/10.1111/j.1524-4733.2007.00230.x

52. 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.

53. Keeley, T. J. H. (2014). *Capability as an outcome measure in randomised controlled trials*.

54. Weinstein, M. C., Torrance, G., & Mcguire, A. (2009). QALYs: The Basics The Facts How Are QALYs Used? *Value in Health*, *12*, S5–S9. https://doi.org/10.1111/j.1524-4733.2009.00515.x

55. Weinstein, M. C., Torrance, G., & Mcguire, A. (2009). QALYs : The Basics. *Value In Health*, *12*(1), S5–S9.

56. Whitehead, S. J., & Ali, S. (2010). Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin*, *96*, 5–21. https://doi.org/10.1093/bmb/ldq033

57. Ryan, M., Gerard, K., & A. (2008). *Using Discrete Choice Experiments to Value Health and Health Care*. (M. Ryan, K. Gerard, & M. Amaya-Amaya, Eds.). Springer.

58. De Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health Economics*, *21*(2), 145–172. https://doi.org/10.1002/hec.1697

59. Harris, J. (1987). QALYfying the value of life. *Library. Protected by copyright. on March*, *13*, 117–123. https://doi.org/10.1136/jme.13.3.117

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

61. Chisholm, D., Healey, A., Knapp, M., & Chisholm -A Healey, D. (1997). *(No Title)*. *Soc Psychiatry Psychiatr Epidemiol* (Vol. 32). Springer-Verlag.

62. 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. https://doi.org/10.1002/hec.1277

63. Payne, K., Mcallister, M., & Davies, M. L. (2013). Valuing the economic benefits of complex interventions: when maximising health is not sufficient. *Health Economics*, *22*, 258–271.

64. Kinghorn, P., Robinson, A., Smith, R. D., Kinghorn, P., Robinson, A., & Smith, R. D. (2015). Developing a Capability-Based Questionnaire for Assessing Well-Being in Patients with Chronic Pain. *Soc Indic Res*, *120*, 897–916. https://doi.org/10.1007/s11205-014-0625-7

65. Donaldson, C., Atkinson, A., Bond, J., & Wright, K. (1988). Should QALYs be programme-specific? *Journal of Health Economics*, *7*(3), 239–257. https://doi.org/10.1016/0167-6296(88)90027-6

66. Wiseman, V. (1997). Caring: The neglected health outcome? Or input? *Health Policy*, *39*(1), 43–53. https://doi.org/10.1016/S0168-8510(96)00848-2

67. Verkerk, M. A., Busschbach, J. J. V, & Karssing, E. D. (2001). Health-related quality of life research and the capability approach of Amartya Sen. *Quality of Life Research*, *10*, 49–55.

68. Ehlers, L. H., & Vestergaard, A. S. (n.d.). *Costing in health economic evaluation - Theory and practice* (1st ed.). Aalborg University Press.

69. Fox-Roshby, J., & Cairns, J. (2005). *Economic Evaluation*.

70. Van Hout, B. A., Maiwenn, A. J., Gilad, G. S., & Frans, R. H. F. (1994). Costs, effects and C/E-ratios alongside a clinical trial. *Health Economics*, *3*(5), 309–319.

71. Robeyns, I. (2017). *Wellbeing, Freedom and Social Justice - The Capability Approach Re-Examined*. Cambridge, UK: Open Book Publisher.

72. Sen, A. (1999). *Commodities and Capabilities*. Oxford India.

73. Robeyns, I. (2005). The Capability Approach: a theoretical survey. *Journal of Human Development*, *6*(1), 93–117. https://doi.org/10.1080/146498805200034266

74. Law, I., & Widdows, A. H. (n.d.). Conceptualising Health: Insights from the Capability Approach. https://doi.org/10.1007/s10728-007-0070-8

75. Coast, J., Smith, R., & Lorgelly, P. (2008). Should the capability approach be applied in Health Economics? *Health Economics*, *17*(6), 667–670. https://doi.org/10.1002/hec.1359

76. Sen, A. (1997). *Inequality Reexamined*. *Quality of Life*. Harvard University Press.

77. Sen, A. (n.d.). *Welfare, Freedom and Social Choice: a Reply*. Retrieved from https://www-jstor-org.zorac.aub.aau.dk/stable/pdf/40723937.pdf?refreqid=excelsior%3Ab27717d3589439c230f7eb7f51107771

78. Nussbaum, M. C. (2003). Capabilities as fundamental entilements: Sen and social justice. *Feminist economics*, *9:2–3*, 33–59. https://doi.org/10.1080/1354570022000077926

79. Robeyns, I. (2006). The capability approach in practice. *Journal of Political Philosophy*, *14*(3), 351–376. https://doi.org/10.1111/j.1467-9760.2006.00263.x

80. Cookson, R. (2005). QALYs and the capability approach. *Health Economics*, *14*(8), 817–829.

81. Anand, P., Hunter, G., Carter, I., Dowding, K., & Guala & Martin Van Hees, F. (2009). The Development of Capability Indicators. *Journal of Human Development and Capabilities*, *10*(1), 125–152. https://doi.org/10.1080/14649880802675366

82. Lorgelly, P. K., Lorimer, K., Fenwick, E. A. L., Briggs, A. H., & Anand, P. (2015). Operationalising the capability approach as an outcome measure in public health: The development of the OCAP-18. *Social Science and Medicine*, *142*, 68–81. https://doi.org/10.1016/j.socscimed.2015.08.002

83. Lorgelly, P. K., Lorimer, K., Fenwick, E., & Briggs, A. H. (2008). *The Capability Approach: developing an instrument for evaluating public health interventions*.

84. Simon, J., Anand, P., Gray, A., Rugkåsa, J., Yeeles, K., & Burns, T. (2013). Operationalising the capability approach for outcome measurement in mental health research. *Social Science & Medicine*, *98*, 187–196.

85. Nussbaum, M. C. (2000). *Women and human development: the capability approach*. Cambridge University Press.

86. Netten, A., Burge, P., Malley, J., Potoglou, D., Towers, A. M., Brazier, J., … Wall, B. (2012). Outcomes of social care for adults: Developing a preference-weighted measure. *Health Technology Assessment*, *16*(16), 1–165.

87. Malley, J. N., Towers, A.-M., Netten, A. P., Brazier, J. E., Forder, J. E., & Flynn, T. (2012). An assessment of the construct validity of the ASCOT measure of social care-related quality of life with older people. *Health and Quality of Life Outcomes*, *10*(21).

88. 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 and Medicine*. https://doi.org/10.1016/j.socscimed.2005.08.023

89. Sutton, E. J., & 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. https://doi.org/10.1177/0269216313489368

90. Canaway, A., Al-Janabi, H., Kinghorn, P., Bailey, C., & Coast, J. (2017). 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. https://doi.org/10.1177/0269216316650616

91. Mitchell, P. M., Venkatapuram, S., Richardson, J., Iezzi, A., & Coast, J. (2017). Are quality-adjusted life years a good proxy measure of individual capabilities? *Pharmaco Economics*, *35*(6), 637–646.

92. Holst-Kristensen, A. W., Fonager, K., & Pedersen, K. M. (2019). Test–retest reliability of ICECAP-A in the adult Danish population. *Quality of Life Research*, (29), 547–557.

93. Al-Janabi, H., Peters, T. J., Brazier, J., Bryan, S., Flynn, T. N., Clemens, S., … Coast, J. (2012). An investigation of the construct validity of the ICECAP-A capability measure. *Quality of Life Research*, *22*(7), 1831–1840.

94. Flynn, T. N., Huynh, E., Peters, T. J., Al-Janabi, H., Clemens, S., Moody, A., & Coast, J. (2015). Scoring the ICECAP-A capability instrument. Estimation of a UK genrel population tariff. *Health economics*.

95. Mühlbacher, A. C., Kaczynski, A., Zweifel, P., & Johnson, F. R. (2016). Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. *Health Economics Review*, *6*(1), 1–14. https://doi.org/10.1186/s13561-015-

0079-x

96. Louviere, J. J., Flynn, T. N., & Marley, A. A. . (2015). *Best-worst scaling, Theory method and applications* (1st ed.). Cambridge University Press.

97. Frost, M. H., Reeve, B. B., Liepa, A. M., Stauffer, J. W., & Hays, R. D. (2007). What Is Sufficient Evidence for the Reliability and Validity of Patient-Reported Outcome Measures? *Value in Health*, *10*.

98. Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales, a practical guide to their development and use* (5th ed.). Oxford University Press.

99. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., … de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*. https://doi.org/10.1016/j.jclinepi.2010.02.006

100. De Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in Medicine: A Practical Guide*. Cambridge University Press.

101. Carmines, E. G., & Zeller, R. A. (1980). *Reliability and Validity Assessment*. SAGE university press.

102. Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163.

103. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

104. Bland M., J., & Altman G, D. (1999). Measuring agreement in method comparison studies. *Statistical Methods in medical Research*, *8*(99), 135–160.

105. Kane, M. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement*, *38*(4), 319–342.

106. Brazier, J., & Deverill, M. (1999). A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health economics*, *8*, 41–51.

107. Jayadevappa, R., Malkowicz, S. B., Wittink, M., Wein, A. J., & Chhatre, S. (2012). Comparison of distribution- and anchor-based approaches to infer changes in health-related quality of life of prostate cancer survivors. *Health Services Research*, *47*(5),

1902–1925.

108. Wyrwich, K. W., Norquist, J. M., Lenderking, W. R., & Acaster, S. (2013). Methods for interpreting change over time in patient-reported outcome measures. *Quality of Life Research*, *22*, 475–483.

109. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

110. Mariann Helter, T., Coast, J., Łaszewska, A., Stamm, T., & Simon, J. (n.d.). Capability instruments in economic evaluations of health-related interventions: a comparative review of the literature. *Quality of Life Research*, *1*, 3. https://doi.org/10.1007/s11136-019-02393-5

111. Al-Janabi, H., Flynn, T. N., Peters, T. J., Bryan, S., & Coast, J. (2015). Test-retest reliability of capability measurement in the UK general population. *Health Economics*, *24*(5), 625–630.

112. Keeley, T., Al-Janabi, H., Nicholls, E., Foster, N. E., 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*, 2319–2331.

113. Chen, G., Ratcliffe, J., Billingsley Kaambwa, •, Mccaffrey, • Nikki, & Richardson, J. (2018). Empirical Comparison Between Capability and Two Health-Related Quality of Life Measures. *Soc Indic Res*, *140*, 175–190. https://doi.org/10.1007/s11205-017-1788-9

114. Engel, L., Mortimer, D., Bryan, S., Lear, S. A., & Whitehurst, D. G. T. (2017). An Investigation of the overlap between the ICECAP-A and five preference-based health-related quality of life instruments. *Pharmaco Economics*, *35*(7), 741–753.

115. Keeley, T., Al-Janabi, H., Lorgelly, P., & Coast, J. (2013). A Qualitative Assessment of the Content Validity of the ICECAP-A and EQ-5D-5L and Their Appropriateness for Use in Health Research) A Qualitative Assessment of the Content Validity of the ICECAP-A and EQ-5D-5L and Their Appropriateness for Use. *PLoS ONE*, *8*(12), 85287.

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

117. Goranitis, I., Coast, J., Day, E., Copello, A., Freemantle, N., Seddon, J., … Frew, E. (2016). Measuring Health and Broader Well-Being Benefits in the Context of Opiate Dependence: The Psychometric Performance of the ICECAP-A and the EQ-5D-5L.

Value in Health, *19*(6), 820–828.

118. Engel, L., Bryan, S., Noonan, V. K., & Whitehurst, D. G. T. (2018). Using path analysis to investigate the relationships between standardized instruments that measure health-related quality of life, capability wellbeing and subjective wellbeing: An application in the context of spinal cord injury. *Social Science and Medicine*. https://doi.org/10.1016/j.socscimed.2018.07.041

119. Engel, L., Chen, · Gang, Richardson, · Jeffrey, & Mihalopoulos, C. (1936). The impact of depression on health-related quality of life and wellbeing: identifying important dimensions and assessing their inclusion in multi-attribute utility instruments. *Quality of Life Research*, *27*, 2873–2884. https://doi.org/10.1007/s11136-018-1936-y

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

121. Peak, J., Goranitis, I., Day, E., Copello, A., Freemantle, N., & Frew, E. (n.d.). Predicting health-related quality of life (EQ-5D-5 L) and capability wellbeing (ICECAP-A) in the context of opiate dependence using routine clinical outcome measures: CORE-OM, LDQ and TOP. https://doi.org/10.1186/s12955-018-0926-7

122. McLachlan, H. L., Forster, D. A., Davey, M. A., Farrell, T., Gold, L., Biro, M. A., … Waldenström, U. (2012). Effects of continuity of care by a primary midwife (caseload midwifery) on caesarean section rates in women of low obstetric risk: the COSMOS randomised controlled trial. *BJOG : an international journal of obstetrics and gynaecology*, *119*(12), 1483–92. https://doi.org/10.1111/j.1471-0528.2012.03446.x

123. Mitchell, P. M., Al-Janabi, H., Richardson, J., Iezzi, A., & Coast, J. (2015). The Relative Impacts of Disease on Health Status and Capability Wellbeing: A Multi-Country Study. https://doi.org/10.1371/journal.pone.0143590

124. Tang, C., Xiong, Y., Wu, H., & Xu, J. (2018). Adaptation and assessments of the Chinese version of the ICECAP-A measurement. *Health and Quality of Life Outcomes*, *16:45*.

125. Shahtaheri, R. S., Nikfar, S., Sari, A. A., & Yekani Nejad, M. S. (2020). Cross-Cultural Adaptation and Psychometric Analysis of the Persian Version of the ICEpop CAPability Measure for Adults Capability Measure in the Iranian General Population. *Value in Health Regional Issues*. https://doi.org/10.1016/j.vhri.2020.01.001

126. Coast, J., Peters, T. J., Lucy, A. E., Ae, N., Ae, K. S., & Flynn, T. (2008). An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people. *Quality of Life Research*, *17*, 967–976.

127. Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*(24), 3186–3191.

128. Flynn, T. N., Huynh, E., Peters, T. J., Al-Janabi, H., Clemens, S., Moody, A., & Coast, J. (2015). Scoring the Icecap-a capability instrument. Estimation of a UK general population tariff. *Health economics*, *24*(3), 258–69. https://doi.org/10.1002/hec.3014

129. Holst-Kristensen, A. W., Linton, M.-J., Mitchell, P. M., Coast, J., Fonager, K., & Pedersen, K. M. (2020). *Construct validity and responsiveness of the Danish ICECAP-A*.

130. The municipality of Aalborg. (2019). *Aalborg i tal 2019*. Retrieved from http://www.e-pages.dk/aalborgkommune/3204/html5/

131. Hæsum, L., Hansen, L., & Ehlers, L. (2017). *Oversigt over data på sundhed- og ældreområdet i Aalborg kommune*. Retrieved from https://www.dchi.aau.dk/digitalAssets/285/285895_aak-rapport-2017_final.pdf

132. Pedersen, C. B., Gøtzsche, H., Møller, J. Ø., Mortensen, P. B., & Med Bull, D. (n.d.). *The Danish Civil Registration System A cohort of eight million persons*. Retrieved from http://ugeskriftet.dk/files/scientific_article_files/2018-10/dmb3816.pdf

133. Handberg, C., & Nielsen, C. V. (2013). Tilfredshed med kommunal rehabilitering af borgere med kronisk sygdom. *Danish Journal of nursing*, *113*(11), 74–79.

134. Worcester, M. U. C., Murphy, B. M., Mee, V. K., Roberts, S. B., & Goble, A. J. (2004). Cardiac rehabilitation programmes: predictors of non-attendance and drop-out. *European Journal of Cardiovascular Prevention and Rehabilitation*, *11*, 328–335. https://doi.org/10.1097/01.hjr.0000137083.20844.54

135. Fischer, M. J., Scharloo, M., Abbink, J. J., van 't Hul, A. J., van Ranst, D., Rudolphus, A., … Kaptein, A. a. (2009). Drop-out and attendance in pulmonary rehabilitation: The role of clinical and psychosocial variables. *Respiratory Medicine*, *103*(10), 1564–1571. https://doi.org/10.1016/j.rmed.2008.11.020

136. Selzler, A.-M., Simmonds, L., Rodgers, W. M., Wong, E. Y. L., & Stickland, M. K. (2012). Pulmonary Rehabilitation in Chronic Obstructive Pulmonary Disease: Predictors of Program Completion and Success. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, *9*(5), 538–545. https://doi.org/10.3109/15412555.2012.705365

137. Wittmer, M., Volpatti, M., Piazzalonga, S., & Hoffmann, A. (2011). Expectation, satisfaction, and predictors of dropout in cardiac rehabilitation. *European Journal of*

Preventive Cardiology, *19*(5), 1082–1088.

138. Mikkelsen, T., & Thomsen, Kristian Korsgaard Tchijevitch, O. (2014). Non-attendance and drop-out in cardiac rehabilitation among patients with ischaemic heart disease. *Danish Medical Journal*, *61*(10).

139. Cassidy, S., Turnbull, S., Gardani, M., & Kirkwood, K. (n.d.). Attendance at pulmonary rehabilitation classes: An exploration of demographic, physiological and psychological factors that predict completion of treatment. https://doi.org/10.1177/1479972314527469

140. Pardaens, S., De Smedt, D., De Bacquer, D., Willems, A. M., Verstreken, S., & De Sutter, J. (2017). Comorbidities and Psychosocial Characteristics as Determinants of Dropout in Outpatient Cardiac Rehabilitation. *Journal of Cardiovascular Nursing*, *32*(1), 14–21. https://doi.org/10.1097/JCN.0000000000000296

141. Abebaw, Y., Ali, Y., Noel, D., & Edna, B. (2007). Predictors of drop-out from an outpatient cardiac rehabilitation programme. *Clinical Rehabilitation*, 222–229. https://doi.org/10.1177/0269215506070771

142. Schwennesen, N., Henriksen, J. E., & Willaing, I. (2016). Patient explanations for non-attendance at type 2 diabetes self-management education: A qualitative study. *Scandinavian Journal of Caring Sciences*, *30*(1), 187–192. https://doi.org/10.1111/scs.12245

143. Ruano-Ravina, A., Pena-Gil, C., Abu-Assi, E., Raposeiras, S., van 't Hof, A., Meindersma, E., … González-Juanatey, J. R. (2016). Participation and adherence to cardiac rehabilitation programs. A systematic review. *International Journal of Cardiology*, *223*, 436–443.

144. Holst-Kristensen, A. W., Fonager, K., & Pedersen, K. M. (2020). *Determinants of Participation and Drop-out in Municipal Rehabilitation*.

145. Holst-Kristensen, A. W., Fonager, K., & Pedersen, K. M. (2020). *Three perspectives on economic analysis of health-related quality of life and healthcare utilisation in municipal rehabilitation programmes – a comparison of socioeconomic groups*.

146. The Health Coordination Committee for the Northern Jutland Region. (2017). *Status på sundhedsaftale 2015-2018*. Retrieved from https://rn.dk/sundhed/til-sundhedsfaglige-og-samarbejdspartnere/sundhedsaftalen-2015-2018/-/media/Rn_dk/Sundhed/Til-sundhedsfaglige-og-samarbejdspartnere/Sundhedsaftaler/Sundhedsaftale-2015/Status-på-sundhedsaftalen-2015-2018.ashx

147. Sørensen, J., Davidsen, M., Gudex, C., Pedersen, K. M., & Brønnum-Hansen, H. (2009). Danish EQ-5D population norms. *Scandinavian Journal of Public Health*, *37*, 467–474. https://doi.org/10.1177/1403494809105286

148. *Developing NICE guidelines: the manual Developing NICE guidelines: the manual Process and methods guides Developing NICE guidelines: the manual*. (2014).

APPENDICES

Appendix A. Test–retest reliability of ICECAP-A in the adult Danish population

Quality of Life Research (2020) 29:547–557
https://doi.org/10.1007/s11136-019-02331-5

Test–retest reliability of ICECAP-A in the adult Danish population

Annette Willemoes Holst-Kristensen[1] · **Kirsten Fonager**[2,3] · **Kjeld Møller Pedersen**[1]

Accepted: 3 October 2019 / Published online: 18 October 2019

Abstract

Purpose To investigate the test–retest reliability of Investigating Choice Experiments Capability measure for Adults (ICECAP-A) in the adult Danish population.

Methods The original English ICECAP-A was translated into Danish by forward–backwards translation using the guidelines by Beaton et al. Three hundred and-thirty-two participants with mean age of 57 years participated in a Web-based study. Data concerning relative and absolute agreement were analysed by the intra-class correlation coefficient and Bland–Altman plot with limits of agreement. The overall and item consistency was investigated by weighted kappa statistics from baseline to 2-week follow-up. Logistic regression was used to study the effect of the sociodemographic characteristics with inconsistent responses as the dependent binary variable. The independent variables were age, sex, education, income, and region of residence at baseline.

Results The baseline ICECAP-A preference-based index score was 0.84, and at follow-up, 0.83. The ICC was 0.86 (95% CI 0.826–0.884), and limits of agreement were 0.164 and −0.151. The kappa coefficient ranges from 45 to 65%, between random and perfect agreement. The logistic regression to analyse inconsistent responses showed no significant association between the overall index score and sociodemographic characteristics, and no clear pattern was found concerning the individual item inconsistency.

Conclusions Evidence regarding the reliability of the Danish version of ICECAP-A is satisfactory for both the index score agreement and the individual item consistency and is a reliable measure to be used in a Danish context and future health economic evaluations.

Keywords Capability approach · ICECAP-A · Reliability · Quality of life · Outcome measurement · Test–retest

Introduction

Whenever a new alternative preference-based instrument for use in health economic evaluation is developed, it is important to investigate reliability and validity in other populations than where it originated. This also holds for Investigating Choice Experiments Capability measure (ICECAP) developed in England and here is looked at from a Danish perspective in terms of test–retest reliability [1].

ICECAP is based on Sen's capability approach [1] and it is an open question whether ICECAP is a substitute for Quality Adjusted Life Years (QALY) or a supplement. Sen sees the capability approach as an alternative to the standard "welfarist" and also an alternative approach to the extra welfarist approach of QALYs [2, 3].

Within health economics, cost-effectiveness and cost-utility analysis is the dominant economic evaluation paradigm [4]. QALY is the standard "extra welfarist" approach to the benefit side in the cost-utility analysis. QALY captures the qualitative and quantitative impact of an intervention by combining the length of life and the impact on Health-related Quality of Life (HrQoL) [4–7]. In order to generate QALYs, health utilities are necessary. Utilities are preference weights that can be equated with a value or desirability.

✉ Annette Willemoes Holst-Kristensen
awh@id.aau.dk

Kirsten Fonager
k.fonager@rn.dk

Kjeld Møller Pedersen
kmp@sam.sdu.dk

[1] Department of Business and Management, Danish Center for Healthcare Improvements, Aalborg University, Fibigerstraede 11, 9220 Aalborg, Denmark

[2] Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

[3] Department of Social Medicine, Aalborg University Hospital, Aalborg, Denmark

The concept is that individuals move through different health states over time, and each health state has a value on a scale from 0 and 1. With the value zero being dead and one perfect health. States worse than death can occur and have a negative value [8].

However, there are some limitations to the QALY paradigm because it limits the comparison of health interventions to those interventions that result in outcomes commonly captured by QALYs, i.e. health related but not broader outcomes. The QALY-based approach excludes other useful pieces of information besides health like feeling safe and secure. QALY was not developed to capture non-health-related aspects and hence relevant information beyond health [4]. As an example, the aim for an outpatient rehabilitation programme for chronically ill patients who are in risk of mental, physical, and social limitations is to improve their chance for an independent and meaningful life based on the patient's entire life situation [9, 10]. It is questionable whether QALY is adequate as the primary outcome measure in such interventions [1, 11]. Instead, it is argued that a broader generic measure of well-being going beyond health may be a supplement and useful instrument for comparing the outcomes of a diverse range of interventions in the fields of public health and social care, interventions aimed at helping individuals maintain independence, dignity, comfort, and social relations. Such outcomes are neglected by only measuring health changes in terms of QALY [1, 11]. For these reasons, alternative preference-based instruments for use in health economic evaluation have been developed, such as the ICECAP-A discussed here and the Adult Social Care Outcomes Toolkit (ASCOT) [11].

This study aims to investigate the test–retest reliability of the ICECAP-A in the adult Danish population.

The capability approach

Amartya Sen pioneered the capability approach in economics as an alternative framework for assessing individual well-being. Sen sees the approach as an alternative to the standard "welfarist" approaches. Welfarist approaches assess the state of affairs in terms of individual welfare or utility, while the capability approach assesses one's state in terms of the individual's freedom to pursue valuable outcomes or reach valuable states of being and is therefore thought of as an "extra welfarist" approach [3, 12]. The idea of the capability approach is to assess well-being in terms of people's functionings and capabilities and let this reflect Quality of Life (QoL) in a broader sense [13]. Functionings and capabilities are essential aspects of an individual's well-being. Functionings are the things a person actually 'does' or 'is' and can be various activities from simple functionings, for example, going to work, or eating, to more complex functioning, such as being happy, having a family, and being healthy. According to Sen, well-being should furthermore include freedoms to achieve, and the individual's capabilities represent these freedoms. Capabilities represent a person's freedom, opportunity, and ability to generate valuable outcomes; they essentially provide a set of potential combinations of functionings available to an individual [13, 14]. A common example of the difference between capabilities and functionings is the difference between starving and fasting. The functioning is starving in both cases, but the capability to obtain an adequate amount of food is only available for the person fasting [2]. The distinction between functionings and capabilities is between achievements on one side and freedoms or valuable opportunities on the other. The combination of a person's functioning and capabilities represents their capability set, and the capability set represents their opportunity freedom, their freedom to choose between alternative combinations of functionings [3].

ICECAP-A questionnaire

The ICECAP family of questionnaires consists of ICECAP-A for adults, ICECAP-O for elderly above 65 years of age, and the ICECAP-SCM for end-of-life treatment. They were all designed to measure a particular set of capabilities related to the ability to achieve valuable functionings in life. The ICECAP-A represents the only attempt so far to develop a generic capability instrument that can be used for economic evaluation across a broad range of patient groups and the general population. It is a self-completion questionnaire developed using qualitative methods [15], designed to capture capability across five attributes of life, each of which has four levels ranging from the full capability (level 4) to no capability (level 1). The five attributes were identified through in-depth qualitative interviews identifying capabilities that are important to people. The five attributes cover an individual's capability and freedom to have stability, attachment, autonomy, achievement, and enjoyment in their life [1]. Stability refers to the ability to feel settled and secure, attachment to the ability to have love, friendship, and support, autonomy to the ability to be independent, and achievement to an individual's ability to achieve and progress, to move forward in life; enjoyment refers to people's ability to have enjoyment and pleasure [1, 14].

The ICECAP-A attributes can be turned into a preference-based index score using a best–worst scaling (BWS) approach, which is a multiattribute approach to measure preferences [16, 17]. It was used in a UK study based on 413 randomly sampled individuals. Respondents were interviewed and presented with a set of hypothetical scenarios with different best–worst options representing all five attributes of the ICECAP-A. Respondents were then asked to

choose within each profile which attribute is best and which is worst given the hypothetical scenario. The estimated capability values are then a function of the choice frequencies. The study shows that all five attributes are of importance, especially the values for stability and attachment, which are somewhat stronger than the other three. The individual responses are scored on a 0–1 scale (index), and ICECAP-A values are anchored to the "no capability" state, which is the zero point with 1 being "full capability". A zero index score is not anchored as dead as a QALY score is. However, it is possible to interpret it in this way; hence, "no capabilities" provide a meaningful lower anchor [1, 18].

Method

Translation

The original ICECAP-A questionnaire was translated to Danish following modified principles adapted from Beaton et al., involving a forward–backwards translation and a pilot study [19].[1] The original ICECAP-A questionnaire was translated from its original language English into Danish by two different translators with Danish as a native language and high level of English skills resulting in two different versions, A and B. To assess which of the two versions were best suited and to establish face validity, a pilot test was undertaken. The purpose was to investigate whether the questionnaire on the surface appears to be relevant to the respondents, their willingness and ability to answer, and possible doubts about the meaning of questions. As the questionnaire is to be used in a rehabilitation centre in a later study, the A and B versions were tested on persons in rehabilitation. The two respondent groups (10 in group A and 11 in group B) were generated as a purposive sample, by showing up at the rehabilitation centre in Aalborg and handing out the questionnaire randomly to the 21 participants. The rehabilitation centre is the place where the ICECAP-questionnaire subsequently would be used. They were randomly given version A or B and asked to comment on phrasing/wording, possible misunderstanding and misinterpretation of the questions, and suggestions for alternative phrasing. The corresponding author was present if there were any questions, but no face-to-face interviews were conducted. After this version, we had a reconciliation process where version B was selected since it had no remarks on phrasing or wording, but in A, there was doubt about the Danish phrasing for independence.

After the pilot test, the Danish version B was back-translated from Danish to English by two translators. One of which was a new translator with English as their native language and high level of Danish and one of the translators from the first version with Danish as their native language. The back-translation resulted in two different versions C and D. The authors of this study discussed these versions and chose the version that seemed to represent the original phrasing most closely. To ensure the Danish version reflected the same meaning as the original English version, the ICECAP team at the University of Birmingham gave feedback and approved the back-translation. The ICECAP team felt that the translation for the attributes "Love, friendship and support", "Being Independent", and "Achievement and progress" were accurate and appropriate in the Danish back-translation. The most substantive piece of feedback was related to the attribute "feeling settled and secure". The concern was related to the use of the term "thrive". This term was changed to "settled" in the English back-translation, but it did not influence the Danish translation since there is only one word for this phrasing in Danish. This dialogue resulted in the final Danish version [20]—available in "Appendix 1".

Data and participants

Data for the Danish reliability study came from a Web-based study conducted by the professional survey agency EPINION in December 2017 with 800 participants at baseline, 18 years of age or more. Respondents were recruited among EPINION online panel members representing the general Danish population. Respondents self-completed the electronic ICECAP-A questionnaire on two occasions, 2 weeks apart. This interval is believed to be long enough for the respondents not to remember their previous answer and short enough to not expect a real change in their quality of life and general health [21]. EPINION requested the respondents provide sociodemographic information, age, sex, education, annual income, and region of residence. Education was defined in three levels according to the International Standard Classification of Education (ISCED). The annual taxable income was predefined by EPINION and divided into six categories ranging from 26,810 euro to 67,027 euro and a "no reply" category. To the ICECAP-A, Danish version was added a self-rated health question, "How is your overall health" from Short Form 36 (SF36) on a 1–5 scale, where one is excellent, and five is bad. Furthermore, the respondents in the 2-week follow-up questionnaire were asked "Has your health changed over the past 2 weeks? yes/no". Respondents with a change in health status were excluded from further analysis because this could have an influence on the answers but have no relation to the reliability of the questions. A 100% reliable answer to a question was defined as giving the same answer at baseline and follow-up [21].

[1] The corresponding author obtained permission to translate and use ICECAP-A into the Danish version by the ICECAP team, University of Birmingham.

Statistics

Test–retest reliability of responses was analysed both for the index score as a whole and for the individual items.

The ICECAP-A scores were computed using the British index scores of the original ICECAP-A based on the algorithm for calculating the index score provided by the ICECAP team. We examine the relative reliability of the index score as a whole with the intra-class correlation coefficient (ICC). ICC was calculated using a two-way mixed model between baseline and the 2-week follow-up. We choose this model because there are multiple scores from the same rater, and the data are continuous [22]. To examine the absolute reliability, we calculated and presented a Bland–Altman plot with bias and upper and lower limits of agreement (LoA). We calculated the difference in the index score from baseline to follow-up for each respondent and the 95% LoA of the mean difference for the whole group. The 95% LoA was estimated by the mean difference $\pm$ 1.96 standard deviations of the differences. As recommended by Bland and Altman [23], 95% of the differences between measurements at baseline and follow-up are expected to lie within the LoA. This difference is visualised in the Bland–Altman plot, where the individual differences are plotted against the mean of the baseline to follow-up [23].

To calculate the individual item consistency, we used the linear weighted kappa statistic because the data are categorical with more than three ordered categories [24]. The kappa coefficient is a chance-adjusted agreement coefficient, and the weighted version accounts for the fact that inconsistent responses could vary in their level of inconsistency. Kappa can take any value from -1 to $+1$, where negative values indicate that the observed agreement is less than the expected from chance alone, the value of 0 indicates exact chance agreement, and positive values indicate that the observed agreement is higher than expected from chance. Values ranging from 0.41 to 0.60 are considered moderate, values from 0.61 to 0.80 indicate substantial agreement, and values from 0.81 to 1 stand for almost perfect agreement [21, 25]. The ICC estimates, their 95% confident intervals, and the weighted kappa coefficients were calculated using STATA 14.1.

Logistic regression was used to study the effect of the sociodemographic characteristics with inconsistent responses as the dependent binary variable. The binary variable for inconsistency between the answers from baseline to follow-up was a yes/no variable with zero indicating consistency and one if inconsistent. The variable was created for both the overall index score and each of the ICECAP-A attributes. The independent variables were age, sex, education [26], income, and region of residence at baseline.

Results

Eight hundred and four individuals were invited and participated in the first round of the reliability test–retest in December 2017, and out of these, 397 completed both the baseline and follow-up questionnaire 2 weeks later. During the 2-week follow-up, 65 respondents (16%) reported that their health had changed in the past 2 weeks and therefore were excluded from further analysis. As a result, the sample size used in the analysis is 332 individuals. The study population was broadly representative regarding sex, age, region of residence, and self-rated health when compared to national statistics—available on request.

Most of the respondents were men (55%) and had a high level of education, there was an even distribution concerning income, and 30% of the respondents were resident in the Capital Region of Denmark. Most of the respondents reported "good" or "very good" self-rated health at baseline, Table 1. After 2 weeks, 105 (31.6%) respondents rated self-rated health differently at follow-up compared to the baseline, most of whom only changed one level, Table 2. Inconsistency concerning the ICECAP-A questions was evenly distributed in all five questions with a frequency of 24.6–30.7%, Table 2. A total of 82 individuals (24.6%) had no inconsistency in any of the questions, but the remaining 250 had 462 inconsistent answers divided between all five ICECAP-A attributes. Fifteen (4.5%) of the respondents made a change resulting in an answer two levels higher or lower than the first answer. The baseline mean ICECAP-A index score was 0.84, with the follow-up index score of 0.83 (Fig. 1), representing 79 different capability states at baseline and 83 at follow-up. The frequencies are displayed in Table 3 and show that the majority of respondents place their answer in the second highest level except for the question about enjoyment, where the majority answer in the highest. The ceiling effect, defined as the highest possible score, is 6%. The mean individual ICC agreement was 0.86 (95% CI 0.826–0.884) and 0.92 (95% CI 0.905–0.938) for the group average representing the relative reliability. Respondents with poor and fair general health had an ICECAP-A index score of 0.55–0.70, those with good general health had a mean score of 0.85, and the respondents with very good or excellent had a mean score between 0.894 and 0.896. The absolute reliability resulted in a Bland–Altman plot (Fig. 2); the upper and lower LoA were 0.164 and -0.151, respectively. All except 20 (6%) respondents were within the 95% LoA. The Bland–Altman plot indicates that there is no systematic errors and no systematic correlation between bias and the size of the measure. Moreover, there is no correlation between the differences and the measured value. The weighted kappa coefficient ranges from 0.45 to 0.65 in agreement for the ICECAP-A items, 45–65% of the way

Table 1 Baseline characteristics and self-rated health state

Baseline characteristics ($n=332$)	Mean/frequency (%)
Age	57 (SD 13.19)
Sex	
Male	182 (55)
Female	150 (45)
Education	
Low (< 11 years)	25 (7.53)
Medium	129 (38.86)
High	178 (53.61)
Annual taxable (€)	
26,810	47 (14)
26,811–40,215	66 (20)
40,216–53,621	56 (17)
53,622–67,026	51 (15)
67,027+	67 (21)
No reply	45 (13)
Region	
Capital Region of Denmark	100 (30)
Region Zealand	58 (18)
Region of Southern Denmark	79 (24)
Central Denmark Region	62 (18)
The North Denmark Region	33 (10)
Self-rated health	
Excellent	26 (7)
Very good	115 (34)
Good	135 (41)
Fair	53 (16)
Poor	3 (0.9)

between random and perfect agreement (Table 4). When the inconsistent answers concerning self-rated health were excluded, the weighted kappa coefficient increases to 0.48–0.68.

A logistic regression model was used to investigate if the inconsistency between baseline and follow-up index scores and inconsistency for each ICECAP-A question differed according to sociodemographic characteristics as sex, age, education, income, and region of residence. The base case (based on highest frequency) is male of mean age, a high education level, an income of more than 67,027 Euro (€), and lives in the Capital Region of Denmark. For the overall index score, no significant association was found between the inconsistency in the score and the sociodemographic characteristics. Concerning the individual items, a significant association between inconsistency and attachment, autonomy, achievement, and enjoyment were found. For attachment, the odds of being inconsistent were 2.79 times higher if the respondent had an income of 26,811–40,215 € (p value 0.01). Concerning autonomy, low education was associated with 2.49 higher odds (p value 0.02); however, having an income of 26,811–40,215 € and 40,216–53,621 € was associated with significantly lower odds (p value 0.02 and 0.05, respectively) Achievement was associated with significantly lower odds for those with medium education (p value 0.03), and for enjoyment, low education was significantly associated with inconsistency (p value 0.03) (Table 5).

Table 2 Frequency of inconsistent answers in self-rated health and in each of the five ICECAP-A attributes

Number of inconsistent answers (%)					
Self-rated health					105 (31.6)
ICECAP-A	Stability	Attachment	Autonomy	Achievement	Enjoyment
	82 (24.6)	92 (27.7)	102 (30.7)	88 (26.5)	98 (29.5)

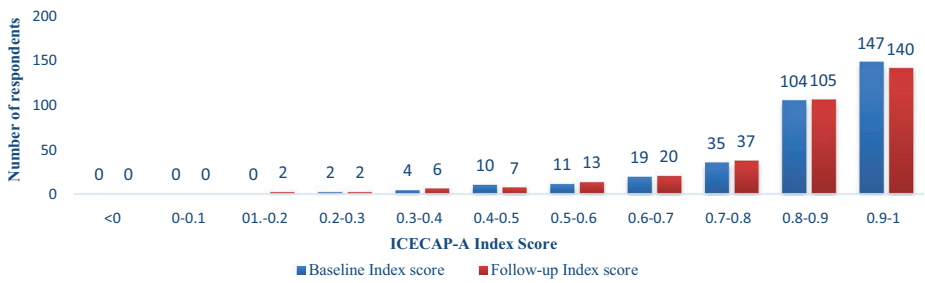


Fig. 1 Distribution of ICECAP-A index scores at baseline and follow-up ($n=332$)

Table 3 The frequency of ICECAP-A answers at baseline and follow-up

	Frequency at baseline (%)	Frequency at follow-up (%)
Stability		
I am able to feel settled and secure in all areas of my life	111 (33.4)	97 (29.2)
I am able to feel settled and secure in many areas of my life	192 (57.8)	205 (61.7)
I am able to feel settled and secure in a few areas of my life	25 (7.5)	27 (8.1)
I am unable to feel settled and secure in any areas of my life	4 (1.2)	3 (1.0)
Attachment		
I can have a lot of love, friendship and support	117 (35.2)	118 (35.5)
I can have quite a lot of love, friendship and support	152 (45.8)	150 (45.2)
I can have a little love, friendship and support	52 (15.7)	53 (16.0)
I cannot have any love, friendship and support	11 (3.3)	11 (3.3)
Autonomy		
I am able to be completely independent	120 (36.1)	108 (32.5)
I am able to be independent in many things	193 (58.1)	200 (60.2)
I am able to be independent in a few things	17 (5.1)	20 (6.1)
I am unable to be at all independent	2 (0.6)	4 (1.2)
Achievement		
I can achieve and progress in all aspects of my life	60 (18.1)	61 (18.4)
I can achieve and progress in many aspects of my life	212 (63.8)	206 (62.0)
I can achieve and progress in a few aspects of my life	57 (17.2)	60 (18.1)
I cannot achieve and progress in any aspects of my life	3 (0.9)	5 (1.5)
Enjoyment		
I can have a lot of enjoyment and pleasure	154 (46.4)	157 (47.3)
I can have quite a lot of enjoyment and pleasure	137 (41.3)	130 (39.2)
I can have a little enjoyment and pleasure	37 (11.1)	40 (12.0)
I cannot have any enjoyment and pleasure	4 (1.2)	5 (1.5)

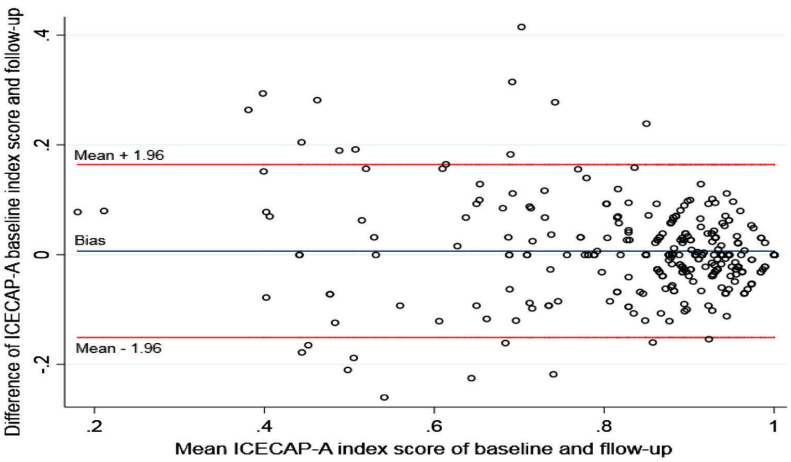


Fig. 2 Bland–Altman plot between test and retest. The red lines represent upper and lower LoA, and the blue line is the bias, representing the mean of differences

Discussion

In this study, the test–retest reliability of the ICECAP-A capability measure for the general adult Danish population was investigated. We found that the reliability test of the ICECAP-A indicated that the index score has good test–retest reliability in terms of ICC (0.86) and moderate agreement for each item (45–65%), using the weighted

Table 4 Kappa statistic of the ICECAP-A attributes

ICECAP-A	Kappa statistics			
	Agreement (%)	Expected agreement (%)	Weighted kappa (κ)	Standard error
Stability	91.37	79.54	0.578	0.042
Attachment	90.56	72.31	0.659	0.039
Autonomy	89.26	80.31	0.455	0.044
Achievement	90.96	79.08	0.568	0.038
Enjoyment	89.96	74.99	0.599	0.042

kappa. The moderate agreement could be explained by the respondents defined as outliers in the Bland–Altman plot. Outliers were the respondents that changed their answers more than one level, for example, answered a level 4 at baseline but a level 2 at follow-up. These changes result in lower kappa coefficients because of the use of weighted kappa statistics. The use of ICC, a Bland–Altman plot, and weighted kappa statistics provided different evidence about the test–retest reliability of ICECAP-A, which gave a better picture of its reliability.

This study managed to enrol more respondents than the original ICECAP-A reliability study by Al-Janabi et al. [26]. The Al-Janabi study was based on 237 individuals, answering both ICECAP-A and the European Quality of Life 5 Dimensions (EQ-5D) and resulted in a baseline index score of 0.78 and 0.80, respectively, slightly lower than our study. The study by Al-Janabi et al. showed that there were 84% inconsistent responses concerning the ICECAP-A measure, and for the EQ-5D, there were 38% inconsistent responses. The authors point out that this may be explained by the vast difference in the number of capability and health states (82

Table 5 Logistic regression results of odds for inconsistency defined as a binary yes/no variable for the index score and all five ICECAP-A attributes

	Odds ratio					
	ICECAP-A index score	Stability	Attachment	Autonomy	Achievement	Enjoyment
Age	1.00	1.00	1.01	0.99	1.01	0.98
Sex						
Male	1	1	1	1	1	1
Female	0.94	1.09	1.17	0.67	1.09	0.86
Education[a]						
Low	1.92	2.03	0.55	**2.49**	1.02	**2.82**
Medium	0.92	0.93	0.87	1.60	**0.54**	1.01
High	1	1	1	1	1	1
Annual taxable (€)						
Under 26,810	0.65	0.66	1.49	0.62	1.60	0.59
26,811– 40,215	1.13	1.23	**2.79**	**0.39**	0.78	0.93
40,216 53,621	1.17	0.45	1.26	**0.44**	0.70	1.30
53,622–67,026	2.17	0.92	1.40	0.79	1.09	1.43
67,027+	1	1	1	1	1	1
No reply	0.60	1.04	1.88	0.54	0.75	0.69
Region						
Capital Region of Denmark	1	1	1	1	1	1
Region Zealand	1.34	0.77	1.72	1.16	0.95	1.40
Region of Southern Denmark	1.18	1.50	1.48	1.55	0.67	1.13
Central Denmark Region	0.85	0.77	1.79	0.79	0.70	1.22
The North Denmark Region	0.72	1.09	0.62	0.71	1.17	0.63
Self-rated health	1.60	1.21	1.33	1.26	1.11	0.90

[a]According to International Standard Classification of Education (ISCED). Statistically significant with p values < 0.05 are indicated in bold

and 25, respectively) partly because the ceiling effect is less likely in ICECAP-A (3% vs. 35%) and the fact that ICECAP-A has four response categories versus EQ-5D-3L's three levels. This was similar in our study, where 75% of the respondents had one or more inconsistent answers, and 6% selected the top state of the ICECAP-A at baseline. The frequency of inconsistency occurs across all the ICECAP-A questions. The inconsistency concerning question one and two have similar inconsistency rates as the study by Al-Janabi et al. [26]. Our study, however, had relatively higher inconsistency in the remaining three questions: 26–30% compared to 2–13% in the study by Al-Janabi et al. [26]. The inconsistency in our study may be explained by the differences in self-rated health. The respondents were asked if their health had changed within the last 2 weeks, and they were asked to self-rate their health. However, some of the respondents who tick 'no' to any changes in health, within the last 2 weeks had changes in their self-rated health (31.6%). Were the inconsistent answers concerning self-rated health excluded, the weighted kappa coefficient increases to 0.48–0.68. However, the respondents may have had a change in their well-being between baseline and follow-up which could explain the inconsistency and therefore not directly related to the reliability of the Danish version of the ICECAP-A.

The present study is limited because the respondents were only asked if their health had changed during the 2 weeks. They should have been asked if their health or their well-being had changed. This means that we do not know if any other changes in the respondent's life, for example, well-being, which has resulted in possible inconsistency at follow-up. Another limitation is the overweight of respondents with "good" or "very good" health (75%). This may indicate better agreement due to less variability in good health status, and not related to the reliability of the questionnaire. However, the general health status is representative of the general population, and in order to capture this issue, further research into the correlation between lower state health states and variability is needed. The study could also be limited, not by the number of respondents, but by the fact that the 332 respondents were from a panel that could result in biased answers and underrepresentation compared with the general population since some groups could be more willing to participate than others. There was an overrepresentation of respondents with a low level of education and low income compared to the general population. Reminders could have been sent out and could have resulted in a larger and representative sample. However, the time perspective was important. The optimum solution would have been a representative sample drawn by Statistic Denmark. The EPINION agency assures that they are aware of the limitation with panels and attempts to counter this by using intelligent targeting and invitation systems to ensure representativity along with a weighting system. The system is also designed to reinvite under-represented participants based on age, sex, education, occupation, and region of residence. Methodologically, the study is limited by the fact that one of the translators translated on both the forward and the backwards translation. According to the guidelines [19], it is preferable to use translators with no prior knowledge of the original questionnaire and any of the translations. However, the translations here have been widely discussed and pilot-tested. Hence, this limitation is believed to not influence the final results in this study. Lastly, a methodological limitation is the lack of using qualitative methods in the pilot test. Face-to-face interviews would have provided to buttress both reliability and face validity.

The logistic regression analysis showed no association between sociodemographic characteristics and inconsistency in the overall index score. However, there were significant results for the individual questions, but no consistent patterns of significant sociodemographic differences were found. The study by Al-Janabi et al. likewise found no association between inconsistency and age, sex, or education [26]. It is debatable if income is a relevant parameter to influence one's reliability, but in this study, it was assumed to have a possible impact in the same sense as education level. However, income did not show clear patterns across respondents.

The purpose of the ICECAP-A is to have a valid preference-based instrument to use in health economic evaluations that go beyond the QALY health instrument. However, no health economic evaluations with ICECAP-A as the sole outcome measure has been identified. The reason may be that ICECAP-A is relatively new, and only a few registered, validated, and reliable translations exist [20]. According to Flynn et al., ICECAP-A is being used in clinical studies across the UK, the USA, Australia, and New Zealand, indicating international interest for a well-being measurement

[18]. Were ICECAP-A to be used alongside HrQoL instruments, the potential implication of double counting must be considered, hence the possible overlap described by Engel et al. [27] who investigated the overlap between ICECAP-A and five preference-based HrQoL instruments. They conclude that ICECAP-A provides additional complementary information and has a certain overlap with the Assessment of Quality of Life 8-dimension (AQoL-8D) questionnaire [27]. ICECAP-A might stand alone as a broader well-being instrument without leading to false claims, but the same broad information to some degree could potentially be captured by the AQoL-8D. However, the two instruments are not interchangeable.

The sensitivity of QALYs to broader non-health outcomes is being questioned [1, 28], but so is the capability approach and the different attempts to measure well-being [29]. The capability approach, in general, has been criticised for endorsing one particular conception of a good life, for emphasising choice rather than welfare, and for being too individualistic [4]. In a commentary by Karimi et al. [29], the capability-based questionnaires are critically reviewed. It is argued that the questionnaires' questions may be inaccurate in descriptions of the individual's exact capability set. Karimi et al. believe that the measured capability sets represent only that one combination, lacking the value of choice and that one combination may not be achievable, they may be answering unrealistic hence the questioning technic "I am able to…". Also, the values are being questioned as inadequate since it is not considered as a set. Karimi et al. suggest that the capability set should be measured more indirectly [29]. When thinking of one's capabilities, individuals might vary in their time frame and their relevant limitations when identifying their capability set. This could be a limitation for any questionnaire because questionnaires usually endorse one particular concept and are answered in relation to the time frame one has in mind at that particular moment and reflect that one specific state. If all of the HrQoL and capability measurements want to reflect QoL, then all of them represent that one combination—the combinations of the questions asked.

Had the ICECAP-A been developed in Denmark, the five attributes might have been different, since stability, attachment, autonomy, achievement, and enjoyment may not be the five most important areas for the Danes' understanding of well-being [30]. However, the use of validated questionnaires is preferable to self-developed Danish versions of measurement for capabilities and well-being. In future research, the internal reliability and validity of the Danish ICECAP-A version should be investigated to gain more knowledge of its application in Danish interventions, as a Danish value-set should be a consideration.

Conclusion

The Danish version showed satisfactory test–retest reliability for both the index score agreement and the individual item consistency and hence is a reliable measure to be used in a Danish context and future health economic evaluations. The investigation of responsiveness and construct validity is an essential task in future work.

Acknowledgements We are grateful for the respondents and the staff at EPINION for their time and assistance with the survey. Also, we thank Joanna Coast, Paul M. Mitchell, and Myles-Jay Linton for their helpful comments.

Funding This study was funded by the municipality of Aalborg, and The Health Foundation (Helse Fonden) (Grant No. 463019).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The study has been carried out in accordance with the General Data Protection Regulation (2015-509-00007). Moreover, The National Committee on Health Research Ethics has assured us that no approval is needed.

Informed consent Informed consent was obtained from all individual participants included in the study by the professional survey agency EPINION.

Appendix 1

Spørgeskema omhandlende din livskvalitet

Marker hvilke udsagn, der bedst beskriver din generelle livskvalitet på nuværende tidspunkt, ved at sætte ÉT kryds (X) i ÉN kasse for hvert spørgsmål.

1. Tryghed og trivsel	
Jeg er i stand til at føle tryghed og kan trives i **alle** dele af mit liv	
Jeg er i stand til at føle tryghed og kan trives i **store** dele af mit liv	
Jeg er i stand til at føle tryghed og kan trives i **få** dele af mit liv	
Jeg er **ikke i stand til** at føle tryghed og **kan ikke** trives i nogen dele af mit liv	

2. Kærlighed, venskab og opbakning	
Jeg har mulighed for at opnå **al** den kærlighed, venskab og opbakning jeg vil	
Jeg har mulighed for at opnå **meget** af den kærlighed, venskab og opbakning jeg vil	
Jeg har mulighed for at opnå **lidt** af den kærlighed, venskab og opbakning jeg vil	
Jeg har **slet ikke** mulighed for at opnå den kærlighed, venskab og opbakning jeg vil	

3. Selvstændighed	
Jeg har mulighed for at være **fuldstændig** selvstændig i mit liv	
Jeg har mulighed for at være selvstændig i **mange** situationer i mit liv	
Jeg har mulighed for at være selvstændig i **få** situationer i mit liv	
Jeg har **slet ikke** mulighed for at være selvstændig i mit liv	
4. Præstation og udvikling	
Jeg kan præstere og udvikle mig i **alle** dele af mit liv	
Jeg kan præstere og udvikle mig i **mange** dele af mit liv	
Jeg kan præstere og udvikle mig i **få** dele af mit liv	
Jeg kan **hverken** præstere eller udvikle noget i mit liv	

5. Glæde og tilfredsstillelse	
Jeg har mulighed for at opnå **meget** glæde og tilfredsstillelse	
Jeg har mulighed for at opnå en **del** glæde og tilfredsstillelse	
Jeg har mulighed for at opnå **lidt** glæde og tilfredsstillelse	
Jeg har **slet ikke** mulighed for at opnå glæde eller tilfredsstillelse	

Kontroller, at du kun har sat ÉT kryds for hvert af de fem spørgsmål

References

1. Al-Janabi, H., Flynn, T. N., & 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.
2. Sen, A. (1999). *Development as freedom*. Oxford: Oxford University Press.
3. Sen, A. (2009). *The idea of justice* (1st ed.). Cambridge: Harvard University Press.
4. Smith, R., Lorgelly, P., Al-Janabi, H., Venkatapuram, S., & Coast, J. (2012). The capability approach: An alternative evaluation paradigm for health economics? In A. M. Jones (Ed.), *The Elgar companion to health economics* (2nd ed., pp. 415–424). Cheltenham, UK: Edward Elgar Publishing Limited.
5. Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes* (4th ed.). Oxford: Oxford University Press.
6. Culyer, A. J. (1989). The Normative economics finance and of health provision. *Oxford Review of Economic Policy, 5*(1), 34–58.
7. Gold, M. R., Siegel, J. E., Russell, L. B., & Weinstein, M. C. (1996). *Cost-effectiveness in Health and Medicine* (1st ed.). New York: Oxford University Press.
8. Weinstein, M. C., Torrance, G., & Mcguire, A. (2009). QALYs: The basics. *Value In Health, 12*(1), S5–S9.
9. Sundhedsstyrelse. (2005). *Terminologi—forebyggelse sundhedsfremme og folkesundhed*. Retrieved from http://sundhedsstyrelsen.dk/publ/publ2005/cff/termpjece/termpjece3jun05.pdf.
10. Johansen, J., Rahbek, J., Møller, K., & Jensen, L. (2004). *Hvidbog om rehabiliteringsbegrebet. Rehabilitering i Danmark*. Retrieved from http://www.marselisborgcentret.dk/fileadmin/filer/Publikationer/PDF_er/Hvidbog.pdf.
11. Makai, P., Brouwer, W. B. F., Koopmanschap, M. A., Stolk, E. A., & Nieboer, A. P. (2014). Quality of life instruments for economic evaluations in health and social care for older people: A systematic review. *Social Science and Medicine, 102,* 83–93.
12. Cookson, R. (2005). QALYs and the capability approach. *Health Economics, 14*(8), 817–829.
13. Sen, A. (1999). *Commodities and capabilities*. Oxford: Oxford India.
14. Al-Janabi, H., Peters, T. J., Brazier, J., Bryan, S., Flynn, T. N., Clemens, S., et al. (2013). An investigation of the construct validity of the ICECAP-A capability measure. *Quality of Life Research, 22*(7), 1831–1840.
15. Mitchell, P. M., Venkatapuram, S., Richardson, J., Iezzi, A., & Coast, J. (2017). Are quality-adjusted life years a good proxy

measure of individual capabilities? *Pharmaco Economics, 35*(6), 637–646.
16. Mühlbacher, A. C., Kaczynski, A., Zweifel, P., & Johnson, F. R. (2016). Experimental measurement of preferences in health and healthcare using best-worst scaling: An overview. *Health Economics Review, 6*(1), 1–14. https://doi.org/10.1186/s13561-015-0079-x.
17. Louviere, J. J., Flynn, T. N., & Marley, A. A. (2015). *Best-worst scaling, theory method and applications* (1st ed.). Cambridge: Cambridge University Press.
18. Flynn, T. N., Huynh, E., Peters, T. J., Al-Janabi, H., Clemens, S., Moody, A., et al. (2013). Scoring the ICECAP-A capability instrument. Estimation of a UK genrel population tariff. *Health Economics, 24*(3), 258–269.
19. Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186–3191.
20. Holst-Kristensen, A. W. (2018). Danish ICECAP-A. Retrieved January 29, 2019, from https://www.birmingham.ac.uk/Documents/college-mds/research/Danish-version-of-ICECAP-A.pdf.
21. Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales, a practical guide to their development and use* (5th ed.). Oxford: Oxford University Press.
22. Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163.
23. Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*(99), 135–160.
24. Brenner, H., & Kliebsch, U. (1996). Dependence of weighted kappa coefficients on the number of categories. *Epidemiology, 7*(2), 199–202.
25. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.
26. Al-Janabi, H., Flynn, T. N., Peters, T. J., Bryan, S., & Coast, J. (2015). Test-retest reliability of capability measurement in the UK general population. *Health Economics, 24*(5), 625–630.
27. Engel, L., Mortimer, D., Bryan, S., Lear, S. A., & Whitehurst, D. G. T. (2017). An Investigation of the overlap between the ICECAP-A and five preference-based health-related quality of life instruments. *Pharmaco Economics, 35*(7), 741–753.
28. Edward Brazier, J., Rowen, D., Lloyd, A., & Karimi, M. (2019). Future directions in valuing benefits for estimating QALYs: Is time up for the EQ-5D? *Value in Health, 22,* 62–68.
29. 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.
30. Kulturministeriet. (2018). Danmarkskanon. Retrieved January 24, 2019, from https://www.danmarkskanon.dk/.

Appendix B. An Investigation of Construct Validity and Responsiveness of the Danish ICECAP-A

An Investigation of Construct Validity and Responsiveness of the Danish ICECAP-A

Annette Willemoes Holst-Kristensen[1], Paul Mark Mitchell[2], Myles-Jay Linton[2], Joanna Coast[2], Kirsten Fonager[3], Kjeld Møller Pedersen[1]

1 Danish Center for Healthcare Improvements, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
2 Health Economics Bristol, Population Health Sciences, University of Bristol, Bristol, UK
3 Department of Clinical Medicine, Aalborg University, Denmark; Department of Social Medicine, Aalborg University Hospital, Denmark

Abstract

Purpose: This study aimed to provide the first assessment of construct validity of ICECAP-A in patients with cardiovascular disease, chronic obstructive pulmonary disease and diabetes, and to assess the responsiveness of the measure in this group. **Method**: Data were provided from patients attending rehabilitation in the municipality of Aalborg, Denmark, from March 2018 to March 2019. Patients answered a questionnaire from the healthcare centre and the ICECAP-A at baseline and 12 weeks follow-up. To assess construct validity, *a priori* hypotheses were developed. Based on these hypotheses, associations between sociodemographic characteristics, 'general health', a freedom dimension, and ICECAP-A were analysed through chi-squared tests and Spearman rank correlations for categorical and ordinal variables, respectively. To investigate responsiveness, the anchor-based method was used. Patients were divided into improved, worsened or no change, based on changes between baseline and follow-up on the anchor measures ('general health' and 'freedom'). To quantify responsiveness, both the weighted and un-weighted ICECAP-A scores' effect sizes, standardised response means and t-tests were used. Findings were explored across different age groups. **Result**: Of all the hypothesised associations, 16 of 26 (62%) were in the expected direction. The expected relationships were found between ICECAP-A scores and general health and freedom to do things. ICECAP-A was responsive in terms of capturing the effects on general health and the freedom to do things. Differences were found across age groups, with greater responsiveness to change in those aged under 65 years. The item-by-item analysis showed that capability was mainly driven by stability and autonomy. **Conclusion**: This study has shown that the Danish ICECAP-A is a valid and responsive measure of the effects of an exercise and education-based rehabilitation programme.

Background

The ICECAP-A is a measure of wellbeing with a theoretical basis in Amartya Sen's work. The capability approach assesses wellbeing in terms of individual 'functionings' and 'capabilities'. Functionings refer to the things an individual 'is' or 'does', ranging from fundamental aspects of life such as 'being healthy' to more complex aspects such as 'having self-respect'. Capabilities represent an individual's freedom to carry out functionings. This is important, because a person may be able to function in a particular way, but may choose not to utilise that functioning [1–3].

The ICECAP-A conceptualises wellbeing as the capability of an individual to achieve valuable functionings. ICECAP-A has five attributes: stability, attachment, autonomy, achievement and enjoyment [4]. The initial aim of the ICECAP instruments was to develop a broad measure of quality of life (QoL) for use in economic evaluation [4]. Several other capability measures have been developed, such as the OxCAP and ASCOT [5–7]. However, the ICECAP measures are distinct as they provide a generic measure of capability wellbeing for use in the economic evaluation of interventions in areas such as health and social care, where a broader aim like empowerment is to be explored [4, 8].

Some evidence is available on the reliability [9–11], content validity [12], construct validity [13–15] and responsiveness of the ICECAP-A measure [14, 16] in various populations, but so far, most evidence relates to the original UK version. In the Danish context, only one reliability study of the ICECAP-A in the general population exists [11]. This study aims to provide the first assessment of construct validity in patients with cardiovascular disease (CVD), chronic obstructive pulmonary disease (COPD) and diabetes, and to assess the responsiveness of the ICECAP-A for this group in a Danish municipal rehabilitation setting.

Method

Data collection and setting

Data were collected on a routine basis from patients attending rehabilitation in the municipality of Aalborg from March 2018 to April 2019. Patients were referred by their general practitioner or the hospital to the rehabilitation programme after an acute event necessitating a hospital stay related to their CVD, COPD or diabetes. Sociodemographic characteristics included age, gender (female or male), cohabitation (binary), education (defined in three levels according to the International Standard Classification of Education (ISCED): low <11 years of schooling, medium 11–16 years of schooling, high >16 years of schooling) and socioeconomic status (employed, unemployed or other benefits, or retired). All attending patients were asked to complete a questionnaire developed by the healthcare centre (the Aalborg questionnaire, available on request) and the ICECAP-A questionnaire at baseline and 12 weeks follow-up after the completion of the rehabilitation programme. It was the patient's choice as to whether they wished to complete the questionnaire on each occasion.

Municipal rehabilitation

In Denmark, the 98 municipalities offer rehabilitation programmes to chronically ill patients with, for example, CVD, COPD, and/or diabetes. The programmes are situated at the healthcare centre in Aalborg and at times in 'satellite' centres in varied locations across the municipality. The programmes provide exercise and education to groups of varying size.

The exercise sessions take place one to two times a week and are of low to moderate intensity. The education component covers knowledge of the disease; dietary advice; the importance of physical activity, smoking cessation and medicine consumption; and goals and motivation. The programmes usually commence within a few weeks after discharge from the hospital and continue for 8–12 weeks [17]. They are not offered routinely to chronically ill persons.

Measuring rehabilitation outcomes

The municipality of Aalborg, Denmark, decided in 2018 to develop a self-completion questionnaire to evaluate their rehabilitation programme. The full questionnaire consists of 33 questions, including background information of gender, employment status, education level and cohabitation. Additional questions concerning training level and satisfaction with the program were collected at follow-up. The healthcare centre uses six of the questions to interpret and evaluate the rehabilitation programmes: (1) 'general health', (2) 'improvement of quality of life', (3) 'feeling fit to do the things I want to', (4) 'better at handling everyday life after programme', (5) 'know how to sustain health in the future' and (6) 'able to be more physically active after programme'. Questions 1 and 3 were the only questions asked at both baseline and follow-up; the rest were only asked at follow-up. Questions 1–5 have four to five possible response categories (where higher scores indicate greater levels of general health, for example). Question 6 had a binary response option (yes or no).

Construct validity

Construct validity is the degree to which an instrument (such as a questionnaire) measures what it is hypothesised to be measuring. It can be assessed by considering the degree to which expected relationships between a measure and other factors are confirmed [18, 19]. Best-practice guidance on psychometric analyses highlights the importance of *a priori* statement of hypotheses on the anticipated relationship between the constructs explored [20]. Drawing on Sen's theoretical framework for the establishment of capabilities, capability can be limited by reduced socioeconomic status and improved by good circumstances [3]. For the assessment of construct validity, *a priori* hypotheses were developed based on existing evidence about the ICECAP measures in other contexts [13, 14]. Table 1 indicates the expected direction between the five attributes of ICECAP-A, and indicators of socioeconomic status, general health and freedom in terms of 'feeling fit to do the things I want to' included in the Aalborg questionnaire.

Table 1 Hypothesised positive relationships between ICECAP-A attributes and the Aalborg questionnaire

ICECAP-A	Stability	Attachment	Autonomy	Achievement	Enjoyment	Total score
General health	+		+	+	+	+
'Feeling fit to do the things I want to'	+	+	+	+	+	+
Employment	+	+	+	+	+	+
Education level	+		+	+		+
Cohabitation	+	+		+	+	+

The interpretation of Table 1 is as follows. The stability attribute is initially expressed as being able to feel settled and secure, and relates to the absence of significant changes in life and stress. It is therefore hypothesised that significant negative life changes were likely to be associated with reduced capability (such as changes in general health). The validity study by Al-Janabi et al. found that, among other factors, employment, education and relationship status were associated with stability in a positive direction [13]. Therefore, this study expected an association between stability and employment, education and cohabitation in a positive direction, despite the different definitions of relationship status and education

level. The attachment attribute is stated in terms of being able to have love, friendship and support, and relates to the ability to interact with others and have good relationships. Al-Janabi et al. found an positive association between attachment, employment and relationship status [13]. This study therefore anticipated finding an association between attachment, employment and cohabitation in a positive direction. The autonomy attribute is defined as being able to be independent and relates to looking after oneself and making one's own decisions. Previously, positive associations between autonomy and employment and education have been found [13]. It was therefore anticipated that higher capability level for autonomy would be associated with higher level of employment and education in this study. The achievement attribute is defined as being able to achieve and progress, and reflects individuals' abilities to move forward and achieve their goals. Previously, positive associations between achievement and employment, education and relationship status have been found [13]. It was therefore anticipated that capability for achievement would be associated with employment, education and cohabitation in a positive direction in this study. The enjoyment attribute is defined as being able to have enjoyment and pleasure in life. It reflects opportunities for the small pleasures in life, as well as things that are perceived to be enjoyable or exciting. As such, an association with employment and cohabitation was anticipated in a positive direction [13].

The ICECAP-A measure was developed to measure the effectiveness of health and social care interventions. The degree of variation in health and healthcare usage is reflected in individuals' capabilities, and therefore is essential and of interest, because poor health and disabilities affect one's capabilities [4, 13]. Previous studies concerning ICECAP-A have found that impairments to physical health reduce the capability for stability, autonomy, achievement and enjoyment [13, 21]. Therefore, this study anticipated an association between general health and stability, autonomy, achievement and enjoyment. Here, it was anticipated that the question focusing on general health would be interpreted by participants as a question about physical health only, given the reasons that they were accessing the service, and thus would not be associated with attachment. 'Feeling fit to do the things I want to' was hypothesised to be associated with all five attributes of the ICECAP-A, and high levels of capability were anticipated to relate to a high level of this question of freedom. This hypothesis is based on the findings by Al-Janabi et al. where a similar question was asked, 'I can do the things in life I want to do', and an association was found with all attributes [13].

Statistical analysis

Based on these hypotheses (Table 1), associations between selected variables and the ICECAP-A attributes at baseline were analysed using chi-squared tests for categorical variables and Spearman rank correlation for ordinal variables. A correlation was considered strong if the coefficient was higher than 0.5, moderate if the coefficient was between 0.3 and 0.5, and weak if the coefficient was below 0.3 [22].

Responsiveness

Outcome measures being able to detect meaningful changes is central to their usefulness in health and social care interventions. Two core ideas in the assessment of evaluative instruments are sensitivity to change and responsiveness. Sensitivity to change refers to the ability of instruments to measure change statistically. Responsiveness addresses the detection of the clinically relevant change [18, 23].

To assess responsiveness, some criterion is needed to ascertain where patients have changed over time. The two main methods for assessing responsiveness are the distribution- and anchor-based approaches. The distribution-based method

uses the effect size of the difference between groups to measure variability, standard response means, standard error of measurement and responsive statistics. The anchor-based method is sample-independent and examines the relationship with an anchor, such as a QoL measure, to explain the meaning of a particular degree of change [24]. The anchors can either be cross-sectional or longitudinal. An anchor-based analysis aims to assess whether scores on the target measure change in an anticipated way, as indicated by changes in the scores on the anchor [25]. Distribution methods alone do not provide information about the clinical relevance of the observed change. Therefore, this study assessed responsiveness, using anchor-based methods to investigate the association between change over time in the ICECAP-A scores and change over time in the anchors. An exploratory analysis of the correlation between the change scores of longitudinal outcome measures was used to support the choice of anchors for this study.

Using Cohen's rule, correlations were considered strong when the coefficients were >0.50, moderate when ≥0.30, and weak when <0.30. Therefore, 0.30 was used as a correlation threshold to define an at least moderate association between an anchor and outcome measure change score [26]. General health and 'feeling fit to do the things I want to' were the only two questions for which there were longitudinal data, but they were only used if they reached a threshold of baseline correlation of 0.3 (at least moderate correlation). For appropriate anchors, patients were divided into three groups depending on the changes in scores in general health and 'feeling fit to do the things I want to': (1) those who had worsened between baseline and follow-up scores, (2) those who had improved between baseline and follow-up scores, and (3) those with no change in scores between baseline and follow-up.

When assessing the responsiveness of a weighted measures such as ICECAP-A [8], consideration needs to be given independently to both the descriptive system [4] and the value weighting of the descriptive system. It is essential that the descriptive system can detect a change in a construct for the weighted measure to reflect meaningful change. If the analysis only uses the weighted tariffs scores, a misleading conclusion could be made, that is, a conclusion whereby the measure is thought not to be responsive, when, in fact, the descriptive system of the measure shows change, but the value weightings suggest that these changes are not highly valued [27]. The weighted tariffs scores are also reflective of the UK population and not those of the Danish public. Therefore, for each anchor, two analyses are presented: (1) an analysis of the 'un-weighted' descriptive system of the ICECAP-A and (2) an analysis of the 'weighted tariff scores'. For the un-weighted and weighted analysis, change was calculated in groups that improved and worsened. Un-weighted scores were calculated by summing ICECAP-A item response levels, with four indicating full capability on an item and one indicating no capability on an item. The weighted tariff scores were calculated using the UK general population tariff from Flynn et al. [28]. Findings were explored across different age groups (<65 versus ≥65 years of age).

Responsiveness of the ICECAP-A scores was assessed using the Cohen's effect size (ES) and standardised response mean (SRM). Additionally, a paired t-test was applied to test the null hypothesis, that no change in the response means between baseline and follow-up had occurred. These indices were calculated separately for patients who reported improved, worsened or no change in the anchors [18, 26]. The effect size was calculated by dividing the mean difference between baseline and follow-up scores by the standard deviation (SD) of baseline scores; SRM was calculated by dividing the mean score change (follow-up minus baseline) by the standard deviation of the change [25]. For all indices, a value of <0.2 was considered small, 0.2-0.5 moderate and >0.5 large responsiveness [26]. The range of the un-weighted score was 16 (5–20), and for the weighted, the tariff scores were 1 (0–1) with higher scores on both representing higher

capability. Age differences in responsiveness were investigated by subgroup analysis using a group <65 years of age and a group ≥65 years of age.

To assess the responsiveness of the individual ICECAP-A items, a response profile (frequency of participants answering each level for each item, at baseline and follow-up) was completed for the two anchors. Change in response profiles between baseline and follow-up was analysed for each item to indicate which items were the 'drivers' of change in the overall measure.

Statistical analysis

The investigation of construct validity was based on all baseline data. The responsiveness analysis was based on complete cases in terms of questionnaire data because of high rates of missing data (78%); hence, imputation was not considered. The type of missing was anticipated to be missing completely at random because in all cases the entire questionnaire was missing. The reason for the amount of missing is that there was voluntary completion of the questionnaire, both at baseline and follow-up. Therefore, complete case analysis was performed for the responsiveness analysis. All analyses were carried out in Stata version 15 with a significance level set at 1% and 5%.

The study was carried out in accordance with the General Data Protection Regulation (2015-509-00007). In accordance with the Danish National Committee on Health Research Ethics, this research satisfies the criteria of being 'questionnaire and register-based research excluding human biological material', and thus was not required to undergo a formal ethics procedure [29].

Results

A total of 729 patients were registered at baseline as having completed the rehabilitation programme. At baseline, 454 patients completed the ICECAP-A. Of these, 155 completed the ICECAP-A at follow-up, and this population was used in the following analyses. The baseline characteristics for the complete cases and for the whole sample are presented in Table 2. More men were included, and just over half were aged over 65 years, with a similar proportion being retired. Around two thirds were living with a spouse and approximately half had a medium level of education, with a similar number having a low as a high level of education.

Table 2 Baseline characteristics

Characteristics	**Category**	**Frequency (%) whole sample n=729**	**Frequency (%) included in construct validity n=454**	**Frequency (%) included in responsiveness n=155**
Gender				
	Female	305 (42%)	183 (40%)	61 (39%)
	Male	424 (58%)	271 (60%)	94 (61%)
Age				
	18–29	4(1%)	2(0.5%)	-
	30–44	31 (4%)	21(4.5%)	6 (4%)
	45–64	284 (39%)	186(41%)	72 (46%)
	65+	410 (56%)	245(54%)	77 (50%)
Occupation				
	Retired	407 (56%)	242(53%)	84 (54%)
	Employed	196 (27%)	142(31%)	51 (33%)
	Unemployed/ other benefits	126 (17%)	70(16%)	20 (13%)
Education (based on ISCED classification)				
	Low	193 (26%)	106(23%)	38 (24%)
	Medium	356 (49%)	222(49%)	63 (41%)
	High	180 (25%)	126(28%)	54 (35%)
Cohabiting				
	Cohabiting	483 (66%)	301(34%)	104 (67%)
	Non-cohabiting	246 (34%)	153(66%)	51 (33%)
Diagnosis				
	Diabetes	297 (41%)	166(36%)	56 (36%)
	Cardiovascular	215 (29%)	148(33%)	57 (37%)
	COPD	217 (30%)	140(31%)	42 (27%)
Baseline scores	**Measure range**			
Un-weighted score	5–20	16.63	16.63	16.65
Weighted tariff scores	0–1	0.87	0.87	0.88
General health	1–5	2.86	2.86	2.85
'Feeling fit to do the things I want to'	1–4	3.47	3.50	3.52

Patients' responses (complete cases) at baseline and follow-up are listed in Table 3. The baseline weighted tariff scores was 0.87 and the follow-up weighted tariff scores was 0.89, thus a change of 0.02. The majority of responses had the highest or second-highest level of capabilities for each of the five attributes. Nevertheless, some patients indicated that their capability level was limited (little or no capability) in most of the five attributes. However, the proportion was small (<5 patients), and in the autonomy attribute, there were no responses at the lowest level at follow-up. The percentage of patients reporting the highest response level increased for each of the attributes between baseline and follow-up data collection.

Table 3 Patient rehabilitation responses to ICECAP-A measure at baseline and 12 weeks follow-up. Baseline weighted tariff scores was 0.87 and follow-up weighted tariff scores was 0.89, thus a change of 0.02.

ICECAP-A attributes (n=155)	Baseline frequency (%)	Follow-up frequency (%)
Stability		
I am able to feel settled and secure in all areas of my life	56 (36%)	67 (43%)
I am able to feel settled and secure in many areas of my life	89 (57%)	79 (51%)
I am able to feel settled and secure in a few areas of my life	6 (4%)	8 (5%)
I am unable to feel settled and secure in any areas of my life	4 (3%)	1 (1%)
Attachment		
I can have a lot of love, friendship and support	82 (53%)	83 (54%)
I can have quite a lot of love, friendship and support	57 (36%)	58 (37%)
I can have a little love, friendship and support	15 (10%)	14 (9%)
I cannot have any love, friendship and support	1 (1%)	-
Autonomy		
I am able to be completely independent	77 (50%)	83 (54%)
I am able to be independent in many things	70 (45%)	67 (43%)
I am able to be independent in a few things	4 (2.5%)	5 (3%)
I am unable to be at all independent	4 (2.5%)	-
Achievement		
I can achieve and progress in all aspects of my life	30 (19%)	40 (26%)
I can achieve and progress in many aspects of my life	105 (68%)	106(68%)
I can achieve and progress in a few aspects of my life	16 (10%)	8 (5%)
I cannot achieve and progress in any aspects of my life	4 (3%)	1 (1%)
Enjoyment		
I can have a lot of enjoyment and pleasure	88 (57%)	99 (64%)
I can have quite a lot of enjoyment and pleasure	59 (38%)	49 (32%)
I can have a little enjoyment and pleasure	5 (3%)	7 (4%)
I cannot have any enjoyment and pleasure	3 (2%)	-

Construct validity

Table 4 shows the associations between selected variables and ICECAP-A attributes at baseline. Of the 26 hypothesised associations, 16 (62%) were in the expected direction. Hypothesised associations that did not meet our a priori tests were (1) education, cohabitation and the stability attribute, (2) employment and the attachment attribute, (3) employment, education (negative correlation, but close to zero -0.0005) and the autonomy attribute, (4) education, cohabitation and the achievement attribute, (5) employment and the enjoyment attribute, and (6) employment, education and the weighted tariff score. In contrast, the associations between general health and the attachment attribute, were not hypothesised.

Table 4 Construct validity: Test of association by p-values along with the correlation between ICECAP-A, baseline characteristics and the questionnaire from the healthcare centre, using the chi-squared test and correlation matrix

	Stability	Attachment	Autonomy	Achievement	Enjoyment	Un-weighted score	Weighted tariff score
Association							
General health	**0.00***	0.03**	**0.00***	**0.00***	**0.00***	**0.00***	**0.00***
'Feeling fit to do the things I want to'	**0.00***	**0.00***	**0.00***	**0.00***	**0.00***	**0.00***	**0.00***
Employment	**0.00***	**0.51**	**0.70**	**0.00***	**0.13**	**0.26**	0.15
Education	**0.25**	0.08	**0.17**	**0.66**	0.11	**0.23**	0.31
Cohabitation	**0.09**	**0.00***	0.08	**0.13**	**0.03****	**0.00***	**0.00***
Gender	0.78	0.28	0.14	0.25	0.35	0.06	0.07
Age	0.21	0.11	0.46	0.27	0.36	0.43	0.40
Correlation							
General health	0.51	0.33	0.31	0.43	0.46	0.54	0.50
'Feeling fit to do the things I want to'	0.50	0.28	0.34	0.39	0.44	0.52	0.52

* Statistically significant (in expected direction) with p-values <0.01
**Statistically significant (in expected direction) with p-values <0.05
Bold= hypothesised

Responsiveness

Based on the correlations, analyses of general health and 'feeling fit to do the things I want to' were chosen as anchors, as both reached strong correlation (0.54 and 0.52) and were therefore appropriate to use as anchors (see Table 4). Table 5 shows the change in un-weighted and weighted tariff scores in groups that reported improved (n=70) and worsened (n=16) general health scores. In groups that reported improved general health scores, ICECAP-A scores increased (0.05), and in the groups that reported a worsening of general health scores, ICECAP-A scores decreased (-0,06). The ES and SRM for those reporting an improvement in general health were small for both the un-weighted and weighted tariff scores; for those who reported a worsening in general health scores, the ES and SRM were moderate to strong. The ES and SRM in ICECAP-A scores were more substantial in the groups that reported a worsening of general health than improvement.

Table 5 Responsiveness: Mean changes in un-weighted scores and weighted tariff scores by anchor change groups (n=155)

Anchor group	Baseline ICECAP-A (SD)	Follow-up ICECAP-A (SD)	Mean ICECAP-A change (95% CI)	Difference in SD	ES[a]	SRM[b]
General health						
Un-weighted scores						
Improved (n=70)	16.39 (2.64)	17.53 (2.07)	1.14 (0.69;1.560)*	1.90	0.43	0.6
No change (n=69)	16.76 (2.54)	16.88 (2.32)	0.12 (-0.22;0.46)	1.41	0.05	0.09
Worsened (n=16)	17.25 (1.24)	16.19 (1.97)	-1.06 (-1.69;-0.43)*	1.18	-0.85	-0.90
Weighted tariff score						
Improved (n=70)	0.87 (0.16)	0.92 (0.09)	0.05 (0.03;0.08)*	0.11	0.32	0.45
No change (n=69)	0.88 (0.02)	0.89 (0.12)	0.01 (-0.01;0.03)	0.08	0.5	0.13
Worsened (n=16)	0.92 (0.05)	0.86 (0.12)	-0.06 (-0.10;-0.01)*	0.09	-1.2	0.67
'Feeling fit to do the things I want to'						
Un-weighted scores						
Improved (n=37)	15.73 (2.75)	16.84 (2.29)	1.11 (0.54;1.67)*	1.70	0.40	0.65
No change (n=103)	16.94 (2.45)	17.32 (2.17)	0.38 (0.04;0.72)*	1.74	0.16	0.22
Worsened (n=15)	16.87 (1.50)	16.27 (2.09)	-0.6 (-1.51;0.31)	1.64	-0.40	0.37
Weighted tariff score						
Improved (n=37)	0.83 (0.18)	0.89 (0.12)	0.06 (0.03;0.10)*	0.10	0.33	0.6
No change (n=103)	0.89 (0.14)	0.91 (0.11)	0.02 (-0.001;0.04)	0.09	0.14	0.22
Worsened (n=15)	0.90 (0.05)	0.87 (0.10)	-0.03 (-0.08;0.01)	0.08	-0.60	0.38

*Statistically significant with p-values < 0.05
[a]ES (effect size) – mean ICECAP-A change/SD of baseline scores
[b]SRM (standardised response mean) – mean change/SD of the difference

Table 5 shows the change in un-weighted and weighted tariff scores in groups that reported improved (n=37) and worsened (n=15) 'freedom' scores. In groups that reported improved freedom scores, ICECAP-A scores increased (0.06), and in the groups that reported a worsening of freedom scores, ICECAP-A scores decreased (-0,03). The change in ICECAP-A scores was more substantial in the groups that reported an improvement of freedom. The ES and SRM for those reporting an improvement in freedom were small to moderate for both the un-weighted and weighted tariff scores; for those who reported a worsening in freedom scores, the ES and SRM were small.

Subgroup analysis of responsiveness in different age groups

The results concerning responsiveness in the different age groups (Table 6) showed small differences, with the younger age group having a higher mean change, ES and SRM than the older group. In anchor group GH <65 the improved patients had a weighted tariff score of 0.85 at baseline and 0.91 at follow-up – mean change 0.06. The worsened group <65, had a weighted tariff score of 0.91 at baseline and 0.85 at follow-up – mean change -0.06. In the group >65 the patients had a weighted tariff score of 0.88 at baseline and 0.93 at follow-up – mean change 0.05 in the improved group. The worsened group ≥65, had a weighted tariff score of 0.91 at baseline and 0.86 at follow-up – mean change -0.05. In the anchor group 'Feeling fit to do the things I want to' the patients that improved had a weighted tariff score of 0.82 at baseline and 0.88 at follow-up – mean change 0.06. The worsened group had a weighted tariff score of 0.89 at baseline and 0.82 at follow-up – mean change -0.07. In the group ≥65 with GH as anchor the patients that improved had a weighted tariff score of 0.83 at baseline and 0.91 at follow-up – mean change 0.08. The worsened group ≥65, had a weighted tariff score of 0.92 at baseline and 0.93 at follow-up – mean change -0.0003. The sample size was small; however, the distribution was 50/50 between groups. More respondents improved they general health (n=35) compared with those improving in 'Feeling fit

to do the things I want to' (n=15). The ES and SRM were larger in the <65 groups. In the <65 group, both the improved and worsened mean change were statistically significantly different between baseline and follow-up. This was only the case with the improved group in the ≥65 subgroup. Results concerning freedom showed small ES and SRM in both age groups, but smallest in the ≥65 subgroup.

Table 6 Responsiveness: Mean changes in un-weighted scores and weighted tariff scores by age groups

Anchor group	Baseline ICECAP-A (SD)	Follow-up ICECAP-A (SD)	Mean ICECAP-A change (95% CI)	Difference in SD	ES[a]	SRM[b]
General health						
Age group under 65						
Un-weighted scores						
Improved (n=35)	16.03 (2.97)	17.46 (2.42)	1.43* (0.80;2.06)	1.85	0.48	0.77
No change n=33)	16.57 (2.62)	16.60 (2.52)	0.03 (-0.39;0.44)	1.16	0.01	0.02
Worsened n=10)	17.3 (1.34)	16.1 (1.91)	-1.2* (-2.08;-0.32)	1.23	-0.90	-0.98
Weighted tariff score						
Improved (n=35)	0.85 (0.19)	0.91 (0.12)	0.06* (0.02;0.10)	0.11	0.32	0.55
No change (n=33)	0.86 (0.15)	0.87 (0.14)	0.01 (-0.01;0.03)	0.55	0.07	0.02
Worsened (n=10)	0.91 (0.05)	0.85 (0.11)	-0.06* (-0.12;-0.004)	0.08	-1.2	0.75
Age group 65+						
Un-weighted scores						
Improved (n=35)	16.74 (2.27)	17.6 (1.68)	0.85* (0.20;1.52)	1.93	0.37	0.44
No change (n=36)	16.94 (2.48)	17.14 (2.11)	0.19 (-0.35;0.74)	1.62	0.08	0.12
Worsened (n=6)	17.17 (1.17)	16.33 (2.25)	-0.83 (-2.06;0.40)	1.60	-0.71	-0.52
Weighted tariff score						
Improved (n=35)	0.88 (0.12)	0.93 (0.05)	0,05* (0.01;0.08)	0.10	0.42	0.5
No change (n=36)	0.89 (0.13)	0.90 (0.11)	0.01 (-0.02;0.04)	0.11	0.08	0.1
Worsened (n=6)	0.91 (0.04)	0.86 (0.14)	-0.05 (-0.16;0.06)	0.09	-1.3	-0.6
'Feeling fit to do the things I want to'						
Age group under 65						
Un-weighted scores						
Improved (n=22)	15.55 (2.84)	16.68 (2.71)	1.14* (0.55;1.72)	1.32	0.40	0.86
No change (n=48)	16.81 (2.65)	17.31 (2.25)	0.50 (-0.02;1.02)	1.80	0.19	0.28
Worsened (n=8)	16.5 (1.93)	15.25 (2.12)	-1.25 (-2.57;0.07)	1.58	-0.65	-0.79
Weighted tariff score						
Improved (n=22)	0.82 (0.19)	0.88 (0.14)	0,06* (0.02;0.08)	0.07	0.32	0.71
No change (n=48)	0.88 (0.16)	0.90 (0.12)	0.02 (-0.00;0.05)	0.1	0.13	0.1
Worsened (n=8)	0.89 (0.07)	0.82 (0.12)	-0.07 (-0.14;0.00)	0.08	-1	-0.38
Age group 65+						
Un-weighted scores						
Improved (n=15)	16(2.67)	17.06 (1.53)	1.06 (-0.14;2.28)	2.19	0.40	0.48
No change (n=55)	17.05(2.12)	17.33(2.12)	0.27 (-0.19;0.73)	1.69	0.13	0.39
Worsened (n=7)	17.29 (0.76)	17.43 (1.40)	0.14 (-1.21;1.50)	1.46	0.18	0.1
Weighted tariff score						
Improved (n=15)	0.83 (0.16)	0.91 (0.06)	0.08 (-0.00;0.16)	0.14	0.50	0.57
No change (n=55)	0.90 (0.11)	0.91 (0.10)	0.01 (-0.01;0.03)	0.09	0.50	0.11
Worsened (n=7)	0.92 (0.03)	0.93 (0.04)	0.003 (-0.03;0.04)	0.04	0.10	0.08

*Statistically significant with p-values <0.05
[a]ES (Effect size) – mean ICECAP-A change/SD of baseline scores
[b]SRM (Standardised response mean) – mean change/SD of the difference

The item-by-item analysis (Table 7) showed that in the group of patients reporting an improvement in general health, the largest increase was in stability and in the patients reporting worsening of general health, the biggest decrease was in autonomy. In the group of patients reporting an improvement in 'feeling fit to do the things I want to', the increase was

comparable across attributes with increases in attachment lowest, and in the patients reporting worsening in 'feeling fit to do the things I want to', the biggest decreases were seen in autonomy.

Table 7 Item-by-item analysis: Distribution of changed response according to anchor

Anchor		Change between baseline and follow-up (%)				
		Stability	Attachment	Autonomy	Achievement	Enjoyment
General health						
	⇧	15 (21)	3 (4)	9 (13)	12 (17)	10 (14)
Improved (n=70)	=	55 (79)	67 (96)	61 (87)	68 (83)	60 (86)
	⇩	-	-	-	-	-
	⇧	-	2 (12.5)	-	-	-
Worsened (n=16)	=	13 (81)	12 (75)	11 (69)	14 (88)	13 (81)
	⇩	3 (19)	2 (12.5)	5 (31)	2 (12)	3 (19)
'Feeling fit to do the things I want to'						
	⇧	5 (14)	3 (8)	6 (16)	6 (16)	5 (14)
Improved (n=37)	=	32 (86)	34 (92)	31 (84)	31 (84)	32 (86)
	⇩	-	-	-	-	-
	⇧	-	-	-	-	-
Worsened (n=15)	=	14 (93)	13 (87)	10 (67)	14 (93)	15 (100)
	⇩	1 (7)	2 (13)	5 (33)	1 (7)	-

Discussion

This is the first study to assess the construct validity and responsiveness of the Danish ICECAP-A measure. To achieve this, it used longitudinal data from a rehabilitation setting in a population of chronically ill patients. The findings indicate that scores on the Danish ICECAP-A are associated with indicators of freedom and general health. The results provide evidence about the instrument's ability to respond to differences in socioeconomic characteristics such as employment, education and cohabitation. The responsiveness analysis explored changes in the ICECAP-A scores in response to general health and freedom, and the results indicate that the ICECAP-A is responsive and that patients younger than 65 years of age appear more responsive than older patients. The Danish ICECAP-A, therefore, demonstrated encouraging construct validity and responsiveness in a rehabilitation setting among chronically ill patients. The item-by-item analysis showed that those reporting an increase in general health and 'Feeling fit to do the things I want to' scores the largest change in Achievement and autonomy respectively, and those reporting an decreased general health and 'Feeling fit to do the things I want to' score the largest change was found in autonomy in both.

The overall findings are consistent with previous studies that found the ICECAP-A to be promising in terms of validity [13–15] and responsiveness [14, 16] in different populations and health conditions. The most comparable is the study by Al-Janabi et al. [13], where the ICECAP-A was found to be associated with various socioeconomic variables, the EQ-5D, and questions concerning freedom and opportunities. The most noticeable result was that the present study found an association between general health and the attribute attachment where Al-Janabi et al. found the opposite. Al-Janabi et al. did, however, find an association between anxiety and depression and attachment. This could indicate that the participants in this study considered mental health to be a part of general health, which could relate to differences in the setting, but could also reflect the increasing focus on mental health across society more generally since the Al-Janabi research was published in 2013.

The study benefits from the available Danish ICECAP-A translation (discussed elsewhere [11]) that made it possible to investigate the psychometric properties of ICECAP-A. Further, this study extends our academic knowledge around accurate outcomes assessment in the context of rehabilitation medicine among chronically ill patients. ICECAP-A is still a relatively new questionnaire, and so developing a better understanding of the tool's validity and responsiveness across populations is essential for its further use in health economic evaluations. Previous studies have demonstrated construct validity in different populations, including the general British population [13], women with irritable lower urinary tract symptoms [14] and a population with depression [15].

One methodological limitation of the study is the small number of possible anchors and lack of clinical anchors. While the use of general health as an anchor was driven by methodological considerations when considering a capability measure's suitability for use in health interventions, it is essential to identify how the instrument responds to changes in health. Health is one of many factors that affect the capability of a person and a relevant factor in this study population in particular. A smaller change in capability scores would, therefore, be expected in response to changes in health, and could have been useful to investigate with more anchors than general health. A previous study used EQ-5D as an anchor, in a population with depression, resulting in a correlation between all attributes of the ICECAP-A [15]. This study had a large proportion of missing data in term of patients not having both a baseline and follow-up measures. The missing was anticipated to be missing completely at random because the entire questionnaires was missing. The amount of missing may be due to that fact that it was voluntary completion of the questionnaire, both at baseline and follow-up. This could influence the results if the sample is different from the missing data and decrease the power of the sample. However, the proportion of missing was assumed too large (78%) to impute.

The evidence of validity and responsiveness presented in this study adds to the psychometric profile of the ICECAP-A measure, and the results provide an initial indication that the ICECAP-A may be responsive in public health research and chronically ill populations. In the Danish municipal rehabilitation setting, no national outcome measurement procedures exist, so a more extensive study with more participating municipalities would be interesting to explore the implications further. Establishing the psychometric performance of a measure is a continuous process, and further research is needed to explore how well the ICECAP-A performs in different public health and social care settings, such as in interventions regarding self-care. Ideally, capability measures could be incorporated into future health agreements and clinical guidelines. More importantly, it is necessary to show personnel in healthcare centres and decision-makers the benefits of implementing ICECAP-A in everyday work as a tool in public health and social care interventions, and not just as a scientific instrument.

Conclusion

This study provides the first investigation into construct validity and responsiveness to change for the Danish translation of the ICECAP-A and the first investigation into responsiveness to change for any ICECAP measure in the context of CVD, COPD and diabetes. The Danish ICECAP-A has demonstrable potential for accurately measuring the effect of rehabilitation. Furthermore, it appears to be responsive in terms of capturing the effects on general health and the freedom to do things. Future research into the psychometric properties of the Danish ICECAP-A would be beneficial to clinicians and decision-makers in Denmark interested in capturing broader benefits to patients, beyond just health.

References

1. Sen, A. (1997). *Inequality reexamined*. *Quality of life*. Cambridge, MA: Harvard University Press.

2. Robeyns, I. (2017). *Wellbeing, freedom and social justice – The capability approach re-examined*. Cambridge, UK: Open Book Publisher.

3. Sen, A. (1993). Capability and well-being. In M. Nussbaum & A. Sen (Eds.), *Quality of life* (pp. 30–53). Oxford, UK: Oxford University Press.

4. Al-Janabi, H., Flynn, T. N., & 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.

5. Lorgelly, P. K., Lorimer, K., Fenwick, E. A. L., Briggs, A. H., & Anand, P. (2015). Operationalising the capability approach as an outcome measure in public health: The development of the OCAP-18. *Social Science and Medicine*, *142*, 68–81. https://doi.org/10.1016/j.socscimed.2015.08.002

6. Simon, J., Anand, P., Gray, A., Rugkåsa, J., Yeeles, K., & Burns, T. (2013). Operationalising the capability approach for outcome measurement in mental health research. *Social Science & Medicine*, *98*, 187–196.

7. Netten, A., Burge, P., Malley, J., Potoglou, D., Towers, A. M., Brazier, J., Wall, B. (2012). Outcomes of social care for adults: Developing a preference-weighted measure. *Health Technology Assessment*, *16*(16), 1–165.

8. Flynn, T. N., Huynh, E., Peters, T. J., Al-Janabi, H., Clemens, S., Moody, A., & Coast, J. (2015). Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. *Health Economics*, *24*(3), 258–69. https://doi.org/10.1002/hec.3014

9. Al-Janabi, H., Flynn, T. N., Peters, T. J., Bryan, S., & Coast, J. (2015). Test–retest reliability of capability measurement in the UK general population. *Health Economics*, *24*(5), 625–630.

10. Tang, C., Xiong, Y., Wu, H., & Xu, J. (2018). Adaptation and assessments of the Chinese version of the ICECAP-A measurement. *Health and Quality of Life Outcomes*, *16*, 45.

11. Holst-Kristensen, A. W., Fonager, K., & Pedersen, K. M. (2019, October). Test–retest reliability of ICECAP-A in the adult Danish population. *Quality of Life Research*. https://doi.org/10.1007/s11136-019-02331-5

12. Keeley, T., Al-Janabi, H., Lorgelly, P., & Coast, J. (2013). A qualitative assessment of the content validity of the ICECAP-A and EQ-5D-5L and their appropriateness for use in health research. *PLoS ONE*, *8*(12), 85287.

13. Al-Janabi, H., Peters, T. J., Brazier, J., Bryan, S., Flynn, T. N., Clemens, S., … Coast, J. (2012). An investigation of the construct validity of the ICECAP-A capability measure. *Quality of Life Research*, *22*(7), 1831–1840.

14. Goranitis, I., Coast, J., Al-Janabi, H., Latthe, P., & Roberts, T. E. (2016). The validity and responsiveness of the ICECAP-A capability-well-being measure in women with irritative lower urinary tract symptoms. *Quality of*

Life Research, *25*, 2063–2075.

15. Mitchell, P. M., Al-Janabi, H., Byford, S., Kuyken, W., Richardson, J., Iezzi, A., & Coast, J. (2017). Assessing the validity of the ICECAP-A capability measure for adults with depression. *BMC Psychiatry*, *17*, 46.

16. Keeley, T., Al-Janabi, H., Nicholls, E., Foster, N. E., 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*, 2319–2331.

17. Johansen, J., Rahbek, J., Møller, K., & Jensen, L. (2004). *Hvidbog om rehabiliteringsbegrebet. Rehabilitering i Danmark*. Retrieved from http://www.marselisborgcentret.dk/fileadmin/filer/Publikationer/PDF_er/Hvidbog.pdf

18. Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales, a practical guide to their development and use* (5th ed.). Oxford, UK: Oxford University Press.

19. Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302.

20. Kane, M. (2001.). Current concerns in validity theory. *Journal of Educational Measurement*, *38*(4), 319–342.

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

22. Cohen, J. (1988). Set correlation and contingency tables. *Applied Psychological Measurement*, *12*(4), 425–434.

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

24. Jayadevappa, R., Malkowicz, S. B., Wittink, M., Wein, A. J., & Chhatre, S. (2012). Comparison of distribution- and anchor-based approaches to infer changes in health-related quality of life of prostate cancer survivors. *Health Services Research*, *47*(5), 1902–1925.

25. Wyrwich, K. W., Norquist, J. M., Lenderking, W. R., & Acaster, S. (2013). Methods for interpreting change over time in patient-reported outcome measures. *Quality of Life Research*, *22*, 475–483.

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

27. Brazier, J., & Deverill, M. (1999). A checklist for judging preference-based measures of health related quality of life: Learning from psychometrics. *Health Economics*, *8*, 41–51.

28. Flynn, T. N., Huynh, E., Peters, T. J., Al-Janabi, H., Clemens, S., Moody, A., & Coast, J. (2015). Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. *Health Economics*, *24*(3), 258–269.

29. Ministry of Health. (2017). *Lov om videnskabsetisk behandling af sundhedsvidenskabelige forskningsprojekter*. Retrieved from https://www.retsinformation.dk/pdfPrint.aspx?id=192671

Appendix C. Determinants of Participation and Drop-out in Municipal Rehabilitation

Determinants of Participation and Drop-out in Municipal Rehabilitation

Annette W Holst-Kristensen[1], Kirsten Fonager[2], Kjeld Møller Pedersen[1]

1 Danish Center for Healthcare Improvements, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
2 Department of Clinical Medicine, Aalborg University, Denmark; Department of Social Medicine, Aalborg University Hospital, Denmark

Abstract

Background: Rehabilitation can be a central factor for patients with chronic illness when it comes to regaining and maintaining functional levels and thereby a meaningful and independent life. Despite the evidence of the effects, numerous patients fail to adhere to and complete rehabilitation programmes. This study, therefore, aims to investigate; the rates of attendance, drop-out and non-attendance, and the sociodemographic and clinical predictors for attendance, drop-out and non-attendance and to analyse possible gender differences in a Danish municipal rehabilitation programme. The objective is to give healthcare professionals at the healthcare centre a better understanding of why patients fail to attend and/or complete a rehabilitation programme, and potentially enable healthcare centres to develop targeted activities that encourage attendance. **Methods**: The study uses a multinomial logistic regression to investigate attendance, drop-out and non-attendance as dependent variables. The interpretation is based on relative risk ratios. A subgroup analysis of gender is carried out and tested using seemingly unrelated estimations. **Results**: The results indicate that the risk of dropping out is significantly higher if the patients are employed or unemployed compared with retirement, have a low household income and have 1–4 comorbidities compared with no comorbidities. The strongest predictors for non-attendance is being single, having a low level of education and being unemployed. Comparing the two genders' regressions models indicates that there is no significant difference between females and males in the drop-out group and no significant difference between genders in the non-attendance group compared with attenders. **Conclusion**: The study found significant differences between sociodemographic variables and attendance status. However, there is no significant difference across predictor variables between genders as hypothesised.

Keywords: Municipal rehabilitation, prediction, attendance, participation, drop-out

Introduction

There is solid evidence-based knowledge about the positive effects of different rehabilitation programmes for patients with chronic illness [1, 2]. Increased life expectancy, and an increasing number of patients living longer with chronic illness increases the importance of preventive initiatives, including municipal rehabilitation programmes as tertiary initiatives [3]. The programmes are known to facilitate psychological and physical recovery following acute events and actively rehabilitate patients with chronic illness [3–8]. In Denmark, approximately 500,000 people suffer from cardiovascular disease (CVD), chronic obstructive pulmonary disease (COPD) and/or diabetes (DM) [9], all of which are candidate conditions for municipal rehabilitation.

Denmark has a network of municipal rehabilitation programmes offered by the 98 Danish municipalities to chronically ill patients with, for example, CVD, COPD and DM. The programmes provide exercise, education and support. They are offered to groups of varying size and often across several disease groups (e.g. COPD and DM). The exercise sessions take place one to two times a week and are of low to moderate intensity. The education component covers knowledge of the disease, dietary advice and the importance of physical activity, smoking cessation, medicine consumption and help participants to set goals and stimulate motivation. The programmes usually commence within a few weeks after discharge from the hospital and continue for 8–12 weeks [5] and are not offered routinely to chronically ill persons.

Rehabilitation can be a crucial factor for people with chronic illness when it comes to regaining and maintaining functional levels and thereby a meaningful and independent life [10]. Despite the evidence of the effects, numerous patients fail to adhere to and complete rehabilitation programmes. Participation among CVD and COPD patients is usually low (20–50%) [11, 12]. Specifically, patients who are smokers, physically inactive, unmarried, unemployed and low educated are less likely to attend [11, 13–15]. Additional predictors for drop-out and non-attendance for CVD are being female, older than 70 years, having depression and low perception of illness. Similar reasons were found for COPD along with worsening in other medical reasons and low baseline health status [11–18].

Although men and women achieve the same effects from rehabilitation, studies have shown that drop-out and non-attendance are more common among women [11, 19], and the reasons for drop-out and non-attendance are different between gender [11, 15]. However, the degree to which the sociodemographic predictors are the same for men and women are less explored, and no studies concerning rehabilitation for chronic illness in general as one non-disease-specific group were found.

This study aims to investigate; (i) the rates of attendance, drop-out and non-attendance to a municipal rehabilitation programme (ii) the sociodemographic and clinical predictors for drop-out and non-attendance and (iii) whether any possible gender differences in predictor variables are present. The objective is to give healthcare professionals at the healthcare centre a better understanding of why patients fail to attend and/or complete a rehabilitation programme. The results of this research may potentially enable healthcare centres to develop targeted activities that encourage attendance for all referred patients.

Methods

Study population

The study population is from the municipality of Aalborg, Denmark, with a population of app. 200,000 inhabitants. The inclusion criteria for this study were: patients referred to municipal rehabilitation over the period from 2007 to 2014, patients in the target group (CVD, COPD or DM), patients with attendance status and patients with residence in the municipality of Aalborg. The attendance status is registered by the healthcare centre and aggregated to three categories: attendance, drop-out and non-attendance. Attenders are the patients with a baseline and follow-up interview, drop-outs are the ones with a baseline interview, but no follow-up and lastly, the non-attenders are referred patients who never showed up for rehabilitation. The municipal rehabilitation programme is not necessarily disease specific. The municipality of Aalborg has sufficiently large classes to divide the participants by disease, but this is not the case for all Danish municipalities.

Data and model structure

Data from the rehabilitation centre (from 2007-2014) included personal identification number, attendance status, diagnoses, gender and age (n=2,655). These data were combined with registry data from Statistics Denmark one year before the referral date to rehabilitation programmes. The registry data included marital status, education, socioeconomic status (employment status), household income and healthcare utilisation. The study uses a multinomial logistic regression (mlogit, STATA 15) method to estimate three minus one models with attendance as the reference case (e.g. the probability of drop-out compared to attendance and the probability of non-attendance compared to attendance). The interpretation is based on the relative risk ratio (RRR). The dependent variable is a multinomial variable with three classes: attendance; drop-out and non-attendance. All independent variables were characteristics identified *a priori* based on the literature [11–18, 20]. Sociodemographic characteristics included age, gender (female/male), marital status (married, separated/divorced, widowed, never married), education level (low <11 years of schooling, middle 11–16 years of schooling, high >16 years of schooling), socioeconomic status (employed, unemployed/other benefits, retired), taxable

income (Euro, €) as a household variable (family income and personal income if no family income is registered). Clinical information included diagnosis, and Charlson comorbidity index [21] created as a category variable with 0, 1–2, 3–4, >4 comorbidities. Furthermore, health utilisation including the number of visits at the GP/specialists visits, outpatient visits and hospital admissions one year prior to the referral date. These variables were included as a rough proxy for health status.

As a subgroup analysis, gender is investigated using two separate regressions—one for female and one for male.

Statistical analysis

Descriptive statistics (frequencies, mean, standard deviation and median) are used to present patients' baseline-characteristics, attendance, drop-out and non-attendance rates. For this purpose, t-tests and Chi-square tests are used when appropriate. The significance level was set at p-value<0.05. Group comparison of gender is tested by a mlogit post estimate test and with a Wald chi-square test with the use of seemingly unrelated estimation commands (suest, STATA 15) The data were analysed using STATA 15.

The study was carried out in accordance with the General Data Protection Regulation (2015-509-00007). Moreover, according to The National Committee on Health Research Ethics, no approval was required.

Results

Sociodemographic characteristics

A total of 4,361 patients were referred to the municipal rehabilitation programme in Aalborg from 2007–2014. Of those, 2,655 met the inclusion criteria, and the remaining 1,706 had no attendance status. The baseline characteristics are given in Table 1. The mean age is 65–66 years across attendance groups and gender. Most patients in the study were married (44–61%), and of these more men than women were married (61% vs 47%); in contrast, more women were widowed (25%) than men (8%). The rate of divorce was higher among the patients who dropped out or were non-attenders (21–22%). The majority of the patients had a medium length of education (43–48%) and were more often men than women (56% vs 38%). Socioeconomic status showed that most patients were retired (52–62%) and of these, most were women (62%). The rate of employment was even in the attendance groups, but more men (30%) than women (19%) were employed. The rate of unemployed patients was highest in the attendance group (15%). Looking at the diagnoses across attendance groups, patients with COPD were represented to a higher degree in the drop-out group 62%, and diabetes was less represented in the drop-out group (16%). Concerning gender, men were more frequently represented in the diagnoses group CVD. Lastly, 34–56% of the patients had no comorbidities, and 1–2% had more than four comorbidities.

Attendance and non-attendance are similar in proportion with 50% of the patients; in contrast, a larger proportion (65%) of the drop-out group had comorbidities (Table 2).

Table 1 Sociodemographic characteristics of all patients referred to the rehabilitation programme in the municipality of Aalborg, separately for attendance groups and gender.

	Attendance N=1626	Drop-out N=545	Non-attendance N=484	Male N=1277	Female N=1378
Age; mean (SD)	65.5 (10.5)	65.3 (11.3)	65.9 (11.7)	65.1 (10.9)	65.9 (10.8)
Civil status					
Married	*955 (59%)*	*263 (48%)*	*216 (44%)*	*780 (61%)*	*654 (47%)*
Separated/divorced	*251 (15%)*	*121 (22%)*	*100 (21%)*	*200 (16%)*	*272 (20%)*
Widowed	*251 (15%)*	*96 (18%)*	*97 (20%)*	*106 (8%)*	*338 (25%)*
Never married	*169 (11%)*	*65 (12%)*	*71 (15%)*	*191 (15%)*	*114 (8%)*
Education level					
Low	*607 (37%)*	*243 (45%)*	*220 (45%)*	*397 (31%)*	*673 (49%)*
Medium	*763 (47%)*	*240 (44%)*	*199 (41%)*	*693 (54%)*	*509 (37%)*
High	*228 (14%)*	*43 (8%)*	*47 (10%)*	*153 (12%)*	*165 (12%)*
Missing	*28 (2%)*	*19 (3%)*	*18 (4%)*	*34 (3%)*	*31 (2%)*
Socioeconomic status					
Employed	*431 (26%)*	*110 (20%)*	*103 (21%)*	*378 (30%)*	*266 (19%)*
Unemployed/ on benefits	*238 (15%)*	*143 (26%)*	*110 (23%)*	*228 (18%)*	*263 (19%)*
Retired	*956 (59%)*	*289 (54%)*	*271 (56%)*	*669 (52%)*	*847 (62%)*
Missing	*<5*	*<5*		*<5*	*<5*
Income (taxable €)					
<26,000	*236 (15%)*	*141 (26%)*	*102 (21%)*	*172 (13%)*	*307 (22%)*
26,000–40,000	*510 (31%)*	*201 (36%)*	*172 (35%)*	*407 (32%)*	*476 (35%)*
40,000–55,000	*328 (20%)*	*74 (14%)*	*98 (20%)*	*257 (20%)*	*243 (18%)*
55,000–70,000	*214 (13%)*	*44 (8%)*	*26 (5%)*	*154 (12%)*	*130 (9%)*
>70,000	*331 (20%)*	*76 (14%)*	*85 (18%)*	*281 (22%)*	*211 (15%)*
Missing	*7 (1%)*	*9 (2%)*	*1(<1%)*	*6 (<1%)*	*11 (1%)*

Italic cells are those where significant level is <0.05

Table 2 Clinical characteristics of all patients referred to a rehabilitation programme in the municipality of Aalborg separately for attendance groups and gender

	Attendance N=1626	Drop-out N=545	Non-attendance N=484	Male N=1277	Female N=1378
Diagnoses					
Cardiovascular disease	*393 (24%)*	*119 (22%)*	*120 (25%)*	*386 (30%)*	*246 (18%)*
COPD	*792 (49%)*	*338 (62%)*	*246 (51%)*	*564 (44%)*	*812 (59%)*
Diabetes	*441 (27%)*	*88 (16%)*	*118 (24%)*	*327 (26%)*	*320 (23%)*
Comorbidities					
0	*905 (56%)*	*185 (34%)*	*244 (50%)*	*672 (53%)*	*662 (48%)*
1–2	*594 (37%)*	*275 (51%)*	*206 (43%)*	*466 (36%)*	*609 (44%)*
3–4	*110 (6%)*	*77 (14%)*	*29 (6%)*	*117 (9%)*	*99 (7%)*
>4	*17 (1%)*	*8 (1%)*	*5 (1%)*	*22 (2%)*	*8 (1%)*
Health utilisation one year before referral, Median [IQR]					
Contact with GP/ Special GP/ Therapist	*93[65;134]*	*104[65;149]*	*93[62;146]*	*89[58;132]*	*101[71;145]*
Outpatient visit	7[3;17]	9[4;22]	8[3;17]	8[3;19]	8[3;17]
Hospital Admission	*5[3;9]*	*7[4;12]*	*6[3;11]*	6[3;10]	6[3;10]

Italic cells are those where significant level is <0.05
IQR – Interquartile range

Attendance rates

In total, 2,655 patients had an attendance status. Of those, 1,626 were registered with the status 'Attendance' meaning that they had completed the rehabilitation programme. The attenders were evenly distributed across gender—52% were

female and 48% male. Across attendance groups, the attenders accounted for 60% of all referred patients. No difference in the distribution across gender was found. The drop-out and non-attenders had the same pattern. Furthermore, there was no difference between genders, and both groups account for 20% of the referred patients. In other words, there is no difference in the distribution of female and males in the three attendance groups, and 60% of the referred patients completed the programme, 20% dropped out, and 20% of referred patients never showed up.

Sociodemographic and clinical differences across attendance groups

Table 3 shows the multinomial regression model comparing the attendance group (the reference group) with the drop-out group and with the non-attendance group. The results indicate that the risk of dropping out is significantly higher if the patient is employed or unemployed compared to retirement status, having a low household income and having 1–4 comorbidities. Having a household income >40,000 € decreases the risk of dropping out. The results indicate that the risk of non-attendance is associated with being single, low education level, being unemployed and having more hospital admissions than the attendance group. Having a household income >55,000 € decreases the risk of non-attendance. Testing the association between the groups (i.e., whether the models as a whole are different from each other and not just the individual predictors) showed that there is a significant difference (p-value 0.000) between all three groups: attendance, drop-out and non-attendance.

Table 3 Prediction variables for drop-out and non-attendance compared with the reference group – attendance.

	Drop-out model n=545		Non-attendance model n=484	
	RRR	95 % CI	RRR	95 % CI
Gender (male)	0.91	0.71–1.16	1.18	0.87–1.44
Age	1.00	0.99–1.02	1.02	0.99–1.03
Civil status				
Married (ref. group)	-	-	-	-
Separated/divorced	1.05	0.74–1.48	*1.47*	1.03–2.09
Widowed	0.73	0.50–1.36	*1.46*	1.01–2.12
Never married/single	0.88	0.57–1.36	*1.78*	1.18–2.67
Education level				
Low	1.14	0.88–1.49	*1.34*	1.03–1.75
Medium (ref. group)	-	-	-	-
High	0.66	0.43–1.01	0.84	0.56–1.25
Socioeconomic status				
Employed	*1.60*	1.06–2.39	1.17	0.78–1.74
Unemployed/on benefits	*2.08*	1.39–3.11	*1.84*	1.22–2.79
Retirement (ref. group)	-	-	-	-
Household income (taxable €)				
<26,000	*1.49*	1.07–2.09	1.01	0.71–1.45
26,000–40,000 (ref. group)	-	-	-	-
40,000–55,000	*0.55*	0.38–0.78	1.12	0.80–1.57
55,000–70,000	*0.59*	0.38–0.91	*0.53*	0.32–0.89
>70,000	*0.65*	0.42–0.99	1.26	0.83–1.90
Diagnoses				
CVD	0.98	0.72–1.30	1.01	0.74–1.36
COPD (ref. group)	-	-	-	-
Diabetes	0.86	0.62–1.20	0.93	0.68–1.30
Comorbidities				
0 (ref. group)	-	-	-	-
1–2	*1.85*	1.40–2.45	0.99	0.75–1.30
3–4	*2.74*	1.79–4.20	0.71	0.42–1.20
>4	1.58	0.62–4.04	0.80	0.27–2.38
Health utilisation				
Contact with GP/ Special GP/ Therapist	0.99	0.99–1	0.99	0.99–1.01
Outpatient visit	0.99	0.99–1	0.99	0.99–1.01
Hospital Admission	*1.03*	1.01–1.05	*1.03*	1.01–1.05

Italic cells are those where significant level is <0.05

Subgroup analysis of gender differences in sociodemographic and clinical predictors

The subgroup analysis for females and males separately, as shown in Table 4, demonstrated that the significant predictors for women in the drop-out group were that they were two times more at risk of dropping out if they were unemployed and 2–3 times more at risk if they had 1–4 comorbidities. In contrast, women were of less risk of dropping out if they were widowed. Compared with men in the drop-out group, the women were less likely to drop-out if they were living with a spouse, and the risk of dropping out because of comorbidities was higher for women. Comparing the women in the drop-out group with the non-attendance group again underlines the marital status and comorbidities as the distinctive predictor.

The significant predictors for men indicated that men were at two times higher risk of dropping out if they had 3–4 comorbidities and at lower risk of dropping out when having an income of >40.000€. The significant predictors for men in the non-attendance group indicated that being unmarried and having low education level were associated with higher

risk of non-attendance. Comparing with men in the drop-out group the non-attenders had lower risk if being employed on the contrary the men in the drop-out group were at higher risk if being employed.

Comparing the two regression models statistically with a suest test indicated that there was no overall significant difference (p-value=0.33) between females and males in the drop-out group and no overall significant difference between gender in the non-attendance group (p-value=0.32).

Table 4 Subgroup analysis: differences across predictor variables for men and women separately.

	Drop-out model				**Non-attendance model**			
	Female n=297		**Male n=248**		**Female n=239**		**Male n=245**	
	RRR	95 % CI	RRR	95 % CI	RRR	95 % CI	RRR	95 % CI
Age	1.00	0.97–1.03	1.01	0.99–1.03	*1.03*	1.00–1.06	1.00	0.98–1.03
Civil status								
Married	-	-	-	-	-	-	-	-
Separated/divorced	0.79	0.49–1.32	1.48	0.88–2.47	1.41	0.86–2.31	1.35	0.80–2.30
Widowed	*0.55*	0.34–0.90	1.15	0.60–2.19	1.37	0.84–2.23	1.41	0.74–2.68
Never married	0.55	0.27–1.11	1.32	0.74–2.36	1.56	0.82–2.96	*1.92*	1.18–3.32
Education level								
Low	1.12	0.78–1.62	1.11	0.75–1.63	1.04	0.72–1.53	*1.72*	1.18–2.51
Medium	-	-	-	-	-	-	-	-
High	0.66	0.37–1.21	0.68	0.36–1.28	0.81	0.47–1.42	0.83	0.46–1.48
Socioeconomic status								
Employed	1.21	0.65–2.24	1.73	1.00–2.99	1.63	0.89–2.99	0.81	0.47–1.84
Unemployed/other benefits	*2.05*	1.17–3.61	2.15	1.18–3.93	*2.79*	1.55–5.00	1.50	0.82–2.76
Retirement	-	-	-	-	-	-	-	-
Household Income (taxable €)								
<26,000	*1.79*	1.14–2.81	1.20	0.71–2.03	1.10	0.69–1.76	1.03	0.57–1.84
26,000–40,000	-	-	-	-	-	-	-	-
40,000–55,000	0.64	0.39–1.06	*0.47*	0.27–0.81	1.07	0.66–1.74	1.17	0.72–1.90
55,000–70,000	0.54	0.29–1.00	0.61	0.32–1.17	*0.43*	0.21–0.91	0.67	0.32–1.38
>70,000	0.76	0.41–1.40	0.56	0.30–1.02	0.96	0.51–1.79	1.66	0.94–2.95
Diagnoses								
CVD	1.05	0.68–1.62	0.81	0.53–1.23	0.94	0.60–1.46	1.07	0.70–1.64
COPD	-	-	-	-	-	-	-	-
Diabetes	1.03	0.65–1.63	0.68	0.53–1.23	0.93	0.60–1.45	0.94	0.58–1.55
Comorbidities								
0	-	-	-	-	-	-	-	-
1–2	*2.23*	1.51–3.31	1.49	0.98–2.27	0.94	0.64–1.38	1.17	0.77–1.77
3–4	*3.14*	1.71–5.78	*2.48*	1.34–4.60	0.44	0.19–1.02	1.12	0.55–2.28
>4	3.50	0.70–17.43	1.16	0.35–3.82	0.57	0.06–5.49	1.20	0.34–4.3
Health utilisation								
Contact with GP/ Special GP/ Therapist	1.00	0.99–1.02	1.0	0.99–1.0	1.0	0.99–1.0	0.99	0.99–1.01
Outpatient visit	1.00	0.99–1.01	1	0.99–1.0	1.0	0.99–1.0	*0.98*	0.97–0.99
Hospital Admission	*1.04*	1.02–1.07	1.02	0.99–1.0	*1.04*	1.01–1.06	*1.03*	1.01–1.06

Italic cells are those where significant level is <0.05

Discussion

Predicting which patient characteristics are associated with drop-out and non-attendance in rehabilitation programmes is essential in order to optimise the use of resources within municipalities and the health care system in general, and ultimately to address the needs of patients with CVD, COPD or DM. The most significant variables associated with drop-out were being employed or unemployed compared to being pensioners and having more than one comorbidity. The strongest predictors for non-attendance were living without a spouse, having a low level of education and being unemployed. Lastly, the gender subgroup analysis explored potential predictors for gender differences in rehabilitation.

The overall findings are consistent with previous studies[11, 13–15] that found that employment, comorbidities, marital status and education level are predictors for drop-out and non-attendance. Smoking, physical activity and depression have previously been investigated as predictors for drop-out and non-attendance. These variables were not available in this study but should be categorised as possible unobserved predictors.

The subgroup analysis found higher rates of attendance among women (52%) compared with past studies that found low (15–49%) attendance rates among women[11, 19]. The suest test comparing the two regressions models found no overall statistical difference between genders for either drop-out or non-attendance. However, the results from the subgroup analysis found a few significant differences between the individual prediction variables, hence, pointing towards a difference between females and males. The few significant predictor variables in the separate regressions are probably due to random variation in the data. However, looking at the marital status and comorbidities, these predictors stand out, where women are at lower risk of dropping out if they are single and men are at higher risk if they are single. Furthermore, women are at higher risk of dropping out when having comorbidities compared to women in the non-attending group, who are at lower risk if having comorbidities – both compared to women who attend. We found some significant separate predictors but no significant difference overall potentially because the unconstrained model only tests eight predictors, whereas the constrained model testes 16 predictors (8 for female and 8 for male). Hence, if all the coefficients across gender were the same, the chi-square statistic would not be significant, as in this case. Therefore, the apparent differences we found in the coefficients for each gender were significant but sufficiently small to be attributed as sampling errors.

Past studies of participation in rehabilitation have been concerned with one specific disease[11–18, 20]. When our study combines more diseases, it enables us to say something generally about the municipal rehabilitation programmes for chronically ill patients and makes it more transferable to other municipalities, where rehabilitation is often offered for the disease groups combined. The study shows that there were no differences across diseases, and hence the specific disease

had no impact on attendance rate. In Denmark, this is relevant because many of the 98 municipalities are too small to provide disease-specific programmes, and the content of classes offered is more or less the same. This is a useful finding from a resource utilisation view because the rehabilitation programmes can be a mix of patients with different diseases, and as a result of this, optimal take up in terms of filling up the classes could be achieved.

The definition of drop-out and non-attendance from rehabilitation is somewhat subjective. The municipality of Aalborg has a well-defined in-house system as used in this study, whereas previous studies usually defined drop-out as attending less than 50% of the programme, measured as the number of services, e.g. number of training classes. The Danish definition does not focus on registration of services and hence could not follow the previous definition. This, of course, affects the estimation of attendance rates and makes it difficult to compare with previous studies. It would be preferable if all municipalities implemented the same attendance status registration, and moreover, strive towards lowering the percentage of missing data in attendance status, to as low a level as possible. Consistent registration in all municipalities would enable making up national attendance rates. Ideally, the aim is that 85% of all referred patients start the rehabilitation programme[22]. This was almost the case; 82% of the patients with an attendance status started the programme.

The strength of this study lies in the use of Danish registries. The registries were used for socioeconomic and clinical predictors and then combined with the attendance status registered in the healthcare centre. This approach made it possible to investigate all referred patients and having an almost complete dataset. Moreover, the registries enabled us to investigate the non-attendance patients, information the healthcare centre has no possibility of collecting. The amont of missing data in the study was limited, thus imputation should have been considered. Furthermore the models are longitudinal and take no account for time and person-time.

The study is limited by the lack of appropriate data concerning referral. The authors know that the patients are referred to rehabilitation. However, details about who referred, and when the patient was referred in relation to hospital discharge date are lacking. The knowledge of who referred the patients was collected by the healthcare centre, but unfortunately, with many missing data points. An investigation of the referral frequency could give a better idea of the referral patterns, and the possible impact GPs and the hospital could have on participation. Therefore, it would potentially reveal whether there is inequality in the referral patterns. Additionally, the current practice in registering attendance status complicates good analyses for decision-making because of the substantial amount of missing data on attendance status. Furthermore, the study is limited by the narrow perspective of only investigating the municipality of Aalborg. It would have been

relevant to include more municipalities; however, this was not possible in the current study. Another limitation worth mentioning is the time period. The data were gathered from 2007–2014, during which time organisational changes within the healthcare centre might have influenced the results of this study. This research assumes that all patients received the same programme in the same setting; however, this might not have been the case and is, therefore, a limitation of the study.

The practical value of this work lies in the way the results may be applied. Health professionals working in healthcare centres can use these findings to encourage attendance amongst these high-risk profiles. The study is the first step in finding optimal and personalised rehabilitation programmes. Some of the significant predictors for both men and women for not attending rehabilitation were investigated, but the healthcare centres cannot act on the predictors alone. We now know that specific profiles are at higher risk of dropping out than others, but we do not know what would make them stay in the programme, and what would make the non-attenders show up in the first place. To investigate retainment of the referred patients in the programme, a qualitative study would be the next step to investigate why, for instance, unemployed patients drop-out and why unmarried patients are more reluctant to show up despite being referred and to investigate what would make them attend.

Conclusion

The results of this study indicate that there are significant differences across sociodemographic and clinical variables regarding attendance status. However, there is no overall significant difference between the predictor variables between genders, in contrast to the initial hypothesis. The study also demonstrated that there is no significant difference between the three disease groups in terms of attendance.

Reference list

1. Anderson L, Oldridge N, Thompson DR, Zwisler A-D, Rees K, Martin N, et al. Exercise-Based Cardiac Rehabilitation for Coronary Heart Disease. J Am Coll Cardiol. 2016;67:1–12.

2. McCarthy B, Casey D, Devane D, Murphy K, Murphy E, Lacasse Y. Pulmonary rehabilitation for chronic obstructive pulmonary disease (Review). Cochrane Database Syst Rev. 2015.

3. Handberg C, Nielsen C. Evaluering af rehabiliteringsindsatsen overfor borgere med kronisk sygdom i Brønderslev Kommune 2012. 2012. http://www.marselisborgcentret.dk/siteassets/viden-om-rehabilitering/publikationer/publikationer-om-rehabilitering/evaluering_af_rehabiliteringsindsatsen.pdf. Accessed 6 Feb 2019.

4. Clark AM, Hartling L, Vandermeer B, Mcalister F a. Meta-Analysis: Secondary Prevention Programs for Patients with Coronary Artery Disease. Ann Intern Med. 2005;143:659–72.

5. Marselisborg Centret. Rehabilitering i Danmark Hvidbog om rehabiliteringsbegrebet. http://www.rehabiliteringsforum.dk/contentassets/519c219e42a3454da34db60a8454de58/hvidbog.pdf. Accessed 9 Nov 2018.

6. Pedersen PM, Olsen TS. National klinisk retningslinje for Rehabilitering af patienter med KOL. 2018. https://www.sst.dk/-/media/Udgivelser/2018/National-Klinisk-Retningslinje-for-Rehabilitering-af-patienter-med-KOL.ashx?la=da&hash=81976AC77E8FCEF659052AE0A1ED378C18AEAC47.

7. The Danish Health Authority. Nationale klinisk retningslinjer for udvalgte sundhedsfaglige indsatser ved rehabilitering til patienter med type 2 diabetes. 2015. https://www.sst.dk/da/nyheder/2015/~/media/CD075EE6B0144C6E806097921B63B98A.ashx.

8. Axel S, Gade H. National klinisk retningslinje for hjerterehabilitering. 2015. https://www.sst.dk/da/udgivelser/2015/~/media/401919781c684ee9aae544eb5e76847b.ashx. Accessed 6 Feb 2019.

9. Flachs E, Eriksen L, Koch M, Ryd J, Dibba E, Skov-Ettrup L, et al. Sygdomsbyrden i Danmark - Sygdomme. 2015. https://www.sst.dk/da/sygdom-og-behandling/~/media/00C6825B11BD46F9B064536C6E7DFBA0.ashx.

10. Handberg C, Nielsen CV. Tilfredshed med kommunal rehabilitering af borgere med kronisk sygdom. Danish J Nurs. 2013;113:74–9.

11. Worcester MUC, Murphy BM, Mee VK, Roberts SB, Goble AJ. Cardiac rehabilitation programmes: predictors of non-attendance and drop-out. Eur J Cardiovasc Prev Rehabil. 2004;11:328–35. doi:10.1097/01.hjr.0000137083.20844.54.

12. Fischer MJ, Scharloo M, Abbink JJ, van 't Hul AJ, van Ranst D, Rudolphus A, et al. Drop-out and attendance in pulmonary rehabilitation: The role of clinical and psychosocial variables. Respir Med. 2009;103:1564–71. doi:10.1016/j.rmed.2008.11.020.

13. Selzler A-M, Simmonds L, Rodgers WM, Wong EYL, Stickland MK. Pulmonary Rehabilitation in Chronic Obstructive Pulmonary Disease: Predictors of Program Completion and Success. COPD J Chronic Obstr Pulm Dis. 2012;9:538–45. doi:10.3109/15412555.2012.705365.

14. Wittmer M, Volpatti M, Piazzalonga S, Hoffmann A. Expectation, satisfaction, and predictors of dropout in cardiac rehabilitation. Eur J Prev Cardiol. 2011;19:1082–8.

15. Mikkelsen T, Thomsen, Kristian Korsgaard Tchijevitch O. Non-attendance and drop-out in cardiac rehabilitation among patients with ischaemic heart disease. Dan Med J. 2014;61.

16. Cassidy S, Turnbull S, Gardani M, Kirkwood K. Attendance at pulmonary rehabilitation classes: An exploration of demographic, physiological and psychological factors that predict completion of treatment. doi:10.1177/1479972314527469.

17. Pardaens S, De Smedt D, De Bacquer D, Willems AM, Verstreken S, De Sutter J. Comorbidities and Psychosocial Characteristics as Determinants of Dropout in Outpatient Cardiac Rehabilitation. J Cardiovasc Nurs. 2017;32:14–21. doi:10.1097/JCN.0000000000000296.

18. Abebaw Y, Ali Y, Noel D, Edna B. Predictors of drop-out from an outpatient cardiac rehabilitation programme. Clin Rehabil. 2007;:222–9. doi:10.1177/0269215506070771.

19. Ruano-Ravina A, Pena-Gil C, Abu-Assi E, Raposeiras S, van 't Hof A, Meindersma E, et al. Participation and adherence to cardiac rehabilitation programs. A systematic review. Int J Cardiol. 2016;223:436–43.

20. Schwennesen N, Henriksen JE, Willaing I. Patient explanations for non-attendance at type 2 diabetes self-management education: A qualitative study. Scand J Caring Sci. 2016;30:187–92.

21. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J, et al. Coding Algorithms and for Defining Comorbidities Data in ICD-9-CM Administrative. Med Care. 2005;43:1130–9.

22. Region Nordjylland. Den Politiske Sundhedsaftale 2015-2018. 2015. https://rn.dk/sundhed/til-sundhedsfaglige-og-samarbejdspartnere/sundhedsaftalen-2015-2018/-/media/Rn_dk/Sundhed/Til-sundhedsfaglige-og-samarbejdspartnere/Sundhedsaftaler/Sundhedsaftale-2015/Status-på-sundhedsaftalen-2015-2018.ashx. Accessed 14 Aug 2019.

Appendix D. Three perspectives on economic analysis of health-related quality of life and healthcare utilisation in municipal rehabilitation programmes – a comparison of socioeconomic groups

Three perspectives on economic analysis of health-related quality of life and healthcare utilisation in municipal rehabilitation programmes – a comparison of socioeconomic groups

Annette Willemoes Holst-Kristensen[1], Kirsten Fonager[2], Kjeld Møller Pedersen[1]

1 Danish Center for Healthcare Improvements, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
2 Department of Clinical Medicine, Aalborg University, Denmark; Department of Social Medicine, Aalborg University Hospital, Denmark

Abstract

Objective: to evaluate the effects of a rehabilitation programme in a Danish municipality. Effects are measured by the Short Form 36 (SF-36) questionnaire and by healthcare utilisation. Analysis will be developed as a cost-utility analysis. The aim is to investigate possible differential gains across socioeconomic groups and to investigate healthcare utilisation using difference-in-difference analysis. **Background**: The average lifespan is increasing, but this comes with a concomitant increase in the rate of people living with chronic illness. This calls for efficient resource use and preventive initiatives such as rehabilitation programmes. In Denmark, the municipalities have responsibility for rehabilitation programmes offered to patients with cardiovascular disease, chronic obstructive pulmonary disease, and diabetes, among others. **Method**: Data from the municipality of Aalborg, Denmark for 2007–2014 were linked with data from the Danish National Registers. The following outcomes were analysed: health-related quality of life (HrQoL), hospital admissions, outpatient visits, and GP/specialist visits. Subgroup analysis compared three socioeconomic groups. The effects are assessed by applying a crosswalk utility score to SF-36 scores using a regression algorithm. Difference-in-difference (DID) analysis is used to examine healthcare utilisations before and after rehabilitation between attenders and non-attenders/dropouts. **Results**: 481 patients were included in the intervention group and 84 in the control group. The CUA resulted in an ICER of €19,056 per extra quality-adjusted life year (QALY) gained. The subgroup analysis shows that the employed group gained the most. The difference-in-difference analyses showed no significant difference in healthcare utilisation between the 'register intervention' group and 'register control' group. **Conclusion:** Attenders of municipal rehabilitation programmes have a better health-related quality of life compared to non-attenders. The subgroup analysis found that the employed may be gaining more from rehabilitation than the unemployed and retired. No significant DID in healthcare utilisation was found.

Introduction

The average lifespan is increasing, but this is accompanied by an increase in the rate of people living with chronic illness. This creates a need for systematic and effective municipal rehabilitation programmes (1). In Denmark (DK), the municipalities are responsible for the rehabilitation of chronically ill patients. Patients with cardiovascular disease (CVD), chronic obstructive pulmonary disease (COPD) and/or diabetes are referred to rehabilitation programmes in their home municipality. All three diseases are common, with increasing incidence rates. According to the Danish burden of disease report, the overall prevalence of cardiovascular diseases, COPD and diabetes mellitus is 169,099, 247,570 and 67,733 persons, respectively, across all age groups (2). Moreover, the three diseases accounted for more than 5% of all outpatient visits and more than 4% of all hospitalisations in 2012 (2). Compared with other Nordic countries, the mortality rate is higher in DK for all three diseases (2).

Rehabilitation is a recommended standard of care for CVD, COPD and diabetes patients in Denmark after an acute event (3–5) – e.g., hospitalisation. The effectiveness of the rehabilitation programmes for all three diseases is well known, but

in different settings and only within specific disease areas. Studies have found that rehabilitation is effective in improving health-related quality of life (HrQoL), decreasing the need for hospital admission, decreasing the number of outpatient visits, and improving clinical outcomes (6–9). While economic evaluations of rehabilitation have been published (10–12), little is known about the cost-effectiveness of municipal rehabilitation, and to the authors' knowledge, none considers socioeconomic subgroups to demonstrate how cost-effectiveness varies.

This study aims to: (1) evaluate the cost-effectiveness of a rehabilitation programme offered in a large Danish municipality, (2) as a novelty, investigate whether any socioeconomic groups gain more from rehabilitation than others in terms of HrQoL, and (3) compare patients' resource use in the primary and secondary sectors.

Method

Study design and population

A cost-utility analysis (CUA) on rehabilitation was performed from a municipal payer perspective, and QALYs were obtained by crosswalking the SF-36 scores. The analysis was conducted with a 12-week time horizon consistent with the length of the rehabilitation programme. Patient-level data from the municipality of Aalborg (HrQoL data) were obtained for 2007–2014 and linked with data from the Danish National Registers (marital status, income, socioeconomic status in terms of employment, and healthcare utilisation). The patients suffered from either CVD, COPD or diabetes. The CUA is a comparison between an intervention group and a control group, with the intervention group being the patients attending and completing the rehabilitation programme and having both baseline and follow-up SF-36 scores, and the control group consisting of non-attenders and drop-outs who did not complete rehabilitation but did have an SF-36 baseline score. The two groups were reasonably comparable (Table 1). The DID analysis was a registry-based analysis of healthcare utilisation between all those completing rehabilitation and those who were non-attenders or drop-outs.

Danish Municipal Rehabilitation programme

The Danish definition of a rehabilitation programme is a concentrated and time-limited collaboration between a patient, relatives and health professionals. It is offered to patients who have, or are at risk of having, significant physical, mental and/or social limitations after an acute event – typically a hospitalisation episode. Rehabilitation revolves around the patient's entire life situation, and treatment is based on coordinated, coherent and knowledge-based efforts. The overall aim of the Danish rehabilitation programmes for patients with health-impaired functioning levels is ideally to achieve a meaningful and independent life (or prevent deterioration or relapse). It is a process that enables the patients to maintain and promote quality of life and regain previous functioning levels, or the highest possible functioning level, and learn to live with chronic illness (13). The rehabilitation programmes consist of a start-up interview, 12 weeks of education and exercise/training, and an interview at the end of the programme. The start-up interview included collection of information on baseline characteristics, completion of the SF-36 and a variety of physical tests. The education encompasses optional courses such as knowledge of disease, dietary advice, the importance of physical activity, smoking cessation, and medicine consumption. At the end of the programme, patients completed the SF-36 and performed the physical test if possible (14).

Cost

Data related to the costs of providing the rehabilitation programme were collected by the healthcare centre. The direct cost of providing the rehabilitation programme consisted of salaries for the staff related to the rehabilitation teams allocated to each programme (for each disease – CVD, COPD and diabetes). Rent and depreciation were excluded. The staff comprised dieticians, therapists and nurses, who spent varying numbers of hours in the rehabilitation programmes depending on their profession and the programme to which they were assigned. Therapists were assigned the most hours, and dieticians were assigned the least. The average salary represents the time that staff spent on the programmes, including overtime. The salary represents total earnings, including public holiday payment, pension, and overtime payment. The salaries were €36.50, €39.42 and €42.11 per hour for a dietician, a therapist and a nurse, respectively and the hours spent on each 12-week programme varied from 16–42 for dieticians, 77–100 for therapists and 60–71 for nurses per week, depending on the disease area. The direct costs per patient were calculated by dividing total yearly salary costs with the total number of referred patients (N=4,361) over the study period. All costs were calculated as 2017 prices using the general consumer price index and a currency conversion rate of 745DKK=€100. This resulted in an average cost per patient referred for 12 weeks of rehabilitation of €362.70, €201.70 and €268.60 for CVD, COPD and diabetes, respectively. For the control group, it is assumed that there are zero costs, as the baseline interview is the only municipal service for this group. The baseline interview is, however, disregarded in the analysis because it does not differ between groups and is therefore not relevant.

Outcome Measurement

Health-related quality of life was measured at baseline and after 12 weeks using the SF-36. In order to conduct a CUA, the SF-36 scores were crosswalked to a single 'preference-based' utility score indicating the value that would be given to the health state by the general population. The crosswalk was done by extracting the appropriate SF-36 responses and using them to develop a six-item health state classification, the SF-6D, using the Brazier algorithm and the SF-6D methodology (15). The SF-6D comes with a set of weights obtained from the British population using a standard gamble method for utility-elicitation, with scores of 0 and 1 representing the worst and best possible health states, respectively. The QALYs were calculated as the area under the curve. To adjust for baseline differences and improve precision, regression analyses were applied to estimate incremental QALYs. The following baseline covariates were used: socioeconomic status, sex, marital status, and comorbidities. For the control group, it is assumed that baseline HrQoL scores are unchanged after 12 weeks.

Cost-utility analysis

A cost-utility analysis was undertaken to assess the cost-effectiveness of the rehabilitation programme compared with the control group (base case). The mean number of QALYs gained by completing rehabilitation and the costs expended were used to calculate the incremental cost-effectiveness ratio (ICER):

$$\text{ICER} = \frac{\text{Cost}_{rehabilitation} - \text{Cost}_{control}}{\text{Effect}_{rehabilitation} - \text{Effect}_{control}} = \frac{\Delta \text{C}}{\Delta \text{E}}$$

Cost_{I} and Effect_{I} are the rehabilitation cost and effects, in QALYs. The control group expresses the cost and effect for patients who did not complete rehabilitation. As noted above, $\text{Cost}_{control}$ is zero and $\text{Effect}_{control}$ is the effect at baseline, which is assumed to be constant over the 12 weeks. For the subgroup analysis, the intervention and control groups were split into three socioeconomic groups (employed, unemployed / other cash benefits, and retired) and compared with each

other, resulting in three different ICERs: one for the employed, one for the unemployed / other cash benefits, and one for the retired subgroup. The ICERs provide a point estimate of the mean cost per QALY gained by attending rehabilitation.

Difference-in-difference analysis

A DID analysis was used to analyse the healthcare utilisation differences. DID analysis is an appropriate method when randomisation is not possible. The analysis was performed by comparing the average change over time in the outcome variable between the patients being rehabilitated and the non-attenders and drop-outs. The analyses were based on registry-linked data. Therefore, the study population was higher than in the CUA, as it included all referred patients in the study period, divided into attendees (register intervention) and non-attenders/drop-outs (register control). A pre-rehabilitation period was defined as one year prior to the referral date to rehabilitation, and the post-rehabilitation period was defined as one year after the completion date of rehabilitation. The analysed outcomes were primary and secondary healthcare use. Primary healthcare use includes all contacts in the primary healthcare sector recorded in the Danish National Health Service Register for primary care, including visits to GPs, office-based specialists, psychologists and physiotherapists. Utilisation is measured as expenditure. Resource use in primary healthcare was valued using the tariffs of the national agreements between the Danish National Health Service and the professional associations of medical specialists. Secondary healthcare use includes contact registered with hospitals (both inpatient and outpatient) from the National Patient Register. Resource use was valued using DRG tariffs for inpatient services and the Danish Ambulatory Grouping System (DAGS) tariffs for outpatient visits (16). Logistic regression was used to calculate the differences in healthcare utilisation and adjusted for age, sex, education, marital status and socioeconomic status. The results of the analysis are presented as the average cost of healthcare utilisation per patient, as well as the differences between the register intervention and register control groups.

Statistical analyses

Baseline characteristics were analysed using Student t-tests for continuous variables and χ^2 tests for categorical variables. The ICER value, being the ratio of two differences which may not have a normal distribution, has an unknown sample distribution. It was, therefore, necessary to estimate the sample distribution around the point estimate non-parametrically. This is most appropriately done using the “bootstrap” technique. By this method, 5,000 hypothetical incremental costs and effects are modelled. The bootstrap method estimates the sample distribution of a statistic through a large number of simulations, based on sampling with replacement from the original data. This allows estimation of confidence intervals for the ICER in order to summarise the uncertainty due to sampling variations (17). The bootstrap can be used to represent the joint distribution in the incremental cost-effectiveness plane (CE-plane). This is illustrated graphically in a scatterplot, where each dot represents an incremental cost and effect. All analyses were carried out in Stata version 15 with a significance level set at 5%.

The study has been carried out in accordance with the General Data Protection Regulation (2015-509-00007). Also, in accordance with the Danish National Committee on Health Research Ethics, this research satisfies the criteria of being ‘questionnaire and register-based research excluding human biological material’, and thus was not required to undergo a formal ethics procedure [29].

Results

Cost-utility analysis

Of the referred patients (N=4,361), 481 had a complete SF-36 at baseline and follow-up and had an SF-6D score after the crosswalk, and 87 patients were in the control group and had only a baseline SF-36 score. The baseline characteristics show a significant difference in education level, with higher rates of a high education level in the intervention group and higher rates of a low education level in the control group. The intervention group had a higher proportion of patients with diabetes and a lower proportion of patients with COPD compared with the control group. The patients in the control group had more comorbidities and higher healthcare use before rehabilitation, and the SF-36 baseline physical score was significantly lower than the intervention group.

Table 1 Baseline characteristics of the intervention and control groups

Baseline characteristics	**Intervention group (n=481)**	**Control group (n=87)**
Sex, female/male	239/242	44/43
Age (SD)	66 (SD 8.9)	67.4 (SD 9.2)
Marital status		
Widow/widower	64 (13%)	14 (16%)
Divorced	72 (15%)	22 (25%)
Married	303 (63%)	46 (53%)
Never married	42 (9%)	5 (6%)
Highest Education *		
Low (<11 years)	165 (34%)	40 (47%)
Medium	237 (50%)	39 (46%)
High	75 (16%)	6 (7%)
Socioeconomic status		
Employed	119 (25%)	13 (15%)
Unemployed / other cash benefits	67 (14%)	17 (20%)
Pension	295 (61%)	57 (65%)
Annual taxable income, € (IQR)		
Personal	22,000 (17,100;30,900)	20,000 (16,200;24,700)
Diagnosis*		
COPD	279 (58%)	63 (72%)
CVD	56 (12%)	13 (15%)
Diabetes	146 (30%)	11 (13%)
Comorbidity (numbers) *		
0	239 (50%)	28 (32%)
1–2	212 (44%)	44 (51%)
3–4	23 (5%)	13(15%)
>4	7 (1%)	2 (2%)
Healthcare utilisation before rehabilitation, median (IQR)		
GP/specialist visits*	90 (61;134)	107 (75;148)
Outpatient visits	7 (3;18)	9 (5;23)
Hospital admissions*	6 (3;9.5)	9 (5;14)
Baseline QALY score (n=5,000)	0.16 (−0.05;0.04)	0.15 (0.09;0.21)

*Statistically significant with p-values <0.05

The incremental cost-effectiveness ratio for the base-case analysis suggests that the rehabilitation programme provides an incremental cost of €258 (CI 258.3;258.5) and an incremental effect of 0.014 QALYs (CI 0.0136;0.0137), resulting in an ICER of €19,056 per extra QALY gained. This places the base-case ICER in the upper-right quadrant of the incremental cost-effectiveness scatter plot, making rehabilitation more costly and more effective, Figure 1.

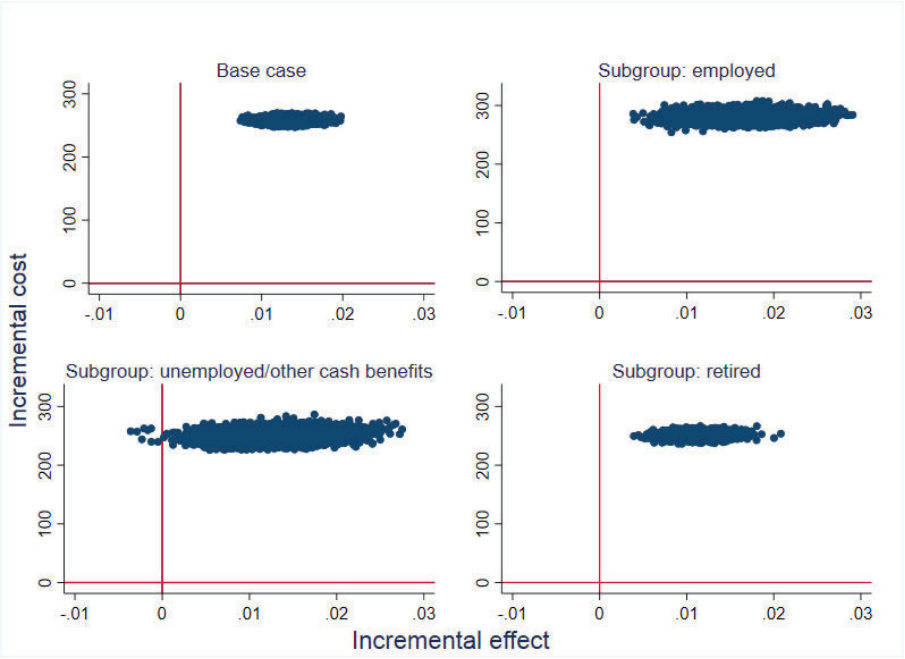


Figure 1 Cost-effectiveness plane of base-case and subgroups.

Subgroup analysis: socioeconomic differences in ICER

The intervention and control groups were split into three subgroups: employed, unemployed / other cash benefits, and retired. Even though the numbers in the control groups were relatively small, they were still meaningful. There were no significant differences between the intervention and control groups in baseline characteristics within each subgroup. However, as expected, there were differences across the subgroups. Among the patients completing rehabilitation, there was a significant difference across socioeconomic status in terms of marital status, education level, comorbidity and baseline level of the SF-36 physical domain.

The ICER of the employed analysis suggests that the rehabilitation programme provides an incremental cost of €282 (CI 281.6;282) and an incremental effect of 0.016 QALYs (CI 0.0164;0.0166), resulting in an ICER of €17,547 per extra QALY gained. The ICER of the unemployed / other cash benefits analysis suggests that the rehabilitation programme provides an incremental cost of €250 (CI 250;251) and an incremental effect of 0.013 QALYs (CI 0.0133;0.0135), resulting in an ICER of €19,100 per extra QALY gained. The ICER for the retired analysis suggests that the rehabilitation programme provides an incremental cost of €251 (CI 250;251) and an incremental effect of 0.011 QALYs (CI 0.0111;0.0113), resulting in an ICER of €22,209 per extra QALY gained.

This places all three subgroups in the upper-right quadrant of the incremental cost-effectiveness scatter plot, making rehabilitation more costly and more effective. However, the unemployed / other cash benefits group is also represented in the upper-left quadrant, meaning some of the patients may have reduced QALYs (Figure 2).

Table 2 *Baseline characteristics of the subgroups*

Baseline characteristics	**Subgroups**					
	Employed		Unemployed / other cash benefits		Retired	
	Intervention n=119	Control n=13	Intervention n=67	Control n=17	Intervention n=295	Control n=57
Sex, female/male	62/75	4/9	41/34	9/8	177/166	31/26
Age (SD)	58 (8.19)	57 (3.64)	58 (4.62)	58 (6)	71 (5.90)	72
Marital status						
Widow/widower	<5	<5	<5	<5	59 (20%)	12 (21%)
divorced	15 (13%)	<5	17 (25%)	8 (47%)	40 (14%)	11 (19%)
Married	82 (69%)	8 (62%)	38 (57%)	6 (35%)	183 (62%)	32 (56%)
Never married	20 (17%)	<5	9 (13%)	<5	13 (4%)	<5
Highest Education						
Low (<11 years)	24 (20%)	5 (38%)	22 (33%)	5 (29%)	119 (41%)	30 (55%)
Medium	66 (56%)	6 (46%)	39 (58%)	10 (59%)	132 (45%)	23 (42%)
High	28 (24%)	<5	6 (9%)	<5	41 (14%)	<5
Annual taxable income (€) (IQR)						
Personal	28,700 (23,000;36,800)	24,900 (22,500;32,000)	22,300 (17,048;26,740)	23,600 (20,600;24,700)	19,800 (16,397;26,103)	18,900 (15,100;22,900)
Diagnose						
COPD	51 (43%)	8 (62%)	42 63%	14 (82%)	186 (63%)	41 (72%)
CVD	15 (13%)	<5	8 12%	<5	33 (11%)	8 (14%)
Diabetes	53 (44%)	<5	17 25%	<5	76 (26%)	8 (14%)
Comorbidity						
0	74 (62%)	8 (62%)	30 45%	3 (18%)	135 (46%)*	17 (30%)*
1–2	44 (37%)	5 (38%)	32 48%	13 (76%)	136 (46%)*	26 (46%)*
3–4	-	-	<5	<5	19 (6%)*	12 (21%)*
>4	<5	-	<5	-	5 (2%)*	<5*
Healthcare utilisation before rehabilitation, median (IQR)						
GP/specialist visits	75 (51;108)	85 (60;113)	90 (52;129)	98 (71;127)	99 (69;149)	113 (84;153)
Outpatient visits	6 (2:17)	6 (1;9)	7.5 (4;19.5)	24 (10;39)	7 (3:18)	8.5 (5;19.5)
Hospital admissions	5 (4;8)	3 (1.5;6.5)	7(4;11)	8.5 (5.5;12.5)	6(3:10)	9.5 (7;14)
Baseline QALY score (n=5000)	0.17 (0.16;0.18)	0.15 (0.14;0.16)	0.15 (0.14;0.16)	0.14 (0.13;0.15)	0.16 (0.16;0.17)	0.15 (0.15;0.16)

*Statistically significant with p-values <0.05

Difference-in-difference analysis

In the DID analysis, all observations for the programme are used because the focus is on the utilisation of services, leading to a larger number of patients in both the intervention and control groups than in the above analyses. The control group consists of non-attenders and drop-outs who did not complete rehabilitation.

The pre- to post-rehabilitation change in the patients' outpatient visits, hospital admissions, and GP/office-based specialist visits in the rehabilitation group was not significant for the intervention group. All the outcomes were lower in the post-intervention period compared with the pre-intervention period for control patients. However, the DID analysis showed no significant differences in healthcare utilisation.

Table 2 Difference-in-difference analysis of healthcare utilisation before and after exercise-based rehabilitation

	Register intervention group (n=2,171)			Register control group (n=484)			DID in €	Adjusted DID in €	P-value	95% CI
	Before	After	Difference	Before	After	Difference				
Outpatient visits	1,269	1,158	−111	1,261	1,164	−97	−18	4	0.99	−551;559
Admissions	5,417	3,295	−2,122	5,746	4,434	−1,312	−810	−617	0.39	−2,023;794
GP/specialist visits	643	482	−161	685	482	−203	42	29	0.39	−37;95

Discussion

This study investigated, from a narrow municipal financial perspective, the cost-effectiveness of the rehabilitation programme in the municipality of Aalborg, Denmark in patients with CVD, COPD and diabetes. Overall, we found the rehabilitation programme to be cost-effective, with an incremental cost of €258 and an incremental QALY gain of 0.014, giving an ICER of €19,056 per QALY gained. No economic evaluations or cost analyses have previously investigated the economic implications of municipal rehabilitation of chronically ill patients. The subgroup analysis found that the employed gained the most. The results should be interpreted with caution because the control group is small and data originates from an observational study. The control group, however, is relatively similar to the intervention group. Furthermore, there are trends here indicating that there could be socioeconomic differences.

The extent to which an intervention is cost-effective depends on the threshold value. In Denmark, there is no fixed threshold, and therefore it is unknown whether an ICER of €19,056 per QALY is deemed cost-effective. The small effect may be blurred if you only consider the ICER, since it seems cost-effective, but the effect alone seems minimal. Therefore, it is relevant to discuss if the QALY gain is of a minimal important difference (MID). According to Walters and Brazier, the MID of the SF-6D ranges from 0.010–0.048 across nine different patient groups – e.g., COPD with an MID of 0.010 (18). If this were applicable for this study, the effect of rehabilitation could be interpreted as a clinically important difference, but a gain of 0.014 still seems minimal. The reason for the small QALY effect may be the outcome measure. Is SF-36 the most relevant outcome measure in regards to the aim of rehabilitation? Outcome measures for complex interventions such as rehabilitation should be broad and capture more than health, and SF-36 does this imperfectly. This line of thinking is in line with the aim of rehabilitation, where health and quality of life are not the only aims (19). Furthermore, the patients are referred to rehabilitation after an acute event, and therefore their HrQoL has already changed during treatment in hospital. Therefore it would have been preferable if the data collection started at the beginning of the hospital stay.

A strength of the present study is that it reflects an ongoing and everyday municipal rehabilitation programme. Most studies regarding rehabilitation are in an outpatient setting, why little is known of the HrQoL effects in a municipal setting. This study also serves as an example for other municipalities. Furthermore, the study is register-linked, giving more precise baseline characteristics than self-reported data.

The study has several limitations. One is the narrow cost perspective using only the direct salary costs. However, in most public welfare programmes, salary costs make up 60–80% of total costs – and in rehabilitation programmes, this is probably closer to 80% than 60%. Other costs that ideally should have been included are rent and the maintenance and

depreciation costs of the exercise equipment. For the perspective to be societal, the patient cost should be included as well – e.g., transportation and time. However, this study had a narrow perspective because we wanted to explore the cost for the healthcare centre alone and the effect in the short run. Salary cost is the only cost that can be saved if rehabilitation is not performed. This supports the decision to have only salary cost in the analysis. The control group was not ideal, as the size was small, and the baseline QALY was assumed to be constant. This could have been further investigated with sensitivity analysis, to see how sensitive the QALY score is to the assumption of a zero effect – e.g., by using random follow-up values.

The sample used for the analysis of the study is small: 481, compared to the 4,361 patients who were referred. The large proportion of missing participants is due to the lack of SF-36 completions and because of the large percentage of non-attenders and drop-outs, at 18% and 21%, respectively. The missing SF-36 data was considered missing completely at random, hence the all questions were missing. A reason for this could be that the patients never had the opportunity to fill out the questionnaire or were not capable of completing it. In future studies, healthcare centres should be aware of how to obtain good-quality data. It should be noted that patients are often very willing to answers questionnaires if they are given the right instructions and have the purpose of the questionnaire explained to them.

Conclusion

The CUA of the rehabilitation programme in the municipality of Aalborg resulted in an increased incremental effect; however, with no official threshold in Denmark, it is not possible to make concrete conclusions regarding the cost-effectiveness. The results from the subgroup analysis are similar to the base-case CUA. However, the employed may be gaining more from rehabilitation than the unemployed and retired. No significant DID in healthcare utilisation was found.

References

1. The Danish Health Authority. Ældres Sundhed Og Trivsel [Internet]. 2019. Available from: https://www.sst.dk/da/udgivelser/2019/~/media/388EDD0F865045F1BABA4542FA83FCD2.ashx

2. Flachs E, Eriksen L, Koch M, Ryd J, Dibba E, Skov-Ettrup L, et al. Sygdomsbyrden i Danmark - Sygdomme [Internet]. Sundhedsstyrelsen. 2015. Available from: https://www.sst.dk/da/sygdom-og-behandling/~/media/00C6825B11BD46F9B064536C6E7DFBA0.ashx

3. Axel S, Gade H. National klinisk retningslinje for hjerterehabilitering [Internet]. 2015 [cited 2019 Feb 6]. Available from: https://www.sst.dk/da/udgivelser/2015/~/media/401919781c684ee9aae544eb5e76847b.ashx

4. Pedersen PM, Olsen TS. National klinisk retningslinje for Rehabilitering af patienter med KOL [Internet]. Sundhedsstyrrelsen. 2018. Available from: https://www.sst.dk/-/media/Udgivelser/2018/National-Klinisk-Retningslinje-for-Rehabilitering-af-patienter-med-KOL.ashx?la=da&hash=81976AC77E8FCEF659052AE0A1ED378C18AEAC47

5. The Danish Health Authority. Nationale klinisk retningslinjer for udvalgte sundhedsfaglige indsatser ved rehabilitering til patienter med type 2 diabetes [Internet]. 2015. Available from: https://www.sst.dk/da/nyheder/2015/~/media/CD075EE6B0144C6E806097921B63B98A.ashx

6. Jacobsen R, Rusch E, Andersen PK, Adams J, Jensen CR, Frølich A. The effect of rehabilitation on health-care utilisation in COPD patients in Copenhagen. Clin Respir J [Internet]. 2014 Jul [cited 2015 Aug 13];8(3):321–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24279723

7. Godtfredsen NS, Grann O, Dalsgaard LS, Lavesen M, Pors B, Dollerup J, et al. Effects of community-based pulmonary rehabilitation in 33 municipalities in Denmark – results from the KOALA project. Int J Chron Obstruct Pulmon Dis. 2018;Volume 14:93–100.

8. Anderson L, Oldridge N, Thompson DR, Zwisler A-D, Rees K, Martin N, et al. Exercise-Based Cardiac Rehabilitation for Coronary Heart Disease. J Am Coll Cardiol. 2016 Jan;67(1):1–12.

9. Vadstrup ES, Frølich A, Perrild H, Borg E, Røder M. Effect of a group-based rehabilitation programme on glycaemic control and cardiovascular risk factors in type 2 diabetes patients: The Copenhagen Type 2 Diabetes Rehabilitation Project. Patient Educ Couns. 2010;84:185–90.

10. Papadakis S, Oldridge NB, Coyle D, Mayhew A, Reid RD, Beaton L, et al. Economic evaluation of cardiac rehabilitation: a systematic review. Eur Soc Cardiol. 2005;

11. Wong WP, Feng J, Pwee KH, Lim J. A systematic review of economic evaluations of cardiac rehabilitation. BMC Health Serv Res. 2012;12:243.

12. Griyths TL, Phillips CJ, Davies S, Burr ML, Campbell IA. Cost effectiveness of an outpatient multidisciplinary pulmonary rehabilitation programme. Thorax. 2001;56:779–84.

13. Marselisborg Centret. Rehabilitering i Danmark Hvidbog om rehabiliteringsbegrebet [Internet]. [cited 2018 Nov 9]. Available from: http://www.rehabiliteringsforum.dk/contentassets/519c219e42a3454da34db60a8454de58/hvidbog.pdf

14. Johansen J, Rahbek J, Møller K, Jensen L. Hvidbog om rehabiliteringsbegrebet. Rehabilitering i Danmark [Internet]. 2004. Available from: http://www.marselisborgcentret.dk/fileadmin/filer/Publikationer/PDF_er/Hvidbog.pdf

15. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002 Mar;21(2):271–92.

16. Lynge E, Lynge Sandegaard J, Rebolj M. The Danish National Patient Register. Scand J Public Health [Internet]. 2011 [cited 2019 Aug 5];39(7):30–3. Available from: www.sst.dk

17. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. Health Econ. 1997;6(4):327–40.

18. Walters SJ, Brazier JE. Health and Quality of Life Outcomes What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D [Internet]. 2003 [cited 2019 Nov 11]. Available from: http://www.hqlo.com/content/1/1/4

19. Holst-Kristensen AW, Fonager K, Pedersen KM. Test–retest reliability of ICECAP-A in the adult Danish population. Qual Life Res. 2019;(29):547–57.

ISSN (online): 2246-1256
ISBN (online): 978-87-7210-616-8

AALBORG UNIVERSITY PRESS