Corvinus University of Budapest

A NEW METHOD FOR THE MEASUREMENT OF ACCEPTABLE HEALTH STATES

PhD Dissertation

Supervisor: Prof. Márta Péntek, PhD

dr. Zsombor János Zrubka

Budapest, 2019

dr. Zsombor János Zrubka

Corvinus University of Budapest
Department of Health Economics

Supervisor: Prof. Péntek Márta, PhD

Corvinus University of Budapest

Doctoral School of Management

A NEW METHOD FOR THE MEASUREMENT OF ACCEPTABLE HEALTH STATES

PhD Dissertation

dr. Zsombor János Zrubka

Budapest, 2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The ideas and analyses in the dissertation are the products of my own research activity, however, in the planning, implementation and analysis of the research, I received continuous support from my supervisor, Prof. Márta Péntek, from Prof. László Gulácsi, the head of the Department of Health Economics, and from the colleagues of the Department of Health Economics. Furthermore, dr. Zoltán Hermman, associate professor at the Corvinus University of Budapest, and Áron Papp, a student receiving scholarship contributed to certain phases of the research. Also, in the collection of data I received assistance from Balázs Szilágyi and Anett Hodos. Therefore, acknowledging the teamwork, I present the results in first-person plural. I would like to thank the joint work and experiences to all of them.

I. INTRODUCTION

Due to technological innovation, demographic change and the growth of income, GDP-proportional health expenditure has regained intense growth in the OECD member states after a stagnation that followed the economic crisis of 2008. Sustainable financing of healthcare became one of the key challenges of countries with developed economies. Decision-makers may need to introduce new decision-methods in order to maintain equitable and legitim allocation of resources despite the growing fiscal pressure. While in the past decades the QALY concept has been adopted by an increasing number of countries for the measurement of health outputs, its amendment has been suggested both in the measurement of individual well-being or in the measurement of societal preferences, for supporting health resource allocation decisions. Amending the QALY concept and new ways of measuring health outcomes became recently intense areas of research in the field of health economics.

The research of acceptable health states (AH) started in the Netherlands as well as at the Corvinus University of Budapest, approximately 10 years ago. The measurement of acceptable health problems applied together with sufficientarian theory – a novel theory of justice in the field of health economics – aims to provide a transparent picture for decision-makers about the society's age and disease severity related preferences. However, due to the methodological difficulties, indicators have not yet been developed, which would make the practical application of the AH concept feasible.

In PhD studies I have been searching for answers to the problems of sustainable healthcare financing, especially in the field biosimilars, which represent cost-effective alternatives to the expensive biological therapies, as well as in the field of health outcomes measurement for economic analysis, the results of which I will introduce in later sections of my dissertation.

In the empirical research detailed in my dissertation, the main goal was to further develop the methods that enable the practical use of indicators based on acceptable health states. After theoretical planning, the Department of Health Economics launched in early 2018 the „Health and Ageing" study, under the leadership of Prof. Márta Péntek, during which I had opportunity to test the new measurement methods of acceptable health. In the dissertation I will introduce results of this research in the steps detailed below:

a) In the literature review I introduce the challenges of health financing decision making, state-of-the-art methods of resource allocation, the theoretical background of acceptable health and the possible areas of its further development.
b) After stating the research goals, I introduce the new methods of measuring acceptable health from the theoretical basics through introducing results of my research as well as the critical interpretation of the findings.
c) In the last part of the dissertation I will analyse the association between subjective wellbeing (happiness) and acceptable health, and I expand on further research areas and practical applicability of acceptable health.

The main results of my research are the following:

a) We introduced two new methods of measuring acceptable health. We developed an adaptive testing algorithm for the joint measurement of acceptable health states, which provides as more accurate and reliable measure when compared to previous methods. For measuring acceptable health, we used the EQ VAS questionnaire for the first time. Both methods were easy to implement in practice.
b) Using the new methods, we confirmed the results of previous research: people accept more health problems with age, and mild problems are considered more acceptable than severe ones.
c) We have demonstrated that acceptable health from the individual's point of view shows similar priorities in terms of age and severity, as from a societal perspective.
d) Backed with data, we pointed out the main areas of developing the acceptable health concept further, which are mainly related to increasing the reliability of data collection and results.

The E-matrix obtained through the new measurement method (joint evaluation) of acceptable health makes possible the practical application of results, by adapting QALY-based health economic models. At the end of the dissertation I introduce possibly ways of applying acceptable health measurements in health technology assessment, while relying on sufficientarian and utilitarian as well as egalitarian theories of justice.

As a summary, despite its potential areas for further development detailed in the dissertation, acceptable health is a promising new area in the measurement of health outcomes, which may open new research avenues on an international scale.

Note: terms indicated with asterisk (*) are included the glossary at the end of the dissertation in alphabetical order.

II. LITERATURE REVIEW

II.1. Challenges of sustainable healthcare financing

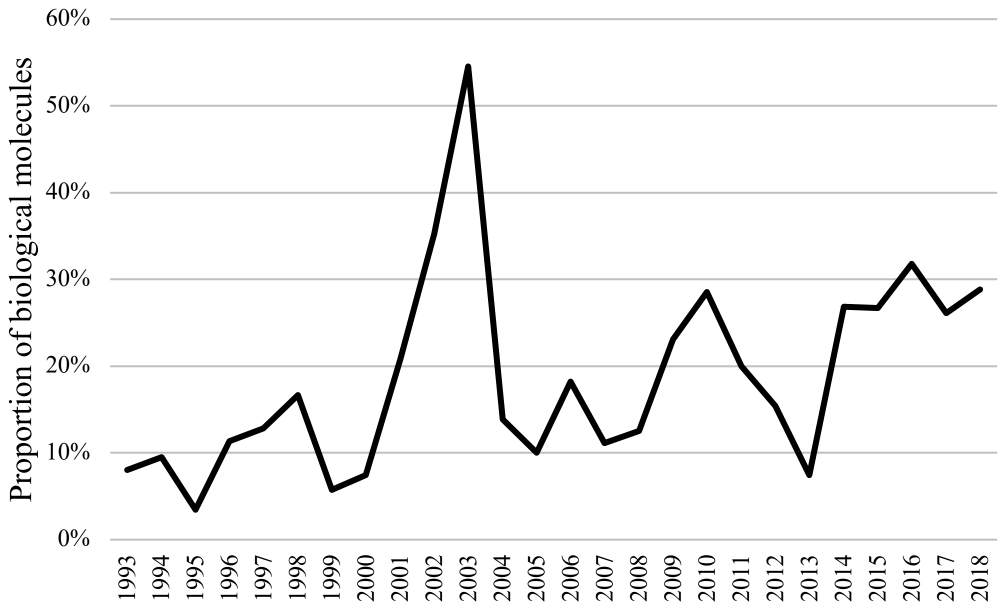
Sustainable financing is one of the key challenges of OECD countries. (OECD, 2015a) Growth of healthcare expenses has surpassed GDP growth over the past 20 years, and after the temporary diminution due to the 2008 economic crisis, since 2014 they have been growing at a speed above GDP again. (OECD, 2018) According to the European Commission, the definition of sustainable financing is the following: *"the ability to continue now and in the future current policies (with no changes regarding public services and taxation) without causing public debt to rise continuously as a share of GDP".* (European Commission, 2014) Sustainable financing does not preclude the increase of health expenditure in case it reflects the willingness to pay of the society, and it is possible to regroup resources from other budget items. However, if the necessary resources are not available, then it may be required to constrain otherwise desirable expenditure. (OECD, 2015a) In countries with low to medium income, growth of health expenditure was explained by the response to unmet health needs and action to provide universal coverage. In countries with high income, development of medical technology, demographic shift, increase of salaries and anomalies of the health care systems are the key causes of the increasing expenses. (OECD, 2015a) The key driver of costs if technological innovation, which, according several estimates, makes up as high as 50% of expenditure growth in healthcare. (de la Maisonneuve and Martins, 2013, Willeme and Dumont, 2015)

In addition to radically innovative medical devices, mainly innovative medicines contribute to the growth of expenses, which can only partly be balanced by the use of so-called generic medicines. (Willeme and Dumont, 2015) In the past few years, specialty medicines became the drivers of pharmaceutical expenditure. (OECD, 2018)

The industrial production of specialized protein medicines enabling targeted therapeutic interventions (enzymes, antibodies, hormones, cells) is not based on chemical synthesis any more, but happens in living organisms, which we call biological therapies. (FDA, 2018) The yearly cost of specialty medicines used for the treatment of oncology disorders have surpassed 10000 USD per year by the early

2000's. In 2012, for most of the newly registered oncology therapies, the average yearly treatment cost was more than 100000 USD. (OECD, 2015b) Most of these specialty medicines are biological molecules. The trends are well illustrated by the fact, that the first CAR-T cell therapy, Kymriah® (tisagenlecleucel), which was registered in 2017 in the USA for the treatment of childhood leukaemia was introduced with a list price of 475000 USD. Owing to its radical efficacy, the product was cost-effective. However, the extreme costs are incurred in the first year of therapy, while the life-year gains or societal benefits are accumulated subsequently over many years. (Whittington et al., 2018) The proportion of biological therapies among newly registered innovative drug molecules is increasing steadily. (Figure 1) In 2018 from the 59 new innovative medicines registered in the USA, 17 were biological molecules. (Mullard, 2019) From the 10 global top selling drugs in 2017 seven were biologicals, all with yearly sales over 5 billion USD. The turnover of biological molecules made up 25% of the global pharmaceutical market in 2016 and will reach 30% by 2022. (EvaluatePharma, 2017)

Figure 1 Proportion of biological molecules among newly registered medicines



Source: Adapted from Mullard, FDA data from 1993 to 2018 (Mullard, 2019)*

II.2. Interventions for sustainable healthcare financing

II.2.1. Main areas of intervention for sustainable healthcare financing

The OECD report from sustainable financing (OECD, 2015a) identified three main areas of intervention: (1) reallocation of budget items in favour of health expenses (2) increasing the efficiency of public expenditure, (3) re-evaluation of the ratio of public and private health expenses. Reallocation of resources is primarily the role of fiscal policy. According to the OECD, increasing the share of private expenditure (either in the form of private insurance or increasing the co-pay of services) can result in decreasing efficiency of healthcare systems. (OECD, 2015a) In the OECD member states, health products (medicines and medical devices) make up the third biggest item (19%) among healthcare costs after outpatient care (33%) and hospital care (28%). In Hungary, medical products make up the biggest category (32%), and in terms of pharmaceutical expenditure (28.8%) Hungary is the second among the OECD members states, and first among member states of the European Union. (OECD, 2017)

The efficiency of public expenditure can be improved mainly be the following interventions:

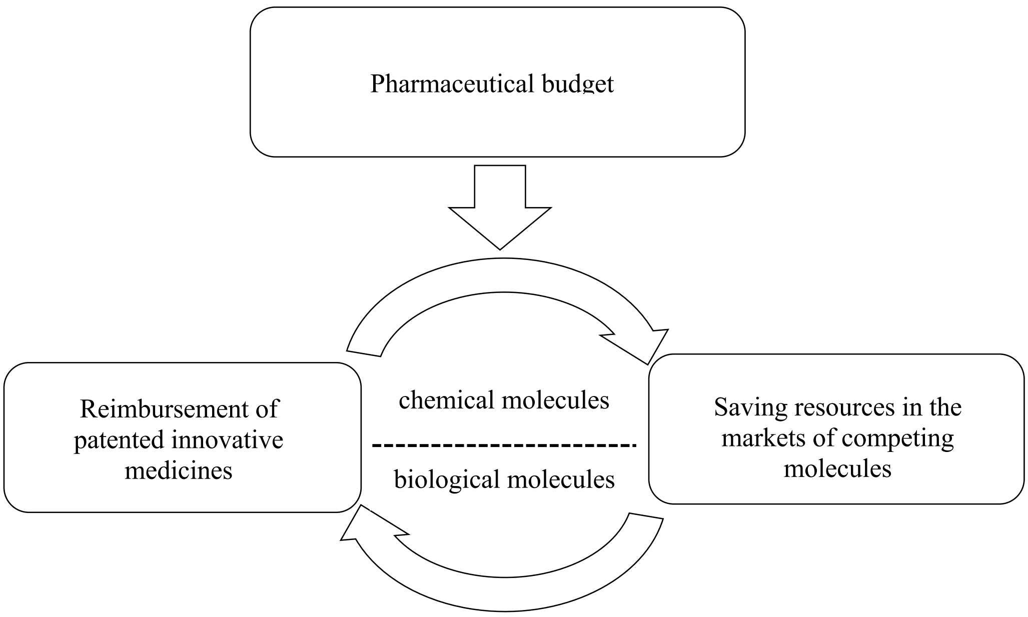
(a) Containment of pharmaceutical expenditure
(b) Standardising heterogenous medical practice
(c) Efficient management of patient-pathways
(d) Reforming the financing of health-service providers
(e) Supporting health promotion and prevention programs.

Pharmaceutical expenses can be curbed by increasing the penetration of generic medicines and by efficient procurement practices, while increasing patient co-pay is a less efficient intervention with negative consequences on access to care. (OECD, 2015a)

II.2.2. The reimbursement policy cycle

The growth of pharmaceutical expenses and the equilibrium of the pharmaceutical budget can be managed by balancing the savings in competitive markets of off-patent molecules and the expenses of reimbursing new, innovative patented medicines. (Figure 2) Fostering price competition in competitive markets releases resources, which can be used for the reimbursement of innovative molecules with market exclusivity. The reimbursement of cost-effective therapies ensures the efficient use of the released resources. (Dankó and Molnár, 2013) The examination of the cost-effectiveness and budget impact of new therapies is the role of health technology assessment* (HTA), which will be introduced in chapter II.5. With proper long-term planning, the sustainability of reimbursement policies can be ensured. Just as the entry of new biological molecules posed a challenge on reimbursement policies and practices, the loss of their patent protection also raised new medical and regulatory questions, which required the development of new reimbursement policies in order to enhance competition in the biological markets. (Gulacsi, 2014) The case studies illustration the economic impact of biosimilar medicines will be presented in chapters II.2.4. and II.2.5.

Figure 2 Reimbursement policy cycle



Source: Adapted from p180. Dankó and Molnár (2013), (Dankó and Molnár, 2013)

II.2.3. The impact of austerity measures on the efficiency of health systems

The analysis of austerity measures on health expenditure following the 2008 financial crisis provided mixed results in terms of healthcare indicators and was not clearly separable from the general effects of overall economic downturn. Nevertheless, the utilisation of healthcare services decreased unambiguously, the long-term effects of which are not yet foreseeable. (van Gool and Pearson, 2014) According to the latest OECD analysis, widening the range of services that are available in primary care with full coverage, as well as decreasing the proportion of co-pay are the two policies that contribute most to the improved efficiency of healthcare, while health technology assessment, which informs decision-makers about the value and societal effects of new health technologies, in parallel with improved health indicators, has contributed to growing health expenditures as well. (Lorenzoni et al., 2018) In the future, for sustainable financing, increasing selectivity may be warranted during health financing decision making, including value-based reimbursement programs supported by health technology assessment, and the diminishing of public reimbursement in the case of not cost-effective technologies. (OECD, 2015a) In forthcoming decades, in supply-driven markets, decision-makers managing constrained resources will work under increasing financial pressure. (Dankó and Molnár, 2013)

II.2.4. Case study: the economic effects of biosimilars

In this chapter, I will briefly introduce the results of my research supporting the utilisation of biosimilar medicines. (Pentek et al., 2017, Vezer et al., 2016, Zrubka, 2017a, Zrubka, 2017b)

After the patent expiry of traditional, chemical molecules, lower-priced competitors containing the same active ingredient are called generic medicines. (GaBi Online, 2012) Generic medicines make up 52% in volume and 25% in value the pharmaceutical markets of the OECD member states. (OECD, 2017) Biosimilar medicines are highly similar to authorised originator* reference biological products and introduced to the markets at a lower price. (European Medicines Agency, 2017) With the increasing sales of biological molecules, the role of biosimilars in moderating pharmaceutical expenses is increasing. However, in the OECD member states, the penetration of biosimilars compared to generics shows a much more diverse picture. While first-generation biosimilar erythropoietin (a hormone used for the treatment of the anaemia in chronic renal insufficiency or oncology disorders) (EMEA, 2018) had 68% market penetration in 2015 (highest: 100%, lowest: 2%), the second-generation infliximab (a TNF-α inhibitor monoclonal antibody* used for the treatment of chronic immune-mediated inflammatory disorders) (EMEA, 2019) reached only 27% market-share. (highest: 82%, lowest: 2%). (OECD, 2017) The originator reference brand of infliximab, Remicade™ was ranking 6-7th on the list of top-selling drugs globally between 2012 and 2018. (PMLive, 2018). Biosimilar infliximab was the first biosimilar monoclonal antibody (mAb). Compared to the first-generation biosimilars, biosimilar mAbs can be characterised by much greater molecular complexity and economic significance.

Despite the ambitious expectations about savings, in most European countries, after the launch of biosimilar infliximab, the sales of the molecule continued to increase in terms of value, and the utilisation of more expensive originator molecules increased at a greater pace compared to the biosimilar products. (Pentek et al., 2017) The market penetration below expectation can be explained by multiple factors (Zrubka, 2017a, Pentek et al., 2017, Simmons, 2018):

a) Among the regulations of generic and biosimilar medicines, interchangeability has the greatest significance with regards to market competition. (Zrubka, 2017a) While according to the common European regulations, traditional originator and generic medicines are freely substitutable in pharmacies without the consent of the prescribing physician – giving way to price competition – in case of biological medicines, switching from the originator reference product to a biosimilar can happen with the consent of the prescribing physician. The interchangeability of biological medicines from professional or economic reasons falls under national regulations. (EGA, 2016) Despite the fact that the development of biosimilar medicines is based on the quality controls required after the manufacturing process modifications of originator reference products, which has been performed by the European Medicinal Agency 404 times before the approval of the first biosimilar mAb, (Vezer et al., 2016) due to the lingering concerns about the safety and quality of biosimilars, the interchangeability similar to generic medicines was not allowed by the financing protocols in many countries, which impeded the implementation of effective measures enhancing the uptake of biosimilar medicines. (Moorkens et al., 2017)
b) Savings with biosimilars depended on the choices of prescribing physicians, who were neither enforced, nor incentivised for prescribing biosimilars. (Moorkens et al., 2017, Zrubka, 2017a)

c) Due to the non-transparent agreements between the producers of originator reference products and health authorities, the net prices of originator biological products were lower than their list prices, therefore the expected price-advantage of biosimilar was substantially lower compared to the expectations. In Hungary, according to the analysis of published patient-data and tender agreements of originator products, the price of infliximab and its competitor originator brands were procured at prices approximately 50% below the official list prices. (Zrubka, 2017a)

The other barrier of savings with biosimilars (with infliximab in our own research) was that due to the high costs of originator molecules, access to TNF-α was very low in many countries (Pentek et al., 2014a), therefore, the expectation from the biosimilar molecule was improved access, re-investment of the savings into the same therapeutic area in order to improve access to therapy. (Brodszky et al., 2014)

II.2.5. Case study: new therapeutic opportunities with biosimilars

In the chapter below, I will shortly introduce the published (Zrubka et al., 2018a, Gulacsi et al., 2019) as well as submitted (Zrubka et al., 2019c) results of my research in the field of the health technology assessment of biosimilar medicines.

The complex economic effect of biosimilars is illustrated by the fact that these molecules can have a role in both sides of the reimbursement cycle. In addition to generating savings via price-competition, they may represent cost-effective alternatives among the innovative therapies. An example is the timing of the start of biological therapy in chronic immune-mediated disorders. In rheumatoid arthritis* (a disorder characterised by chronic inflammation of the joints), according to the European treatment guidelines, biological therapies are recommended only after the treatment failure of conventional anti-rheumatic therapies. (Smolen et al., 2017) However, an increasing number of publications have shown that the (even intermittent) use biologics in the earliest phase of the disease may prevent the development of a chronic progressive condition (Nagy and van Vollenhoven, 2015, van Nies et al., 2014), preventing considerable disease burden for the individual and costs for the society. Interestingly, due to the cost-effectiveness arguments, most therapeutic guidelines do not support biologics in the earliest phase of the disease. Our systematic literature review and meta-analysis* has shown, that in randomized controlled studies immediate TNF-α inhibitor therapy was more efficacious on the long-term in the treatment of early rheumatoid arthritis, than strategies applying a delayed start of biological treatment. Our meta-analysis has also shown that the differences between the results of clinical studies are mainly explained by methodological differences, while the efficacy of individual TNF-α molecules do not differ from each-other. Re-examination of the evidences about the early use of TNF-α inhibitors in rheumatoid arthritis may lead to the update of therapeutic guidelines, and the more widespread use of biologic therapies. (Gulacsi et al., 2019)

Our other research projects pinpointed the difficulties of synthesising the evidence available for biological therapies, and therefore highlighted the challenges of making evidence-based health financing decisions. In case of infliximab, studies examining the effect of first-line biological therapy in early rheumatoid arthritis published results in 200 different combinations of outcomes and time-points. Out of these we found in only 17 cases (8.5%) two identical end-points, which enabled the synthesis of the available evidence. (Zrubka et al., 2018a). The studies examining the cost-effectiveness of first-line infliximab therapy in early rheumatoid arthritis were performed using the originator reference infliximab. The therapeutic guidelines published after the market entry of biosimilar infliximab referred to the unfavourable cost-effectiveness first-line biological therapy despite that cost-effectiveness studies were not performed with biosimilars priced at considerably lower levels compared to the originator molecules. (Zrubka et al., 2019c) The cost-effectiveness analysis of early biological therapy in rheumatoid arthritis is part of the ongoing research at the Department of Health Economics of the Corvinus University of Budapest. The example of infliximab illustrates, that although in the future the popularity and utilisation of biosimilars will increase, due to the deficiencies of the available evidence and regulations, biosimilars will probably not be able to fully offset the growth of expenses generated by newly registered innovative biological molecules entering the market.

II.3. Ethical decision-making under constrained resources

The allocation of scarce resources (rationing) is a subject of several theories of justice*, grounded in various ethical political or philosophical backgrounds. The levels of allocation, as well as the means of implementation are diverse, as the opinions about the topic. (Tanyi and Kollányi, 2008) The views that reject the restriction of provision of health services due to economic reasons are not acceptable neither on ethical, nor economic grounds. (Tanyi and Kollányi, 2008) Designing an efficient health service package (and thereby allocation of scarce resources based on economic arguments) is one of the main pillars of universal health coverage, according to the WHO. (WHO, 2010) Williams argued that cost-effectiveness analysis is ethical, since all costs are associated with consequences, and therefore they have negative implications on the bearers. Therefore, ignoring costs during any decision can be regarded as unethical practice. He also points out that every decision is made by hidden priorities, and the benefit of systematic cost-effectiveness analysis is derived exactly from revealing those hidden preferences so that the most appropriate ethical stance for the given context can be deliberately chosen. (Williams, 1992) Norman Daniels and James Sabin (Daniels and Sabin, 2008, Daniels, 2000) see fairness and legitimacy of resource allocation decisions in the nature of the process leading to the decisions, and not in the applied principles of justice per se. According to the theory of „accountability for reasonableness" the characteristics of a fair and legitimate process are the following:

a) Transparency: the principles serving as the basis for decisions need to be transparent.
b) Relevance: decisions have to be made by principles which are acceptable for all stakeholders of the decisions
c) Revision: there must be a mechanism, which allows the revision of past decisions on the grounds of new evidences, and settle disputes
d) Enforcement: there must be a process that guarantees the implementation of the above three elements.

The authors stress that decisions must follow one of the general theories of justice and they should not be discriminative.

II.4. Theories of fair allocation of health resources

Theories of justice differ from each other in whether they are concerned primarily with the fair allocation of health, healthcare resources, access or the satisfaction of needs. (Culyer and Wagstaff, 1993) Health cannot be redistributed, and health is affected by several societal factors. Therefore, practically applicable principles rather concentrate on the fair distribution of the societal determinants of health, as there is no consistent theory, which would guide the fair allocation of health. (Tanyi and Kollányi, 2008, Gulácsi, 2012) The following theories are concerned with the fair allocation of health resources:

b) Utilitarianism* focuses on maximising individual utilities, its practical implementation is based on cost-effectiveness analysis. (Tanyi and Kollányi, 2008)

c) Communitarianism* focuses on the principles of deliberative democracy, in which individuals choose the health programs according to their own preferences, and collectively define its structure, rules, principles of distribution etc. (Tanyi and Kollányi, 2008)

d) Egalitarianism* is the most frequently cited theory of justice, with many interpretations, but the most common starting point is that people are morally equal, therefore their life chances cannot be determined by morally arbitrary factors (Tanyi and Kollányi, 2008) The following theories can be linked to egalitarian principles:
    - The needs-based approach states that people with equal needs need to be treated equally. (Gulácsi, 2012)
    - Equal access suggests equal chance for the access to services, irrespective of their results. (Gulácsi, 2012)
    - Minimum essential health coverage in each disease, defined via wide societal consensus. (Gulácsi, 2012)
    - The fair innings* principle suggests that everyone is entitled a fair length of life (e.g. 70-75 years), and we need to provide equal chances for people to reach it. (Gulácsi, 2012)

e) Prioritanism is the critique of the equal needs principle and suggests that people in more severe conditions should receive greater priority. (Parfit, 1995, Nord et al., 1999)

f) Libertarianism is the least commonly cited principle, which derives its healthcare model from the autonomy of the individual and the refusal of external authority. (Gulácsi, 2012)

g) Sufficientarianism is a new theory of justice, according to which it is not the equal distribution of goods that matter, but that everyone should have enough. (Gosseries, 2011, Wouters et al., 2017) Sufficientarianism can be further divided to a positive and a negative thesis (Casal, 2007):
   - Positive thesis: it is morally important that individuals live above the sufficientarian threshold.
   - Negative thesis: no other theories of justice (especially egalitarianism and prioritarianism) are morally irrelevant.

When interpreting theories of justice, it is important to emphasize, that parallelly with their development, technological innovation has shaped the relationship between the demand and supply of health services. At the establishment of the British National Health Service in 1948, regardless of age, gender, religion, place of living or societal class, provision of the same health services was the guiding principle, which is driven by the individuals' needs and not their ability to pay. This principle was soon replaced by the problem of the distribution of scarce resources (rationing). (Crisp, 2002) All theories of justice have elements that can be criticised from an alternative viewpoint, and most of them contradict the principle of maximum cost-effectiveness. The Oxfordshire Health Authority applies the accountability of reasonableness principle in practice, during which cost-effectiveness is evaluated initially, followed by a multiple-step deliberative process, gradually evaluating additional criteria of justice, and eventually decisions are made about the need extent to which cost-effectiveness principles are overridden. (Hope et al., 2002)

II.5. Health Technology Assessment (HTA)

According to the definition of the European Network for Health Technology Assessment (EUnetHTA) HTA *"is a multidisciplinary process that summarises information about the medical, social, economic and ethical issues related to the use of a health technology in a systematic, transparent, unbiased, robust manner. Its aim is to inform the formulation of safe, effective, health policies that are patient focused and seek to achieve best value."* (EUnetHTA, 2007) The recommended modules of HTA by EUnetHTA are the following (EUnetHTA):

a) Relative effectiveness assessment
   - Current use of the technology
   - Technical specifications
   - Safety
   - Clinical efficacy
b) Local (national) assessment
   - Costs and economic evaluation
   - Ethical evaluation
   - Organisational effects
   - Patient/ and society-level effects, legal implications

Health economic evaluation involves most frequently cost-utility or budget impact analysis. Both methods compare the new technology with the usually applied standard care. Cost-utility analysis* (CUA) matches health outcomes measured in quality adjusted life years* (QALYs) with costs, while budget impact analysis* (BIA) analyses the financial effects of introducing a new technology on the healthcare system. (Gulácsi, 2012)

II.6. Multiple criteria decision analysis (MCDA) in healthcare

Due to the multitude of sometimes conflicting distributional principles and the complexity of the information affecting decisions, health financing decision making is a complex, mentally demanding process, in which, without rational decision-analytical techniques, ad-hoc decisions may result in system-level flaws. (Baltussen and Niessen, 2006) The structure provided by Multiple-Criteria Decision Analysis

(MCDA) can provide the transparency required by the accountability of reasonableness principle. (Thokala et al., 2016) MCDA can be used in many layers of health decision making but compared to the complexity of the decisions being taken, it is a rarely used methodology. MCDA techniques can be manifold depending on the decision, available data and the underlying theories. According to their two main definitions, we differentiate techniques which aggregate multiple criteria in a single metric, (Keeney and Raiffa, 1993) or techniques which enable deliberation without aggregating the results (Belton and Stewart, 2002). The International Society of Pharmacoeconomics and Outcomes Research (ISPOR) has issued its guidelines for applying MCDA in evaluation of the value of health technologies. The main steps of the process are the following:

(f) Defining the problem: defining the goals, type of problem, alternatives, stakeholders and types of expected outcomes
(g) Criteria: choosing relevant criteria for the evaluation of the alternatives
(h) Evaluation of performance: gathering data based on the selected criteria and creation of a performance-matrix
(i) Scoring: within each criterion, determining the levels of performance
(j) Weighting: determining the relative importance of individual criteria
(k) Aggregation: based on the performance rating of alternatives, ranking alternatives based on their aggregated results
(l) Sensitivity analysis: evaluating the robustness of results.
(m) Interpretation of results and reporting

In the 35 English-language publications about MCDA between 2005 – 2015, 93% contained criteria regarding the technology, and 81.8% contained criteria regarding patient-centric aspects as well as budget impact, which highlights the importance of integrating economic aspects with other quality-related criteria in a transparent manner. (Kim et al., 2017)

II.7. Measuring health outcomes using quality adjusted life years (QALYs)

II.7.1. Introduction to the QALY concept

The QALY is the most frequently used measure of health outcomes in health economic analyses in countries with developed economies. (Rios-Diaz et al., 2016) The QALY expresses the length of life and the quality of life in a single measure, where 1 denotes full health, while death is denoted by 0. Health states worse than death are indicated by negative numbers. The QALY does not differentiate between the length or quality of life, the severity of conditions or the individuals. A year in full health is considered equal to ten years spend in 0.1 quality of life, and the improvement of the health from 0.2 to 0.4 in a 30-year-old person is equal to the improvement from 0.8 to 1.0 at the age of 60 (Gulácsi, 2012).

The utility values of each health state are determined by measuring the preferences of the general population. It is still a matter of active debate, whether utilities should be derived from the normal population or affected patients, and whether the evaluated health states should be imagined, or actually experienced by respondents (Brazier et al., 2018). According to one reasoning, patients can provide more accurate information abouth the health states they experience, and they are the stakeholders of the decisions. According to the other reasoning, members of the general population can evaluate conditions impartially, and they are the bearers of the costs of care. Patients usually adjust to their health states, and therefore they evaluate the imagined health states as less severe, compared healthy individuals. HTA authorities in most countries rely on utilities derived from the normal population evaluating imagined health states, but interestingly the Swedish authority prefers, and some countries accept health economic evaluations using utilities derived from patients. (Brazier et al., 2018) Health-related utilities can be determined via direct and indirect methods. Direct methods use valuate concrete disease-descriptions, while indirect methods render utilities to health states described by generic quality of life instruments. The common feature of both methods is that they evaluate the utility of discrete health states (Brazier et al., 2019).

II.7.2. The EQ-5D questionnaire

In health-economic analyses the EQ-5D* is the most frequently applied generic quality of life instrument (Brazier et al., 2019). The so-called EQ-5D-3L version measures 3 levels of health problems (1: no, 2: moderate, 3: severe) in five health dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), thereby describing 243 (3^5) discrete health states. We denote health states by a five-digit number constructed from the problem levels in each dimension, called and EQ-5D-3L profile*. For example, the 21121 profile indicates a health state characterised by moderate problems in the mobility and pain/discomfort dimensions (EuroQoL Group, 1990). There is version using five problem-levels, and thereby describing 3125 (5^5) distinct health states (EQ-5D-5L) (Herdman et al., 2011), however, this dissertation focuses on the EQ-5D-3L questionnaire and the term EQ-5D will refer to the EQ-5D-3L instrument. The sum of the problems levels in each dimension are called the „misery index” (Augustovski et al., 2013), which can be used to estimate the severity of problems, but it is not adequate measure for determining QALYs. In case of the EQ-5D-3L instrument, the misery index can take values between 5 (1+1+1+1+1) and 15 (3+3+3+3+3). The EQ VAS* thermometer is also part of the EQ-5D instrument, which records the current health status of respondents between the imaginable worst and best (0-100) health states. Utilities can be rendered to the discrete EQ-5D health states by using the standard gamble (SG)*, time trade-off (TTO)*, rating scale (VAS)*, or recently the discrete-choice experiment (DCE)* methods. The common feature of all techniques is that the valuation is performed only on a selected health states (frequently based on considerations not necessarily optimal for econometric evaluation) (Yang et al., 2018). The displayed health states contain combinations of different levels of problems in the five dimensions, which are evaluated jointly by the individuals. The full value-set is determined using econometric models.

II.7.3. Valuation of health states

The methods of health valuation are the following:

a) SG: subjects need to choose from two situations: A) he/she stays in the displayed health state until the end of his/her life, B) due to a therapy, with *p* probability he/she returns to perfect health, or with *1-p* probability he/she dies immediately. The utility of the health state is calculated from the p value where the individual becomes neutral between the two alternatives. (Gulácsi, 2012)

b) TTO: the subject needs to choose from two alternatives: A) he/she stays in the displayed hypothetical health state for *x* years. B) he/she gets perfect health for t years and dies immediately thereafter. The utility can be calculated from the t value, where the individual becomes neutral between the two alternatives. (Gulácsi, 2012)

c) DCE: subjects need to choose between two alternatives: A) he/she spends x time in health state „A" or B) he/she spends *y* time in health state „B". The combination of problems and the length of time periods always requires evaluation, there is no straightforward choice. The utility can be derived using multivariate (e.g. multinomial logit) regression (Bansback et al., 2012, Baji, 2012, Hauber et al., 2016).

d) VAS: the subject marks the place of death on the health thermometer between the best and worst imaginable health states. Then he/she marks several predefined health states. The utility is derived from the relative place of health states between full health (100) and death (0) (Gulácsi, 2012, Brooks et al., 2003).

The utilities rendered to each EQ-5D profile are called EQ-5D index values. In different countries the index values determined with specific valuation methods are called value sets*. For instance, the EQ-5D-3L index values of health states 11111, 21121 and 33333 using the United Kingdom's TTO value set (UK-TTO) are 1, 0.727 and -0.594, respectively (Dolan, 1997).

The theoretical background of measuring health-related utilities is rooted in the Von-Neumann-Morgenstern expected utility theory (Von Neumann and Morgenstern, 1944). With the exception of the VAS rating scale, valuation methods involve a hypothetical choice situation, an exchange involving sacrifice and different levels of uncertainty. However, due to the different cognitive evaluation tasks, the valuation results of the different methods are also different. (Bleichrodt and Johannesson, 1997, Bansback et al., 2012) Furthermore, several studies have demonstrated that during the evaluation of utilities, respondents diverged from perfectly rational choices, according to patterns predictable by Kahneman and Tversky's prospect theory (Bleichrodt and

Johannesson, 1997, Kahnemann and Tversky, 1979). Utility measurements are influenced by the following systematic distortions (van Osch et al., 2004):

a) Time preference: people value more recent health gains as more valuable than distant ones in time. This results the under-valuation of utility values in TTO exercises.
b) Estimation of probability: people overestimate the likelihood of events with low probability (<0.33) and undervalue the likelihood of events with high probability (>0.33). This results in the overestimation of utilities in SG experiments.
c) Loss-aversion: people are more sensitive to losses than to gains, therefore both TTO and SG result in over-valuation of health states.
d) Measurement scale effect: the interaction of the instrument and the responses have been described for both SG, TTO and VAA (Parkin and Devlin, 2006, van Osch et al., 2004, Matejka et al., 2016).

II.7.4. Critical analysis of the QALY concept

In the following chapter I will briefly refer to the results of my research conducted with the EQ-5D and CarerQoL questionnaires, which complement the QALY concept or focus on its measurement properties (Zrubka, 2017b, Zrubka et al., 2019a, Baji et al., 2019, Zrubka et al., 2019b).

From systematic distortions of the valuation methods we can conclude that health related utilities cannot be viewed as time-independent linear concepts, while we apply this assumption during health economic analyses.

The inaccuracies of models based on QALY maximisation are not explained only by the systematic distortions of the valuation methods, but also by their differences from societal priorities about health gains*. After reviewing 64 empirical studies, Dolan and colleagues have identified the following typical societal preferences (Dolan et al., 2005):

a) People give priority to treating more severe patients
b) People give priority to young people (especially young adults) over older individuals

c) People would attach negative priorities to those who are (may be (*own remark*)) considered responsible for their health problems.
d) People would give priority to those who bear responsibility for others (e.g. parents of small children)
e) People prefer to balance societal inequalities

The most typical pattern in the terms of time and quality of life is the diminishing marginal utility: the societal preferences to treat healthier and older individuals tend to be lower. (Dolan et al., 2005) Furthermore, threshold-dependent changes of health preferences were also observed in some studies: individuals would distribute equally health gains that are greater than a certain threshold. It this threshold is not achievable, then people would focus the resources on a smaller group of individuals in need (Olsen, 2000).

A further critique of the QALY model is that certain variables that are important for the health or wellbeing of the individual are not measured sensitively enough or not measured at all by generic quality of life instruments. The most important elements that are missing from the current QALY model are the following (Brazier et al., 2019):

a) Health dimensions: sleep, vision, hearing, cognition
b) Dimensions of social and subjective wellbeing: autonomy, social relationships, feeling in control, positive affect
c) The quality of life of informal caregivers

Most dimensions that are missing from the QALY model have validated measures, but the value sets required for health economic analyses have not been determined, with some exceptions:

a) ICECAP: this measure of wellbeing is based on Amartya Sen's capability approach, which was developed to measure the effect of health and social care interventions. The questionnaire measures full-capabilities / lack of capabilities along several dimensions, and the value set has been determined by best-worst scaling method. ICECAP has adapted versions for several situations (adults, older individuals, individuals with terminal illness, caregivers) (Al-Janabi et al., 2012, Flynn et al., 2015)
b) CES (Caregiver Experience Scale): member of the ICECAP „family", this instrument measures the quality of life of informal caregivers. Its valuation

was performed using the best-worst scaling method, based on the preferences of caregivers of elderly patients (Al-Janabi et al., 2011).

c) CarerQoL measures the quality of life (both positive and negative experiences) of informal caregivers in a similar format to the EQ-5D. The developers rendered a happiness-based value set to CarerQoL reflecting the preferences of the normal population. The value set has been determined in several countries (Brouwer et al., 2006, Hoefman et al., 2017). The Hungarian adaptation of CarerQoL is already ongoing at the Department of Health Economics of the Corvinus University of Budapest (Zrubka, 2017b, Baji et al., 2019).

Overall, models that maximise QALYs do not provide a solution for the optimal allocation of scarce resources, neither from a utilitarian, nor from an egalitarian point of view, which explains, why it is so important to consider other qualitative perspectives in addition to results of cost-effectiveness analyses during health resource allocation decision making. However, the EQ-5D based QALY concept is supported by the deep knowledge about the measurement properties of the instrument due to very thorough psychometric validation. Furthermore, due to the widespread use of the instrument, data allowing the standard comparison of the broadest range of interventions are available (Brazier et al., 2019). However, comparative studies do not solve the problem of measurement biases. In our own research, we compared the EQ-5D value sets of four countries for 18 diseases. The results pointed out the differences of the value sets that arise from different valuation methodologies. During the pairwise comparisons of disease burden (the difference between the utilities of patients and those of the age and gender-matched normal population) between all diagnoses, change of the value set changed the priority order of diseases in 23% of the cases (Zrubka et al., 2019a, Zrubka et al., 2019b). Measuring QALYs and the application of EQ-5D is an area under continuous methodological development, and financing decisions based exclusively on QALY-based calculations can be just as flawed as the ones, which ignore the health-economic evidences obtained through the use of currently available instruments.

II.8. Measuring acceptable health states (AH)

Theories of justice concerning the fair allocation of health resources consider perfect health as the point of reference, where health losses or gains are compared to perfect health with utility 1. The concept of acceptable health states (AH) is based on the assumption that certain health problems are taken by people as a natural consequence of aging, which is worthwhile taking into consideration during the allocation of scarce health resources, and use them as a reference-point instead of perfect health (Wouters et al., 2017). Empirical studies have shown that people consider more health problems acceptable with age. (Brouwer et al., 2005, Pentek et al., 2014b, Wouters et al., 2015)

The measurement of acceptable health is based on the prospect theory of Kahneman and Tversky, according to which people evaluate gains and losses compared to an internal reference point. In turn, the reference point is influenced by the options and the expectations of individuals (Kahnemann and Tversky, 1979). A loss below the reference point (not acceptable health state) is more important for people, than the gain above the reference point (better than acceptable health). According to the AH concept, not acceptable health states would receive greater, while acceptable health states would receive smaller priorities during decision making, so the willingness to pay of the society would be greater for health gains below the acceptability threshold than for ones above (Wouters et al., 2017).

The measurement of AH is a new field, apart from my research, so far three studies were concerned with the measurement of AH. All three studies have confirmed that there is an internal reference point, according to which people evaluate the acceptability of health states in different ages (Brouwer et al., 2005, Pentek et al., 2014b, Wouters et al., 2015). It was a common feature of all three studies that the acceptability of health states was measured using the EQ-5D instrument, which is applied for QALY calculations as well.

The first study was conducted by Brouwer et al. at the Erasmus University Rotterdam by interviewing 226 individuals between 18-65 years of age from the Dutch general population (Brouwer et al., 2005). The key questions of the study were:

a) Are certain non-perfect health states acceptable as age advances?
b) Is the acceptability of problems different in certain health dimensions?

The question was asked separately for each EQ-5D dimension, according to the following wording: „please indicate with X,

The question was asked by each dimension separately, according to the following wording: *„Can you indicate beyond what age do you consider the specified level of problems with 'mobility' to be acceptable?"* The version of this question used in our research is shown in Figure 3). Age groups were asked between 30 and 80 years in 10-year intervals. Participants were instructed to imagine perfect health in other situations. In addition to EQ-5D dimensions, questions relating to sexual activity and fertility were asked as well. The results confirmed that the acceptability of problems was different by dimensions and increased with age. The level of acceptable health was somewhat rated lower by respondents than the health status of the normal population for each age group. Moderate problems were more frequently acceptable than severe ones, and problems were least frequently acceptable in the anxiety / depression dimension. As a weakness of the research, the authors noted that the evaluation of the acceptability of problems happened by dimension, separately, while in reality problems happen in combination, and their joint evaluation would probably influence their acceptability. Their other question was whether people mean the same problems under the same problem levels in different ages? If a young or older respondent evaluates the same objective functional state with different severity labels in different ages, that may influence the assessment of acceptability (Brouwer et al., 2005).

Figure 3 Separate evaluation of acceptable health problems

Mobility
Could you please indicate from which age onwards you consider problems with mobility acceptable?

	Age (years)						
	From 30 onwards	From 40 onwards	From 50 onwards	From 60 onwards	From 70 onwards	From 80 onwards	Never
Some problems in walking about				X			
Confined to bed						X	

The question used in our research during the separate evaluation of the acceptability of health states

The second study was conducted by Péntek et al. in Hungary on 77 patients with rheumatoid arthritis starting biological therapy. (Pentek et al., 2014b) The research objective and methods were similar to the Dutch study, but in addition to EQ-5D, acceptable health states were measured using the HAQ-DI disease-specific quality of life instrument (Maska et al., 2011). The study confirmed in both the EQ-5D and HAQ-DI dimensions that chronic patients consider more problems acceptable with age, and the level of acceptable problems increases with age. Severe problems were less acceptable than moderate ones, and patients considered problems with usual activities the least acceptable. Patients rated level of acceptable health above 60 years of age as lower than that of the normal population, while higher than their own health status. Among the limitations of the study the small sample size, the lack of a validated instrument for the measurement of acceptable health, and the problem, which was already pinpointed by Brouwer et al: the evaluation of acceptable health states happened separately by dimensions and joint jointly (Pentek et al., 2014b).

The third study was also conducted by the Erasmus University Rotterdam team (Wouters et al., 2015). Data were collected from a representative online sample of 1067 individuals in the Netherlands. The goals of the research were the following:

a) Reinforcing the concept of AH: are less-than-perfect health states acceptable and does their acceptability depend on age?
b) From acceptable health states and acceptable length of life constructing a „measure of acceptable health”, which, similarly to the fair innings concept, would measure a fair lifetime amount of acceptable health.
c) Analysis of the association of acceptable health with socio-demographic variables and health related experiences.

Acceptable health was measured in three ways:

a) Separate evaluation: in each EQ-5D dimension, from what age the different levels of problems are acceptable? (Between 40 and 90 years, in 10-year intervals).
b) Joint evaluation: what are the ages, from which the below mentioned three EQ-5D health profiles (21211, 22221, 33312) are acceptable?
c) Acceptable length of life: from what age is death acceptable?

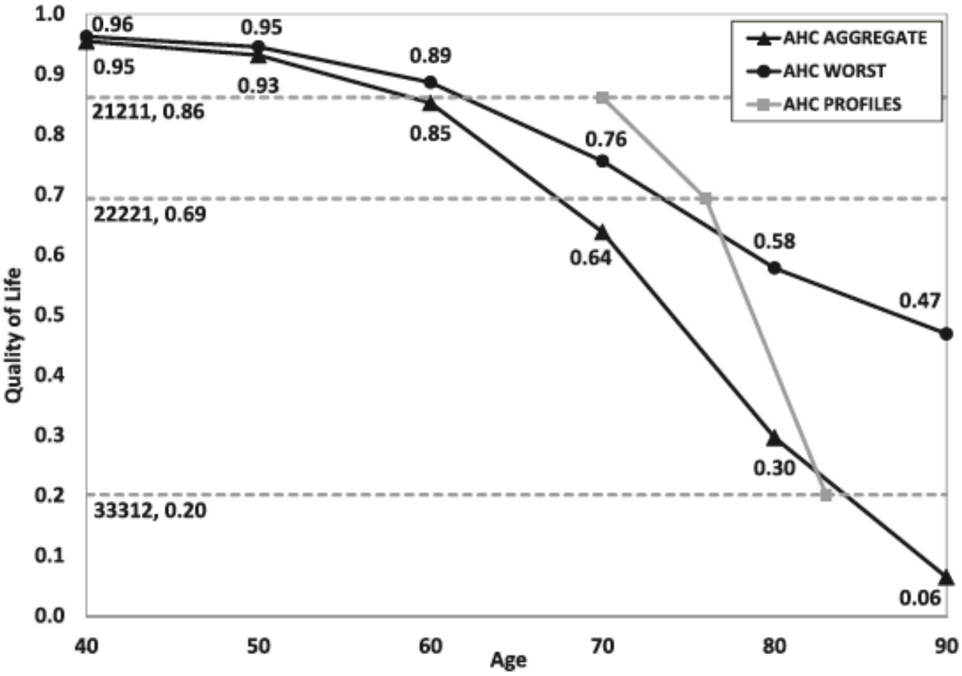
Subjects of the separate evaluation did not receive further instructions regarding how to imagine the health problems in each dimension: a) with full health in other dimensions, b) in combination with other problems. For summarizing their results, the authors constructed the following measures of AH:

a) Aggregate Acceptable Health Curve ($AHC_{aggregate}$): the main assumption is that the health state aggregated from the problems that are acceptable separately in each dimension would also be acceptable by the respondent. The sample's $AHC_{aggregate}$ is constructed from the mean EQ-5D-3L index values of the individual $AHC_{aggregate}$ health states.
b) Worst Acceptable Health Curve (AHC_{worst}): the main assumption is that problems in each EQ-5D-3L dimension are acceptable only in the presence of perfect health in other dimensions. Since health problems in several dimensions may be indicated as acceptable in a certain age, the sample's AHC_{worst} is constructed from the lowest individual AHC_{worst} values in each age.
c) Profiles' Acceptable Health Curve ($AHC_{profiles}$): the mean age from which EQ-5D-3L profiles are considered acceptable by the sample is plotted against their EQ-5D-3L index values.

The three acceptable health curves ($AHC_{aggregate}$, AHC_{worst}, $AHC_{profiles}$) are depicted in Figure 4. Main findings of the study were the following:

a) The concept of AH was reinforced: people consider more non-perfect health states acceptable with age. Similarly to the previous Dutch study, severe problems are less acceptable as moderate ones, and problems were least acceptable in the anxiety / depression dimension.
b) Age, health status, lifestyle and experiences concerning health influence the amount of acceptable health. Older respondents and ones who had chronic or severe disease considered more health problems acceptable. People following a healthy diet considered less problems acceptable. The lifespan of the closest relative also influenced the amount of acceptable health.

Figure 4 Acceptable health curves



Source: (Wouters et al., 2015) We used the figure with permission from the authors under the Creative Commons 4.0 (CC BY 4.0) agreement. The vertical axis indicates quality of life measured by the EQ-5D-3L index values. Aggregate Acceptable Health Curve, ($AHC_{aggregate}$): mean index values of health states aggregated from all acceptable health problems in the five dimensions. Worst Acceptable Health Curve, (AHC_{worst}): mean of the lowest index values of acceptable health states constructed from isolated health problems in each dimension assuming full health in others. Profiles' Acceptable Health Curve, (AHCprofiles): constructed from the index values of the jointly evaluated health profiles and the mean age from which they were considered acceptable

The limitation of the study was the following:

a) There were no respondents above 65 years of age, therefore the preferences of the elderly population could not be analysed.
b) AH was measured via separate evaluation. The three AHCs showed inconsistent picture. While the difference was rather big between the levels of $AHC_{aggregate}$ and AHC_{worst} in older age groups, the jointly evaluated $AHC_{profiles}$ crossed the curves gained with separate evaluation. (When evaluated jointly, moderate states were acceptable in older ages, while more severe states in younger ages, than expected from the results of separate evaluation.) Finding the exact location of the acceptable health curve was not successful, and the authors proposed the joint testing of more EQ-5D profiles as a future research field in order to improve the accuracy of measurement.

c) Acceptable health was not defined exactly: therefore, it could be interpreted in multiple ways by respondents (it is the health state still enough for wellbeing, or the still tolerable level of suffering?)
d) The reference person was not defined: respondents could evaluate acceptable health referring to themselves or others as well.

The authors noted, that the normative framework of applying AH in health financing decision making has not yet been established, therefore they could just conceptually illustrate the application of AH.

II.9. The normative framework for using AH in financial decision making

Based on sufficientarian reasoning, Wouters et al. examined a possible normative framework for applying AH in financial decision making (Wouters et al., 2017). Their main conclusions were the following:

a) Deterioration of health with age is a natural phenomenon, it affects everyone, therefore AH may be a feasible reference point for health financing priority setting.
b) The concept of AH fits the sufficientarian principle, insofar as it does not aim to either achieve perfect health or equality of health. Its aim is that everyone gets into acceptable health.
c) The positive thesis of sufficientarian reasoning is also acceptable: it is morally important to live above the acceptability threshold. If individuals who are not acceptable health receive greater, while ones in acceptable health receive lower priorities, that can be in harmony with the prioritarian principle, the fair innings principle and also with preferences of the general population. The negative thesis (the invalidity of other theories of justice) is not acceptable.
d) The financing importance weight *(w)* rendered to the health gains can be defined with several versions of the AH-based value-function ($w=f(U)$), where K_E denotes the starting health state in acceptable health, K_N denotes a not acceptable starting health state, E_E denotes an acceptable health state as outcome, E_N denotes a not acceptable health state as outcome. U is the utility of health states, T_E is the acceptability threshold, and c_1, c_2, p_1, p_2 and q are arbitrary parameters, where $0<c_1, c_2, p_1, p_2, q<1$.

- Sufficientarian base value function (Figure 5.a):

$$f(U) = \begin{cases} 1, if\ U \leq\ T_E \\ 0, if\ U > T_E \end{cases} *$$

- Strict sufficientarian value function (Figure 5.b):

$$\mathrm{f(U)} = \begin{cases} 0, \text{if } K_N \wedge E_N \\ 1, \text{if } K_N \wedge E_N \wedge U \leq T_E \\ 0, \text{if } (K_N \wedge E_N \wedge U > T_E)\ \vee (K_E \wedge E_E) \end{cases}$$

- Modest sufficientarian value function (Figure 5.c):

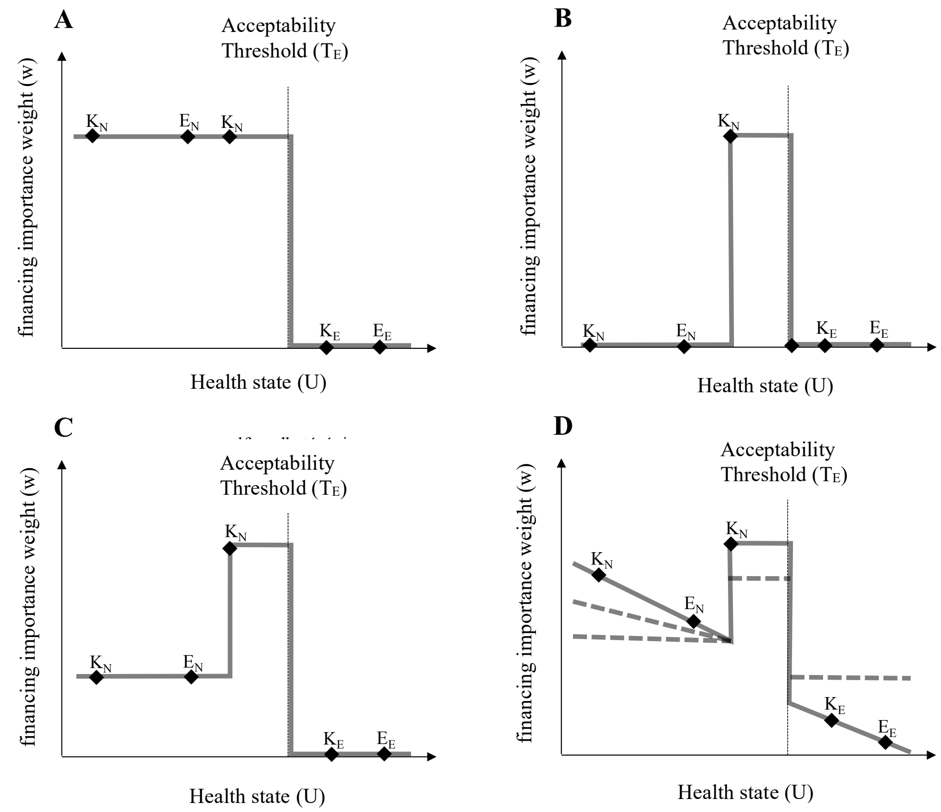
$$\mathrm{f(U)} = \begin{cases} q, \text{if } K_N \wedge E_N \\ 1, \text{if } K_N \wedge E_N \wedge U \leq T_E \\ 0, \text{if } (K_N \wedge E_N \wedge U > T_E)\ \vee (K_E \wedge E_E) \end{cases}$$

- Sufficientarian and prioritarian hybrid value function (Figure 5.d):

$$f(U) = \begin{cases} c_1 - p_1 U, if\ K_N \wedge E_N \\ q, if\ K_N \wedge E_N \wedge U \leq T_E; where\ q <\ c_1 - p_1 U \\ c_2 - p_2 U, if\ (K_N \wedge E_N \wedge U > T_E)\ \vee (K_E \wedge E_E);\ -p_2 U <\ c_1 - p_1 U \end{cases}$$

e) The principle of AH is acceptable morally if there is considerable difference between the utilities of acceptable and not acceptable health states. However, the precise measurement of the AH threshold has not been feasible so far. Therefore, the authors suggested a soft application of the AH concept: in combination with other theories of justice, they consider the application of rounded utility functions feasible instead of using ones with sharp threshold values.

Figure 5 Value functions based on the principle of acceptable health (AH)



Source: adapted from (Wouters et al., 2017). The horizontal axis depicts the utility of health states, while the vertical axis shows the financing priority weights (min 0, max 1) attached to the starting (K) and outcome (E) health states during the technology assessment of a technology. A health gain is illustrated by a move on the horizontal axis from left to right. Health states left from the acceptability threshold are not acceptable, while ones on the right are acceptable. K_N denotes a not acceptable starting health state, E_E denotes an acceptable health state as outcome, E_N denotes a not acceptable health state as outcome. A) Sufficientarian base function: the priority weight (w) of health states in not acceptable health is 1, while in acceptable health states is 0. B) Strict sufficientarian value function: the priority weight of health gains in not acceptable health states is 1, while health gains in acceptable health, or health gains that do not take the individual to acceptable health have priority weight 0. C) Modest sufficientarian value function: in not acceptable health states the priority weight falls between 0 and 1, the weight of health gains in acceptable health states is 0, and the weight of health gains leading from not acceptable health states to the acceptability threshold is 1. D) Sufficientarian and prioritarian hybrid value function: the priority weight of severe health states is greater than that of milder health states, the weight of health gains from non-acceptable health states until the acceptability threshold is 1.

Wouters et al. (Wouters et al., 2017) proposed the following areas of research for the future development of the AH concept:

a) Embedding the AH concept in the normative framework of alternative theories of justice
b) Developing reliable methods of measurement
c) Adapting the AH concept to health outcomes other than QALYs

d) Developing AH framework for interpreting health losses,
e) Determine the interrelationship between AH and time

II.10. Avenues for the further development of the AH concept

The pressure from technological innovation in healthcare requires the fair and effective harmonisation of societal priorities and economic constraints in healthcare decision making. Prioritanism and the fair innings principle are the two widely accepted societal priorities, which could be considered during priority setting in healthcare financing. However, these two principles do not provide answers about what weight the decision maker should attach to each of them during their simultaneous application. Which should receive higher priority from a mild condition in a young individual or a severe one in an older patient? The benefit of the AH principle is that it unites people's preferences about the severity of disease and age – since the acceptability of a health state is defined simultaneously by both its severity and the age of the individual in question. Despite its merits, the QALY-based measurement of health outcomes can be criticised from multiple grounds, therefore it is necessary to consider further aspects for the sake of effective decision-making. The benefit of the AH concept is that it is based on the EQ_5D questionnaire, therefore it provides an alternative measurement method of health outcomes by recycling EQ-5D data, which have been most frequently used for QALY measurement. Introducing the method can be efficient, while it does not preclude the evaluation of the acceptability of other aspects of health, such as capabilities necessary for broader wellbeing, or the dimensions missing from the current QALY concept. Last but not least, the benefit of the AH concept is that it relies on the opinion of the general population, therefore it is a suitable tool to provide transparent and relevant criteria required for legitimate decision making during the priority setting in healthcare financing.

However, the precondition to the practical application of the AH concept is that its accurate and reliable measurement methods need to be developed. The separate evaluation of acceptable health states was a readily usable simple method during the initial exploratory studies of AH. However, health valuation methods evaluate the health states jointly. Therefore, for the reliable and valid measurement of AH we need to develop the method of joint evaluation.

Measuring AH is not based on traditional utility theories, its psychometric properties and its theoretical background for its application for health economic evaluations have yet to be developed.

Proving the basic assumption, that in acceptable health gains are truly less valuable than in not acceptable health states is an important condition that we accept lower financing priorities for health gains in acceptable health states. Furthermore, for valid application of the method, we need to understand better the psychological processes during the evaluation task, and the factors influencing the evaluation of the acceptability of health states.

The normative framework for the practical application of AH has not yet been firmly established yet. In harmony with known societal preferences, we can only accept the goal to get individuals to acceptable health. This is in accordance with the positive thesis of sufficientarian theory. According to utilitarianism, it is also important to get as many individuals to acceptable health as possible. AH is compatible with egalitarianism insofar as we consider equally important the treatment for all who are in not acceptable health states, while those, who are in acceptable health receive lower priority, regardless their age or the severity of their disease.

III. GOALS AND HYPOTHESES

My thesis has two overarching goals in the development of the AH concept, and I have tested six hypotheses in connection with the primary research goals.

1. Goal 1.: developing new measurement methods for AH, which are more accurate than previous techniques.

   1.1. *Joint evaluation of the acceptability of discrete health states*: instead of asking the acceptability of health states by dimension (separate evaluation), we evaluate the acceptability of joint health states displayed as vignettes containing different levels of the five EQ-5D dimensions, using an adaptive algorithm that selects questions based on the previous answers of respondents. In joint evaluation we aim to obtain a yes/no answers for all elements of the acceptability matrix (E-matrix). The E-matrix consists the acceptability information of 243 health states (constructed from the three levels on the five dimensions of the EQ-5D) across six ages from 30 to 80 years in 10-year intervals.

      1.1.1. *Hypothesis 1 (H_1): Using the adaptive algorithm, all elements of the E-matrix can be unequivocally determined for 90% of respondents. (Methods: page 55, results: page 62.)*

      Hypothesis 2 (H_2): With joint evaluation, people consider less problems acceptable compared to separate evaluation. (Methods: page 55, results: page 75.)

   1.2. Overall assessment of AH using the EQ VAS: the EQ VAS measures AH on a continuous scale on which 0 indicates the worst and 100 indicates the best imaginable health state.

      Hypothesis 3 (H_3): When measured with the EQ VAS, people consider worst health states acceptable at older ages than at younger ages. (Methods: page 55, results: page 83.)

2. Goal 2.: Exploring the association of AH and happiness.

   2.1. Measuring the association of the acceptability of individuals' health states and their happiness.

Hypothesis 4 (H_4): In acceptable states the level of health influences peoples' happiness to a lesser extent than in not acceptable health states. (Methods: page 104, results: page 108.)

2.2. Exploring the factors that affect the acceptability of individuals' health states

Hypothesis 5 (H_5): Older individuals are more likely to consider their health acceptable than younger ones (Methods: page 104, results: page 113.)

Hypothesis 6 (H_6): Individuals with worse subjective health consider their condition less acceptable than ones indicating better subjective health. (Methods: page 105, results: page 113.)

IV. THE MEASUREMENT OF ACCEPTABLE HEALTH STATES

IV.1. Background

The first goal of my dissertation was to develop the methodology of the joint evaluation of acceptable health states. The EQ-5D-3L describes 243 discrete health-states, out of which 232 (95.5%) contains problems in more than one dimension, as opposed to the health states included in separate evaluation. Since decision situations involving multiple attributes simultaneously or separately may result in different preferences (Hsee et al., 1999), our main assumption was that similarly to the valuation tasks of health-state utilities, by displaying EQ-5D profiles (Dolan, 1997), the joint evaluation of multiple helath problems will provide more reliable results than separate evaluation. The uncertainty of interpreting the results of separate evaluation is illustrated by the work of Wouters et al. (Wouters et al., 2015) (Figure 4) In their research, the difference between $AHC_{aggregate}$ and AHC_{worst} was nearly twice as big at age 70 as the 0.074 minimal clinically important difference (MCID) of the EQ-5D-3L UK-TTO value set, and it was nearly four times the MCID at age 80. Therefore, Wouters et al. measure the acceptability of three EQ-5D profiles, but they used a different evaluation logic, during the joint evaluation of profiles than during separate evaluation. Separate evaluation provided answers to the question: what percentage of people consider a health state acceptable in a given age. However, joint evaluation answered the question: what is the average age from which a discrete health state is acceptable (Wouters et al., 2015). The two approaches provided inconsistent acceptable health curves (Figure 4), therefore our goal was to develop a joint evaluation strategy, that is consistent with separate evaluation. As a solution, Wouters proposed the evaluation of the acceptability of several EQ-5D profiles. (Wouters et al., 2015) The 243 profiles in 6 ages provide 1458 health states, for which we apply the term E-matrix. We will refer to the elements of the E-matrix as *health states*, while the 243 different health states will be referred to as *EQ-5D-3L pro*files or simply *profiles*. The first question was: how many and which health states should be evaluated in order to receive a satisfactory picture about the acceptability of the 1458 health states?

Another part of the EQ-5D questionnaire is the EQ VAS health thermometer, which serves for the global evaluation of respondent's current health status. While the EQ-5D profiles enable the comparison between individuals, the EQ VAS provides numerical information about the individual's current global health status (Feng et al., 2014). In addition to the evaluation of the discrete health states of the E-matrix, we wondered, how the EQ VAS could be applied for the measurement of acceptable health.

IV.2. Methods

IV.2.1. Joint evaluation of discrete health states

During planning our research, one potential approach was to ask the acceptability of the 1458 health states in a block-design. In an EQ-5D valuation experiment, university students evaluated all 243 profiles. The average time span was 86 minutes. In the DCE valuation of the EQ-5D 54 profiles are valued in pairs (27 questions), and the most profiles included in valuation studies were 47 (Yang et al., 2018). Since 1458 can be written as the product of 27 and 54, in case of 27 questions per respondent, for measuring acceptability with the 95%CI==±5% precision range we need to interview 400*54=21600 individuals, while for 95%CI=±3% we need to involve 54000 respondents. Therefore, we abandoned the block-design strategy.

The next approach was the assembly of a standard set of questions. From a previous study in the general population, we had data from 9260 $AHC_{aggregate}$ curves obtained via separate evaluation (Péntek et al., 2009). Since the aggregate health state is the possible worst health state acceptable in a certain age, during joint evaluation we just need to focus on whether the aggregate health state is acceptable or not, and if not, then what combination of problems in less severe health states are acceptable. This can narrow down considerable the number of necessary questions. However, data has shown a rather heterogenous pattern: from the sample of 9260, we found 4600 different $AHC_{aggregate}$ curves, and for evaluating all unknown health states, we would need to pose on average 94 questions per respondent. The number of possible questions per individual ranged between 0 and 1392. We concluded that with a standard questionnaire the accurate measurement of acceptable health problems is not feasible. Before developing a prediction strategy from incomplete data, we aimed to gain as much information as possible.

The final approach was the development of an adaptive questioning strategy, which would provide the most possible information about the acceptability of the 1458 health states using the fewest possible questions. Based on the results of EQ-5D studies in the normal population (Szende et al., 2014), we relied on the following assuptions when developing the adaptive questioning strategy:

a) We can treat the dimensions of EQ-5D profiles as ordinal variables: in case of a consistent respondent, if a health state is acceptable in a certain age, then we can consider all health states acceptable, in which any level of any problems is the same or lower than the health state in question (better health states). Likewise: if a health state is not acceptable, then those health states are not acceptable either, in which the level of any problem is the same or higher than that of the health state in question (worse health states). We cannot deduct the acceptability of unknown health states in which both more and less severe problems coexist compared to the health state in question.
b) Health deteriorates monotonously with age: in case of a consistent responder, if a health state is not acceptable in a certain age, then the same or worse health states are not acceptable either in that or younger ages. Vice versa: if a health states is acceptable in a certain age, then the same or better health state is acceptable in the same or older ages.

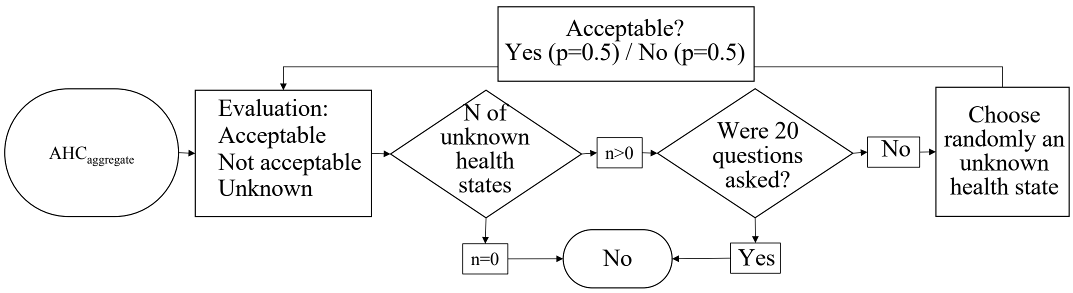
The assumptions above enable us to gain information from the acceptability of more than one health states with a single question. The tested the efficiency of the model via computer simulation. From the $AHC_{aggregate}$ curves gained with separate evaluation from 9260 individuals, our model determined the unknown states according to the following rules:

a) in every age, worse health states than aggregate health states were automatically considered not acceptable
b) if problems were indicated as acceptable in a single dimension, then those problems and milder problems in the same dimension were considered acceptable, if no problems were acceptable in any other dimensions in the same age.

c) all other health states were considered unknown.

Then, unknown health states were chosen randomly by the algorithm, and for each of them it assumed an acceptable / not acceptable answer for a hypothetical question with 50-50% probability. The model posed 20 random questions and re-evaluated the acceptability of all health states after each question – narrowing down the pool of unknown health states before each consecutive question. The algorithm is depicted in Figure 6.

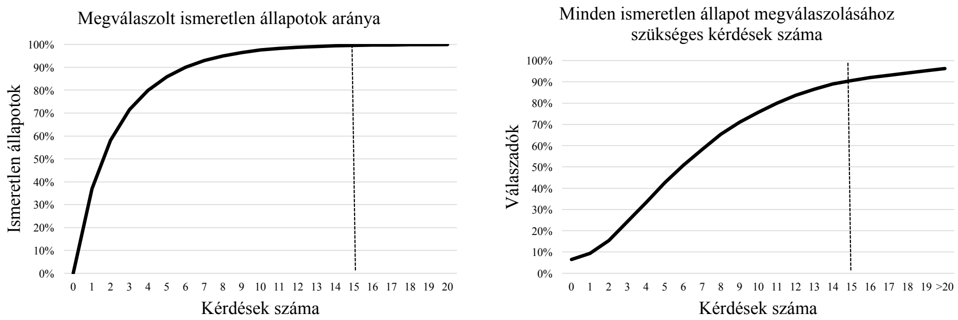
Figure 6 Computer simulation model of the adaptive testing algorithm



From the $AHC_{aggregate}$ data of the 9260 individuals we simulated the acceptability of 13 501 080 health states (9260 individuals*243 profiles*6 ages). Without asking any questions, 92.3% of the health states were not acceptable, 1.2% were acceptable, and 6.5% of them were unknown. After the simulation using 20 questions per respondent, the proportion of unknown states decreased to 0.06%, while for 96% of hypothetical respondents the acceptability of all health states could be determined (Figure 7). The differences between the $AHC_{aggregate}$ and AHCworst curves suggested that results are sensitive to the responses given to the 6.5% unknown health states, therefore we aimed to determine the acceptability of unknown health states as precisely, as possible.

During the fieldwork, we considered the optimal number of health state evaluations during joint evaluation as many as subjects could still pay attention and maintain their collaboration. According to the results of the simulation, with 15 random questions in 90% of subjects the acceptability of the 1458 health states could be fully determined, and the proportion of unknown states remained below 0.4%. Therefore, we decided to use a module of 15 questions during joint evaluation.

Figure 7 Results of the simulated adaptive testing



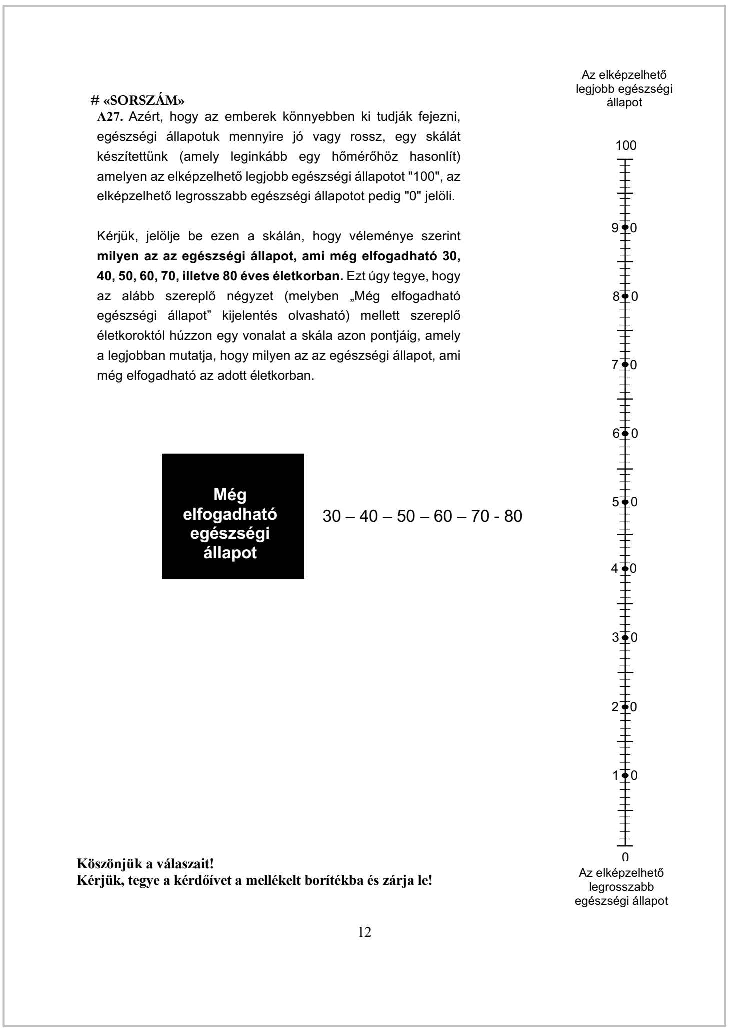
Simulation based on AHC$_{aggregate}$ data of 9260 individuals. For each individual, the algorithm evaluated the acceptability of 1458 (altogether 13 501 080) health states. Left figure: with 15 simulated questions the acceptability of 99.6% of the 13 501 080 health states could be elicited. Right figure: with 15 question the acceptability of all questions could be elicited for 90% of the simulated respondents.

For displaying the results of joint evaluation, on the analogy of AHC$_{aggregate}$ and AHC$_{worst}$, $_{we}$ constructed the Joint Acceptable Health Curve* (AHC$_{joint}$). After joint evaluation, we rendered the EQ-5D-3L index values (UK-TTO value set) to all acceptable health states, and in each age, we calculated from the lowest EQ-5D-3L index calculated for both individuals, and from their averages the sample AHC$_{joint}$.

IV.2.2. Measuring AH with the EQ VAS instrument

In case of electronic questionnaires, the display of visual analogue scales can influence the response pattern (Matejka et al., 2016), therefore we intended to modify the validated EQ VAS instrument as little as possible. On the EQ VAS thermomether respondents indicate their current health between the best (100) and worst (0) possible health states by drawing a straight line from a dot next to the middle of the thermometer. (Oppe and van Reenen, 2015) In VAS valuation studies respondents valued several health states on the same thermometer. Health states were placed on both sides of the VAS scale. (Brooks et al., 2003) In order not to influence responses placed on the vertical VAS thermometer, we arranged ages in a horizontal line in line with the middle of the scale (Figure 8). From acceptable VAS values we constructed the VAS acceptable health curve (AHC$_{vas}$).

Figure 8 EQ VAS adapted to measuring AH



«SORSZÁM»

A27. Azért, hogy az emberek könnyebben ki tudják fejezni, egészségi állapotuk mennyire jó vagy rossz, egy skálát készítettünk (amely leginkább egy hőmérőhöz hasonlít) amelyen az elképzelhető legjobb egészségi állapotot "100", az elképzelhető legrosszabb egészségi állapotot pedig "0" jelöli.

Kérjük, jelölje be ezen a skálán, hogy véleménye szerint **milyen az az egészségi állapot, ami még elfogadható 30, 40, 50, 60, 70, illetve 80 éves életkorban.** Ezt úgy tegye, hogy az alább szereplő négyzet (melyben „Még elfogadható egészségi állapot" kijelentés olvasható) mellett szereplő életkoroktól húzzon egy vonalat a skála azon pontjáig, amely a legjobban mutatja, hogy milyen az az egészségi állapot, ami még elfogadható az adott életkorban.

Még elfogadható egészségi állapot

30 – 40 – 50 – 60 – 70 - 80

Köszönjük a válaszait!
Kérjük, tegye a kérdőívet a mellékelt borítékba és zárja le!

12

Respondents had to link ages arranged horizontally at the middle of the scale with the levels of still acceptable health in respective ages. (0: worst imaginable health; 100 best imaginable health)

IV.2.3. Questions of the „Health and Ageing" study

In early 2018, we performed a cross-sectional survey via personal interviews among 200 members of the Hungarian general population selected by convenience sampling. Subjects provided written consent, and data were collected anonymously. The research plan was approved by the Ethical Committee of the National Research Council (ETT-TUKEB) under the identifier: 5111-2-2018/EKU.

We recorded acceptable health, respondents' health status, their lifestyle and main socio-demographic variables.

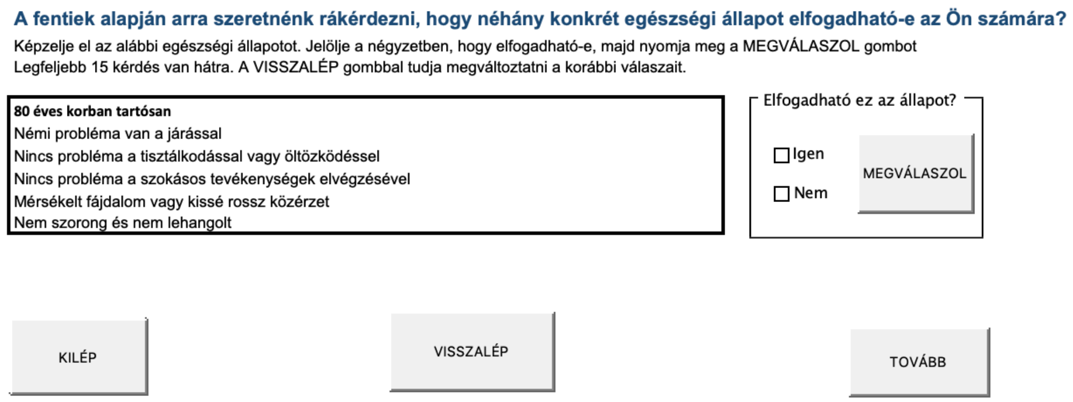
The questionnaire had two parts: we recorded via a computer-based questionnaire the separate and joint evaluation of acceptable health. Other variables, including the acceptable VAS module, were recorded on a paper and pencil questionnaire. We coded all questionnaires and the database was created via joining the respondents' answers anonymously.

We collected the following information via an electronic questionnaire:

a) Participant information and consent
b) Separate evaluation of acceptable health problems
c) Joint evaluation of acceptable health states
d) Persons imagined during the evaluation of acceptable health (from multiple answers we formed three categories, whether respondents imagined only themselves, themselves and others or only others during the joint evaluation.

The joint evaluation module is depicted in Figure 9. Respondents had the opportunity to change their answers provided during separate evaluation.

Figure 9 Joint evaluation module of AH



Respondents evaluated the acceptability of health states presented in the black frame. The adaptive testing algorithm selected 15 questions randomly from the unknown states, based on the previous answers. Source: (Zrubka, 2018b).

In the paper and pencil questionnaire we measured respondents' subjective health with EQ VAS. Acceptable health was measured via the adapted EQ VAS between 30 and 80-year-old age groups, in 10-year intervals. We divided respondents in 7 age groups: 18-24, 25-43, 45-54, 55-74 and 75+.

We also recorded the health status of respondents via the EQ-5D-3L questionnaire (EuroQoL Group, 1990) We analysed separately moderate and severe problems by each dimension, and also created an „any problem" category by counting moderate and severe problems together. In a separate variable, we recorded if respondents indicated any problems in any of the dimensions. We also recorded the health state utilities (EQ-5D-3L index values) for each individual.

For describing the socio-demographic status of respondents, we recorded the following data: age (and age group), gender, family status (married: married or lives in domestic partnership, vs. not married: single, divorced or widowed), education (tertiary: university of college degree, other: primary or secondary education), employment (employed: full or part-time, not employed: pensioner, student, housemaker), household income per capita. Based on data from the Central Statistical Office of Hungary, (KSH, 2015) we grouped respondents in the first two quintiles as having high income and ones in the remaining three quintiles as having low income.

We described the lifestyle of respondents with behaviours associated with health risks: overweight (body mass index (BMI) >25) (Garrow, 1981)), smoking (at any quantity) (Schane et al., 2010)), excessive alcohol intake (men: >14 drinks / week or >4 drinks / occasion, women >7 drinks / week, or >3 drinks/occasion (NIAAA)), lack of exercise ((<150 minutes light exercise / week (WHO)).

Furthermore, we asked questions potentially related to health attitudes: expected life span (own lifespan estimated by the respondent), life span of close relatives (<75 years, >=75 years), informal caregiver status (ones who have provided care for free for at least 6 weeks for relatives or close friends), as well as the use of healthcare services during the 3 months preceding the interview.

We also recorded happiness on a 0-10-point numeric scale, which is one of the simplest and most frequently used measure of subjective wellbeing. (Veenhoven, 2009, Veenhoven, 2012)

IV.3. Statistical methods used for the measurement of AH

IV.3.1. Exploring the measurement properties of AH

We analysed the sample and AH predominantly by descriptive methods, using abundant graphical illustration of the results. When displaying data, depending on the distribution we displayed mean and / or median, and for the description of dispersion, we showed the standard deviation or the interquartile range (IQR). The measurement precision of AH measurements was described by the 95% confidence interval. Subgroups of the sample were compared with non-parametric tests, depending on the sample. We analysed the factors influencing the acceptability of problems via logistic regression, taking into consideration of the multilevel characteristics of the sample: for each respondent, we analysed the acceptability of the 1458 health states, clustering the standard errors on the level of the individual respondents, using robust standard errors. Model fit was checked via the Hosmer-Lemeshow test, its predictive properties were checked by the area under the ROC curve (AUC). The effect of response time on the acceptability of problems was analysed via a linear probability model, in which robust standard errors were clustered on the individual level. The explanatory variables were tested jointly by the Wald test in the model.

IV.3.2. Hypothesis testing

The first research goal was to develop new methods for the measurement of AH, which are more accurate compared to previous methods. We used the following methods for testing the corresponding hypotheses:

H_1**:** *With adaptive testing we can determine the acceptability of all cells of the E-matrix in 90% of subjects.* During the preparatory simulation, we managed to determine the acceptability of all health states in 90% of the 9260 individuals, from whom $AHC_{aggregate}$ curves were available. We expected the same results from the interviews conducted with real respondents. We tested the hypothesis using the 95% exact binomial confidence interval. The expected value was π_0=0,9, π_1 is the proportion of respondents, for whom the acceptability can be determined for all cells of the E-matrix.

We tested the following hypotheses:

H_0: $\pi_1=0,9$

H_{alt}: $\pi_1\neq0,9$

Our H_1 hypothesis can be accepted in case H_0 is accepted. (Results: page 62.)

H_2**:** *When assessed via joint evaluation, people consider fewer problems acceptable compared to separate evaluation.* Due to the non-normal distribution of the data, we tested the hypothesis by comparing median values of the $AHC_{aggregate}$ from separate evaluation ($M_{aggregate}$) and AHC_{joint} from joint evaluation (M_{joint}) via the sign test, using $p=0.05$ significance level. Median values of the curves at all ages were tested jointly.

H_0: $M_{aggregate}=M_{joint}$

H_{alt}: $M_{aggregate}<M_{joint}$

H_2 can be accepted if H_0 is rejected and H_{alt} is accepted. (Results: page 75.)

***H₃*:** *When measured with the EQ VAS, people consider worse health states acceptable in older ages than in younger ages.* Due to the high inter-individual variance of the level as well as slope of AHC_{vas}, we tested the hypothesis using the following multi-level regression model:

$$AHC_{vas_{ik}} = \alpha + \beta age_{AHS_{ik}} + \gamma X_i + \delta age_{AHS_{ik}} * X_i + \mu_i + \tau_i * age_{AHS_{ik}} + \varepsilon_{ik},$$

where age_{AHSik} denotes age k, when respondent i evaluates acceptable health. We centred age_{AHS} at 30 years, therefore the intercept denoted with α represented the mean AHC_{vas} at age_{AHS} 30. The individual variation of the level of AHC_{vas} is denoted by μ, while β indicates the acceptable deterioration rate of health (ADR), and τ indicates the individual component of ADR. The vector X_i indicates the explanatory variables of individual respondent characteristics, the γ and δ coefficients denote the effect of individual explanatory variables on the level and slope of AHC_{vas}, respectively. Individual variance of the level and slope are denoted by μ and τ respectively, which were modelled as random effects. We evaluated H_3 based on the parameter value of β:

H_0: $\beta=0$

H_{alt}: $\beta<0$

H_3 can be accepted if H_0 is rejected and H_{alt} is accepted. (Results: page 83.)

IV.4. Results

IV.4.1. Demography and self-reported health of the sample

The survey was completed by 200 respondents. Mean age was 43,3 years (±SD: 17,3 years). The main socio-demographic characteristics of our respondents are summarized in Table 1. Compared to the Hungarian normal population, the sample was shifted towards individuals with higher education and high income. (KSH, 2011) The weight, and alcohol consumption of respondens was similar to the general population, while lack of exercise and smoking was somewhat less frequent compared to the expected value from the national health survey. (KSH, 2014) Respondents key lifestyle and health variables are summarized in Table 2.

Table 1 Socio-demographic properties of the sample (n=200)

Variable	Category	N	%
Age (n=200)	18-24	24	12.00
	25-34	54	27.00
	35-44	32	16.00
	45-54	43	21.50
	55-64	20	10.00
	65-74	16	8.00
	75+	11	5.50
Gender (n=189)	Male	79	41.80
	Female	110	58.20
Education (n=199)	Primary	9	4.52
	Secondary	73	36.68
	Tertiary	117	58.79
Household income per capita (n=194)	≤ 52 th HUF	5	2.51
	53-74 th HUF	12	6.03
	75-94 th HUF	27	13.57
	95-128 th HUF	35	17.59
	≥ 129 th HUF	120	60.30

Percentages were calculated from the number of respondents with available data

Based on the EQ-5D-3L dimensions, 44% of respondents indicated the presence of any health problems. Severe problems were reported by 4,5% of respondents. The frequency of problems in the respective EQ-5D dimensions were the following:

a) Mobility – any problems: 17.5%, severe problems: 0.5%
b) Self-care - any problems 1.5%, severe problems 0.5%
c) Usual activities - any problems 9.5%, severe problems 1%
d) Pain / discomfort – any problems 28.5%, severe problems 2.5%
e) Anxiety / depression – any problems 33.5%, severe problems 1.5%

The occurrence of health problems increased steeply with age, with the exception of anxiety / depression. Problems in the anxiety / depression dimension were abundant in the younger age groups, and their frequency did not increase as steeply as for other problems (Figure 10).

In every dimension respondents reported fewer problems compared to the national health survey recorded in 2000 (Szende and Nemeth, 2003). The comparison by age groups is shown in Table 3. Compared to the representative national health survey, anxiety / depression occurred more frequently in the youngest (18-24 years old) and mobility problems in the older (65+) age groups. There were fewer problems with self-care in our sample.

Table 2 Lifestyle and health related characteristics of the sample

Variable	Category	N	%
Lifestyle compared to others (n=200)	Healthier	77	38.50
	As healthy	103	51.50
	Less healthy	20	10.00
Body mass index (n=189)	< 25	98	51.85
	≥ 25	91	48.15
Smoker (n=189)	Yes	41	21.96
	No	148	78.31
High-risk alcohol intake (n=189)	Yes	20	10.58
	No	169	89.42
Lack of exercise (n=189)	Yes	105	55.56
	No	84	44.44
Lifespan of closest relatives (n=199)	< 75 years	70	35.18
	≥ 75 years	129	64.82
Problems with health (n=199)	Yes	89	55.28
	No	110	44.72
Have used health services in past 3 months	Yes	117	58.79
(n=199)	No	82	41.21
Informal caregiver (n= 199)	Yes	58	29.15
	No	141	70.85

Percentages were calculated from the number of respondents with available data

Figure 10 Occurrence of health problems in the EQ-5D-3L dimensions

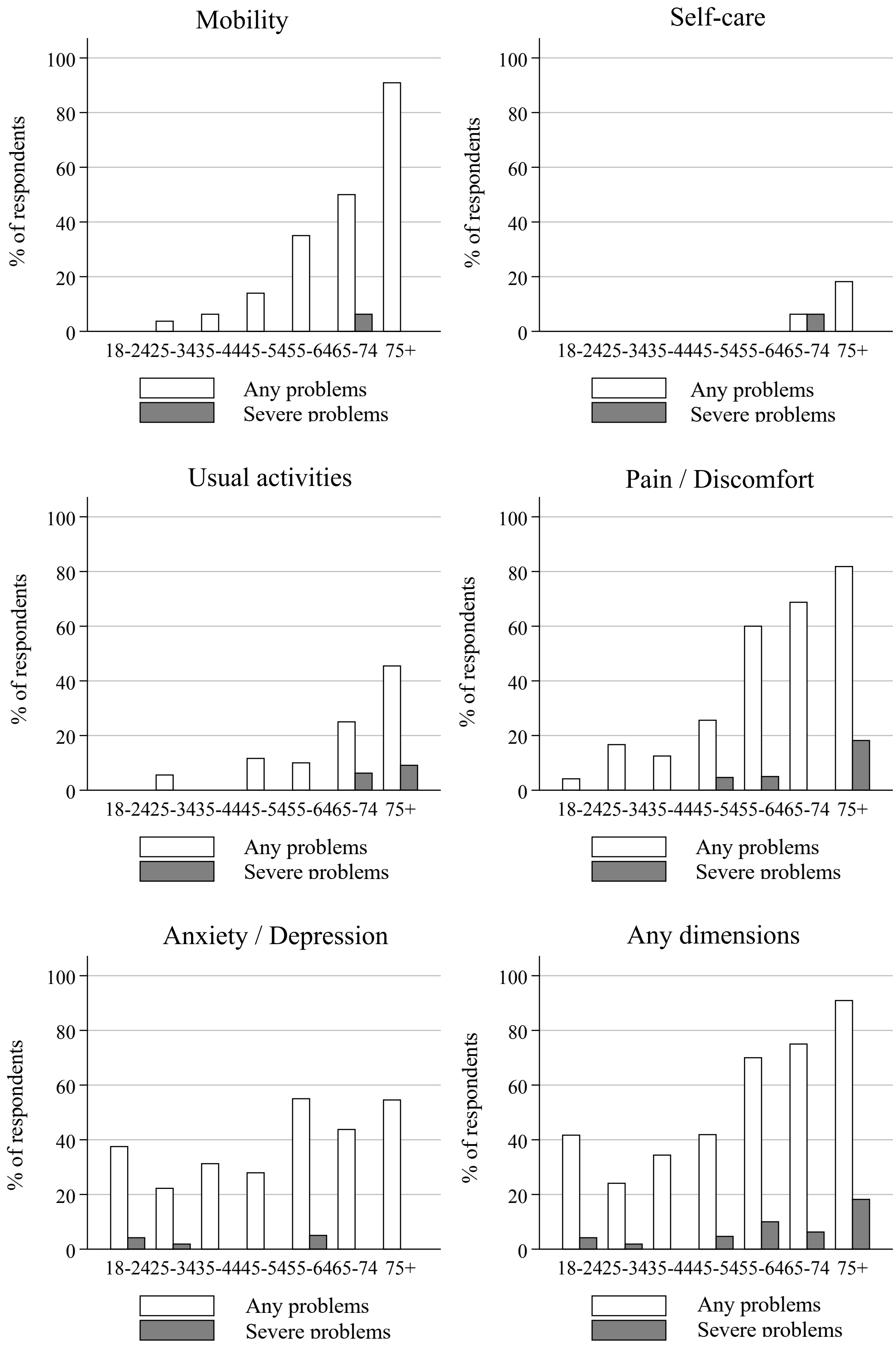


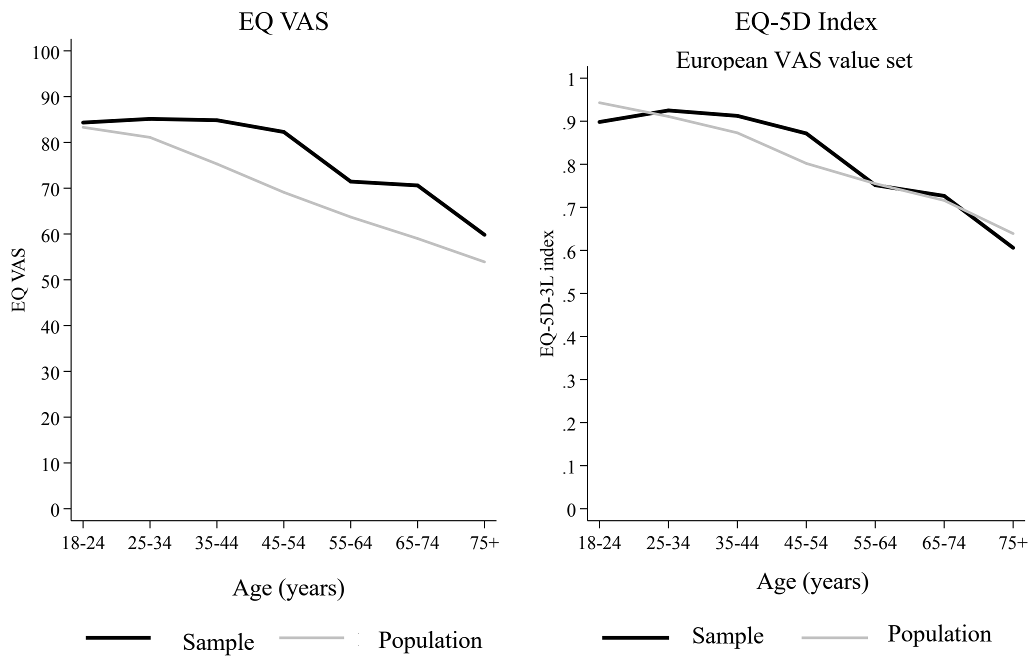
Table 3 Prevalence of health problems compared to the representative national population

		18-24 y		25-34 y		35-44 y		45-54 y		55-64 y		65-74 y		75+ y	
		Sample	Population	Sample	Population	Sample	Population	Sample	Population	Sample	Population	Sample	Population	Sample	Population
Mobility	Any problems	0.0	0.9	3.7	4.6	6.3	8.6	14.0	21.2	35.0	30.3	50.0	39.6	90.9	57.6
	Severe problems	0.0	0.1	0.0	0.6	0.0	1.1	0.0	1.4	0.0	0.5	6.3	1.9	0.0	1.1
Self-care	Any problems	0.0	0.4	0.0	0.7	0.0	1.7	0.0	5.1	0.0	10.0	6.3	14.4	18.2	26.8
	Severe problems	0.0	0.0	0.0	0.3	0.0	0.2	0.0	1.0	0.0	1.3	6.3	2.8	0.0	3.8
Usual activities	Any problems	0.0	1.8	5.6	2.9	0.0	6.8	11.6	16.7	10.0	22.3	25.0	28.8	45.5	44.2
	Severe problems	0.0	0.4	0.0	0.4	0.0	1.2	0.0	2.1	0.0	3.7	6.3	6.0	9.1	8.7
Pain / discomfort	Any problems	4.2	14.5	16.7	18.5	12.5	30.5	25.6	46.1	60.0	54.8	68.8	61.1	81.8	69.6
	Severe problems	0.0	0.1	0.0	0.7	0.0	1.2	4.7	3.3	5.0	4.6	0.0	8.6	18.2	10.1
Anxiety / depression	Any problems	37.5	17.0	22.2	23.9	31.3	29.9	27.9	39.5	55.0	47.1	43.8	47.0	54.5	54.9
	Severe problems	4.2	1.2	1.9	2.3	0.0	1.8	0.0	4.0	5.0	5.5	0.0	6.5	0.0	7.5

Population: National Health Survey, 2000 (OLEF) (Szende and Nemeth, 2003)

Compared with the national health survey from year 2000 (Figure 11), the EQ VAS values in our sample were higher with the exception of the 18-24-year-old group. The EQ-5D index values using the European VAS value set were similar to those in the national health survey (Szende and Nemeth, 2003).

Figure 11 Health status of our sample vs the Hungarian general population



Population: National Health Survey, 2000 (OLEF) (Szende and Nemeth, 2003)

The mean level of happiness was 7,3 (±SD: 2,0), with level 8 reported most frequently. Happiness scores were slightly left-skewed and showed a mild declining trend with age (Figure 12).

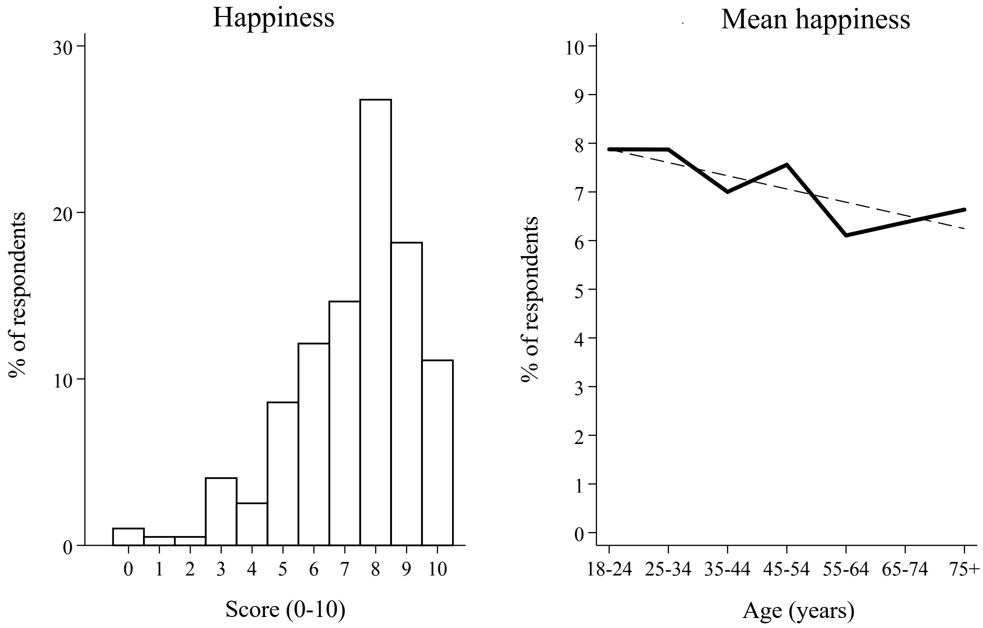
IV.4.2. Efficiency of adaptive testing (H_1)

We evaluated the acceptability of 1458 discrete health states for each of the 200 respondents. Altogether we collected 291600 data points. After separate evaluation, 91,7% of health states were not acceptable (on average 1336.2 per respondent), 1,3% were acceptable (on average 18.8 per respondent) and for 7,1% the acceptability was unknown. Compared to simulation run on the $AHC_{aggregate}$ data from a national sample of 9260 respondents (n=93), in our sample the mean number of unknown states was greater (n=103). During joint evaluation, the algorithm asked on average 11,09 questions from each respondent. For 43,5% of respondents, (n=87) 15 or fewer questions were sufficient to evaluate the acceptability of all health states of their *E-matrix* in question.

The first hypothesis (H_1) of the research was that the acceptability of all health states could be fully determined for 90% of individuals. In case of 200 respondents, the 95% exact binomial confidence interval for the 90% acceptability rate falls between 85,0-93,8%. Our results fell below the lower 95% CI limit, outside the expected range. Therefore, we did not accept our H_1 hypothesis: the efficiency of adaptive testing was inferior compared to our expectations.

During joint evaluation we could elicit the acceptability of 67.4% of the unknown states, on average 69.4 states per respondent. During joint evaluation we could determine the acceptability of 6.25 health states by each question.

Figure 12 Happiness of our respondents



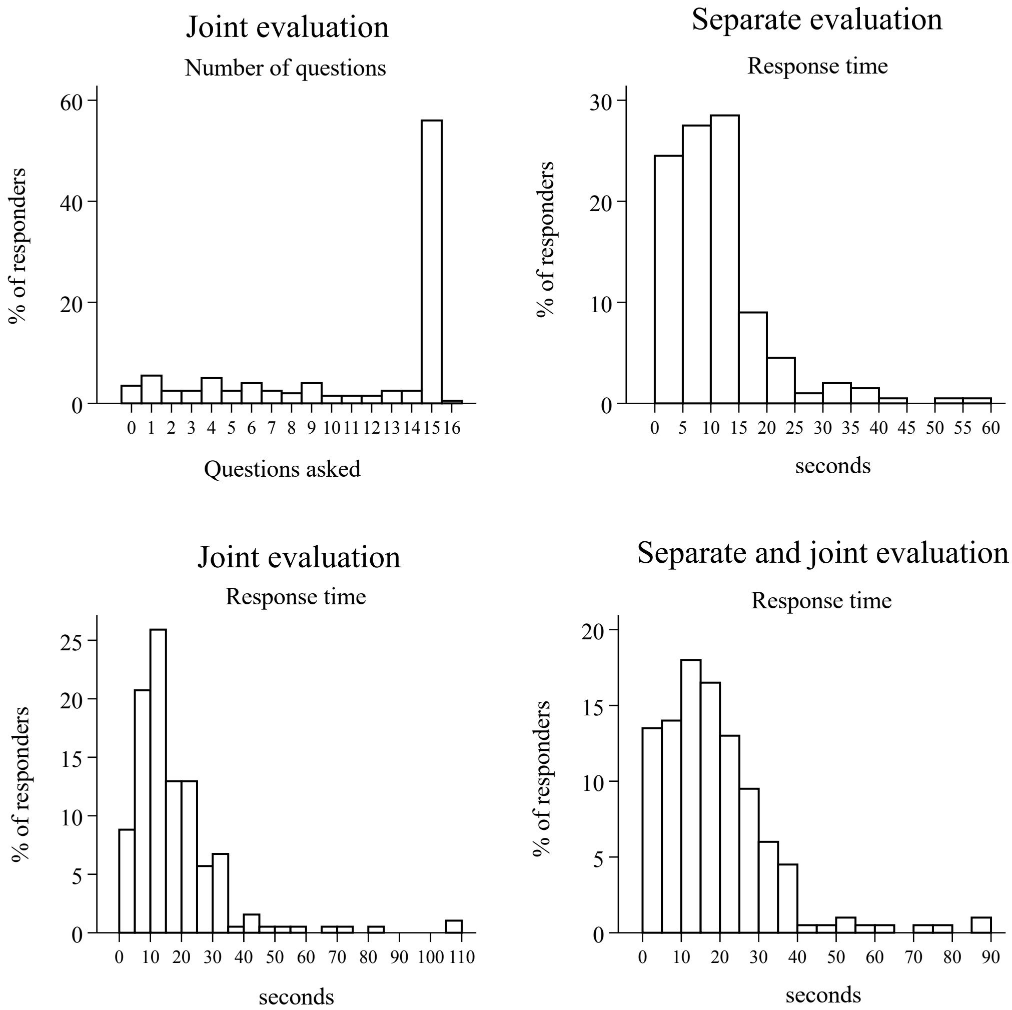
Happiness of respondents was measured on the 0-10 numeral scale. Left figure: distribution of happiness scores. Right figure: bold line – mean happiness with age; dashed line – linear trend of happiness with age

Respondents evaluated 70.2% of the randomly displayed health states during joint evaluation as acceptable, while 29.8% as not acceptable. After joint evaluation, as opposed to the 0.4% estimated value from the simulation assuming 50% probability of acceptance, 2.3% of the health states remained unknown, so the adaptive algorithm in practice performed below the expectation. On average, 33.6 health states per respondent (32.6% of unknown health states after separate evaluation) remained unknown after joint evaluation.

Respondents spent on average 110 seconds answering the questions of separate evaluation (median: 96 sec). The number of questions in separate evaluation was 10 for all respondents (in five dimensions the acceptability of moderate and severe problems), therefore it took on average 11 seconds to provide answer for each question (median: 10 sec). Joint evaluation took on average 175 seconds, (median: 166 sec), mean response time per question was 18 sec (median: 14 sec). The evaluation of discrete health states (separate and joint evaluation with explanations) took on average 392 seconds from respondents (median: 349 sec), the mean time spent answering the survey was 19 seconds per question (median: 16 sec). Figure 13 displays the distribution of response times.

During separate evaluation, questions remained mainly about the acceptability of moderate health problems. The proportion of health states elicited during separate and joint evaluation, and the distribution of unknown health states within elements of the E-matrix are displayed in Figure 14. In case of 71.9% of the elements of the E-matrix, the percentage of unknown states was below 1%. For 9.7% of the E-matrix elements, the percentage of unknown states was over 10%. The maximum percentage of unknown states was 20.0%.

Figure 13 Distribution of the number of questions and response times of health state evaluation



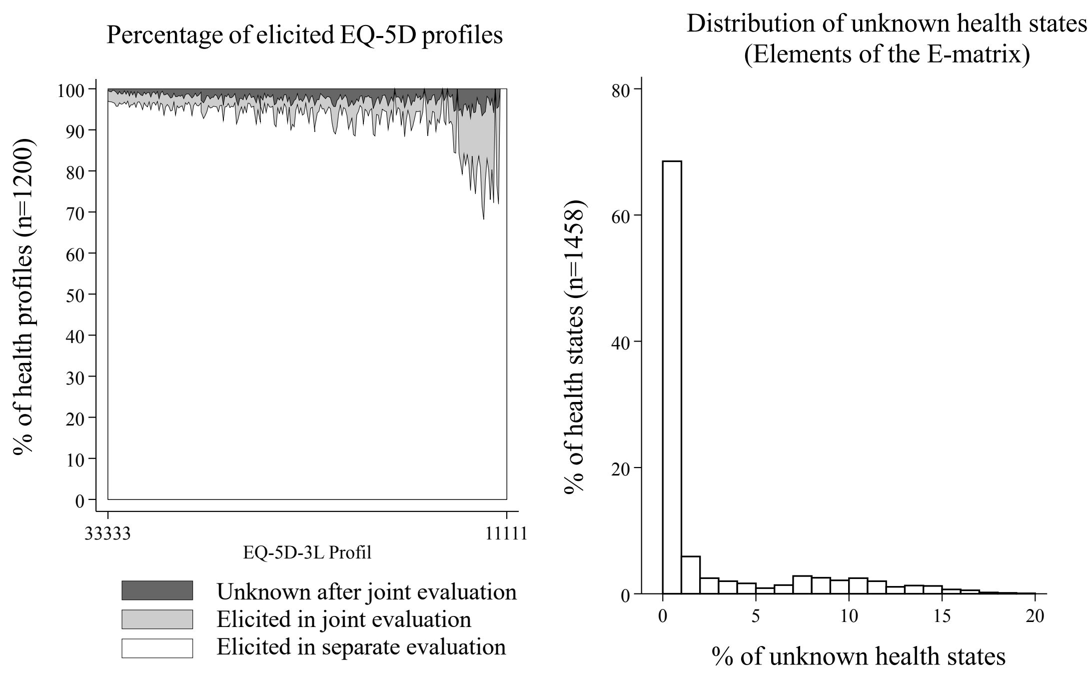
Top left figure: distribution of the number of questions during joint evaluation. In 43.5% of respondents, less than 15 questions were enough for eliciting the acceptability of all health states. Top right and bottom figures: distribution of time spent with separate evaluation, joint evaluation and the total time of the two tasks by the % of respondents

The unknown states remaining after joint evaluation showed association with the severity of the profile and age. The acceptability of eleven EQ-5D-3L profiles was determined unambiguously during separate evaluation (those profiles, in which health problems occurred only in a single dimension and full health). For the remaining *i* health states of combined problems, we evaluated the association between the proportion of unknown health states (u_i) with the severity of problem (s: EQ-5D-3L index, UK-TTO value set), age (k=age-30years), their quadratic terms and interactions with the following linear probability model:

$$u_i=\alpha + \beta_1 s_i^2 + \beta_2 k_i + \beta_3 k_i^2 + \beta_4 s_i * k_i + \beta_5 s_i^2 * k + \beta_6 s_i * k_i^2 + \beta_7 s_i^2 * k_i^2 + \varepsilon$$

The whole model ($F_{(8,653)}=179.5$, $p<0.001$), all main effects and interactions were significant (Wald test), the explanatory power of the model was high ($R^2=0,6874$). The heatmap from the proportion of unknown states is illustrated in Figure 15. Among moderate health sates mainly in ages of 60-70 years, and among more severe problems in the 80-year-old age group remained relatively high the proportion of unknown health states. This association can be used to design more efficient adaptive testing algorithms in the future.

Figure 14 Distribution of unknown health states



*Left figure: on the horizontal axis we ordered the 243 EQ-5D-3L profiles according to their UK-TTO index values, and on the vertical axis we indicated the proportion of health profiles (n=1200, 200 respondents * 6 ages), for which the acceptability could be elicited during separate evaluation (white), joint evaluation (grey) or which remained unknown after joint evaluation (dark grey). Right figure: the horizontal axis shows bands of percentages of health states that remained unknown after joint evaluation (elements of the E-matrix, n=1458, 243 EQ-5D profiles* 6 ages). The vertical axis shows the corresponding percentages of the 1458 health states falling in each band.*

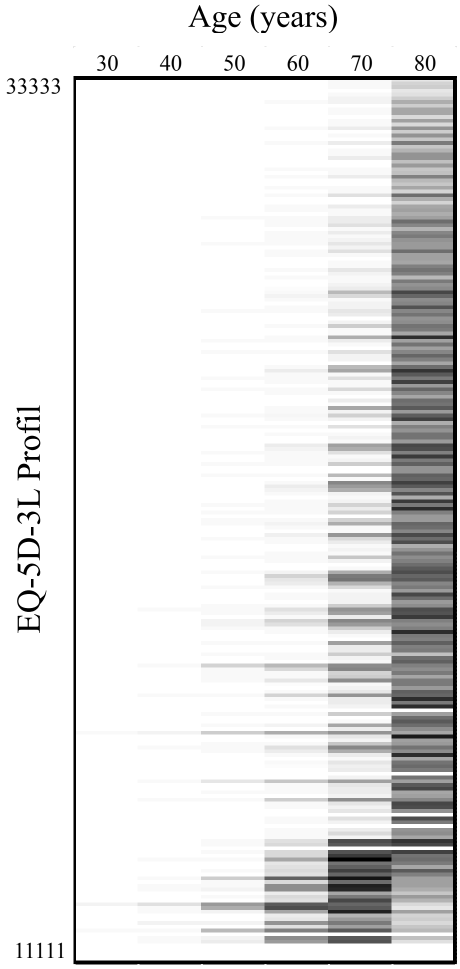
In order to understand, how the proportion of unknown health states influences the precision of measuring AH (proportion of acceptable health states), we determined the 95% confidence interval (*$95\%CI_i$*) for all *i* elements of the E-matrix according to the following:

$$95\%CI_i = 1{,}96 * \sqrt{\frac{p_i * (1 - p_i)}{n * (1 - u_i)}}$$

where p_i denotes the proportion of acceptable health states, and u_i denotes the proportion of unknown health states diminishing the effective sample-size within each health state.

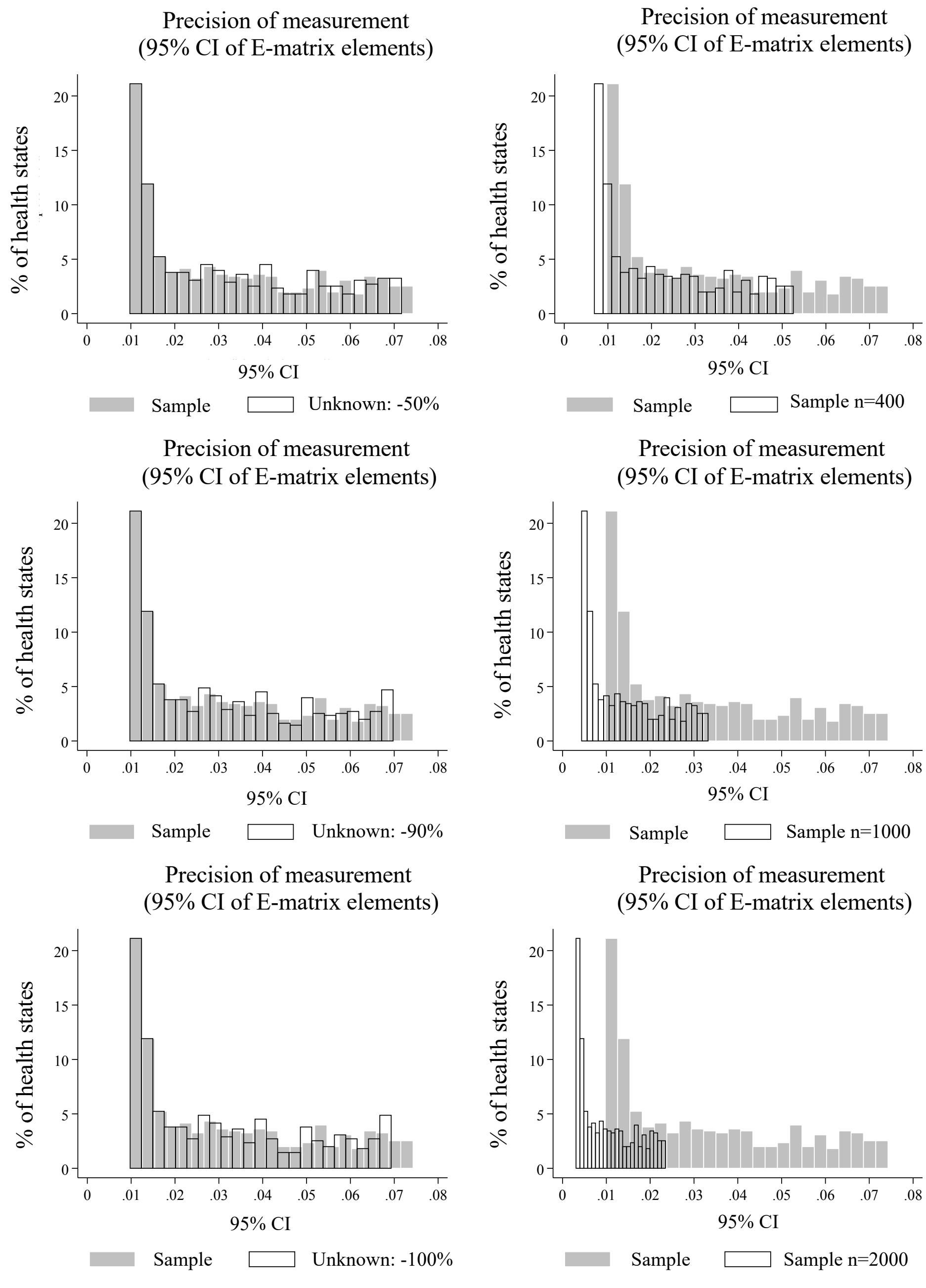
We modelled the effects of decreasing the amount of unknown health states by 50, 90 and 100%, as well as increasing the sample size from 200 to 400, 1000 and 2000. For a sample size of 200, we could determine the acceptability of health states with at least ±7% 95%CI precision. Decreasing the proportion of unknown states improved this only to a small extent. However, increasing the sample size brought dramatic improvement of precision (Figure 16).

Figure 15 Distribution of unknown health states



The heat map illustrates the elements of the E-matrix: the horizontal axis shows six ages, the vertical axis shows 243 EQ-5D-3L health profiles, in increasing order of UK-TTO index values. Colours indicate the proportion of unknown health states. Dark colours indicate greater proportion of unknown health states. The highest percentage of unknown states was 20% (black), in case of white fields the acceptability for all health states could be elicited.

Figure 16 Exploring the precision of AH measurement

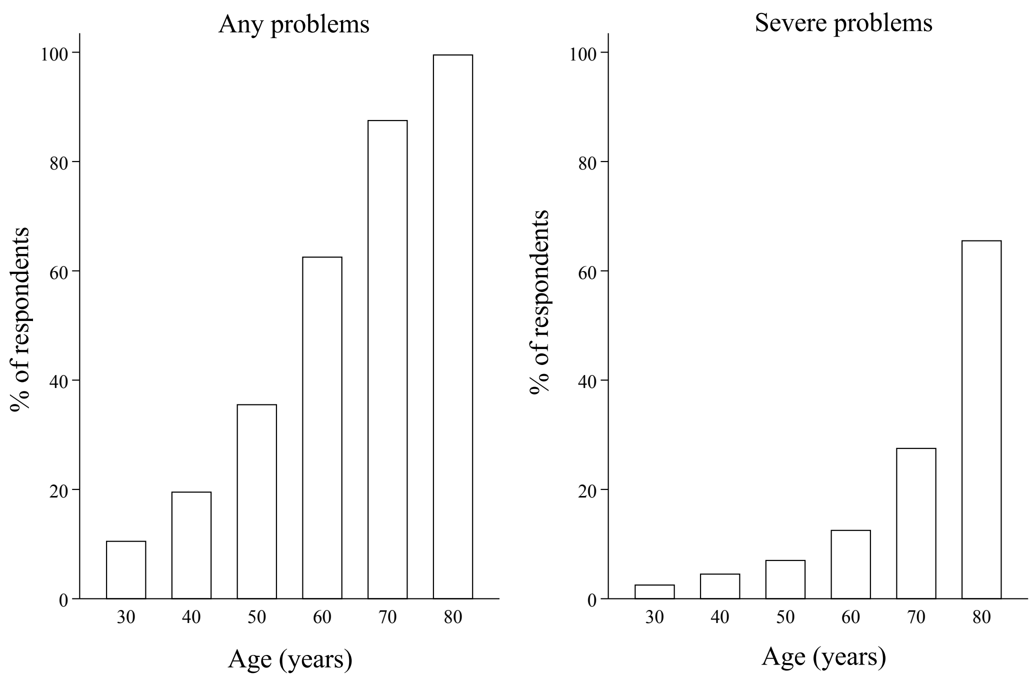


*We characterised the measurement precision of AH with the 95%CI of the proportion of acceptable health states within each element of the E-matrix (1458 health states, 243 EQ-5D-3L profiles*6 ages). Grey bars show the distribution of the measurement precision of the 1458 health states from our sample of 200 respondents. The white bars depict the changes of the measurement precision by changing the number of unknown health states – which can be achieved by improving the efficiency of the adaptive testing algorithm (left column) or increasing the sample size (right column).*

IV.4.3. Measuring the acceptability of health problems via separate evaluation

According to the separate evaluation, the acceptability of moderate and severe health problems increased with age. The acceptability of moderate problems showed dramatic increase after 60 years, and that of the severe problems increased dramatically from 80 years of age. Severe problems were less acceptable than moderate ones. In ages 30, 60 and 80 years, according to 10.5%, 62.5% and 99.5% of respondents was at least one health problem acceptable, respectively. At least one severe problem was acceptable at 30, 60 and 80 years according to 2.5%, 12.5% and 65.5% of respondents. Details are shown in Figure 17.

Figure 17 Acceptability of health problems by separate evaluation



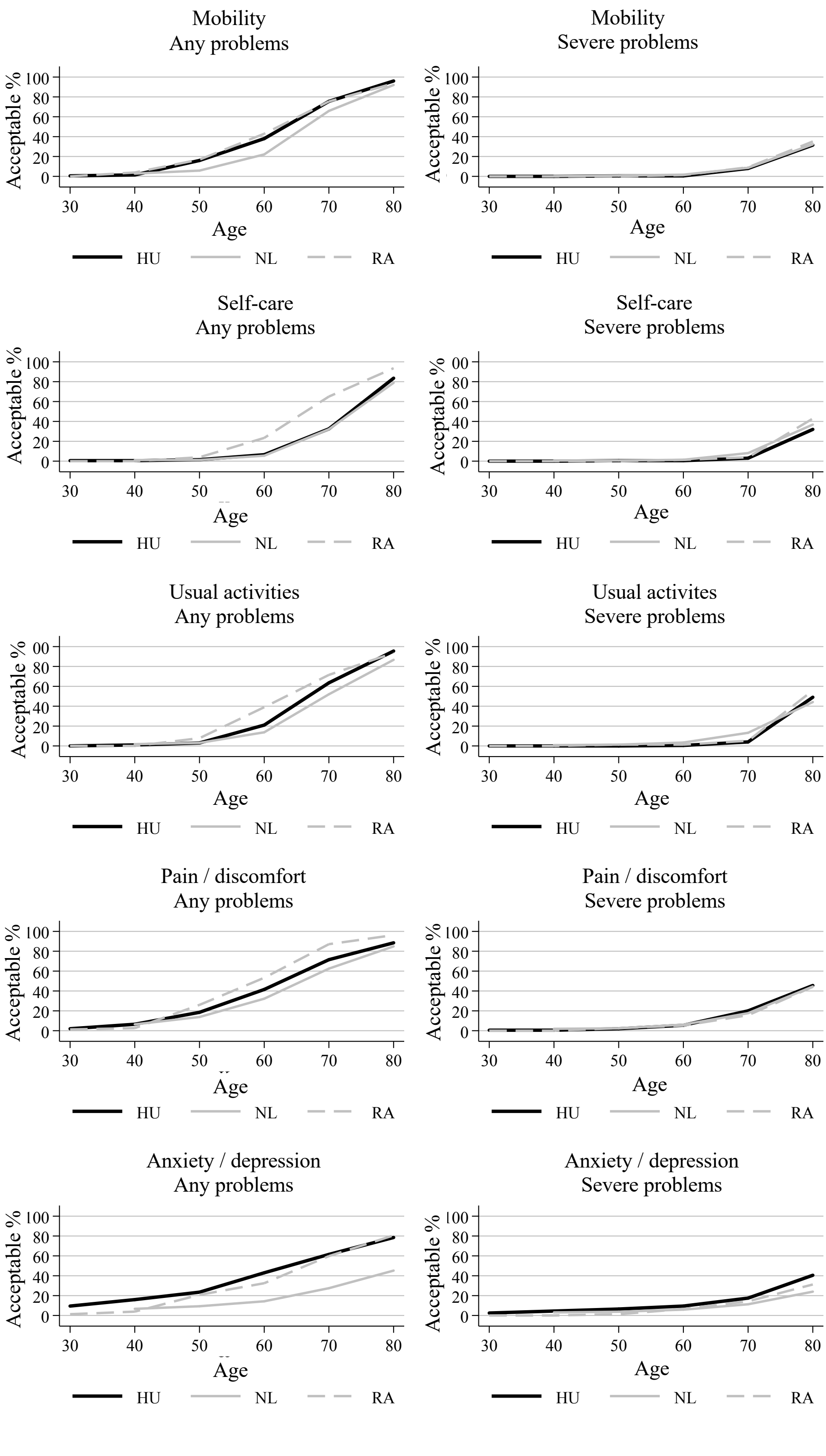
Left figure: the percentage of respondents with at least one problem in any of the five EQ-5D-3L dimensions, Right figure: the percentage of respondents with at least one severe problem in any of the five EQ-5D-3L dimensions,

Table 4 summarizes the acceptability of problems in each EQ-5D-3L dimension. Most frequently moderate problems with mobility were accepted, while severe problems were accepted most frequently in the dimension of usual activities. In younger ages anxiety/depression was the most frequently acceptable problem, but in this group was the highest the proportion of those, who do not accept even moderate problems.

Table 4 Acceptability of health problems in different ages: separate evaluation

		Age (year)						
		30	**40**	**50**	**60**	**70**	**80**	**Never**
Mobility	Any	0.5	1.5	16	38	75.5	96	4
	Severe	0	0	0.5	0.5	8	31.5	68.5
Self-care	Any	0.5	0.5	1.5	6.5	32.5	83.5	16.5
	Severe	0	0	0.5	0.5	3	32	68
Usual activities	Any	0	1	3	21	63.5	95.5	4.5
	Severe	0	0	0	0.5	4	49	51
Pain / discomfort	Any	2	6.5	18.5	41.5	71.5	88.5	11.5
	Severe	0.5	0.5	2	5.5	20	45.5	54.5
Anxiety / depression	Any	9.5	16	23.5	43	61.5	78.5	21.5
	Severe	2.5	4.5	6.5	9.5	17.5	40.5	59.5

Figure 18 Acceptability of health problems in separate evaluation



HU: our sample, NL: Dutch population, RA: Hungarian patients with rheumatoid arthritis.

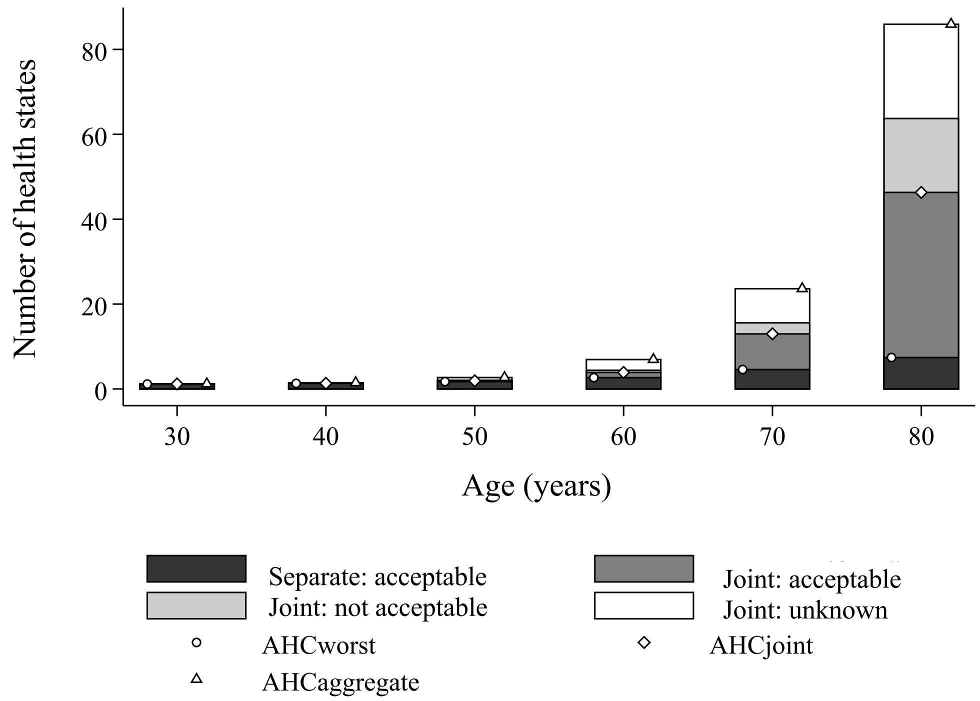
We compared our results of separate evaluation with data measured with the same method in the Dutch general population (Wouters et al., 2015) and among Hungarian patients with rheumatoid arthritis (Pentek et al., 2014b). The acceptability of severe problems was remarkably similar in the three populations. Chronic patients accepted somewhat more problems in the dimensions of self-care, usual activities and pain / discomfort. The Dutch population accepted less anxiety / depression than Hungarians. (Figure 18).

IV.4.4. Acceptability of health states by joint evaluation

The results of joint evaluation were presented in the following international conference poster: (Zrubka et al., 2018b)

The number of health states elicited during joint evaluation are shown in Figure 19. After separate evaluation in each age on average 3.1 health states were acceptable automatically, 17.2 health states were unknown and 222.7 were not acceptable. The share of unknown health states increased with age, in 80 year olds the mean number of automatically acceptable health states was 7.4 and the mean number of unknown states was 78.5. During joint evaluation from the mean 17.2 unknown states by age our respondents considered 3.5 as not acceptable (20.1%), 8.1 as acceptable (47.3%) and the acceptability of 5.6 health states remained unknown (32.6%). The odds ratio between acceptable / not acceptable answers during joint evaluation was 2.35 (95%CI: 2.27-2.44). Figure 19 illustrates the health states which are used for the calculation of the $AHC_{aggregate}$, AHC_{worst} and AHC_{joint} health curves. AHC_{worst} takes the lowest EQ-5D-3L index value of the automatically acceptable health states, $AHC_{aggregate}$ takes the lowest EQ-5D-3L index value from the unknown states after separate evaluation, and AHC_{joint} takes the lowest EQ-5D-3L index value of acceptable health states after joint evaluation.

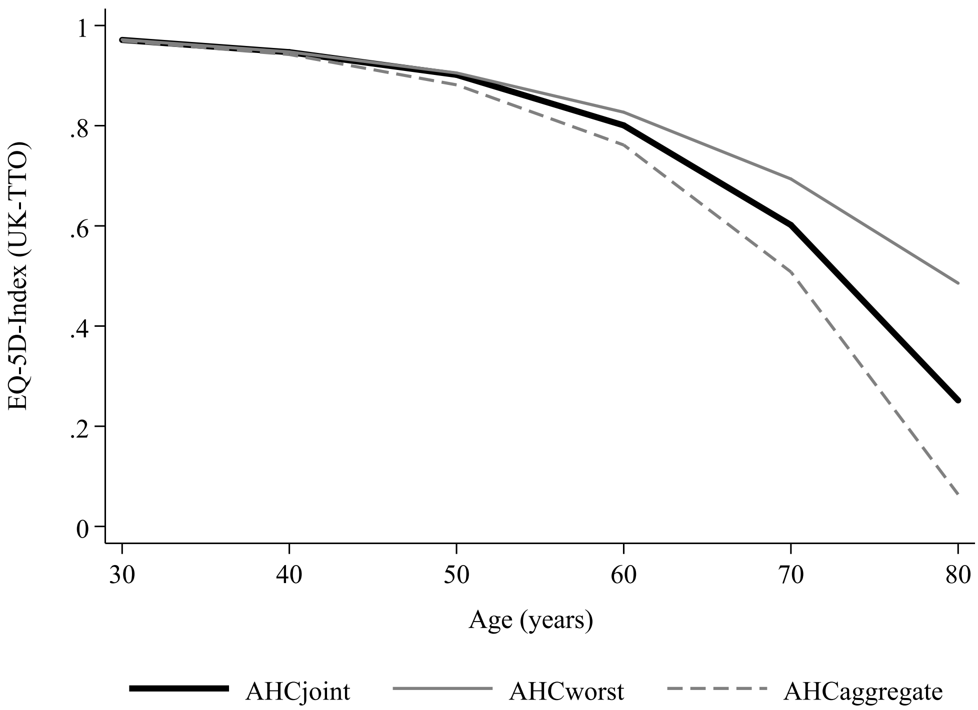
Figure 19 Results of joint evaluation



Columns show the number of elicited and unknown health states by age after joint evaluation. From the 243 EQ-5D-3L profiles in 80 years of age after separate evaluation on average 7.4 health states were acceptable (dark grey) and 78.5 health states were unknown. After joint evaluation, from this 38.9 health states became acceptable (medium grey), 17.4 health states became not acceptable (light grey) and 22.2 health states remained unknown. The markers show the health states used for the calculation of the acceptable health curves. AHCworst (circle) is the mean of the lowest EQ-5D-3L index values of the automatically acceptable health states for each respondent, AHCaggregate (triangle) is the mean of the lowest EQ-5D-3L index values from the unknown states after separate evaluation, and AHCjoint (diamond) is the mean of the lowest EQ-5D-3L index values of acceptable health states after joint evaluation.

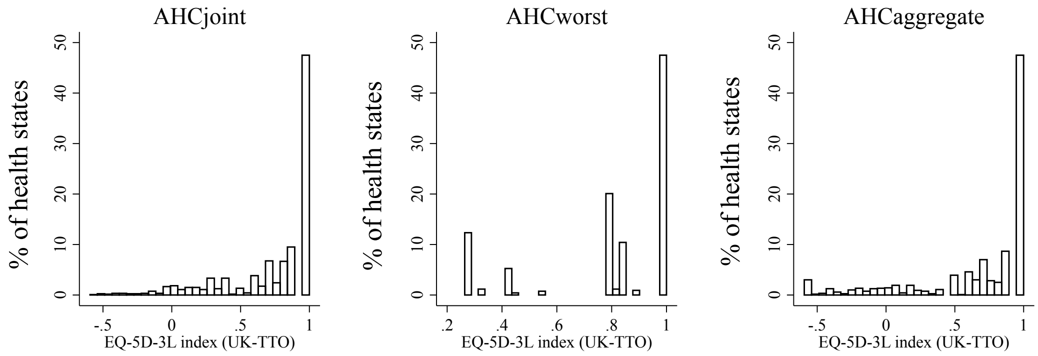
According to our expectations, AHC_{joint} fell among the $AHC_{aggregate}$ and AHC_{worst} curves (Figure 20). All three curves confirmed that people accept more health problems as age advances. The differences between the three curves suggest that the adequate method for measuring AH is the joint evaluation of health states, since the aggregation of health problems after separate evaluation ($AHC_{aggregate}$) overestimates the amount of acceptable health states, while people are willing to accept some combinations of health problems, therefore AHC_{worst} underestimates the amount of acceptable health problems.

Figure 20 AHC_{joint} health curve (H_2)



AHCworst is the mean of the lowest EQ-5D-3L index values of the automatically acceptable health states for each respondent, AHCaggregate is the mean of the lowest EQ-5D-3L index values from the unknown states after separate evaluation, and AHCjoint is the mean of the lowest EQ-5D-3L index values of acceptable health states after joint evaluation.

Figure 21 Distribution of the values of acceptable health curves (AHC)



The three acceptable health curves (AHC_{joint}, AHC_{worst}, $AHC_{aggregate}$) were calculated from the mean of individual AHC_{joint}, AHC_{worst} and $AHC_{aggregate}$ health states. The histograms show the distribution of individual AHC states for the 200 respondents.

Figure 21. shows that the distribution of individual values of all three AHCs are strongly left-skewed, therefore we compared the curves with non-parametric methods.

The second hypothesis (H2) was that by joint evaluation people considered fewer health problems acceptable when compared to separate evaluation, meaning that median values of AHCaggregate are lower than those of the AHCjoint. The results of the sign test of the differences of medians (Maggregate, Mjoint) were summarized in Table 5.

Table 5 Sign test comparing median values of $AHC_{aggregate}$ and AHC_{joint}

M_{joint}-$M_{aggregate}$	Outcome (N)	Expected (N)
>0	148	74
<0	0	74
=0	52	0
Total	200	200

If $M_{joint} > M_{aggregate}$, the exact binomial probability of the result (M_{joint}- $M_{aggregate}$ >0, N=148) is p<0,001, therefore we rejected the null-hypothesis (H_0: $M_{aggregate}$=M_{joint})

and accepted the alternative hypothesis (H_{alt}: $M_{aggregate}$<M_{joint}). *So, we could accept H_2: with joint evaluation people consider fewer health problems acceptable than with separate evaluation.*

The sign test comparing median $AHC_{aggregate}$ and AHC_{joint} values was significant already from age 40 (p=0.002). When comparing with the sign test, the median AHC_{joint} values in all ages jointly were significantly lower than those of AHCworst (p<0.001), and the comparison by ages was significant from age 40 (p=0.0039).

IV.4.5. Accuracy of the joint evaluation of AH

Between the three acceptable health curves created after separate and joint evaluation ($AHC_{aggragate}$, AHC_{worst} and AHC_{joint}) the difference was significant from 40 years of age. The accuracy of AHC_{joint} measurement of determined by the following three factors:

a) Residual inaccuracy from the unknown health states remaining after joint evaluation, which may influence the levels of respondents' individual AHC_{joint} values.
b) The dispersion of Az AHC_{joint} values, which is determined by the size and composition of the sample
c) The uncertainty arising from the measurement problems of the utility values rendered to acceptable health states.

From the above three, during the measurement of AH we can influence only the first two. The measurement properties of EQ-5D-3L utility values falls under the realm of EQ-5D valuation studies (Brooks et al., 2003).

We estimated the magnitude of inaccuracy arising from residual unknown health after joint evaluation by the following steps:

a) We can render to all elements of the E-matrix a single AHC_{joint} value (profile with the lowest EQ-5D-3L value in each age).
b) We calculated the EQ-5D-3L index values for all unknown elements of the E-matrix.
c) We took the difference of AHCjoint values and the EQ-5D-3L index values of the unknown health states. The distribution of the differences is depicted on the left side of Figure 21. The difference was positive, if the EQ-5D-3L index

values were higher than AHC_{joint}. Eliciting these health states does not change the AHC_{joint} curve, because it is calculated from the lowest EQ-5D-3L index values of the elicited health states.

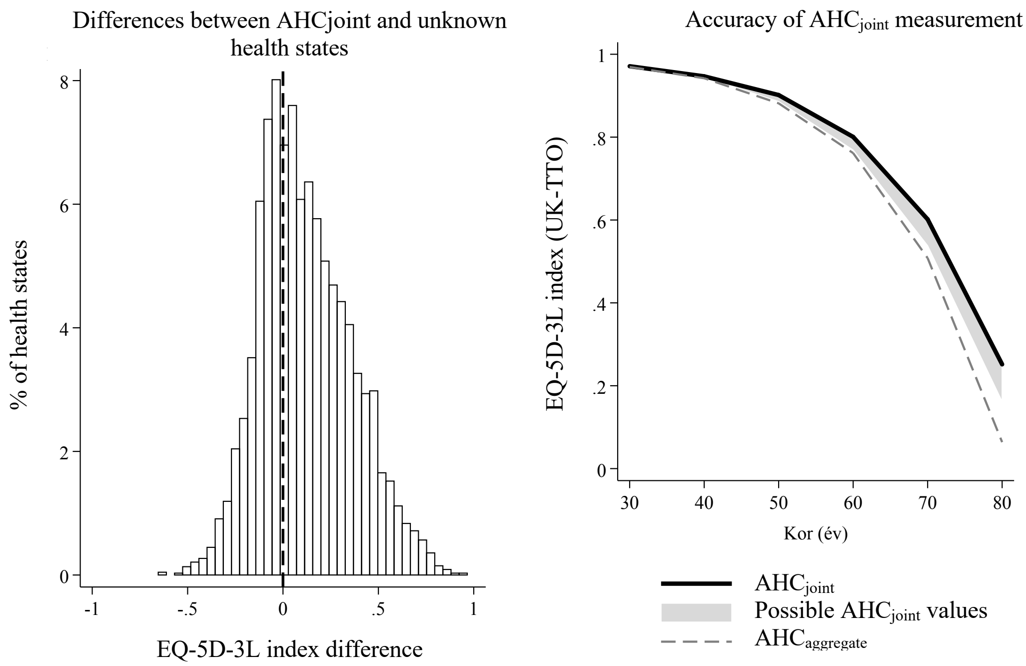
d) The difference was negative, if the EQ-5D-3L index value of known health states was lower than AHC_{joint}. These health states could modify AHC_{joint}, if the respondent evaluated them as acceptable.

e) So, we could get the theoretical minimum of AHC_{joint}, if all residual unknown health states were evaluated as by the respondent as acceptable. Possible values of AHC_{joint} fall anywhere between the theoretical minimum and the known AHC_{joint} curve.

The right graph of Figure 22 illustrates the theoretical minimum and possible range of AHC_{joint} values in addition to the known AHC_{joint} and $AHC_{aggregate}$. The question was whether by eliciting all unknown health states, the difference between AHCjoint and AHCaggregate would remain significant, which could prove without actually eliciting the unknown health states, that a module of joint evaluation with 15 questions results a different AHC_{joint} from the results of separate evaluation. In the opposite case, one could argue, that the difference between the AHCs can be explained by the fact that certain health states remained unknown during the joint evaluation exercise.

We compared the three curves with the sign test, similarly to the method used when testing H_2. The difference between the median values of $AHC_{aggregate}$ and the theoretical minimum of AHC_{joint} was significant for all ages jointly, and from 50 years in each age group ($p \leq 0.001$). The difference between the known AHCjoint and the theoretical minimum of AHC_{joint} was also significant both for all ages jointly, and from 40 years in all ages ($p \leq 0.032$). This result supports the assumption that joint evaluation leads to different result compared to separate evaluation, even without eliciting all unknown health states. During the joint evaluation of AH, compared to the aggregation of problems regarded as acceptable in separate evaluation, people accept fewer problems. However, the EQ-5D-3L index difference between the known AHC_{joint} values and the theoretical minimum of AHCjoint was 0.09, which is greater than the minimally important clinical difference (MCID) of 0.074 for EQ-5D-3L (Pickard et al., 2007). This result draws attention to the importance of the goal to further decrease the remaining number of unknown health states after joint evaluation, in order to avoide statistically and clinically significant difference between the theoretical minimum and the actually measured AHC_{joint} curves.

Figure 22 Accuracy of the measurement of AHC_{joint}



Left figure: the distribution of the differences of AHC_{joint} values and unknown health states in each age for the 200 respondents. If the unknown health states falling on the left side of the dashed line would be acceptable, then the AHC_{joint} curve would change. The EQ-5D-3L index values of unknown states on the right of the dashed line are higher than those of the AHC_{joint}, therefore their acceptability does not influence the AHC_{joint} curve. Right figure: the bold line indicates the AHC_{joint}, the dashed line the $AHC_{aggregate}$, and the grey area the possible values of the AHC_{joint} depending on what proportion of the unknown health states would be evaluated as acceptable by the respondents. The lowest EQ-5D-3L index values of the grey area show the AHC_{joint} in the case if all unknown health states were considered acceptable by the respondents.

The accuracy of AH measurement explained by sampling can be described by the dispersion of AHC_{joint} and $AHC_{aggregate}$ values (Figure 23, left diagram). Given the strong left-skew of the distributions, we displayed median values of the curves, and we depicted the interquartile range (IQR) as a measure of dispersion. From the curve we can read that the dispersion of AHC_{joint} is smaller at older age groups than the dispersion values of $AHC_{aggregate}$. The IQR of AHCjoint at the 70 and 80-year-old age groups was 0.38 and 0.62, while that of AHCaggregate was 0.63 and 0.87, respectively.

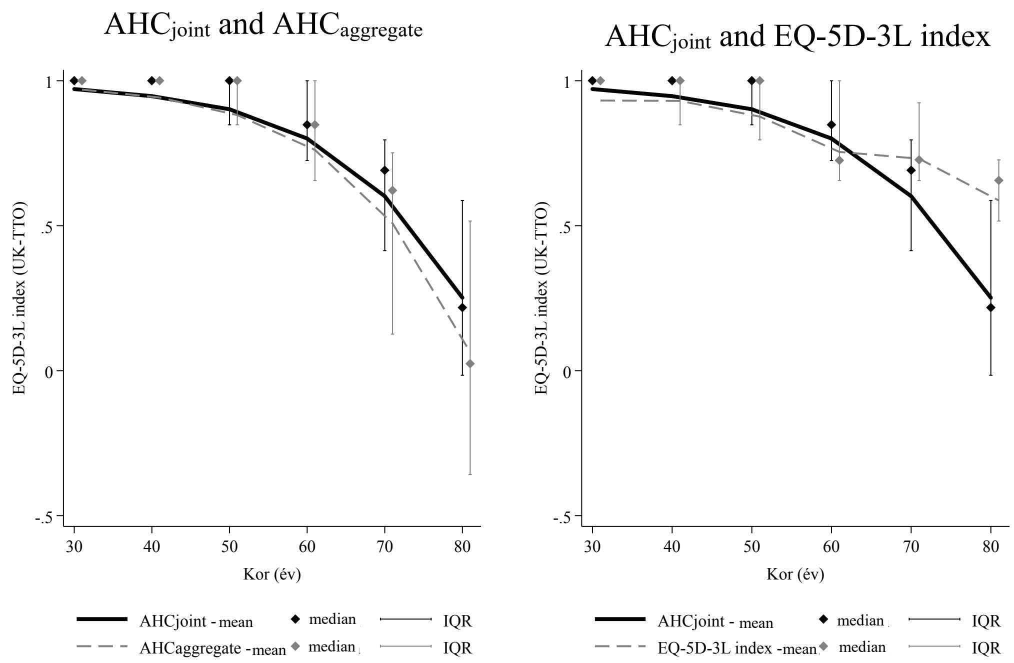
We can treat $AHC_{aggregate}$ as a hypothetical maximum of AHC_{joint} in the case when the number of questions in joint evaluation is 0 instead of 15. So, the difference between the dispersion of $AHC_{aggregate}$ and AHC_{joint} is given by the individual variance of the differences between the latent AH threshold and the artificially constructed $AHC_{aggregate}$ curve given from the responses to separate evaluation questions. Some respondents accepted the problems in combination as well which they considered acceptable during separate evaluation. However, there were respondents, who, having seen the health state vignettes during joint evaluation, asked for the modification of their responses to questions of the separate evaluation module. By decreasing the number of unknown health states, the variance of $AHC_{aggregate}$ due to different response strategies during separate evaluation can be decreased. Furthermore, understanding the factors that influence the individual response strategies, or even estimating them during the evaluation process could support the development of a more efficient adaptive testing algorithm.

IV.4.6. Comparison of AH and respondents self-rated health

We compared AHC_{joint} with respondents self-rated health by the EQ-5D-3L (UK-TTO value set) (Figure 23, right graph) The two curves overlapped until the age of 60. In the 70 and 80-year-old age groups, our respondent rated their health better than the acceptable health curve. The IQR values of AHC_{joint} were smaller until age 50, and larger beyond age 70 than the IQR of self-rated health. These results suggest, that the imagination of respondents about acceptable health in younger ages is relatively homogenous, while the imaginations about acceptable health states in older ages show greater variability than that of the health states actually experienced by our respondents.

We also compared the AHCjoint with the results of the National Health Survey Szende and Nemeth, 2003). Up to age 50 our respondents considered on average better, while above 70 years of age worse health states acceptable, than the health status of the general population. During the comparison, we calculated the AHCjoint curve with the European VAS value set, which was used in the National Health Survey (Figure 26., left graph)

Figure 23 Comparison of AHC_{joint} with health variables of the sample



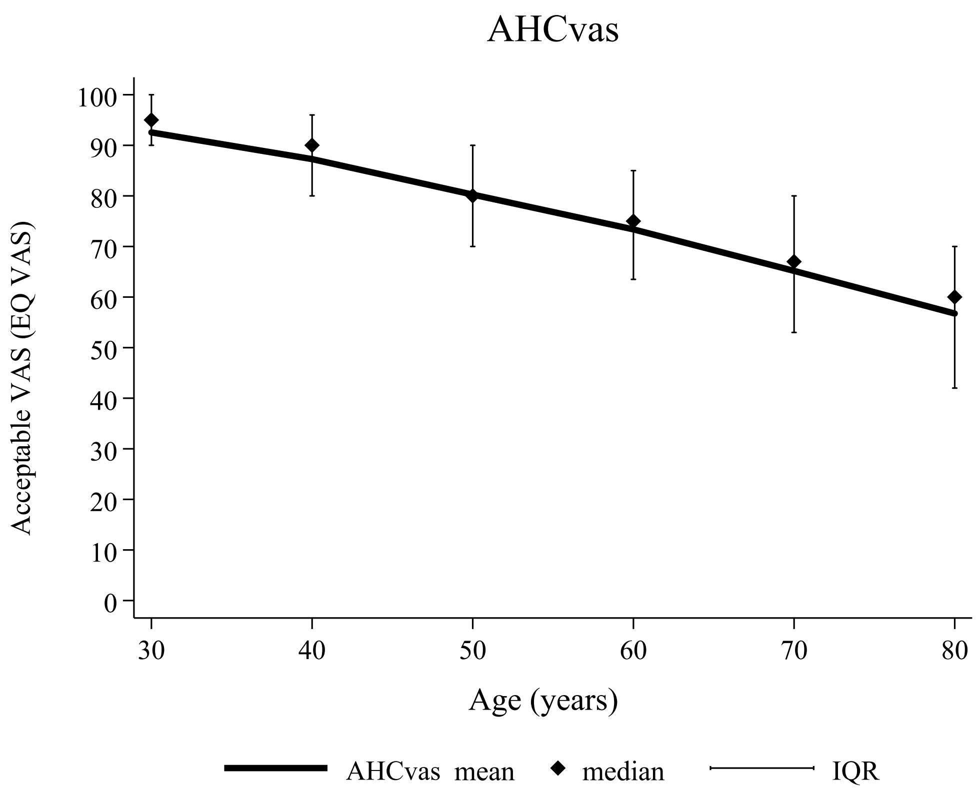
Left figure: comparison of AHC_{joint} (black bold line) and $AHC_{aggregate}$ (grey dashed line). The curves represent the means, diamonds represent the medians and error bars represent the IQR. Right figure: comparison of AHC_{joint} (black bold line) and respondents' corresponding EQ-5D-3L values (grey dashed line). We depicted the 24-34 years age group of respondents next to the AHCjoint measured at 30 years. The curves represent the means, diamonds represent the medians and error bars represent the IQR.

IV.4.7. Measuring AH with the adapted EQ VAS instrument

The two chapters below are based on our published results about using the EQ VAS for the measurement of AH (Zrubka et al., 2018c, Zrubka et al., 2019d).

We had acceptable VAS data from 196 respondents. The acceptable VAS (AHC_{vas}) values at age 30, 40, 50, 60, 70 and 80 were 92.5, 87.3, 80.3, 73.4, 65.2 and 56.7, respectively (Figure 24).

Figure 24 Acceptable VAS health curve (AHCvas)

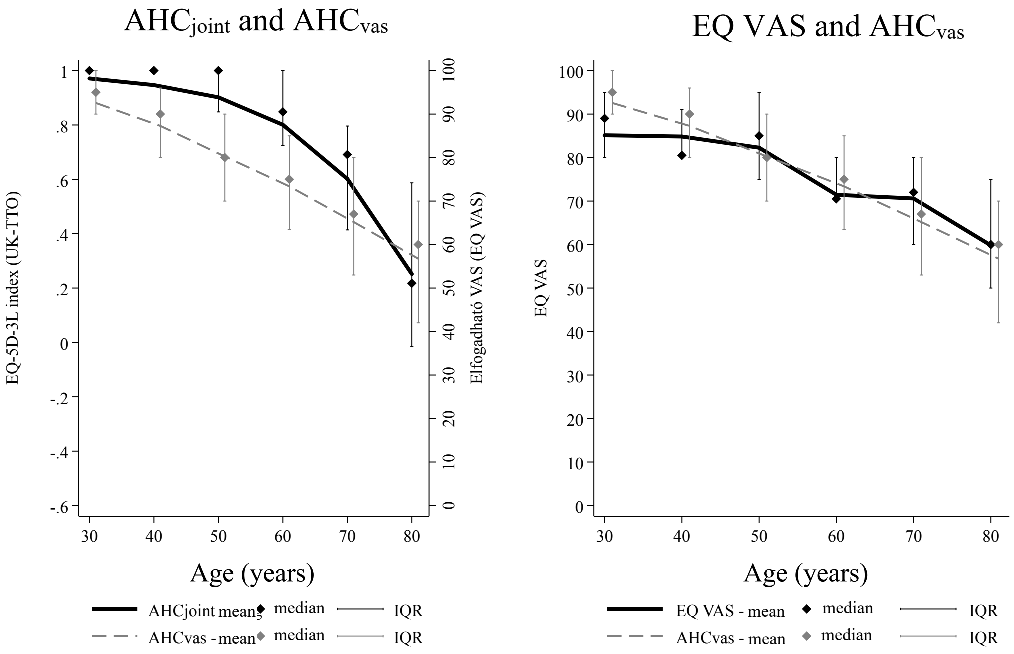


The AHCvas values reinforced the results measured with the discrete health states: people consider more problems acceptable with age. The AHC_{joint} and AHC_{vas} curves within the limits of their measurement range showed similar spread, their dispersion values were similar, however, their shape was different. While AHCjoint obtained via the evaluation of discrete health states indicated minimal amount of acceptable problems up to the age of 50 years, which was followed by a steep decline of AH, the AHC_{vas} showed a quasi-linear course, which suggests that people on average consider an even rate of health deterioration acceptable in each decade (Figure 25, left graph)

Our respondents above 40 years of age considered their own health measured by the EQ VAS similar to the level of acceptable health. The dispersion of the curves was also similar. In the 30-year-old age group, however, the sample's own health was worse compared to the level considered acceptable for their age (Figure 25, right graph).

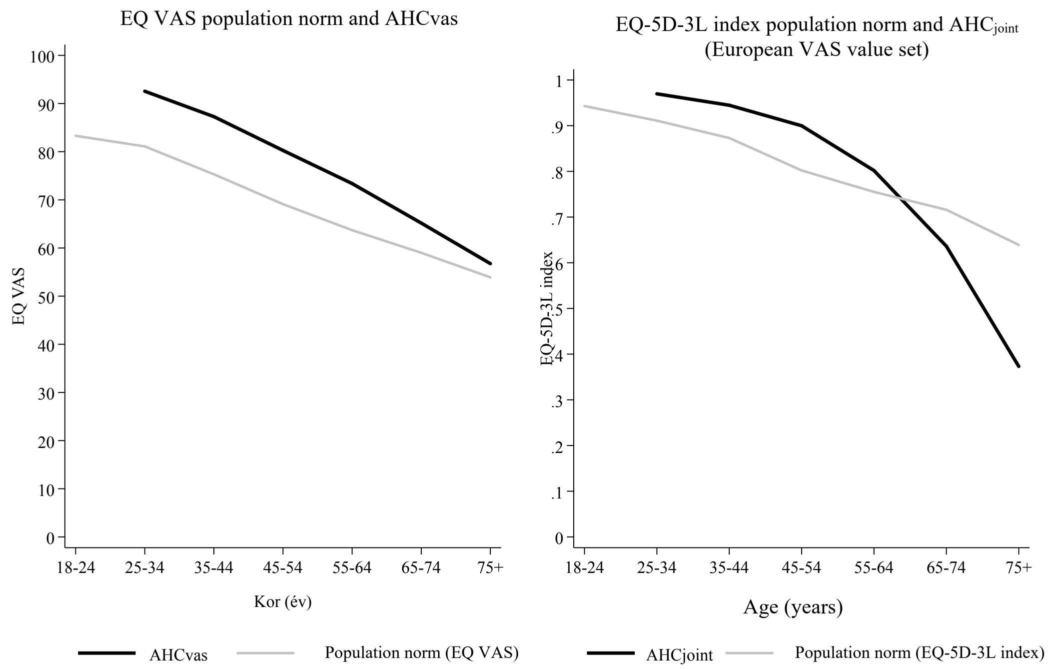
Similarly to the self-reported health of our sample on the EQ VAS, the level of acceptable health was higher compared to the EQ VAS measured in the general population (Szende and Nemeth, 2003) (Figure 26, right graph).

Figure 25 Comparison of AHC$_{vas}$ with health variables of the sample



Left figure: comparison of AHC$_{joint}$ (black bold line) and AHC$_{vas}$ (grey dashed line). The left vertical axis shows the possible extreme values of AHCjoint (EQ-5D-3L index, max:1, min: -0.594), the right vertical axis shows the extreme values of AHCvas (max:100, min:0). The curves represent the means, diamonds represent the medians and error bars represent the IQR. Right figure: AHCvas is depicted next to responders' own EQ VAS scores. The AHCvas measured at 30 years corresponds to the EQ VAS values of the 25-34-year-old age group.

Figure 26 Comparison of AH and population norms



Population norm: National Health Survey (Szende and Nemeth, 2003) The AHCvas measured at 30 years corresponds to the EQ VAS values of the 25-34-year-old age group.

IV.4.8. Factors influencing AHC_{vas} (H_3)

We analysed the AHC_{vas} via multilevel regression. Results are summarized in Table 6. We estimated the model (detailed in IV.2.6.c) in steps. In the baseline model (M1) the intercepts and slopes of individual AHC_{vas} curves differed significantly. The AHCvas at year 30 was 93.4 points (±SD=8.7) (p<0,001). AHC_{vas} showed on average 7,2 points decline in every 10 years. (β_{ageAH}=-0,723, p<0,001). *Since in all further models including explanatory variables the β_{ageAH} was significant and negative, we accepted H_3: people consider worse health states acceptable in older ages compared to younger ages.*

We examined the effects of respondents' individual characteristics in four groups. Individual factors explained only a small amount of the model variance: in the full model containing all explanatory variables (M5), 2/3 of the variance of the intercept and ¾ of the variance of the slope was not explained. Holding all other parameters fixed, the explanation of the M5 coefficients is the following:

a) Base model characteristics: 35-64-year-old man, secondary education, EQ VAS=80.5, not high-risk drinker, non-smoker, body mass index below 25, does exercise regularly, imagined himself during the evaluation of AH, his relatives died before the age of 75 and he is not an informal caregiver.
b) Younger individuals indicated worse health states acceptable at 30 years compared to the 35-64-year-old reference group, while they considered slower deterioration of health acceptable.
c) The self-reported health of respondents influenced significantly both the level of AHCvas and the slope of the curve. Individuals, who rated their health better than average on the EQ VAS, considered higher level of health and slower deterioration acceptable compared to the base model, than those had worse health than average.
d) The lifestyle of respondents influenced acceptable health only to a small extent: ones who do not exercise regularly and who are overweight (body mass index >25) considered a steeper acceptable deterioration rate of health. High risk drinkers considered less steep deterioration of health acceptable.
e) The reference person respondents imagined during the evaluation of acceptable health also influenced the level of acceptable health. Ones who imagined themselves, considered higher level and slower deterioration of health acceptable compared to those, who imagined others.
f) Individuals' experiences with their surroundings also influenced their opinion about AH: the lifespan of close relatives, and informal caregiver status (close relationship with ill or old individuals) also significantly influenced the amount of acceptable health.

Table 6 Multilevel regression of AHC_{vas}

		M1	M2	M3	M4	M5
1st level parameters	Intercept	93.87***	95.76***	92.71***	94.39***	96.71***
		(0.662)	(1.481)	(1.770)	(1.856)	(2.037)
	age_{AH}	-0.723***	-0.757***	-0.751***	-0.696***	-0.734***
		(0.0258)	(0.0587)	(0.0713)	(0.0754)	(0.0822)
2nd level parameters: intercept	Respondents' age: 18-34		-4.143***	-4.654***	-4.564***	-4.636***
			(1.398)	(1.315)	(1.312)	(1.318)
	Respondents' age: 65+		-0.148	3.372*	2.472	2.673
			(2.023)	(1.968)	(1.969)	(2.000)
	Female gender		0.687	0.864	0.660	0.686
			(1.317)	(1.278)	(1.269)	(1.257)
	Tertiary education		-1.053	-1.238	-1.026	-0.856
			(1.321)	(1.212)	(1.204)	(1.193)
	Self-rated health (EQ VAS) [i]			0.280***	0.268***	0.271***
				(0.0468)	(0.0463)	(0.0456)
	High risk alcohol			1.032	0.406	0.0466
				(2.056)	(2.038)	(2.012)
	Smoking			2.114	2.320	2.553*
				(1.557)	(1.538)	(1.527)
	Lack of exercise			1.024	0.796	0.305
				(1.244)	(1.238)	(1.243)
	Body mass index > 25			3.220**	3.455***	3.501***
				(1.312)	(1.295)	(1.289)
	Persons imagined: own and others				-2.844*	-2.825*
					(1.577)	(1.585)
	Persons imagined: others				-3.101**	-2.640*
					(1.377)	(1.368)
	Close relative's life span >75y					-3.061**
						(1.272)
	Informal caregiver					-1.123
						(1.415)
2nd level parameters: slope	Respondents' age: 18-34		0.0832	0.0663	0.0754	0.104*
			(0.0556)	(0.0531)	(0.0534)	(0.0534)
	Respondents' age: 65+		-0.0554	0.0582	0.0360	-0.0180
			(0.0802)	(0.0793)	(0.0799)	(0.0808)
	Female gender		-0.0111	0.00923	0.0116	0.00213
			(0.0523)	(0.0516)	(0.0516)	(0.0509)
	Tertiary education		0.0228	0.00741	0.0123	0.0119
			(0.0524)	(0.0489)	(0.0489)	(0.0483)
	Self-rated health (EQ VAS) [i]			0.00743***	0.00696***	0.00690***
				(0.00189)	(0.00188)	(0.00185)
	High risk alcohol			0.150*	0.132	0.143*
				(0.0828)	(0.0827)	(0.0813)
	Smoking			-0.0219	-0.0229	-0.0421
				(0.0629)	(0.0627)	(0.0620)
	Lack of exercise			-0.125**	-0.132***	-0.140***
				(0.0503)	(0.0503)	(0.0503)
	Body mass index > 25			0.0894*	0.0899*	0.0699
				(0.0529)	(0.0527)	(0.0522)
	Persons imagined: own and others				-0.0558	-0.0343
					(0.0644)	(0.0645)
	Persons imagined: others				-0.138**	-0.142**
					(0.0559)	(0.0552)
	Close relative's life span >75y					0.00713
						(0.0517)
	Informal caregiver					0.166***
						(0.0572)
Random effects parameters	Variance (ageAH)	0.1174***	0.1157***	0.0971***	0.0951***	0.0912***
		(0.0131)	(0.0130)	(0.0110)	(0.0109)	(0.0105)
	Variance (Intercept)	75.483***	70.870***	56.961***	54.746***	52.875***
		(8.697)	(8.257)	(6.813)	(6.633)	(6.446)
	Covariance (Intercept. age_{AH})	0.0414	0.1486	-0.2336	-0.3204*	-0.2828
		(0.2383)	(0.2302)	(0.1965)	(0.1943)	(0.1880)
	AIC	7657.354	7624.635	7522.899	7401.364	7352.859
	Observations (N)	1.145	1.139	1.133	1.115	1.109
	Number of respondents	194	193	192	189	188

Standard errors in parentheses

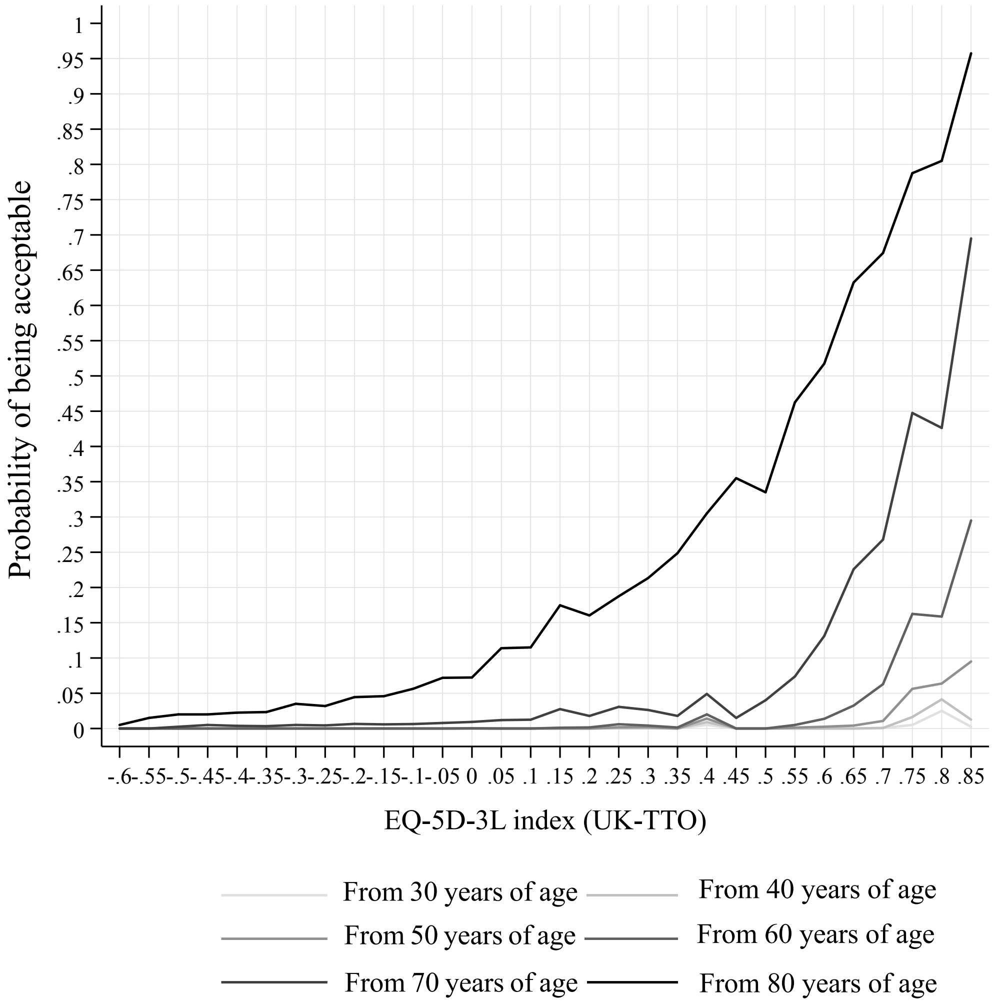
*** p<0.01, ** p<0.05, * p<0.1

i: Self-rated health was measured on EQVAS and was centred to the sample mean

IV.4.9. Az E-matrix

During the joint evaluation we elicited the acceptability of all health states. The E-matrix is shown in Table 7. The most frequently acceptable health state was 21111 (moderate problems with mobility), which was acceptable by 96% of respondents in 80 years of age. The worst health state described by EQ-5D-3L (33333: severe problems in all dimensions) was also acceptable by 0.5% of respondents in 80 years of age (a single respondent from the sample of 200). Figure 27 depicts the acceptability of problems by age and problem severity, measured by the EQ-5D-3L index. Health states with lower EQ-5D-3L index values than 0.5 were acceptable nearly exclusively from 80 years of age. Below 50 years, the acceptability of the health state with highest utility (11211, moderate problem with usual activities) was below 10%. Certain respondents considered worse than dead health states acceptable above 80 years of age. We rounded the EQ-5D-3L index values of problems to 0.05 precision on the graph.

Figure 27 Severity and acceptability of health problems in different ages



We rounded the EQ-5D-3L index values of all 242 EQ-5D-3L profiles to 0.05 precision and ranked in increasing order on the horizontal axis. The vertical axis shows the probability of being acceptable in 10-year age bands.

Table 7 E-matrix

	Age							Age							Age					
Profile	30	40	50	60	70	80	Profile	30	40	50	60	70	80	Profile	30	40	50	60	70	80
11111	100	100	100	100	100	100	21111	0.5	1.5	16	38	75.5	96	31111	0	0	0.5	0.5	8	31.5
11112	9.5	16	23.5	43	61.5	78.5	21112	0	0	2.5	12.5	41	70.5	31112	0	0	0	0	2.5	19
11113	2.5	4.5	6.5	9.5	17.5	40.5	21113	0	0	1.5	2.5	7	31	31113	0	0	0	0	1	9.5
11121	2	6.5	18.5	41.5	71.5	88.5	21121	0	0	3	13.5	50.5	82.5	31121	0	0	0	0	3	23.5
11122	0	0	2.5	17.5	42	67	21122	0	0	1	6.5	28	58.5	31122	0	0	0	0	1.5	15
11123	0	0	0.5	2	7	32	21123	0	0	0	1	5	27	31123	0	0	0	0	1	7
11131	0.5	0.5	2	5.5	20	45.5	21131	0	0	0	0.5	8.5	34	31131	0	0	0	0	1	11.5
11132	0	0	0	0.5	8.5	30	21132	0	0	0	0	5.5	23.5	31132	0	0	0	0	1	8
11133	0	0	0	0.5	3.5	15.5	21133	0	0	0	0	1.5	11.5	31133	0	0	0	0	1	4
11211	0	1	3	21	63.5	95.5	21211	0	0	0.5	8.5	42.5	89.5	31211	0	0	0	0	3	21.5
11212	0	0	0	5.5	34	70.5	21212	0	0	0	4.5	24.5	64.5	31212	0	0	0	0	1.5	15
11213	0	0	0	1	5.5	30	21213	0	0	0	1	3.5	26.5	31213	0	0	0	0	1	8
11221	0	0	1	9	43.5	82.5	21221	0	0	0.5	4.5	33	76.5	31221	0	0	0	0	2	18
11222	0	0	0	3.5	23.5	59.5	21222	0	0	0	2	15	51.5	31222	0	0	0	0	1	11.5
11223	0	0	0	1	5	25.5	21223	0	0	0	1	3.5	20	31223	0	0	0	0	1	5
11231	0	0	0	1.5	7	32.5	21231	0	0	0	0	3	26	31231	0	0	0	0	1	7.5
11232	0	0	0	0	5	24.5	21232	0	0	0	0	2.5	18	31232	0	0	0	0	1	5.5
11233	0	0	0	0	2	11	21233	0	0	0	0	1.5	8.5	31233	0	0	0	0	1	3.5
11311	0	0	0	0.5	4	49	21311	0	0	0	0	2.5	38.5	31311	0	0	0	0	1	13
11312	0	0	0	0	1	32.5	21312	0	0	0	0	1	25	31312	0	0	0	0	0.5	9
11313	0	0	0	0	0.5	16.5	21313	0	0	0	0	0.5	13	31313	0	0	0	0	0.5	4.5
11321	0	0	0	0	2.5	35	21321	0	0	0	0	2	28.5	31321	0	0	0	0	1	9.5
11322	0	0	0	0	1	25.5	21322	0	0	0	0	1	16	31322	0	0	0	0	0.5	5
11323	0	0	0	0	0.5	9	21323	0	0	0	0	0.5	7.5	31323	0	0	0	0	0.5	2.5
11331	0	0	0	0	1.5	15.5	21331	0	0	0	0	1	10.5	31331	0	0	0	0	0.5	4
11332	0	0	0	0	1	8.5	21332	0	0	0	0	1	5	31332	0	0	0	0	0.5	2
11333	0	0	0	0	0.5	2.5	21333	0	0	0	0	0.5	2.5	31333	0	0	0	0	0.5	2
12111	0.5	0.5	1.5	6.5	32.5	83.5	22111	0	0	0.5	3	22	74	32111	0	0	0	0	2.5	20
12112	0.5	0.5	1	3	20.5	61.5	22112	0	0	0	1	13	54.5	32112	0	0	0	0	2	13
12113	0.5	0.5	0.5	1	5	26.5	22113	0	0	0	0	1.5	21	32113	0	0	0	0	1	7
12121	0	0	0.5	2.5	23	69.5	22121	0	0	0.5	1.5	16.5	62.5	32121	0	0	0	0	2	16.5
12122	0	0	0.5	1.5	13.5	48.5	22122	0	0	0	0	8.5	40	32122	0	0	0	0	1.5	10
12123	0	0	0	0	2.5	19	22123	0	0	0	0	1.5	15.5	32123	0	0	0	0	1	6
12131	0	0	0	0	5.5	29.5	22131	0	0	0	0	4	26.5	32131	0	0	0	0	1	8.5
12132	0	0	0	0	2.5	22	22132	0	0	0	0	2	17	32132	0	0	0	0	1	5.5
12133	0	0	0	0	1	10	22133	0	0	0	0	1	8.5	32133	0	0	0	0	1	4
12211	0	0	0.5	2	23	73.5	22211	0	0	0.5	1.5	13.5	68	32211	0	0	0	0	2	16
12212	0	0	0	1	14.5	54.5	22212	0	0	0	0.5	7.5	44.5	32212	0	0	0	0	1.5	10
12213	0	0	0	0	2.5	20	22213	0	0	0	0	1.5	15.5	32213	0	0	0	0	1	6.5
12221	0	0	0.5	1.5	15	61	22221	0	0	0.5	1	9	54.5	32221	0	0	0	0	1.5	11
12222	0	0	0	0.5	8	41.5	22222	0	0	0	0	4	33.5	32222	0	0	0	0	1	5.5
12223	0	0	0	0	2	15.5	22223	0	0	0	0	1.5	11	32223	0	0	0	0	1	4.5
12231	0	0	0	0	2	22.5	22231	0	0	0	0	1.5	19.5	32231	0	0	0	0	1	6
12232	0	0	0	0	1	15.5	22232	0	0	0	0	1	12.5	32232	0	0	0	0	1	4
12233	0	0	0	0	1	7	22233	0	0	0	0	1	6.5	32233	0	0	0	0	1	3.5
12311	0	0	0	0	1	35.5	22311	0	0	0	0	1	26.5	32311	0	0	0	0	1	10
12312	0	0	0	0	0.5	22.5	22312	0	0	0	0	0.5	15	32312	0	0	0	0	0.5	6.5
12313	0	0	0	0	0.5	10	22313	0	0	0	0	0.5	7	32313	0	0	0	0	0.5	4
12321	0	0	0	0	1	24	22321	0	0	0	0	1	17.5	32321	0	0	0	0	1	6.5
12322	0	0	0	0	0.5	15	22322	0	0	0	0	0.5	8.5	32322	0	0	0	0	0.5	3.5
12323	0	0	0	0	0.5	5	22323	0	0	0	0	0.5	4	32323	0	0	0	0	0.5	2.5
12331	0	0	0	0	0.5	9	22331	0	0	0	0	0.5	7	32331	0	0	0	0	0.5	3
12332	0	0	0	0	0.5	4	22332	0	0	0	0	0.5	2.5	32332	0	0	0	0	0.5	2
12333	0	0	0	0	0.5	2	22333	0	0	0	0	0.5	2	32333	0	0	0	0	0.5	2
13111	0	0	0.5	0.5	3	32	23111	0	0	0	0	1	21	33111	0	0	0	0	0.5	11
13112	0	0	0	0	1.5	20	23112	0	0	0	0	1	16	33112	0	0	0	0	0.5	8.5
13113	0	0	0	0	0.5	9.5	23113	0	0	0	0	0.5	9	33113	0	0	0	0	0	5
13121	0	0	0	0	1.5	21	23121	0	0	0	0	1	18	33121	0	0	0	0	0.5	10.5
13122	0	0	0	0	1.5	15	23122	0	0	0	0	1	12.5	33122	0	0	0	0	0	6
13123	0	0	0	0	0.5	8	23123	0	0	0	0	0.5	7.5	33123	0	0	0	0	0	4
13131	0	0	0	0	1	11	23131	0	0	0	0	0.5	9	33131	0	0	0	0	0	5
13132	0	0	0	0	1	7.5	23132	0	0	0	0	0.5	6.5	33132	0	0	0	0	0	3.5
13133	0	0	0	0	0.5	5	23133	0	0	0	0	0.5	4.5	33133	0	0	0	0	0	2
13211	0	0	0	0	0.5	20	23211	0	0	0	0	0.5	15	33211	0	0	0	0	0.5	7
13212	0	0	0	0	0.5	14.5	23212	0	0	0	0	0.5	10.5	33212	0	0	0	0	0.5	4.5
13213	0	0	0	0	0.5	8.5	23213	0	0	0	0	0.5	7	33213	0	0	0	0	0	4
13221	0	0	0	0	0.5	16.5	23221	0	0	0	0	0.5	12	33221	0	0	0	0	0.5	6
13222	0	0	0	0	0.5	10.5	23222	0	0	0	0	0.5	8	33222	0	0	0	0	0	3
13223	0	0	0	0	0.5	6	23223	0	0	0	0	0.5	5	33223	0	0	0	0	0	2.5
13231	0	0	0	0	0.5	8	23231	0	0	0	0	0.5	7	33231	0	0	0	0	0	3
13232	0	0	0	0	0.5	4.5	23232	0	0	0	0	0.5	4	33232	0	0	0	0	0	1.5
13233	0	0	0	0	0.5	4	23233	0	0	0	0	0.5	3.5	33233	0	0	0	0	0	1.5
13311	0	0	0	0	0.5	13	23311	0	0	0	0	0.5	8.5	33311	0	0	0	0	0	5.5
13312	0	0	0	0	0.5	7.5	23312	0	0	0	0	0.5	5	33312	0	0	0	0	0	3.5
13313	0	0	0	0	0.5	5.5	23313	0	0	0	0	0.5	4	33313	0	0	0	0	0	2.5
13321	0	0	0	0	0.5	9	23321	0	0	0	0	0.5	6	33321	0	0	0	0	0	3
13322	0	0	0	0	0.5	5	23322	0	0	0	0	0.5	3	33322	0	0	0	0	0	1.5
13323	0	0	0	0	0.5	3	23323	0	0	0	0	0.5	2.5	33323	0	0	0	0	0	1
13331	0	0	0	0	0.5	4	23331	0	0	0	0	0.5	3	33331	0	0	0	0	0	1.5
13332	0	0	0	0	0.5	1.5	23332	0	0	0	0	0.5	1	33332	0	0	0	0	0	0.5
13333	0	0	0	0	0.5	1	23333	0	0	0	0	0.5	1	33333	0	0	0	0	0	0.5

The data represent the % of respondents who consider the health profiles acceptable

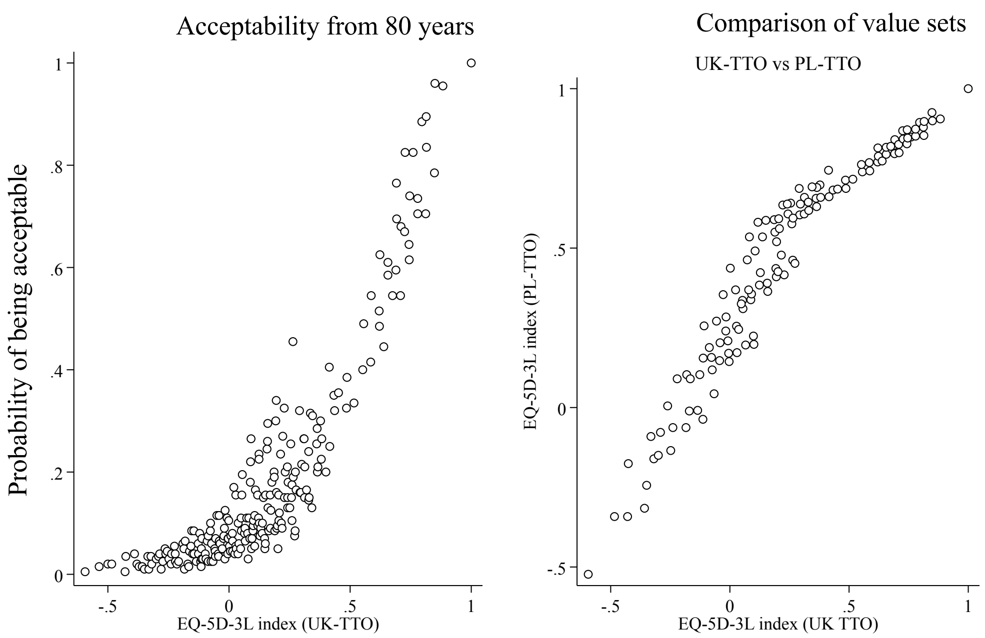
IV.4.10. The association of the acceptability and utility of health states

We also examined, how the acceptability of certain health profiles was associated with their utility levels. Because the EQ-5D-3L questionnaire does not have a Hungarian value set, we used in our calculations the index values of the UK-TTO value (Dolan, 1997) set, which is most frequently applied in Hungary in health economic analyses.

We depicted the EQ-5D-3L index values of the 243 EQ-5D-3L profiles, as well as the probability of their acceptability in the 80-year-old age group (Figure 28, left graph). The association was not linear, the shape of the scatterplot showed a break at approximately 0.2 EQ-5D-3L index values. The acceptability of the profiles below 0.2 EQ-5D-3L index was low, and it depended only a small amount on the change of the utility values. However, above 0.2 EQ-5D-3L index values, the acceptability and utility scatterplot suggested a strong linear association.

We found a similar break when comparing the UK-TTO value set and the Polish TTO value set, in the proximity of 0.2 EQ-5D-3L index values. In the Polish value set, index values of 0.2 in the UK-TTO value set corresponded to values around 0.5, suggesting that in this range the Polish population rated less severe (and probably more acceptable?) than British respondents.

Figure 28 Comparison of the acceptability and utility of EQ-5D profiles



Left figure: EQ-5D-3L index values and acceptability of 243 EQ-5D-3L profiles from 80 years of age. Right figure: comparison of EQ-5D-3L index values calculated with the UK-TTO and Polish-TTO value sets, for the 243 EQ-5D-3L profiles. UK-TTO: British time-trade-off value set, Polish TTO: Polish time-trade-off value set

During the separate evaluation we found that the acceptability of health problems was associated not only with their severity or the age, but also with the dimension the problems were reported in (Table 4.) We wondered, how the results of joint evaluation are affected by the dimensions in which the health problems are reported. We tested the question with a logistic regression model, in which the dependent variable was the acceptability of the EQ-5D-3L profile (yes-no), and as predictors, we entered dummy variables for: age, severity of the profile (misery index), and the occurrence of any problems in the EQ-5D dimensions. We clustered the robust standard errors at the levels of the individuals (Table 8). Greater coefficients represented a greater acceptability of profiles. The explanatory power of the model was good (pseudo-R^2=0.476). Although the Hosmer- Lemeshow test was significant ($\chi^2_{(8)}$=313,19, p<0,0001), which suggests problems with the model fit, in the acceptability range above 0.5% the model fit was good. According to our expectations, in the older age group the acceptability was greater, and it was smaller in case of more severe problems. We compared the problems in different EQ-5D dimensions with the Wald test. The coefficient of self-care problems was significantly smaller than that of any other dimensions, while the pairwise comparison of all other dimensions showed not significant difference. According to the results, in the same age, a health state with the same severity is less likely to be acceptable, if it is characterised by problems in the self-care dimension. Although the evaluation of health-state utilities is based on finding detailed nuances in the differences between people's preferences about each dimension, by analysing the acceptability of health states, we did not find such differences between the effect of EQ-5D dimensions on acceptability, with the exception of the self-care dimension.

Table 8 The influence of health dimensions on the acceptability of health states

Logistic regression		Log-OR	(se)
Age (years)	40	0.66^{***}	(0.15)
	50	1.74^{***}	(0.20)
	60	3.01^{***}	(0.25)
	70	4.81^{***}	(0.28)
	80	6.90^{***}	(0.27)
Severity (misery index)	6	-2.69^{***}	(0.11)
	7	-4.62^{***}	(0.21)
	8	-6.18^{***}	(0.31)
	9	-7.53^{***}	(0.43)
	10	-8.69^{***}	(0.56)
	11	-9.66^{***}	(0.69)
	12	-10.5^{***}	(0.80)
	13	-11.3^{***}	(0.91)
	14	-12.3^{***}	(1.21)
	15	-2.69^{***}	(0.11)
Dimension	Mobility	1.12^{***}	(0.11)
	Self-care	0.65^{***}	(0.099)
	Usual activities	1.09^{***}	(0.12)
	Pain/discomfort	1.21^{***}	(0.12)
	Anxiety/depression	1.05^{***}	(0.12)
Constant		-5.13^{***}	(0.29)
	N	290400	

Robust standard errors were clustered at the individual level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

IV.4.11. The effect of response time on the reliability of AH measurements

Response times of AH evaluation showed considerable variability. We wondered, whether the time spent with answering the questions influences the results of evaluation? During joint evaluation, the number of questions showed negative association with the average response time per question ($r=-0,35$, $p<0,001$). This may mean that respondents learned the evaluation task, but it may also be a signal of tiredness, increasingly superficial and hasty responses. We hypothesised that shorter response time than a certain limit is not enough for the appropriate evaluation of the acceptability of health states, and this may influence negatively the reliability of the measurement. We tested this question using a linear probability model, in which the acceptability of health states was estimated from the age and the severity of health states. In the model, fast responses were indicated by a dummy variable, for which the threshold was changed between 3 and 10 seconds. The main effects as well as the interaction of fast responses with age and severity was tested via a joint Wald test. Response times shorter than 8 seconds per question significantly affected the model coefficients. For response times shorter than 5 seconds, the main effects were significant, faster responses were associated with greater probability of accepting the problems, while the probability of accepting severe problems was lower. We did not store detailed data about the questions and their order asked during the joint evaluation, which could provide information about whether the quick answers were associated with the evaluation situation, the severity of displayed health states or their order (Table 9). However, randomly chosen thresholds around the mean response time (12-19 seconds) did not provide significant results in the model (Wald test $p=0.1479$)

Table 9 Evaluation of response time effects via linear probability model

		Fast responses (sec)							
		<3sec	<4sec	<5sec	<6sec	<7sec	<8sec	<9sec	<1sec
Age	40	0.00065***	0.00064***	0.00065***	0.00069***	0.00073***	0.00074***	0.00070***	0.00076***
	50	0.0031***	0.0031***	0.0031***	0.0033***	0.0034***	0.0035***	0.0036***	0.0038***
	60	0.012***	0.012***	0.012***	0.012***	0.012***	0.012***	0.012***	0.013***
	70	0.051***	0.052***	0.051***	0.052***	0.048***	0.049***	0.048***	0.048***
	80	0.20***	0.20***	0.20***	0.20***	0.20***	0.20***	0.20***	0.20***
Severity	6	-0.66***	-0.66***	-0.66***	-0.66***	-0.65***	-0.65***	-0.65***	-0.64***
(misery index)	7	-0.83***	-0.83***	-0.83***	-0.83***	-0.83***	-0.83***	-0.83***	-0.83***
	8	-0.91***	-0.91***	-0.91***	-0.91***	-0.91***	-0.91***	-0.91***	-0.91***
	9	-0.96***	-0.96***	-0.95***	-0.96***	-0.96***	-0.96***	-0.96***	-0.96***
	10	-0.98***	-0.98***	-0.98***	-0.98***	-0.98***	-0.98***	-0.98***	-0.98***
	11	-0.99***	-0.99***	-0.99***	-0.99***	-0.99***	-0.99***	-0.99***	-0.99***
	12	-0.99***	-0.99***	-0.99***	-0.99***	-0.99***	-0.99***	-1.00***	-0.99***
	13	-1.00***	-1.00***	-1.00***	-1.00***	-1.00***	-1.00***	-1.00***	-1.00***
	14	-1.00***	-1.00***	-1.00***	-1.00***	-1.00***	-1.00***	-1.00***	-1.00***
	15	-1.00***	-1.00***	-1.00***	-1.00***	-1.00***	-1.00***	-1.00	-1.00***
Fast response	Yes	0.014*	0.020***	0.017**	0.010	0.0053	0.0050	0.0011	0.0037
x Ager[a]	40	-0.00024	-0.000056	-0.00017	-0.00036	-0.00049*	-0.00053*	-0.00025	-0.00040
	50	0.0014	0.00040	0.00026	-0.00098	-0.0016	-0.0018*	-0.0018*	-0.0022**
	60	-0.0061**	-0.0068**	-0.0013	-0.0044	-0.0039	-0.0047	-0.0030	-0.0033
	70	-0.013	-0.022*	-0.014	-0.013	0.011	0.0074	0.011	0.0060
	80	-0.068*	-0.089***	-0.087***	-0.043	-0.037	-0.030	-0.013	-0.022
x Severity[b]	6	0.055	-0.0032	0.0087	-0.020	-0.033	-0.037	-0.043	-0.052*
(misery index)	7	0.012	-0.020	-0.0034	-0.0029	-0.0092	-0.012	-0.015	-0.019
	8	-0.017	-0.032*	-0.025*	-0.013	-0.016	-0.016	-0.012	-0.013
	9	-0.021**	-0.027***	-0.026***	-0.015	-0.011	-0.010	-0.0051	-0.0074
	10	-0.015**	-0.017***	-0.017***	-0.010	-0.0028	-0.0017	0.0026	-0.00071
	11	-0.0095**	-0.011***	-0.011***	-0.0064	0.0024	0.0029	0.0071	0.0037
	12	-0.0066**	-0.0069**	-0.0071**	-0.0048	0.0043	0.0042	0.0083	0.0055
	13	-0.0044*	-0.0045*	-0.0046*	-0.0038	0.0042	0.0039	0.0078	0.0057
	14	-0.0026	-0.0026	-0.0027	-0.0028	0.0029	0.0025	0.0063	0.0049
	15	-0.00091	-0.00093	-0.00095	-0.00099	-0.0011	-0.0011	0.0036	0.0029
Constant		0.96***	0.96***	0.96***	0.96***	0.96***	0.96***	0.96***	0.96***
	N	281394	281394	281394	281394	281394	281394	281394	281394
Wald-test	p	<0.0001	<0.0001	0.0002	0.1479	0.0112	0.0070	0.0577	0.1358

IV.5. Discussion

The first part of the dissertation introduced the methodological background and main results of two new measurement methods of AH: joint evaluation and acceptable VAS.

The newly developed adaptive testing algorithm for joint evaluation was feasible in practice, and after the separate evaluation task, it provided a more accurate picture on acceptable health, than separate evaluation.

a) Separate and joint evaluation was feasible within on average 6-7 minutes per respondent
b) Although the efficiency of the adaptive testing algorithm fell below the expected range, instead of 90% of respondents, we could elicit the acceptability of all elements of the E-matrix only for 43.5% of respondents (H_1). However, only 2.3% of all health states remained unknown.
c) During joint evaluation respondents accepted fewer problems than in separate evaluation (H_2). The aggregated health curve from the results of joint evaluation (AHC_{joint}) fell between the values of the acceptable health curves gained after separate evaluation ($AHC_{aggregate}$, AHC_{worst}), and the smaller dispersion of $AHCj_{oint}$ than that of $AHC_{aggregate}$ enabled a more accurate measurement.
d) Measuring AH with the EQ VAS also demonstrated that people in older ages consider more health problems acceptable than in younger ages (H_3).

The results confirmed the conclusions of previous studies, that there is a reference point of acceptable health, based on which people can evaluate, whether certain health states are acceptable in different ages or not (Brouwer et al., 2005, Pentek et al., 2014b, Wouters et al., 2015). Our research confirms previous results, according to which:

a) people accept more health problems with age
b) moderate problems are more acceptable than severe ones
c) the acceptability of problems depends on the type of problem

The first step of joint evaluation is separate evaluation, in which we received strikingly similar results to previous studies, which confirms the validity of the AH concept.

Healthy respondents from another country (Wouters et al., 2015) and chronic patients from Hungary (Pentek et al., 2014b) evaluated acceptable health in a suprisingly similar way. The comparison of acceptable problems reflected previous experiences with the measurement of health outcomes. Patients accomodate the deterioration of their condition, therefore they attach higher utilities to health problems, than healthy individuals (Brazier et al., 2018, Ubel et al., 2003). This was reflected by the opinion of rheumatoid arthritis patiens affected by a painful and destructive inflammatory condition of joints mainly in hands (Smolen et al., 2017), who, in the dimensions of self-care, usual activities and pain / discomfort considered health problems more acceptable than the general population. However, in acceptability of problems in the mobility dimension we found no differences between patients and the participants of our study, despite the fact that there are market differences between the occurrence of problems in this dimension between patients and the normal population (Pentek, 2007). The dutch population considered problems in the anxiety / depression dimension less acceptable than respondents of the Hungarian sample, which reflects the Dutch preferences in EQ-5D-3L valuation studies. Compared to other countries, the Dutch assign greatest disutility to problems experienced in the dimension of anxiety / depression (Szende et al., 2006). Interestingly, pain/discomfort is evaluated similarly negatively to anxiety/depression by the Dutch, but in this field their preferences are similar to other countries', which may explain why acceptability in the pain/discomfort dimension did not differ compared to the Hungarian population. In our own research predominantly problems in the self-care dimension influenced the acceptablilty of health states. By our respondents problems were experienced least frequently in this dimension, and when comparing to the normal population, the difference in the frequency of problems was greatest in in the self-care dimension. Based on these results, it is intrigueing, how certain kinds of health-related experiences, and more narrowly the frequency of problems experienced by oneself or others influence the opinion of individuals about the acceptability of health states.

The acceptability of health states is presumably influenced (at least partly) differently by the preferences of the individuals, than the valuation of health state utilities. The evaluation of AH differs in many regards from the valuation of health utilities. The valuation of AH is not related so closely with assumptions about death and expected life-span, than the valuation of utilities. Also, there is no hypothetical exchange in the evaluation task, and there is no need to estimate probabilities of events (Brazier et al., 2018, Parkin and Devlin, 2006). Therefore, the criteria of the expected utility theory do not hold (van Osch et al., 2004, Bleichrodt and Johannesson, 1997). Alghough we do not understand exaclty the cognivite processes underlying the evaluation of AH, we may assume that those are different from the ones underlying the elicitation of utilities.

During the valuation of utilities respondents need to imagine themselves in health states that span long periods of time. A central question of the debates about the methodology of QALY measurement is how valid this situation is, and whether the respondents momentary state or the opinion of patients who previously experienced problems had is a more authentic foundation for the elicitation of health state utilities (Brazier et al., 2018). The evaluation of AH requires a new hypothetical situation: the acceptability of different health states needs to be imagined in different ages. This, depending on the age and attitude of respondents may be influenced by a mixture of their current and past experiences, as well as their imagination about the future, their own health, and their health-related experiences in their environment. Acceptable health was evaluated significantly differently by those, who imagined themselves or others during the joint evaluation task. Similar differences were found in EQ-5D valuation studies eliciting health state utilities (Mulhern et al., 2013). The individuals health status, lifestyle, the life span of close relatives and informal caregiver status all influenced the evaluation of acceptable health. The dispersion of AHCjoint and the self-rated EQ-5D-3L index values suggested that in older age groups respondents form a more inhomogenous picture about acceptable health problems, than in younger ages. Since the goal of AH measurement is the reflection of societal preferences, it is important to elicit the *general* opinion of individuals. Therefore, building on the learnings from past research, an important aspect of future studies of acceptable health may be the as accurate as possible description of the evaluation task while not losing the merits of its relative simplicity:

a) What do we mean exactly under acceptable health? (Health states without losing wellbeing, or still bearable suffering?)
b) Whose health should be imagined by the respondent. (Himself/herself in different ages, or his/her environment, average or extreme examples?)
c) From whose point of view should respondents evaluate the acceptability of problems. (Their own opinion, or what do people generally consider acceptable?)

We have experienced all above interpretations of our respondents during the fieldwork.

We have discovered a breakpoint in the association between the utility and acceptability of health states in the proximity of EQ-5D-3L index values of 0.2. We have found a break in a similar position in the association between the Polish and British utility values, which were elicited with the very same methodology. Therefore, the question emerges, to what extent the differences between the acceptability and utility of EQ-5D-3L profiles are explained by the differences of the two methods, or the differences between the health-related preferences of Easter-European or British populations. One of the key proceeds of the large number of EQ-5D valuation studies is that they highlight the importance of the differences between country value sets explained by methodological differences between the studies (Brooks et al., 2003, Augustovski et al., 2013, Bernert et al., 2009, Xie et al., 2014). The the benefit of measuring AH lies exactly in its major difference from the health state utilities. The lower the association between AH and utilities, the more adequate it may be for the authentic representation of the society's preferences related to age and the severity of disease in a multiple-criteria decision setting (Marsh et al., 2016, Thokala et al., 2016).

An important aspect of evaluating AH is checking the reliability of the adaptive testing algorithm. Adaptive testing was based on three assumptions, which are logical and seem to be plausible, but have not yet been empirically proven:

a) The monotonicity of health states: if a health state is acceptable, then a less severe health state (same level or lower level of problems in any dimension) is also acceptable and vice versa.
b) The monotonicity of health deterioration: if a health state is acceptable in a certain age, then it is acceptable in older ages and vice versa.
c) Respondents provide reliable answers that are consistent with their priorities.

During our research, unanswered questions emerged concerning all three points above. The first point is contradicted by seemingly illogical health states: let's imagine a health state in which the individual is bedridden, without problems in usual activities and self-care. When evaluating acceptability, respondents may mix what they consider *acceptable* and what *possible*.

Concerning the second point, the anxiety / depression dimension can be mentioned as an example. In our research, we met respondents, who considered anxiety / depression acceptable in younger ages (most frequently mentioning work and family-related reasons), but with the advancement of age, they would consider the attenuation of anxiety / depression symptoms acceptable.

Regarding the third point we have only indirect information from our research. During the joint evaluation we provided the opportunity for our respondents to change their responses of separate evaluation. This happened in case of 12 respondents from the 200, two respondents changed their opinions twice during the joint evaluation task. One may wonder, what was reflected in their first or second opinion, or how reliable were the responses of those, who completed the evaluation task within the shortest time? In EQ-5D valuation studies, due to inconsistent responses, up to one third of respondents are excluded. Milder inconsistencies are accepted, but there were respondents, who valued full health (11111) worse than the worst EQ-5D profile (33333) (Brooks et al., 2003). Our adaptive testing algorithm on one hand protects us from logical inconsistencies, because it automatically fills logically consistent answers. But exactly this feature poses a risk concerning the reliability of the results. The algorithm provides consistency in case of not well-thought-through answers, and it amplifies the effect of random responses. The inclusion of control-questions, and the determination of the acceptable level of inconsistency is an important development task. Another problem is that due to the random sequence of questions starting from the extremely heterogenous $AHCa_{ggregate}$ patterns, the individual properties of questions, and their influence on response patterns is very difficult to study. The analysis of the time spent with answering the questions has highlighted this problem: from the mechanism of the algorithm, it cannot be excluded, that two subjects receive the following randomly selected questions:

a) 33233 in 70 years of age, 11221 in 50 years of age, 33212 in 70 years of age, 12121 in 50 years of age, and so on.
b) 31211 in 60 years of age, 23121 in 60 years of age, 22131 in 60 years of age, 23212 in 60 years of age, and so on.

In the first case, we may assume that the contrast between the questions leads respondents to make faster decisions, while in the second case, the precise interpretation and imagination of the highly similar states may take more time

Although responses shorter than 5-8 seconds influenced significantly the results of evaluation, they could be either positive (confident respondent, easily decidable question) or negative (tiredness, boredom) quality signals, which cannot be controlled according to the current mechanism of the adaptive testing algorithm. The above considerations reinforce the need for amending the joint evaluation module with elements that improve the reliability of the adaptive testing strategy.

In addition to measuring AH by joint evaluation of discrete health states, using the modified EQ VAS was also a novelty of our research. The goal of EQ-5D profiles is to provide cardinal values for public decision-making, that reflect the preferences of the general population, and which make individuals or diseases comparable between each other (d'Aspremont and Gevers, 2002). During the evaluation of discrete khealth states, several acceptable health states may occur in the same age, from which when we construct the AHC_{joint}, we use the lowest utility values that were determined by not reflecting the preferences of the individual in questions, but those of the general society. Therefore the applied utilities may differ from those of the individual in question. However, the EQ VAS provides a single numerical value about the current state of the individual, which can be influenced by health factors, which fall outside the dimensions of EQ-5D (Feng et al., 2014). On the other hand, the EQ VAS does nor provide information about based on what objective criteria do individuals compare the acceptability of health states with their inner reference image during the evaluation task. It merely informs about where AH falls between the best and worst imaginable health states. Alghough the raison d'etre of EQ VAS in the valuation of health has been criticised due to its incompatibility with the expected utility theory, due to its psychometric properties, for the measurement of societal preferences that are not measurable in monetary terms, we can accept the EQ VAS as a suitable instrument. (Parkin and Devlin, 2006) The differences between the courses of the AHC_{vas} and AHC_{joint} curves between the ages 40 and 60 suggests that such problems may influence the evaluation of AH, which are not reflected in the EQ-5D dimensions (sleep, sensory problems, cognition or sexual life) (Brazier et al., 2019, Brouwer et al., 2005), which are indirectly arguments in favour of extending the dimensions used for the elicitation of QALYs. AH may be a feasible concept for understanding the acceptability of health states other than EQ-5D profiles, and thereby it may be used as a more patient-centric tool, or a tool that is capable of the elicitation of societal priorities concerning a broader concept of wellbeing (e.g informal care, capabilities).

In addition to the elaboration of the strengths and limitations of our study, it has to be noted that the sampling of our research was not representative, therefore the reliability and generalisability of our results is limited.

V. THE ASSOCIATION OF ACCEPTABLE HEALTH AND HAPPINESS

V.1. Background

For the validity of the AH concept, a key assumption is that in acceptable health states further health improvement will increase the wellbeing of the individual to a lesser amount, than health gains in not acceptable health states (Wouters et al., 2017). If the individual's wellbeing is closely related to health improvements even beyond the threshold of AH, then it raises ethical concerns against the application of the AH concept, from utilitarian point of view. However, if crossing the AH threshold is not associated with further wellbeing increases, then we can hypothesize that maximising health up to the acceptability threshold also maximises the wellbeing achievable in connection with health gains. In case we consider the treatment of everyone who is not in acceptable health, then it is in harmony with the egalitarian principle. Subjective wellbeing is a complex term, some theories use it synonymously with happiness (Veenhoven, 2012), while other theories view it as a complex construct containing affective (hedonist) and cognitive (eudaimonic) elements (Ryff and Keyes, 1995). In our research the relationship of wellbeing and AH was explored via the association of AH and scores on the 0-10 numeral happiness scale (Veenhoven, 2009, Veenhoven, 1993). Since the measurement of AH according to our method happened from a societal perspective, it was concerned about the acceptability of health states by the society and not about to what extent our respondents considered their own health states acceptable. The priority of the individual versus societal viewpoint in QALY measurement is still a field of active debate among health economists (Brazier et al., 2018). For the better understanding of AH, we felt important to examine it also from the individuals' point of view: what factors influence whether individuals consider their own health acceptable or not? The application of the principle of AH can be fair if individual and societal priorities are in harmony, and those people are actually in acceptable health states, about whom the society assumes that.

Therefore, we felt important to inquire whether the association between AH and subjective wellbeing (happiness) is concordant with the societal assumptions about the association between AH and happiness?

In our research the newly applied EQ VAS measured the level of AH in a single continuous variable, thereby enabling us the more delicate research of the individual factors that influence the acceptability of health problems. In the next sections of the dissertation we were looking for answers for the following questions:

a) How does AH influence the individuals' happiness reflecting the level of individual wellbeing?
b) What factors influence whether individuals consider their own health acceptable?

V.2. Methods

V.2.1. Data

In the previous chapters we measured the level of health that is considered acceptable by our respondents in different ages. However, in the examination of AH and happiness, the key question was whether respondents consider their own health acceptable or not. We intended to answer this question indirectly from the data available from our research, therefore we constructed two new variables and introduced two new terms.

a) In the following we use the term *relative health* for the difference of the respondent's own health and the level of health he/she evaluated as acceptable in his own age group (e.g. for a 35-44-year-old respondent at the age 40). So relative health theoretically a continuous variable that can take values between -100 and +100, which receives 0 value if the respondents' own health and the health state considered acceptable in his/her age group receive the same score.
b) Based on their *relative health*, we formed two groups from respondents: in the *acceptable health* group relative health was positive, respondents evaluated their own health better than the level they consider acceptable in their own age group. In the *non-acceptable health* group respondents had negative relative health values.

V.2.2. Statistical methods

V.2.3. Exploring the relationship between AH and happiness

Happiness is described by a 0-10 numeral scale, which can be analysed as a continuous ordinal or categorical variable. The World Happiness Report analyses happiness data by linear regression (Helliwell et al., 2018) therefore we chose multivariate regression of the analysis of the association between relative health and happiness, in which happiness was the dependent variable, and as explanatory variables we included age, subjective health, relative health as continuous variables, and further categorical variables concerning the demographic status, lifestyle and health problems. In addition to main effects we studied interactions as well, where we used the binary *acceptable health* variable instead of relative health. When comparing multiple models, we considered the explanatory power (R^2 and adjusted R^2), when interpreting model coefficients, we considered $p<0.05$ level of significance. Heteroskedasticity was interpreted visually based on the scatterplot of residuals and predicted values, collinearity of variables was checked based on the VIF value, and outliers were checked based on Cook's D (>4/n), leverage (>2k+2/n), standardised residuals (absolute value >3), or standardised DfBeta (>2/√n) criteria. Heteroscedasticity was also checked by the Breusch-Pagan test, model specification with the Ramsey-RESET test, and the distribution of standard residuals was checked with the Shapiro-Wilk test and the skewness-kurtosis test. We also checked the contribution of individual independent variables to the explanatory power of the model by Shapley-decomposition.

We used logistic regression of study to relationship between relative health and lifestyle. The dependent variable was the acceptable health binary variable, explanatory variables were age, gender, socio-demographic variables, health problems and the lifestyle of individuals. When interpreting model coefficients, we chose $p<0.05$ level of significance. We checked model fit with the Hosmer-Lemeshow test, and the discriminative power of the model was examined by the area under the ROC curve, optimal thresholds were calculated using the Youden-index (sensitivity + specificity-1).

V.2.4. Hypothesis testing

The second goal was to explore the relationship between AH and happiness. The tested the three related hypotheses with the following methods:

***H_4:** In acceptable health status, the level of subjective health influences happiness to a smaller extent than in not acceptable health status.* We tested this hypothesis with the following multivariable regression model:

$$h = \alpha + \beta na + \gamma eqvas + \delta na * eqvas + \mu age + \theta age * eqvas + \boldsymbol{\lambda} X + \varepsilon$$

where h denotes happiness, α is the intercept, na denotes the binary variable of acceptable health with value of 1 if the individual is in not acceptable health status, and 0 if the individual has acceptable health status; *eqvas* is the individuals' self-rated health score on the EQ VAS, *age* is respondents' age centred at 18 years and X is the vector of other explanatory variables. We evaluated H_4 based on the parameter value of δ:

H_0: $\delta=0$

H_{alt}: $\delta>0$

H_4 can be accepted if H_0 is rejected and H_{alt} is accepted. (Results: p108)

***H_5:** Older individuals are more likely to consider their health acceptable than younger ones.*

***H_6:** Individuals with more severe conditions consider their health less acceptable than healthier ones.* We tested H_5 and H_6 hypotheses using the following logistic regression model:

$$logit_{pA} = \alpha + \beta age + \gamma eqvas + \delta X + \varepsilon$$

where $logit_{pA} = log(p_A/(1-p_A))$, p_A is the probability that an individual considers his/her health state acceptable, *age* is the age of respondents centred at 18 years, *eqvas* is the self-rated current health of respondents measured on the EQ VAS scale. We tested the H_5 hypothesis based on the parameter value of β:

H_0: $\beta=0$

H_{alt}: $\beta>0$

H_5 can be accepted if H_0 is rejected and H_{alt} is accepted. (Results: p113)

We tested the H_6 hypothesis based on the parameter value of γ:

H_0: γ=0

H_{alt}: γ>0

H_6 can be accepted if H_0 is rejected and H_{alt} is accepted. (Results: p113)

V.3. Results

V.3.1. Descriptive statistics

From the 200 respondents in case of 32 we did not have data about self-reported health or acceptable health data for the respondents' age group, therefore they were excluded from the analysis. We conducted analyses on a sample of 168. The key characteristics of the sample are summarized below. Mean age was 45.67 years (SD: 15.99, range: 25-93). In the 25-34, 35-44, 45-54, and 65+ age groups were 57 (33.9%), 29 (17.3%), 15 (9.5%) and 24 (14.3%) respondents, respectively. 57.1% of respondents were female, 64.3% were married and 78% was employed, 63% had tertiary education. 77.4% of the respondents came from households in the top two quintiles in terms of net income. Compared to the Hungarian normal population, the sample was younger, and was shifted towards higher education and higher income (KSH, 2011).

Respondents rated on average 1.95 points (SD=13.732) lower their own health compared to the health they considered acceptable for their age. 88 respondents (52.4%) rated their own health as better, while 80 (47.6%) as worse than the acceptable level. 68.4% of the 24-35 year old age group belonged to the not acceptable group, while this decreased to 31% in the 45-54 year old group. In the 55-64 year old group the percentage of not acceptable health was 50%, and it decreased to 29.2% among the 65+ group.

58.9% of respondents were overweight, 20.8% were smokers, 10.1% were high risk drinkers and 58.9% did not exercise regularly. Their own health was rated on the EQ VAS scale 80.51 points from 100 (SD=14.514). Mean happiness was 7.26 points (SD=2.089). Respondents estimated their own expected life-span to 82 years (SD=10.383). The highest estimated life span was 120, the lowest 57 years. The life span of close relatives was shorter than 75 years in 35.7% of the cases. 32.1% of respondents were an informal caregiver during the past 10 years, and 58.3% utilised some health care service.

V.3.2. The association between relative health and happiness

The coefficients and main statistical properties of the models are summarized in Table 10. In a simple regression model relative health showed significant association with happiness (M1), which remained significant after controlling for age and gender (M2). The effect of socio-demographic variables was not significant, and they improved the explanatory power of the model only to a small extent (M3). Adding self-rated health (EQ Vas) increased substantially the explanatory power of the model (M4, R^2=0.318), by adding to the model, the coefficients of relative health and age became non-significant. From EQ-5D-3L dimensions problems in usual activities and anxiety / depression showed significant relationship with happiness (M5). Other variables related to health or lifestyle were not significant (M6), but two variables became significant in this model: pain showed mild positive, while high income mild negative association with the level of happiness. These two effects were contrary to our expectations. After removing non-significant variables, the effect of pain became non-significant (M7). We tested the quality of M7 in detail. The scatterplot of the residuals and predicted values showed asymmetric distribution and outliers. Based on Cook's D (>4/168=0,024), a leverage (>14/168=0,083), standardised residuals (absolute value >3), or standardised DfBeta (>2/√168=0,154) criteria and happiness values of 0, we found 29 (17.3%) outliers. After removing them from the model, the parameters remained practically unchanged.

Table 10 Linear regression analysis of the relationship of happiness and AH

	M1	M2	M3	M4	M5	M6	M7
Relative health	0.028*	0.044**	0.046**	0.007	0.014	0.012	0.012
Age (y)		-0.042**	-0.040**	-0.011	-0.014	-0.016	-0.013
Female		0.286	0.215	0.230	0.172	0.079	
Married			0.274	0.415	0.450	0.462	
Employed			-0.799	-0.319	-0.139	-0.142	
Has diploma			-0.152	-0.04	0.115	0.023	
High income			-0.494	-0.708	-1.173	-0.912*	-1.025**
EQ VAS (0-100)				0.073**	0.045**	0.040**	0.044**
Mobility					-0.684	-0.661	
Self-care					-0.839	-0.825	
Usual activities					-1.427**	-1.424**	-1.396**
Pain / discomfort					0.890	0.931*	
Anxiety / depression					-1.584**	-1.599**	-1.326**
Overweight						-0.409	
Smoking						-0.023	
High risk alcohol						-0.615	
Lack of exercise						-0.344	
Informal caregiver						-0.327	
Used health service						-0.049	
Intercept	7.311**	9.107**	9.566**	2.139	5.356**	6.036**	5.740**
F	5.808	8.080	4.353	9.136	9.282	6.813	18.130
p	0.017	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
R^2	0.034	0.129	0.162	0.318	0.443	0.470	0.406
Adjusted R^2	0.028	0.113	0.125	0.283	0.395	0.401	0.384

* p<0.05, ** p<0.01

The Breusch-Pagan test ($\chi^2_{(df=1)}$ = 7,53, p=0,0061) was significant, which suggested that the model was heteroscedastic. The Ramsey RESET test was also significant ($F_{(3,156)}$=3,22, p=0,0244), which suggested model specification error, therefore the results were not reliable. The residuals showed non-normal distribution based on the Shapiro-Wilk test (p=0.00092) and skewness-kurtosis test ($p_{skewness}$=0.1398, $p_{kurtosis}$=0.0012, joint test: $\chi^2_{(df=2)}$=11.01, p=0.004).

Altogether, the diagnostics of M7 raised questions in several points about the model results, therefore we did not accept the model, and further examined the effect of relative health and age on the relationship of self-rated health and happiness, or in other words, whether being above or below the acceptability threshold in different ages modifies the relationship between happiness and self-rated health?

The next steps of the analysis are summarized in Table 11. Starting from M7, we changed relative health to the binary acceptable health variable, and for the more straightforward interpretation of results, we centred age to 18 years (M8). In the next step we examined the interaction of age and self-rated heatlh, which was not significant (M9). Then we examined the interaction between AH and self-rated health. The explanatory power of the model improved, and the interaction became significant, and the coefficients of self-rated health and AH have changed (M10). Interpretation of M10: if an individual has acceptable health status, then self-rated health does not influence the level of happiness. Based on the coefficients of the model, we could accept H_4, since the δ parameter denoting the interaction of subjective health and acceptable health in explaining happiness was significant and positive ($p<0,05$). However, after excluding outliers based on Cook's D, leverage, dfbeta, and standardized residuals and extremes of happiness, the coefficients did not remain significant. The results were not significant even after excluding 3 cases with the highest leverage values. *Therefore, H_4 could not be accepted with certainty.*

The next question was whether in not acceptable health states, the relationship between happiness and self-rated health depends on age?

Table 11 The relationship between happiness, health and AH

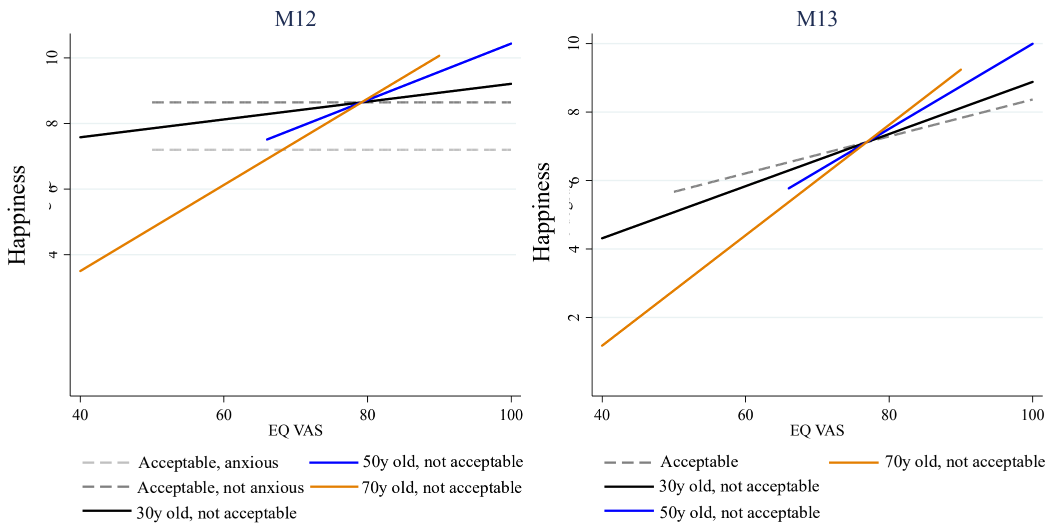
	M8	M9	M10	M11	M12	M13
Not acceptable	0.033	0.026	-3.969*			
Not acceptable * Age				-0.124**	-0.179**	-0.142
Age (centred to 18 y)	-0.007	-0.23	-0.066	-0.010		
EQ VAS (0-100)	0.051**	0.044**	-0.010	0.028		0.054**
Age*EQ VAS		0.000	0.001	0.000		
Not acceptable*EQ VAS			0.049*			
Not acceptable* EQ VAS * Ager				0.002**	0.002**	0.002**
High income	-1.030**	-1.018	-0.953**	-0.948**	-0.789*	
Usual activities	-1.372**	-1.335	-1.279*	-1.264*	-1.716**	
Anxiety / depression	-1.299**	-1.296	-1.271**	-1.273**	-1.446**	
Intercept	4.725**	5.210**	8.698**	6.482**	8.644**	2.973**
F	17.803	15.196	14.63	15.146	22.156	27.487
p	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
R^2	0.402	0.402	0.427	0.436	0.409	0.336
Adjusted R^2	0.379	0.376	0.398	0.407	0.391	0.324

* p<0.05, ** p<0.01

Therefore, we added to AH and the interaction of AH also their interaction with age in the model (M11). The interaction was significant, and the explanatory power of the model improved (R^2=0.436). The interpretation of M11 is the following: in acceptable health self-rated health does not influence happiness, in not acceptable health states in case of young individuals self-rated health influences only to small extent, in older individuals to a greater extent the level of happiness. Shapley decomposition of the R^2 of M11 suggested that AH-related interactions contributed to the explanatory power of the model by approximately 20%. M12 contained only the significant variables of M11. The explanatory power decreased only to a small extent (R^2=0,409). Because usual activities and anxiety / depression are health-related problems, and income also has a strong influence on health (see Table 11 below), we replaced these variables with self-rated health in M13, which became significant. The interpretation of M13: in acceptable health the improvement of health affects happiness only to a moderate extent, while in not acceptable health the effect of self-rated health on happiness is the greater the older the individual is.

The explanatory power of M13 (R^2=0.336) was inferior compared to M12 (R^2=0.409), which was compensated only to a small extent by the simplicity of the model. (M13 adjusted R^2=0,324, M12 adjusted R^2=0,391). Comparing with the World Happiness Report analysis focusing on the individual determinants of happiness (adjusted R^2=0,237) (Helliwell et al., 2018), the explanatory power of models M12 and M13 containing AH as an explanatory variable were good, M12 explained approximately 40%, while M13 explained approximately 1/3 of the variance of happiness. The model was homoscedastic (Breusch-Pagan test: $\chi^2_{(df=2)}$=1,73, p=0,189), its specification was acceptable (Ramsey-RESET test: $F_{(3,160)}$=0,34, p=0,7971), the distribution of standard residuals showed a mild skewness (Shapiro-Wilk test p=0,0011, skewness-kurtosis test: $p_{skewness}$=0,0196, $p_{kurtosis}$=0,0943, p_{joint}=0,0224). However, after the exclusion of outliers based on Cook's D, leverage, DfBeta, standardised residuals and extreme values of happiness, only the effect of health remained significant. The effect of acceptable health did not remain significant even after the exclusion of the 3 observations with the highest leverage values. After the exclusion of the three outliers, according to the Shapley-decomposition, the contribution of AH to the explanatory power of M11 decreased to 7%. Since we could exclude coding error during the data recording (electronic questionnaire), we need to accept data from the outliers as valid responses. Therefore, we can assume that there are subgroups of respondents whose attitudes regarding health and happiness are rather heterogenous, which needs further exploration. Figure 29 illustrates the relationship between health and happiness based on the parameters of M12 and M13. We depicted the relationships in the EQ VAS ranges that were measured in the sample for the given subgroup.

Figure 29 The relationship of AH and happiness



Left figure: based on the M12 parameters (Table 11) the graphical representation of the relationship between happiness and self-rated health (EQ VAS). In acceptable health (grey dashed line) the EQ VAS does not influence the individual's happiness; however, anxiety/depression problems decrease it. In not acceptable health (coloured lines), the relationship between self-rated health and happiness is age-dependent. The older the individual, the steeper the lines are – the greater extent health influences the level of happiness. Right figure: based on the M13 parameters (Table 11) the graph shows the relationship between EQ VAS and happiness. The model is not controlled for specific health problems (anxiety/depression). In acceptable health (grey dashed line) health influences happiness to a smaller extent, while in not acceptable health (coloured lines) self-rated health influences happiness, to a greater extent in older ages.

As a supplementary analysis we explored the relationship between income and the dimensions of health, since health related variables changed the magnitude and significance of the effect of income on happiness (M5). Table 12 summarizes the occurrence of health problems in a split according to income status. Among individuals in the lower income group every health problem (including usual activities and anxiety / depression) occurred more frequently.

Table 12 Occurrence of health problems in different income groups

		Mobility	Self-care	Usual activities	Pain / discomfort	Anxiety / depression	Any problems
High income	N (%)	16 (12.3%)	0 (0.0%)	6 (4.6%)	29 (22.3%)	33 (25.4%)	44 (33.8%)
	Expected N	24.9	1.6	13.2	38.9	40.5	54.4
Low income	N (%)	16 (43.2%)	2 (5.4%)	11 (29.7%)	21 (56.8%)	19 (51.4%)	26 (70.3%)
	Expected N	7.1	0.4	3.8	11.1	11.5	15.5
Fischer's exact p		<0.001	0.048	<0.001	<0.001	0.004	<0.001
$Chi^2_{(1)}$ p		<0.001	0.008	<0.001	<0.001	0.003	<0.001

N=167, the table displays data from the cross-tabulation only for those, who indicated any problems in the dimensions

We conducted Chi^2 and Fischer's exact tests on the cross-tabulation of data. The EQ VAS scores (skewness: -1.13, kurtosis: 2.0 Shapiro-Wilk test, $p<0.001$) were on average 8.9 points higher compared to individuals in the low-income group. The difference was significant. (Independent samples' t-test: $p=0.001$ (equal variances: Levene-test $p=0.220$); Mann-Whitney U test: $p=0.001$). Interpretation of the parameters: the happiness of people in the high- and low-income groups does not differ significantly. This is explained by the greater frequency of health problems among the low-income group, which negatively influence happiness. In case of similar level of problems in usual activities or anxiety / depression, people in the low-income group rated their happiness on average 1.025 points higher than the high-income group. It is important to note that in a cross-tabulation, problems in the anxiety / depression ($chi^2_{(1)}=0.343$, $p=0.558$) and usual activities ($chi^2_{(1)}=0.215$, $p=0.643$) dimensions did not show association with AH.

V.3.3. The relationship between relative health and lifestyle

The results of the logistic regression analysis (M14) are summarized in Table 13. The acceptability of the health status of respondents was influenced significantly by age, self-rated health, overweight and smoking status. Acceptable health status was denoted by the value of 1, while not acceptable health status was denoted by the value of 0. The interpretation of results is the following: *the older individuals were or the* *better their self-rated health was, the greater the chance was that they considered their own health state acceptable. Based on this result we could accept H_5 and H_6.*

Table 13 Factors influencing the probability of acceptable health

Logistic regression	M14
Age (y)	.083**
EQ VAS	.054*
Female	.393
Married	-.139
Employed	-.024
Has diploma	-.510
High income	-.065
Overweight	-1,491**
Smoker	-1,311*
High risk alcohol	.680
Lack of exercise	-.546
Expected life-span	-.027
Close relatives' life span	.052
Informal caregiver	-.517
Healthcare use	-.703
Mobility problems	.538
Self-care problems	.236
Problems with usual activities	.553
Problems with pain / discomfort	-.817
Anxiety / depression	.359
Intercept	2.107*
LR $Chi^2_{(20)}$	62.91
p	<0.001

* p<0.05, ** p<0.01

Holding all other parameters constant, smokers and overweight respondents had greater probability to evaluate their health status as not acceptable. The Hosmer-Lemeshow test after splitting data to 10 deciles (($\chi^2_{(df=8)}$)=4.24, p=0.8346) was not significant, which suggests good model fit. The area under the ROC curve was 0.8377, so the discriminative power of the model was good. The logistic regression model provided accurate prediction in 75.6% of the cases, when using p=0.5 threshold value. The maximum of the Youden-index was at p=0.509. After setting the optimal threshold, the proportion of accurate predictions improved only a small amount to 76.2%.

V.4. Discussion

Acceptable health is an unexplored research field. This was the first study that applied EQ VAS for the measurement of AH. The first question was whether relative health is associated with the individual's happiness? According to our analysis, happiness was determined by the self-rated health of individuals. Relative health influenced the association between health and happiness in an age-dependent manner. In case individuals considered their health better compared to the level considered acceptable in their age, health contributed to a smaller extent to their happiness, while in not acceptable health states health contributed happiness in increasingly with age. Based on these findings *we could accept our H4, but the results were significantly affected by a small number of outliers in the sample, therefore it requires further analysis and repeated measurement in order to ascertain the validity of the finding described above.*

These results can support the positive thesis of the sufficientarian principle applied in health financing decisions, which aims to maximise the number of individuals in acceptable health. The maximisation of acceptable health corresponds with the maximisation of welfare, since individuals' wellbeing is related to their health status unless they are not in acceptable health status (Wouters et al., 2017). According to our results, we need to pay more attention on the treatment of severe diseases in the older age group, and to the treatment of mental healht problems in anyage group and the restoration of usual activities, which influence the happiness, and thereby the subjective wellbeing of the individual to a great extent.

According to our results, health influenced happiness differently than financial wealth. Through the status effect, the financial status relative to others, and not the absolute wealth is the determinant of happiness (Alderson and Katz-Gerro, 2016). However, Hungarian data have not supported this relationship, status effect and the hope generated by others financial growht as a positive information effect have cancelled out each others effects (Hajdu and Hajdu, 2011). In our own sample we demonstrated the the joint effect of self-rated health and the acceptable health on happiness.

However, our data were collected in a cross-sectional study, therefore we could not monitor the changes within individuals, only the differnces between individuals. We need to note that the level of AH showed great variability among individuals.

The second part of our analysis was concerned with the relationship between AH and the lifestyle from the perspective of individual respondents. Higher age and better self-rated health increased the chance that individuals rate their own health better than what they consider acceptable for their age. Based on these findings, we can accept our H_5 and H_6 hypotheses. Two important risk groups: overweight people and smokers indicated their health acceptable with a smaller probability than others, while problems in anxiety / depression and usual activities – which have strong impact on the happiness of individuals – did not influence the acceptability of health states.

Implementing AH-based priority-setting would bring the following effects from the individuals' perspective:

a) According to the AH principle, the treatment of older individuals would receive lower priority. Since older individuals are more likely to consider their health acceptable, therefore their wellbeing would be affected to a lesser amount if their treatment would receive smaller priority. However, this is true only for mild health problems. The treatment of severe health problems contributes to the wellbeing of older individuals to a great extent.
b) According to the AH principle, the treatment of severe conditions would receive higher priority. With the improvement of health, people are more likely to get into acceptable health states, therefore their treatment would influence their wellbeing favourably.
c) Overweight people and smokers consider their health less acceptable than other in case of a similar health condition. According to the AH principle, similar treatment priorities would contribute to the welfare of those individuals, who can be partly held responsible for their health deterioration. This is also in harmony with certain societal priorities (Dolan et al., 2005, Gulácsi, 2012).
d) However, mental problems would remain hidden from the eyes of decision-makers: problems in the anxiety / depression dimension influence greatly the

happiness of individuals, while they are not associated with acceptable health from either the societal or the individuals' perspectives.

The propositions above are partly speculative, because important aspects of the evaluation of AH's societal and individual consequences are not known. However, they highlight the importance, that acceptability should be rendered to the elements of the E-matrix from two directions:

a) The societal perspective shows, to what extent people consider some health states acceptable
a) The individual perspective shows, to what extent individuals consider their own health acceptable

The question is what the relationship is between the two perspectives, and how the application of AH would influence the overall wellbeing of the entire society. AH represents two societal priorities in the allocation of resources: younger individuals should receive priority over older ones and sever conditions should receive priority over moderate ones. Application of the AH principle can lead to the maximisation of the wellbeing of the society, if the maximisation of acceptable health states according to societal priority weights corresponds to the maximisation of acceptable health states that are experienced by individuals.

A key limitation of our research is that we measured acceptable health indirectly, we did ask respondents to evaluate the acceptability of their own health status directly. Furthermore, our cross-sectional sample did not show the fundamental intra-individual relationships between AH and wellbeing, which is a key aspect of health technology assessment studies. Therefore, the associations described by our study have limited value in drawing conclusions about causal relationships. Furthermore, we examined the relationship of AH and wellbeing through the 0-10 numeral happiness scale, while some theories view wellbeing as a different broader concept than mere happiness (Ryff and Keyes, 1995). It is important to note, that sampling was not representative in our research therefore our results are not generalisable, and a small number of outliers influenced our results to a great extent. Therefore we suggest repeating our research in a greater representative sample.

VI. CONCLUSIONS

In search of potential solutions for the sustainability challenges of healthcare financing, my PhD thesis summarizes the research about acceptable health - a new health measurement method reflecting societal priorities concerning age and disease severity - and its association with happiness.

Based on the cross-sectional, non-representative study conducted in the Hungarian population, the following conclusions could be drawn:

a) People consider health problems increasingly acceptable with age
b) Mild health problems are more acceptable than more severe ones
c) AH can be measured more precisely using the joint evaluation method, than with separate evaluation. Joint evaluation is feasible in practice and provides reasonably accurate measures about the acceptability of health states.
d) EQ VAS is a convenient method to measure acceptable health
e) Acceptable health from the individual's perspective reflects similar priorities than from the society's perspective.

Before the adoption of the AH principle in real practice, the reliability of its method needs to be further developed. During joint evaluation we did not evaluate the consistency of responses, therefore the adaptive algorithm needs to be amended with self-control questions. Without control questions, random answers and answers reflecting real preferences cannot be separated.

It is also important to clarify the concept of AH during the evaluation exercise in order to decrease potential errors arising from respondents' misunderstanding or misinterpretation of the evaluation task.

Furthermore, it is important to prove the basic assumption that in acceptable health status a unit health gain has lower utility (or smaller increase of wellbeing) than in not acceptable health status. Our findings concerning the association of AH and happiness are in line with this assumption. However, in our research acceptable health states were measured indirectly, and our results were influenced by a small number of outliers, therefore our measurement needs to be repeated on a sufficiently powered representative sample of the general population, by the direct evaluation of the acceptability of individuals' health status, as well as using more specific validated measures of health-related utility and subjective well-being.

Finally, in addition to determining the *E-matrix* from the society's perspective on AH, we consider important to measure the E-matrix from the individuals' perspective as well, in order to verify the assumption that maximising health states that are considered acceptable from the society's perspective simultaneously result in the maximisation of individual wellbeing.

VII. REFERENCES

AL-JANABI, H., FLYNN, T. N. & COAST, J. 2011. Estimation of a preference-based carer experience scale. *Med Decis Making,* 31**,** 458-68. https://doi.org/10.1177/0272989X10381280

AL-JANABI, H., FLYNN, T. N. & COAST, J. 2012. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Qual Life Res*, 21, 167-76. https://doi.org/ 10.1007/s11136-011-9927-2

ALDERSON, A. S. & KATZ-GERRO, T. 2016. Compared to Whom? Inequality, Social Comparison, and Happiness in the United States. *Social Forces*, 95, 25-54. https://doi.org/10.1093/sf/sow042

AUGUSTOVSKI, F., REY-ARES, L., IRAZOLA, V., OPPE, M. & DEVLIN, N. J. 2013. Lead versus lag-time trade-off variants: does it make any difference? *Eur J Health Econ*, 14 Suppl 1, S25-31. https://doi.org/10.1007/s10198-013-0505-0

BAJI, P. 2012. A diszkrét választás módszere. *Statisztikai Szemle*, 90, 943-963.

BAJI, P., GOLICKI, D., PREVOLNIK-RUPEL, V., BROUWER, W. B. F., ZRUBKA, Z., GULÁCSI, L. & PÉNTEK, M. 2019. The burden of informal caregiving in Hungary, Poland and Slovenia: results from national representative surveys. *The European Journal of Health Economics*. https://doi.org/10.1007/s10198-019-01058-x

BALTUSSEN, R. & NIESSEN, L. 2006. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Eff Resour Alloc*, 4, 14. https://doi.org/10.1186/1478-7547-4-14

BANSBACK, N., BRAZIER, J., TSUCHIYA, A. & ANIS, A. 2012. Using a discrete choice experiment to estimate health state utility values. *J Health Econ*, 31, 306-18. https://doi.org/10.1016/j.jhealeco.2011.11.004

BELTON, V. & STEWART, T. 2002. *Multiple Criteria Decision Analysis: An Integrative Approach*, Massachusetts, Kluwer Academic Publishers.

BERNERT, S., FERNANDEZ, A., HARO, J. M., KONIG, H. H., ALONSO, J., VILAGUT, G., SEVILLA-DEDIEU, C., DE GRAAF, R., MATSCHINGER, H., HEIDER, D., ANGERMEYER, M. C. & INVESTIGATORS, E. S. M. 2009. Comparison of different valuation methods for population health status measured by the EQ-5D in three European countries. *Value Health,* 12**,** 750-8. https://doi.org/ 10.1111/j.1524-4733.2009.00509.x

BLEICHRODT, H. & JOHANNESSON, M. 1997. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *J Health Econ,* 16**,** 155-75. https://doi.org/10.1016/S0167-6296(96)00509-7

BRAZIER, J., ROWEN, D., KARIMI, M., PEASGOOD, T., TSUCHIYA, A. & RATCLIFFE, J. 2018. Experience-based utility and own health state valuation for a health state classification system: why and how to do it. *Eur J Health Econ*, 19, 881-891. https://doi.org/ 10.1007/s10198-017-0931-5

BRAZIER, J. E., ROWEN, D., LLOYD, A. & KARIMI, M. 2019. Future Directions in Valuing Benefits for Estimating QALYs: Is Time Up for the EQ-5D? *Value Health*, 22, 62-68. https://doi.org/10.1016/j.jval.2018.12.001

BRODSZKY, V., BAJI, P., BALOGH, O. & PENTEK, M. 2014. Budget impact analysis of biosimilar infliximab (CT-P13) for the treatment of rheumatoid arthritis in six Central and Eastern European countries. *Eur J Health Econ,* 15 Suppl 1**,** S65-71. https://doi.org/ 10.1007/s10198-014-0595-3

BROOKS, R., RABIN, R. & DE CHARRO, F. 2003. *The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective*, Kluwer Academic Publishers.

BROUWER, W. B., VAN EXEL, N. J. & STOLK, E. A. 2005. Acceptability of less than perfect health states. *Soc Sci Med,* 60**,** 237-46. https://doi.org/10.1016/j.socscimed.2004.04.032

BROUWER, W. B., VAN EXEL, N. J., VAN GORP, B. & REDEKOP, W. K. 2006. The CarerQol instrument: a new instrument to measure care-related quality of life of informal caregivers for use in economic evaluations. *Qual Life Res,* 15**,** 1005-21. https://doi.org/10.1007/s11136-005-5994-6

CASAL, P. 2007. Why sufficiency is not enough. *Ethics,* 117**,** 296-326. https://doi.org/10.1086/510692

CRISP, R. 2002. Treatment According to Need: Justice and the British National Health Service. *In:* RHODES, R., BATTIN, M. P. & SILVERS, A. (eds.) *Medicine and Social Justice.* New York: Oxford University Press.

CULYER, A. J. & WAGSTAFF, A. 1993. Equity and equality in health and health care. J Health Econ, 12, 431-57. https://doi.org/10.1016/0167-6296(93)90004-X

D'ASPREMONT, C. & GEVERS, L. 2002. Social welfare functionals and interpersonal comparability. *In:* ARROW, K., SEN, A. & SUZUMURA, K. (eds.) *Handbook of Social Choice and Welfare.* Elsevier.

DANIELS, N. 2000. Accountability for reasonableness - Establishing a fair process for priority setting is easier than agreeing on principles. *British Medical Journal,* 321**,** 1300-1301. https://doi.org/ 10.1136/bmj.321.7272.1300

DANIELS, N. & SABIN, J. 1997. Limits to health care: fair procedures, democratic deliberation, and the legitimacy problem for insurers. *Philos Public Aff,* 26**,** 303-50. https://doi.org/10.1111/j.1088-4963.1997.tb00082.x

DANIELS, N. & SABIN, J. E. 2008. Accountability for reasonableness: an update. *BMJ,* 337**,** a1850. https://doi.org/ 10.1136/bmj.a1850

DANKÓ, D. & MOLNÁR, M. 2013. *Gyógyszertámogatás,* Budapest, Medicina Könyvkiadó Zrt.

DE LA MAISONNEUVE, C. & MARTINS, J. O. 2013. Public spending on health and long-term care: a new set of projections. *OECD Economic Policy Papers.* OECD.

DOLAN, P. 1997. Modeling valuations for EuroQol health states. *Med Care,* 35**,** 1095-108. https://doi.org/ 10.1097/00005650-199711000-00002

DOLAN, P., SHAW, R., TSUCHIYA, A. & WILLIAMS, A. 2005. QALY maximisation and people's preferences: a methodological review of the literature. *Health Econ,* 14**,** 197-208. https://doi.org/ 10.1002/hec.924

EGA, E. G. M. A. 2016. *In:* ASSOCIATION, E. G. M. (ed.). Brussels: European Generic Medicines Association.

EMEA, E. M. A. 2018. Binocrit: EPAR - Product Information. *06/12/2018 Binocrit - EMEA/H/C/000725 - WS/1507.* European Medicines Agency.

EMEA, E. M. A. 2019. Remsima: EPAR - Product information. *In:* AGENCY, E. M. (ed.) *04/12/2018 Remsima - EMEA/H/C/002576 - IB/0061.*

EUNETHTA. *HTA Core Model* [Online]. Available: https://www.eunethta.eu/hta-core-model/ [Accessed 2019.02.11.].

EUNETHTA 2007. EUnetHTA comments on the Discussion Document: "Health in Europe: A Strategic Approach".

EUROPEAN COMMISSION 2014. Identifying Fiscal Sustainability Challenges in the Areas of Pension, Health Care and Long-term Care Policies. *Occassional Papers* Brussels.

EUROPEAN MEDICINES AGENCY 2017. Biosimilars in the EU - Information guide for healthcare professionals. *In:* AGENCY, E. M. (ed.). London: European Medicines Agency.

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

EVALUATEPHARMA 2017. World Preview 2017, Outlook to 2022. London.

FDA. 2018. *Regulating Biological Products* [Online]. Available: https://www.fda.gov/ForConsumers/ConsumerUpdates/ucm048341.htm [Accessed 2018.02.21.].

FENG, Y., PARKIN, D. & DEVLIN, N. J. 2014. Assessing the performance of the EQ-VAS in the NHS PROMs programme. *Qual Life Res,* 23**,** 977-89. https://doi.org/10.1007/s11136-013-0537-z

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 Econ,* 24**,** 258-69. https://doi.org/10.1002/hec.3014

GABI ONLINE. 2012. *WHO definitions of generics* [Online]. Generics and Biosimilars Initiative (GaBI). Available: http://gabionline.net/Generics/General/WHO-definitions-of-generics [Accessed 2019.02.23.].

GARROW, J. 1981. *Treat obesity seriously: a clinical manual,* Edinburgh, Churchill Livingstone.

GOSSERIES, A. 2011. Sufficientarianism. *Routledge Encyclopedia of Philosophy.* Tailor and Francis. https://doi.org/ 10.4324/9780415249126-S112-1

GULACSI, L. 2014. Biological and biosimilar therapies in inflammatory conditions: challenges for the Central and Eastern European countries. *Eur J Health Econ,* 15 Suppl 1**,** S1-4. https://doi.org/ 10.1007/s10198-014-0588-2

GULÁCSI, L. 2012. *Egészség-gazdaságtan és technológiaelemzés,* Budapest, Medicina Könyvkiadó Zrt.

GULACSI, L., ZRUBKA, Z., BRODSZKY, V., RENCZ, F., ALTEN, R., SZEKANECZ, Z. & PENTEK, M. 2019. Long-Term Efficacy of Tumor Necrosis Factor Inhibitors for the Treatment of Methotrexate-Naive Rheumatoid Arthritis: Systematic Literature Review and Meta-Analysis. *Adv Ther,* 36**,** 721-745. https://doi.org/10.1007/s12325-018-0869-8

HAJDU, T. & HAJDU, G. 2011. A hasznosság és a relatív jövedelem kapcsolatának vizsgálata magyar adatok segítségével. *Közgazdasági Szemle,* LVIII**,** 56-73.

HAUBER, A. B., GONZALEZ, J. M., GROOTHUIS-OUDSHOORN, C. G., PRIOR, T., MARSHALL, D. A., CUNNINGHAM, C., MJ, I. J. & BRIDGES, J. F. 2016. Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Health,* 19**,** 300-15. https://doi.org/10.1016/j.jval.2016.04.004

HELLIWELL, J., LAYARD, R. & SACHS, J. 2018. Statistical Appendix 2 for Chapter 2. *World Happiness Report 2018.* New York: Sustainable Development Solutions Network.

HERDMAN, M., GUDEX, C., LLOYD, A., JANSSEN, M., KIND, P., PARKIN, D., BONSEL, G. & BADIA, X. 2011. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res,* 20**,** 1727-36. https://doi.org/10.1007/s11136-011-9903-x

HOEFMAN, R. J., VAN EXEL, J. & BROUWER, W. B. F. 2017. Measuring Care-Related Quality of Life of Caregivers for Use in Economic Evaluations: CarerQol Tariffs for Australia, Germany, Sweden, UK, and US. *Pharmacoeconomics,* 35**,** 469-478. https://doi.org/10.1007/s40273-016-0477-x

HOPE, T., REYNOLDS, J. & GRIFFITHS, S. 2002. Rationing Decisions: Integrating Cost-Effectiveness with Other Values. *In:* RHODES, R., BATTIN, M. P. & SILVERS, A. (eds.) *Medicine and Social Justice.* New York: Oxford University Press.

HSEE, C. K., LOEWENSTEIN, G. F., BLOUNT, S. & BAZERMAN, M. H. 1999. Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin,* 125**,** 576-590. https://doi.org/10.1037/0033-2909.125.5.576

KAHNEMANN, D. & TVERSKY, A. 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica,* 47**,** 263-291.

KEENEY, R. & RAIFFA, H. 1993. *Decisions with Multiple Objectives: Preferences and Value Tade-Offs. ,* Cambridge, UK, Cambridge University Press.

KIM, H.-J., KIM, Y.-J., PARK, D.-J., LIEW, D. & RHEE, Y. 2017. Multiple Criteria Decision Analysis (MCDA) in Health Technology Assessment: Review of Literature on MCDA Methodology and Decision Criteria. *J Health Tech Assess,* 5**,** 128-142. https://doi.org/10.1016/j.socscimed.2017.06.024

KSH, H. C. S. O. 2011. Population Census 2011.

KSH, H. C. S. O. 2014. ELEF - European Health Interview Survey 2014.

KSH, H. C. S. O. 2015. Time series of annual data - Income and living conditions.

LORENZONI, L., MURTIN, F., SPRINGARE, L.-S., AURAAEN, A. & DANIEL, F. 2018. Which policies increase value for money in healthcare? . *OECD Health Working Papers.* Paris.

MARSH, K., M, I. J., THOKALA, P., BALTUSSEN, R., BOYSEN, M., KALO, Z., LONNGREN, T., MUSSEN, F., PEACOCK, S., WATKINS, J., DEVLIN, N. & FORCE, I. T. 2016. Multiple Criteria Decision Analysis for Health Care Decision Making--Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health,* 19**,** 125-37. https://doi.org/10.1016/j.jval.2015.12.003

MASKA, L., ANDERSON, J. & MICHAUD, K. 2011. Measures of functional status and quality of life in rheumatoid arthritis: Health Assessment Questionnaire Disability Index (HAQ), Modified Health Assessment Questionnaire (MHAQ), Multidimensional Health Assessment Questionnaire (MDHAQ), Health Assessment Questionnaire II (HAQ-II), Improved Health Assessment Questionnaire (Improved HAQ), and Rheumatoid Arthritis Quality of Life (RAQoL). *Arthritis Care Res (Hoboken),* 63 Suppl 11**,** S4-13. https://doi.org/10.1002/acr.20620

MATEJKA, J., GLUECK, M., GROSSMAN, T. & FITZMAURICE, G. 2016. The Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales. *34th Annual Chi Conference on Human Factors in Computing Systems, Chi 2016***,** 5421-5432.

MOORKENS, E., VULTO, A. G., HUYS, I., DYLST, P., GODMAN, B., KEUERLEBER, S., CLAUS, B., DIMITROVA, M., PETROVA, G., SOVIC-BRKICIC, L., SLABY, J., SEBESTA, R., LAIUS, O., KARR, A., BECK, M., MARTIKAINEN, J. E., SELKE, G. W., SPILLANE, S., MCCULLAGH, L., TRIFIRO, G., VELLA BONANNO, P., MACK, A., FOGELE, A., VIKSNA, A., WLADYSIUK, M., MOTA-FILIPE, H., MESHKOV, D., KALABA, M., MENCEJ BEDRAC, S., FURST, J., ZARA, C., SKIOLD, P., MAGNUSSON, E. & SIMOENS, S. 2017. Policies for biosimilar uptake in Europe: An overview. *PLoS One,* 12**,** e0190147. https://doi.org/10.1371/journal.pone.0190147

MULHERN, B., TSUCHIYA, A., DEVLIN, N., BUCKINGHAM, K., ROWEN, D. & BRAZIER, J. 2013. A comparison of three binary choice methods for health state valuation. *Health Economics and Decision Science (HEDS) Discussion Paper.*

MULLARD, A. 2019. 2018 FDA drug approvals. *Nat Rev Drug Discov,* 18**,** 85-89. https://doi.org/10.1038/d41573-019-00014-x

NAGY, G. & VAN VOLLENHOVEN, R. F. 2015. Sustained biologic-free and drug-free remission in rheumatoid arthritis, where are we now? *Arthritis Res Ther,* 17**,** 181. https://doi.org/10.1186/s13075-015-0707-1

NIAAA, N. I. O. A. A. A. A. *Alcohol facts and Statistics* [Online]. Available: https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/alcohol-facts-and-statistics [Accessed 2017.10.01.].

NORD, E., PINTO, J. L., RICHARDSON, J., MENZEL, P. & UBEL, P. 1999. Incorporating societal concerns for fairness in numerical valuations of health programmes. *Health Econ,* 8**,** 25-39. https://doi.org/ 10.1002/(SICI)1099-1050(199902)8:1<25::AID-HEC398>3.0.CO;2-H

OECD 2015a. *Fiscal Sustainability of Health Systems: Bridging Health and Finance Perspectives,* Paris, OECD Publishing.

OECD 2015b. Pharmaceutical spending trends and future challenges. *Health at a Glance 2015, OECD Indicators.* Paris: OECD.

OECD 2017. Health at a glance 2017. *OECD Indicators.* Paris.

OECD 2018. Spending on Health: Latest Trends.

OLSEN, J. A. 2000. A note on eliciting distributive preferences for health. *J Health Econ,* 19**,** 541-50. https://doi.org/10.1016/S0167-6296(00)00035-7

OPPE, M. & VAN REENEN, M. 2015. EQ-5D-3L User Guide (Version 5.1). Rotterdam: EuroQol Research Foundation.

PARFIT, D. 1995. *Equality or priority?* The Lindley Lecture. Kansas, Department of Pilosophy, University of Kansas.

PARKIN, D. & DEVLIN, N. 2006. Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Econ,* 15**,** 653-64. https://doi.org/10.1002/hec.1086

PENTEK, M. 2007. *Rheumatoid arthritises betegek egészségi állapota és betegségterhe Magyarországon.* PhD, Semmelweis Egyetem.

PÉNTEK, M., BRODSZKY, V., GULÁCSI, L., ÉRSEK, K., EXEL, N. J. A., BROUWER, W. B. F. & GULÁCSI, L. 2009. Php82 Assessment of Lifespan and Health Status Expectations in Hungary; Results of a Web-Survey. *Value in Health,* 12**,** A252. https://doi.org/10.1016/S1098-3015(10)74234-0

PENTEK, M., POOR, G., WILAND, P., OLEJAROVA, M., BRZOSKO, M., CODREANU, C., BRODSZKY, N. & GULACSI, L. 2014a. Biological therapy in inflammatory rheumatic diseases: issues in Central and Eastern European countries. *Eur J Health Econ,* 15 Suppl 1**,** S35-43. https://doi.org/10.1007/s10198-014-0592-6

PENTEK, M., ROJKOVICH, B., CZIRJAK, L., GEHER, P., KESZTHELYI, P., KOVACS, A., KOVACS, L., SZABO, Z., SZEKANECZ, Z., TAMASI, L., TOTH, A. E., UJFALUSSY, I., HEVER, N. V., STRBAK, B., BAJI, P., BRODSZKY, V. & GULACSI, L. 2014b. Acceptability of less than perfect health states in rheumatoid arthritis: the patients' perspective. *Eur J Health Econ,* 15 Suppl 1**,** S73-82. https://doi.org/10.1007/s10198-014-0596-2

PENTEK, M., ZRUBKA, Z. & GULACSI, L. 2017. The Economic Impact of Biosimilars on Chronic Immune-Mediated Inflammatory Diseases. *Curr Pharm Des,* 23**,** 6770-6778. https://doi.org/10.2174/1381612824666171129193708

PICKARD, A. S., NEARY, M. P. & CELLA, D. 2007. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health Qual Life Outcomes,* 5**,** 70. https://doi.org/10.1186/1477-7525-5-70

PMLIVE. 2018. *Top Pharma List* [Online]. PMLive. Available: http://www.pmlive.com/top_pharma_list/pharmaceutical_products/remicade [Accessed 2019.02.23.].

RIOS-DIAZ, A. J., LAM, J., RAMOS, M. S., MOSCOSO, A. V., VAUGHN, P., ZOGG, C. K. & CATERSON, E. J. 2016. Global Patterns of QALY and DALY Use in Surgical Cost-Utility Analyses: A Systematic Review. *PLoS One,* 11**,** e0148304. https://doi.org/10.1371/journal.pone.0148304

RYFF, C. D. & KEYES, C. L. 1995. The structure of psychological well-being revisited. *J Pers Soc Psychol,* 69**,** 719-27. https://doi.org/10.1037/0022-3514.69.4.719

SCHANE, R. E., LING, P. M. & GLANTZ, S. A. 2010. Health effects of light and intermittent smoking: a review. *Circulation,* 121**,** 1518-22. https://doi.org/10.1161/CIRCULATIONAHA.109.904235

SIMMONS, C. 2018. Games and greed delay the market enty of money-saving biosimilars. *Stat*, May 9, 2018.

SMOLEN, J. S., LANDEWE, R., BIJLSMA, J., BURMESTER, G., CHATZIDIONYSIOU, K., DOUGADOS, M., NAM, J., RAMIRO, S., VOSHAAR, M., VAN VOLLENHOVEN, R., ALETAHA, D., ARINGER, M., BOERS, M., BUCKLEY, C. D., BUTTGEREIT, F., BYKERK, V., CARDIEL, M., COMBE, B., CUTOLO, M., VAN EIJK-HUSTINGS, Y., EMERY, P., FINCKH, A., GABAY, C., GOMEZ-REINO, J., GOSSEC, L., GOTTENBERG, J. E., HAZES, J. M. W., HUIZINGA, T., JANI, M., KARATEEV, D., KOULOUMAS, M., KVIEN, T., LI, Z., MARIETTE, X., MCINNES, I., MYSLER, E., NASH, P., PAVELKA, K., POOR, G., RICHEZ, C., VAN RIEL, P., RUBBERT-ROTH, A., SAAG, K., DA SILVA, J., STAMM, T., TAKEUCHI, T., WESTHOVENS, R., DE WIT, M. & VAN DER HEIJDE, D. 2017. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis,* 76**,** 960-977. https://doi.org/10.1136/annrheumdis-2016-210715

SZENDE, A., JANSSEN, B. & CABASES, J. 2014. *Self-Reported Population Health: An International Perspective based on EQ-5D,* New York, Springer https://doi.org/10.1007/978-94-007-7596-1

SZENDE, A. & NEMETH, R. 2003. [Health-related quality of life of the Hungarian population]. *Orv Hetil,* 144**,** 1667-74.

SZENDE, A., OPPE, M. & DEVLIN, N. 2006. *EQ-5D value sets: inventory, comparative review and user guide,* Dordrecht, Springer.

TANYI, A. & KOLLÁNYI, Z. 2008. *Etika és egészség-politika,* Budapest, Demos Magyarország Alapítvány.

THOKALA, P., DEVLIN, N., MARSH, K., BALTUSSEN, R., BOYSEN, M., KALO, Z., LONGRENN, T., MUSSEN, F., PEACOCK, S., WATKINS, J. & IJZERMAN, M. 2016. Multiple Criteria Decision Analysis for Health Care Decision Making--An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health,* 19**,** 1-13. https://doi.org/10.1016/j.jval.2015.12.003

UBEL, P. A., LOEWENSTEIN, G. & JEPSON, C. 2003. *Quality of Life Research,* 12**,** 599-607. https://doi.org/10.1023/A:1025119931010

VAN GOOL, K. & PEARSON, M. 2014. Health, Austerity and Economic Crisis: Assessing the Short-term Impact in OECD countries. *OECD Health Working Papers, No 76. .* Paris.

VAN NIES, J. A., KRABBEN, A., SCHOONES, J. W., HUIZINGA, T. W., KLOPPENBURG, M. & VAN DER HELM-VAN MIL, A. H. 2014. What is the evidence for the presence of a therapeutic window of opportunity in rheumatoid arthritis? A systematic literature review. *Ann Rheum Dis,* 73**,** 861-70. https://doi.org/10.1136/annrheumdis-2012-203130

VAN OSCH, S. M., WAKKER, P. P., VAN DEN HOUT, W. B. & STIGGELBOUT, A. M. 2004. Correcting biases in standard gamble and time tradeoff utilities. *Med Decis Making,* 24**,** 511-7. https://doi.org/ 10.1177/0272989X04268955

VEENHOVEN, R. 1993. Happiness in Nations: Overview of happiness surveys using Measure type: 112G / 11-step numeral Happiness. *World Database of Happiness.* Rotterdam: Erasmus University Rotterdam, Happiness Economics Research Organisation.

VEENHOVEN, R. 2009. International scale interval study: Improving the comparability of responses to survey questions about happiness. *In:* MOLLER, V. & HUSCHKA, D. (eds.) *Quality of life and the millenium challenge: Advances in quality of life studies, theory and research.* vol. 35. ed.: Springer. https://doi.org/10.1007/978-94-007-0753-5_3661

VEENHOVEN, R. 2012. Happiness: Also Known as "Life Satisfaction" and "Subjective Well-Being". *In:* LAND, K. C., MICHALOS, A. C. & SIRGY, J. M. (eds.) *Handbook of Social Indicators and Quality of Life Research.* Dordrecht: Springer.

VEZER, B., BUZAS, Z., SEBESZTA, M. & ZRUBKA, Z. 2016. Authorized manufacturing changes for therapeutic monoclonal antibodies (mAbs) in European Public Assessment Report (EPAR) documents. *Curr Med Res Opin,* 32**,** 829-34. https://doi.org/10.1185/03007995.2016.1145579

VON NEUMANN, J. & MORGENSTERN, O. 1944. *Theory of games and economic behavior,* Princeton, Princeton University Press.

WALTERS, S. J. & BRAZIER, J. E. 2005. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res,* 14**,** 1523-32. https://doi.org/ https://doi.org/10.1007/s11136-004-7713-0

WHITTINGTON, M. D., MCQUEEN, R. B., OLLENDORF, D. A., KUMAR, V. M., CHAPMAN, R. H., TICE, J. A., PEARSON, S. D. & CAMPBELL, J. D. 2018. Long-term Survival and Value of Chimeric Antigen Receptor T-Cell Therapy for Pediatric Patients With Relapsed or Refractory Leukemia. *JAMA Pediatr,* 172**,** 1161-1168. https://doi.org/10.1001/jamapediatrics.2018.2530

WHO 2010. Health Systems Financing: the path to universal coverage. *The World Health Report.* Geneva: WHO.

WHO, W. H. O. *Recommended levels of physical activity for adults aged 18 - 64 years* [Online]. Available: https://www.who.int/dietphysicalactivity/factsheet_adults/en/ [Accessed 2017.10.01.].

WILLEME, P. & DUMONT, M. 2015. Machines that Go 'Ping': Medical Technology and Health Expenditures in OECD Countries. *Health Econ,* 24**,** 1027-41. https://doi.org/10.1002/hec.3089

WILLIAMS, A. 1992. Cost-effectiveness analysis: is it ethical? . *Journal of medical ethics,* 18**,** 7-11. https://doi.org/10.1136/jme.18.1.7

WOUTERS, S., VAN EXEL, N. J., ROHDE, K. I. & BROUWER, W. B. 2015. Are all health gains equally important? An exploration of acceptable health as a reference point in health care priority setting. *Health Qual Life Outcomes,* 13**,** 79. https://doi.org/10.1186/s12955-015-0277-6

WOUTERS, S., VAN EXEL, N. J. A., ROHDE, K. I. M., VROMEN, J. J. & BROUWER, W. B. F. 2017. Acceptable health and priority weighting: Discussing a reference-level approach using sufficientarian reasoning. *Soc Sci Med,* 181**,** 158-167. https://doi.org/10.1016/j.socscimed.2017.03.051

XIE, F., GAEBEL, K., PERAMPALADAS, K., DOBLE, B. & PULLENAYEGUM, E. 2014. Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. *Med Decis Making,* 34**,** 8-20. https://doi.org/10.1177/0272989X13480852

YANG, Z., LUO, N., BONSEL, G., BUSSCHBACH, J. & STOLK, E. 2018. Selecting Health States for EQ-5D-3L Valuation Studies: Statistical Considerations Matter. *Value Health,* 21**,** 456-461. https://doi.org/10.1016/j.jval.2017.09.001

ZRUBKA, Z. 2017a. Economic aspects of biosimilar competition in Hungary — The treatment of rheumatic disorders. *Society and Economy,* 39**,** 271-290. https://doi.org/10.1556/204.2017.39.2.6

ZRUBKA, Z. 2017b. [Measurement and health economic evaluation of informal care]. *Orv Hetil,* 158**,** 1363-1372. https://doi.org/10.1556/650.2017.30841

ZRUBKA, Z., BERETZKY, Z., HERMANN, Z., BRODSZKY, V., GULÁCSI, L., RENCZ, F., BAJI, P., GOLICKI, D., PREVOLNIK-RUPEL, V. & PÉNTEK, M. 2019a. A Comparison of European, Polish, Slovenian and British EQ-5D-3L value sets using a Hungarian sample of 18 chronic diseases *The European Journal of Health Economics***,** https://doi.org/10.1007/s10198-019-01069-8

ZRUBKA, Z., BRODSZKY, V., PÉNTEK, M., RENCZ, F. & GULÁCSI, L. 2018a. Pms3 - Infliximab for Disease-Modifying Anti-Rheumatic Drug-Naive Rheumatoid Arthritis Patients: Systematic Review and Descriptive Analysis of Publications of Randomized Controlled Trials. *Value in Health,* 21**,** S288. https://doi.org/10.1016/j.jval.2018.09.1717

ZRUBKA, Z., GOLICKI, D., PREVOLNIK-RUPEL, V., BAJI, P., RENCZ, F., BRODSZKY, V., GULÁCSI, L. & PÉNTEK, M. 2019b. Towards a Central-Eastern European EQ-5D-3L Population Norm: Comparing Data from Hungarian, Polish and Slovenian Population Studies. *The European Journal of Health Economics***,** https://doi.org/10.1007/s10198-019-01071-0

ZRUBKA, Z., GULÁCSI, L., BRODSZKY, V., RENCZ, F., ALTEN, R., SZEKANECZ, Z. & PÉNTEK, M. 2019c. Long-term efficacy and cost-effectiveness of infliximab as first-line treatment in rheumatoid arthritis: systematic review and meta-analysis. *Expert Review of Pharmacoeconomics & Outcomes Research***,** [Submitted article].

ZRUBKA, Z., GULÁCSI, L., RENCZ, F., BRODSZKY, V. & PÉNTEK, M. 2018b. A new approach to assess the acceptability of health problems at different ages: an experimental study using the EQ-5D-3L instrument. *35th EuroQoL Plenary Meeting.* Lisbon, Portugal.

ZRUBKA, Z., HERMANN, Z., GULÁCSI, L., BRODSZKY, V., RENCZ, F. & PÉNTEK, M. 2018c. Prm171 - Determinants of the Acceptability of Health Problems in Different Ages: Exploring a New Application of the EQ VAS. *Value in Health,* 21**,** S385. https://doi.org/10.1016/j.jval.2018.09.2290

ZRUBKA, Z., HERMANN, Z., GULÁCSI, L., BRODSZKY, V., RENCZ, F. & PÉNTEK, M. 2019d. Determinants of the acceptability of health problems in different ages: exploring a new application of the EQ VAS. *The European Journal of Health Economics***,** https://doi.org/10.1007/s10198-019-01060-3

VIII. OWN PUBLICATIONS WITH RELEVANCE TO THE THESIS

A tézishez kapcsolódó saját szakcikkek összesített impact factora: 19,318

BAJI, P., GOLICKI, D., PREVOLNIK-RUPEL, V., BROUWER, W. B. F., ZRUBKA, Z., GULÁCSI, L. & PÉNTEK, M. 2019. The burden of informal caregiving in Hungary, Poland and Slovenia: results from national representative surveys. *The European Journal of Health Economics.* https://doi.org/10.1007/s10198-019-01058-x

GULACSI, L., ZRUBKA, Z., BRODSZKY, V., RENCZ, F., ALTEN, R., SZEKANECZ, Z. & PENTEK, M. 2019. Long-Term Efficacy of Tumor Necrosis Factor Inhibitors for the Treatment of Methotrexate-Naive Rheumatoid Arthritis: Systematic Literature Review and Meta-Analysis. *Adv Ther*, 36, 721-745. https://doi.org/10.1007/s12325-018-0869-8

PENTEK, M., ZRUBKA, Z. & GULACSI, L. 2017. The Economic Impact of Biosimilars on Chronic Immune-Mediated Inflammatory Diseases. *Curr Pharm Des*, 23, 6770-6778. https://doi.org/10.2174/1381612824666171129193708

VEZER, B., BUZAS, Z., SEBESZTA, M. & ZRUBKA, Z. 2016. Authorized manufacturing changes for therapeutic monoclonal antibodies (mAbs) in European Public Assessment Report (EPAR) documents. *Curr Med Res Opin*, 32, 829-34. https://doi.org/10.1185/03007995.2016.1145579

ZRUBKA, Z. 2017a. Economic aspects of biosimilar competition in Hungary — The treatment of rheumatic disorders. *Society and Economy*, 39, 271-290. https://doi.org/10.1556/204.2017.39.2.6

ZRUBKA, Z. 2017b. [Measurement and health economic evaluation of informal care]. *Orv Hetil*, 158, 1363-1372. https://doi.org/10.1556/650.2017.30841

ZRUBKA, Z., BERETZKY, Z., HERMANN, Z., BRODSZKY, V., GULÁCSI, L., RENCZ, F., BAJI, P., GOLICKI, D., PREVOLNIK-RUPEL, V. & PÉNTEK, M. 2019a. A Comparison of European, Polish, Slovenian and British EQ-5D-3L value sets using a Hungarian sample of 18 chronic diseases *The European Journal of Health Economics*, https://doi.org/10.1007/s10198-019-01069-8

ZRUBKA, Z., BRODSZKY, V., PÉNTEK, M., RENCZ, F. & GULÁCSI, L. 2018a. Pms3 - Infliximab for Disease-Modifying Anti-Rheumatic Drug-Naive Rheumatoid Arthritis Patients: Systematic Review and Descriptive Analysis of Publications of Randomized Controlled Trials. *Value in Health*, 21, S288. https://doi.org/10.1016/j.jval.2018.09.1717

ZRUBKA, Z., GOLICKI, D., PREVOLNIK-RUPEL, V., BAJI, P., RENCZ, F., BRODSZKY, V., GULÁCSI, L. & PÉNTEK, M. 2019b. Towards a Central-Eastern European EQ-5D-3L Population Norm: Comparing Data from Hungarian, Polish and Slovenian Population Studies. *The European Journal of Health Economics*, https://doi.org/10.1007/s10198-019-01071-0

ZRUBKA, Z., GULÁCSI, L., BRODSZKY, V., RENCZ, F., ALTEN, R., SZEKANECZ, Z. & PÉNTEK, M. 2019c. Long-term efficacy and cost-effectiveness of infliximab as first-line treatment in rheumatoid arthritis: systematic review and meta-analysis. *Expert Review of Pharmacoeconomics & Outcomes Research*, [Submitted article].

ZRUBKA, Z., GULÁCSI, L., RENCZ, F., BRODSZKY, V. & PÉNTEK, M. 2018b. A new approach to assess the acceptability of health problems at different ages: an experimental study using the EQ-5D-3L instrument. *35th EuroQoL Plenary Meeting*. Lisbon, Portugal.

ZRUBKA, Z., HERMANN, Z., GULÁCSI, L., BRODSZKY, V., RENCZ, F. & PÉNTEK, M. 2018c. Prm171 - Determinants of the Acceptability of Health Problems in Different Ages: Exploring a New Application of the EQ VAS. *Value in Health*, 21, S385. https://doi.org/10.1016/j.jval.2018.09.2290

ZRUBKA, Z., HERMANN, Z., GULÁCSI, L., BRODSZKY, V., RENCZ, F. & PÉNTEK, M. 2019d. Determinants of the acceptability of health problems in different ages: exploring a new application of the EQ VAS. *The European Journal of Health Economics*, https://doi.org/10.1007/s10198-019-01060-3