CORVINUS UNIVERSITY of BUDAPEST
DEPARTMENT OF HEALTH POLICY

Corvinus University of Budapest

Comparative analysis of health state utility measurement methods

PhD Dissertation

Supervisor: Prof. Valentin Brodszky

Candidate: Péter Balázs

Budapest, 2023

Content

List of Tables

List of Figures

Glossary of abbreviations

ABSIS- Autoimmune Bullous Skin Disorder Intensity Score (clinical tool)

ASCOT- Adult Social Care Outcomes Toolkit (questionnaire)

BTD- better than dead (health state)

CET- cost-effectiveness threshold

DLQI- Dermatology Life Quality Index (questionnaire)

EQ-VT- EuroQol Valuation Technology (protocol)

HRQoL- health-related quality of life

HSU- health state utility (measurement score)

HTA- health technology assessment

ICECAP-A/O- Icepop Capability measure for Adults/Elderly (questionnaire)

MAU- multi-attribute utility

mDLQI- mapping based DLQI utility score

MVH- Measurement and Valuation of Health (protocol)

OxCAP-MH- Oxford Capabilities questionnaire–Mental Health

PASI- Psoriasis Area Severity Index (clinical tool)

PRO- patient reported outcomes

QALY- quality-adjusted life years

QoL- quality of life

QWB- Quality of Well-being Index

SARTORIUS- Modified Sartorius Scale (clinical tool)

SCORAD- Scoring Atopic Dermatitis (clinical tool)

SF-6D- Short Form six dimension (questionnaire)

SG- standard gamble

TTO- time trade-off

VAS- visual analogue scale

vDLQI- value set based DLQI utility score

WTD- worse than dead (health state)

WTP- willingness to pay

1. Introduction

The past few decades brought extent focus on measuring the impact of chronic diseases on health-related quality of life (HRQoL) [1]. The interest of certain scientific fields such as medicine, health policy or pharmacology show extent focus on evaluating health interventions that treat various mental or physiological chronic illnesses. There are more than 293.000 publications of 'health economic evaluation' and 174.000+ 'health state utility' publications on a large medical database (PubMed, searched as a combination of: (i) "health gain" OR "health evaluat*" OR "cost" AND ("effective*" OR "utilit*" OR "minimiz*" OR "benefit") and (ii) "health state utilit*" OR "health utilit*" OR HSU OR HSUV OR "utilit*" AND ("score" OR "value*" OR "measure*" OR "elicit*" OR "estimat*") in May 2023).

Health is defined as an infinite demand, which is supplied by the finite health care system. Any health improvement is considered as a public good, while all societies could be a bit healthier in psychological and physiological domains. Many effective health interventions and programs exist that improve the health of population. Economics fundamentally focuses on redistribution of scarce resources − to do it optimally, health economic evaluations are performed for information proposes − allocative decisions are made by the government (or responsible decision maker in healthcare) [2].

Health economics evaluations hold on a primary aim: application of effective redistribution of scarce resources within health care system. Health economics guides decision making by quantifying costs and benefits of an intervention to analyse maximise welfare surplus. Principally the prioritization of societal needs is out of the scope of health economics, such trade-offs are investigated by health policy fields [3].

Most frequently health economic evaluations apply a standard health outcome measure called quality-adjusted life year (QALY) [4]. The QALY combines two components, given as (1) *quality of life* – expressed in terms of life years gained/saved and (2) *quality of life* – measured by a HRQoL instrument, usually a health state utility (HSU) [1] elicitation method [5].

Unforeseen events such as an outbreak of a pandemic (e.g. COVID-19) make extraordinary focus on mental health in medicine as well as in health economic research fields. Mental diseases such as anxiety, bipolar disorder, depression, dysthymia, obsessive-compulsive

[1] The text refers on health state utility as HSU or utility or health utility interchangeably.

disorder, panic disorder, cause deterioration in peoples HRQoL and/or negatively affect their subjective well-being and other non-health related domains such as capabilities. Chronic mental diseases may lead to sleeping and eating dysfunctions, meltdown in work ability and social relations, terminating in disrupted role function, self-assessment or stigmatization. A standardized measurement score is required that express HRQoL uniformly to compare the intervention outcomes across diseases for healthcare resource allocation purposes. More countries set the regulation of financial decision making, suggesting a HSU measurement to quantify the outcome of the intervention [6].

Adequate decisions require comprehensive health economic evaluations that is based on correct estimations [7]. The measurement of HRQoL and within it HSU estimation takes crucial part in evaluation. Utility measurement takes a vital role in health economic evaluation, since being the first step of the analysis. Various direct indirect and disease-specific methods were designed to assess the utility of various health states.

1.1. Key terms and definitions

Quality of life (QoL) is defined as people's perception of their life, that covers multiple domains of life including believes, physical and mental health, social relationships, self-fulfilment and well-being. Theoretically roots in Maslow's human developmental perspective, that is build on hierarchical satisfaction of needs [8].

To quantify health and describe health status in medicine and economic fields − a more narrow − health-specific measurements have evolved out of QoL tools over the past 30 years. *HRQoL* measures have a particular focus on physical, psychological and social domains of health. Each domain consists of more attributes, such as (1) mobility, symptoms, pain are typical physical health components; (2) emotions, mood, self-appraisal are core components of the psychological domain; while (3) ability to function and personal relationships are social domain related components that describe the persons daily role activities [9, 10]. A specific type of HRQoL measurements are the HSU elicitation methods, that solely focus on physical and psychological domains of health.

The *utility* concept is built on the rational decision making theory of Neumann–Morgenstern. The theory assumes that people maximize their utility according to their preferences and weight the components of their health according to that [11]. Generally, the theory is well applicable in health state measurement, therefore the HSU is a widely used approach to assess HRQoL [12, 13]. Utility values express the quality of the health state on a

0-1 interval scale, where zero means death, one is equivalent to full health, while negative values represent health states that are worse than dead.

In direct HSU valuation tasks the health states are often described by vignettes. Health state description also referred as anchoring vignette, is defined as a tool for illustrating a given disease/*imperfect health condition* gravitating towards reality. Direct preference-based tasks health description on the other edge can differ too, *full health state* can be defined in various ways – such as perfect; desired in the future; best possible health state – each likely to modify the HSU results. Direct HSU elicitation methods also may differ in measuring *better than dead* (BTD) when the utility yield is 0-1 or *worse than dead* (WTD) health states that is represent in negative health utilities [14, 15]. Generic or disease specific HRQoL questionnaires are applied as multidimensional instruments which scores can be converted into HSU, and also frequently used in *patientreported outcome* (PRO) as measurement endpoint in clinical studies.

In the past decade a growing interest towards measuring non-health domains of well-being was evident. While the utility measures are rather designed to capture physical and psychological domains of health, the measure of societal elements were marginalized. The *capability* approach emphasizes the individual's ability to achieve their valued objectives. Individual's capability is therefore defined as free to do things that are important in life [16].

Capability and HRQoL domains interact with each other, there is possibly a correlation between the measurement scores. The great variety of health states combined with various individual preferences/capabilities forecasts that people even with same disease will yield different HRQoL and capability scores.

1.2. Addressing the need for research

The impact of chronic diseases on HRQoL explains the need for health interventions. Health economic evaluations ought to advise decision making on redistribution of scarce resources by finding the best incremental value of health gain, while potentially considering the societal welfare, financing thresholds and benefits beyond health.

Evaluations apply QALY combining quality and quantity of life into a universally comparable score. HSU measures are conducted to express the quality component of life, that reflects on peoples' perspectives about physiological (e.g. pain, mobility) and psychological (e.g. emotions, stress, cognition) domains of health states.

This thesis addresses the complexity of HSU assessment and intends to contribute on better understanding it conceptually. Analysing the factors that impact HSU estimates and exploring the relationship between different measurements, complementally with an alternative framework the findings contribute to increase intervention/treatment efficiency in healthcare.

1.3. Overview of the research framework

The dissertation has a primary contribution to *health economic evaluations*, that apply welfarist economic theory principle of maximizing incremental effect per affordable incremental cost. Health effect as outcome of interventions is typically expressed with a unified measure called *QALY*, combining quantity and quality of life referring to health gain. Quality of life in different health states are captured by *HRQoL* measurements, where *health state utility* assessments gain ground due (1) well fitting for cross-comparison of diverse health states, (2) being in line with utilitarian welfare theory. HSU measures can be *direct*, when respondents are challenged to choose between alternatives while revealing preferences for rather quality/quantity of life, or *indirect*, when respondent rate questionnaire items on a rating scale, that cover domains of health (e.g. physical functioning, vitality, social functioning, pain). The dissertation ought to reveal analyse agreement between two certain types of direct and indirect HSU measurements and provide information on which factors have impact on directly assessed utility estimates. Special emphasis was designated to the methodological attributes of the time trade-off tasks, for detecting huge extent of heterogeneity among application.

The QALY based quantification of health outcomes are often criticized for being too narrow and allowing too much variance for preference-weighted component, thus new alternatives (or subsidiary) appeared, like the *capability approach*. Arguments state that besides physiological and psychological health domains, well-being and individual abilities have to be taken into account when evaluating health outcomes of different interventions to counterbalance the lopsided utilitarian analytical framework. (Figure 1)

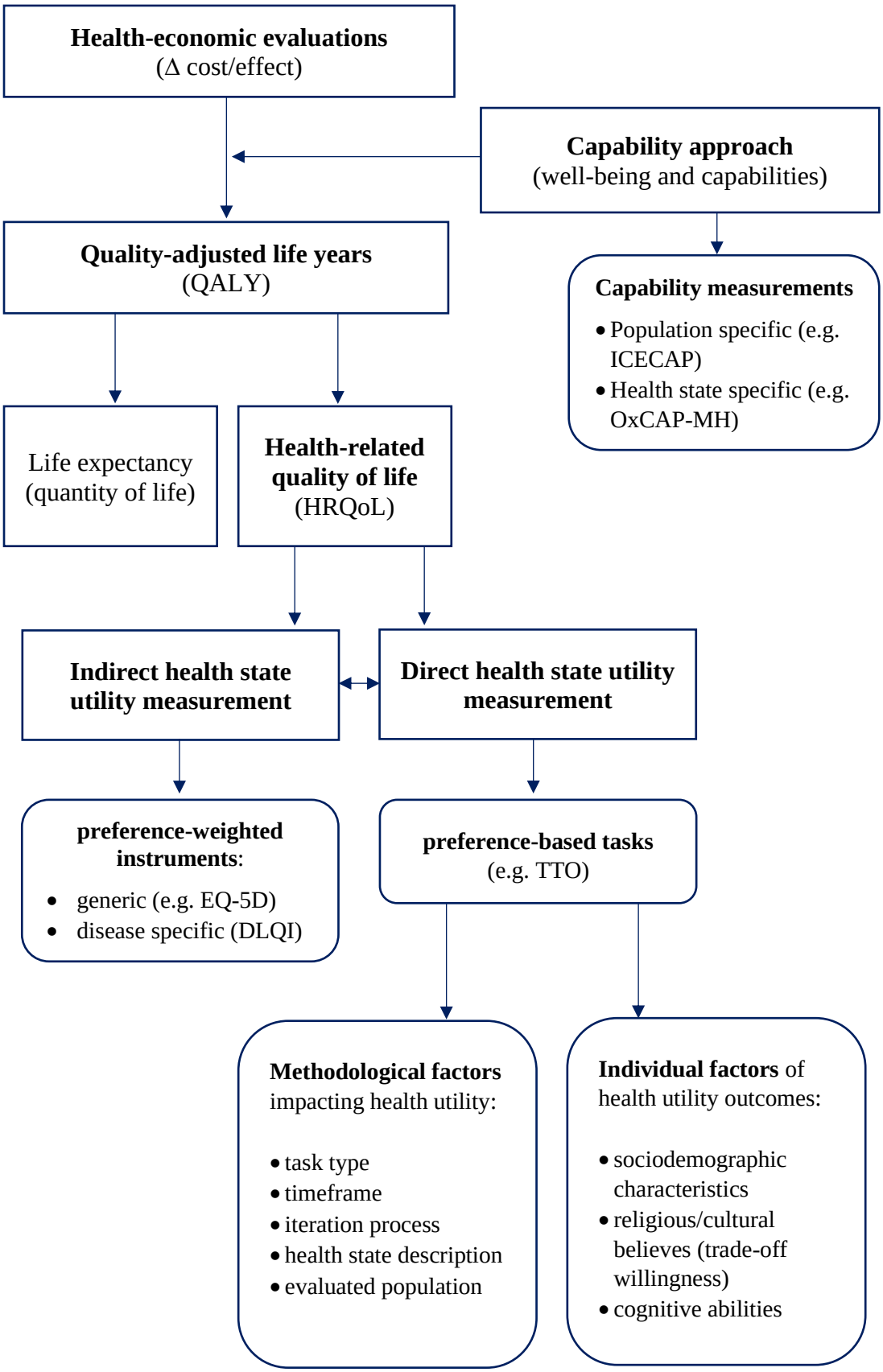

Figure 1 - Illustration of the theoretical framework of the research

2. Theoretical background

2.1. Health economic evaluations

Health economic evaluations cover a broad range of systematic analytical methods that are used to compare the effectiveness of various therapies/interventions by comparing their cost and outcomes. Besides the below introduced cost minimization, cost-benefit, cost-effectiveness and cost-utility analysis program investigators often use other (non-direct comparison) techniques such as cost of illness or cost-consequence analysis [17]. All evaluations are ought to inform the decision-makers of full cost and benefits (often referred as health gain) of health interventions (drug therapy, treatment, new medical device, etc.) while the effective reimbursement of public resources is a priority for the societal welfare.

In the following chapters the four main types of health evaluations are introduced, broadly. They all directly compare costs with outcomes but do so in different ways fitting to the attributes of the intervention. The costs in all examples are defined as total costs summing direct (associated with the intervention) and indirect (associated with reduced productivity due to health state) costs, while different costing methods are disregarded from recent thesis.

2.1.1. Cost-minimization

Cost-minimization analysis (CMA) compares more alternative therapies only in terms of their costs, while their outcome is the same. Such analysis is performed for example if three different depression treatment medications by different producers are compared and the outcome is reduction in weekly depression periods by one.

Antidepressant A	Antidepressant B	Antidepressant C
cost: 5 €	cost: 7 €	cost: 9 €

Table 1 – Cost-minimization analysis example

If the treatment is the same, causing the same outcome than the only the costs are compared. The evaluation would conclude that antidepressant A should be chosen, as being the most cost-effective. The CMA method is unpopular for the reason of not considering the

analysis of treatment effect score confidence intervals, which information is provided when drawing a cost-effectiveness plane [18].

2.1.2. Cost-benefit analysis

Cost-benefit analysis (CBA) is a generic evaluation method, often used for valuing items with no exact market price. Usually both the costs and outcomes of the intervention are measured in monetary terms. CBA accounts for all direct and indirect cost using either a patient, insurer or societal perspective. A variety of tools – such as willingness to pay (WTP), human capital method, estimating friction costs – can be used to calculate (health outcome) benefits [19].

In CBA the health intervention is compared against an alternative, that can be an alternative or no treatment. The result of the analysis in the net benefit (all benefits minus the net cost), which is the decision criterion in CBA. In many cases sensitivity analysis and discounting is part of the CBA, while the data requirements and uncertainty are quite high.

2.1.3. Cost-effectiveness analysis

When different interventions are compared that have a common health outcome – usually expressed as natural units: like life years gained, number of sleeping hours increased, reduction in blood pressure – the advised evaluation type is cost-effectiveness analysis (CEA). For example, if three interventions to treat depression are (A) medication, (B) psychotherapy and (C) medication+psychotherapy. The intervention effect is measured in terms of increase in productive hours/week.

Intervention	A	B	C
Effect	+ 2 hours	+ 2.5 hours	+ 5 hours
Cost	200 €	220 €	400 €

Table 2 - Cost-effectiveness analysis example

Cost effectiveness ratio (CER) – calculated as total cost/effects – enables to directly compare the interventions to choose the one with the lowest number. In this example intervention C is superior to the others while the CER = 400 €/ 5 plus productive hours = 80 that is better than the CER of intervention A and B (100 and 88).

2.1.4. Cost-utility analysis

Cost-utility analysis (CUA) can be considered as the generalized version of cost-effectiveness analysis. The analysis allows to compare different interventions with different effects, though the effects must be measured in a common unit. The health outcome (effect) is measured by HSU and in health economic evaluations often the utility is combined with life expectancy giving the QALY. In the analysis the cost represents the nominator, but instead of natural unit effects the denominator is expressed in terms of QALYs.

Incremental cost-effectiveness ratio (ICER) is calculated of two intervention alternatives to find the more cost-effective choice. The CUA is a very frequently used method of evaluation and considered as the "gold standard" to express cost-effectiveness of health care choices, despite the challenges of health state evaluations and subjective nature of HRQoL.

Let's consider four alternative therapies to treat 100 severe depression patients with different treatment costs and outcomes. The total costs of therapies are given while the outcome have to be measured in terms of health utility*life expectancy. The outcome of treatments is expressed as remaining life expectancy (10 years) multiplied by the utility gain (baseline utility = 0.10). According to the cost-effectiveness plane principle only the less costly but more effective and more costly but more effective interventions are viable choices.

Intervention	A – Advanced medication	B – Psychotherapy	C – Cane pet therapy	D – Lifestyle management
Cost	150 € /patient	230 €/patient	290 €/patient	320 €/patient
Outcome (QALY)	5.2/patient	6.5/patient	6.0/patient	7.4/patient
ICER compared to no treatment	35.7	41.8	58.0	50.0

Table 3 - Cost-effectiveness analysis example

The ICER of intervention A is calculated as: (no treatment: 0 € cost – Cost of Interventionᴀ: 150€*100) / (no treatment: 0.10*10 QALY*100 – Effect of Interventionᴀ: 0.52*10 QALY*100) = 35.7 representing the highest effectiveness. Intervention A could be replaced by Intervention B or D if their cost/QALY is under the financing threshold of the country.

2.1.5. Financing thresholds

Albeit health economic evaluations provide the information about the (intervention) alternatives in terms of ICERs, the policy decision rules differ across jurisdictions. At national level the financing threshold is used to set the maximum financial investment that a government is willing to pay for one additional health gain/patient. Different perception about the function of producing health manifests in 2+2 different approaches to set financing threshold (often cost effectiveness threshold: CET), though at national level the approaches have many variants [20, 21].

a. When a fixed budget is used to maximize health for the population policymakers apply an **explicit maximum threshold**. The health production function can be described as:

$$\Delta c_d/\Delta Q < k\ (cost - effectiveness\ ratio\ of\ interventions)$$

Where Δc_d are only the healthcare (direct) costs, k represents the marginal cost-effectiveness of current spending (that is the previous intervention's $\Delta c_d/\Delta Q$), Q is the health benefit measured in QALYs. The net monetary benefit of the new intervention simply should be better than the displaced intervention, by producing more health per euro spent (optimally k=v).

b. When societal welfare is maximized out of a flexible budget, then **implicit threshold** is used. Production equation, where c_t denote incremental total (direct and indirect) costs and ΔQ the incremental health gain can be written as:

$$\Delta c_t/\Delta Q < v\ (consumption\ value\ of\ health)$$

c. Contingent valuations are also used to set the financing threshold according to the population's preferences, ergo willingness to pay for a QALY.

d. Many argue that health budget related policy decisions should be made by experts, without involving the taxpayer's opinion. Investment decisions estimate the opportunity cost of health benefits forgone, by measuring the impact of disinvestment [22].

Threshold type	**Perspective**	**Country example**
Explicit – fixed budget	Healthcare (supply-side)	United Kingdom, Hungary
Implicit – flexible budget	Societal (demand-side)	Netherlands, Sweden
Willingness to pay – preference based	Societal (demand-side)	United States

Opportunity cost based	Healthcare (supply-side)	Australia

Table 4 - Financing threshold types

The societal value of health gain varies greatly among countries, the national-level regulation strongly depends on more *policy factors*, like (1) setting implicit/explicit or no threshold, (2) consideration of health economic evaluations, or (3) politicization of reimbursement decisions [20]. Vast amount of studies apply contingent valuations to estimate the monetary value of health gain, although the diversity of *methodological aspects*, (1) as the cost perspective, (2) basis of preferences, or (3) the measurement of QALY components complicate the comparison of WTP/QALY values [23, 24]. The following table summarizes information from WTP reviews to demonstrate difference in methods when estimating the societal value of health [25-27]:

Cost perspective	**mean WTP/ QALY PPP $ in 2022**	**Base of preferences**	**Estimation of quality and quantity of life**	**Country example**
Societal/individual	~ 6,300	general population/patients	EQ-5D	China
Individual	~ 30,130	general population	EQ-5D or time trade-off	Hungary
Individual	~ 33,345	general population	EQ-5D or time trade-off	France
Individual	~ 38,410	general population	EQ-5D or time trade-off	Denmark
Societal	~ 42,215	general population	time trade-off or standard gamble or EQ-5D	Sweden
Individual/societal	~ 52,740	patients/general population	time trade-off	USA
Societal	~ 70,905	general population	discrete choice experiment	Australia
Societal	~ 160,555	general population	discrete choice experiment	Taiwan

Table 5 - Different societal values of health (gain) production

Currently the same term is used for both supply and demand-side thresholds, disregarding whether it refers on optimizing for consumption value of health or incremental cost effectiveness. Moreover, the financing threshold does not reveal the division of welfare surplus between the technology producer and population. When the impact of health economic

evaluations on decision-making is stressed, a further gap is also argued between political realities and societal requirements.

2.1.6. Welfarist and extra-welfarist approach of evaluations

The welfarist approach favours population's preferences on deciding value of health production (e.g. measured with QALY gain), while they finance the healthcare system, the CET shall be based on their preferences (expressed as the quality of life weight). The extra-welfarist approach stresses to use the opportunity cost method when setting CETs (practically, financial decisions should be based experts opinion on direct/indirect healthcare investments) [22, 28].

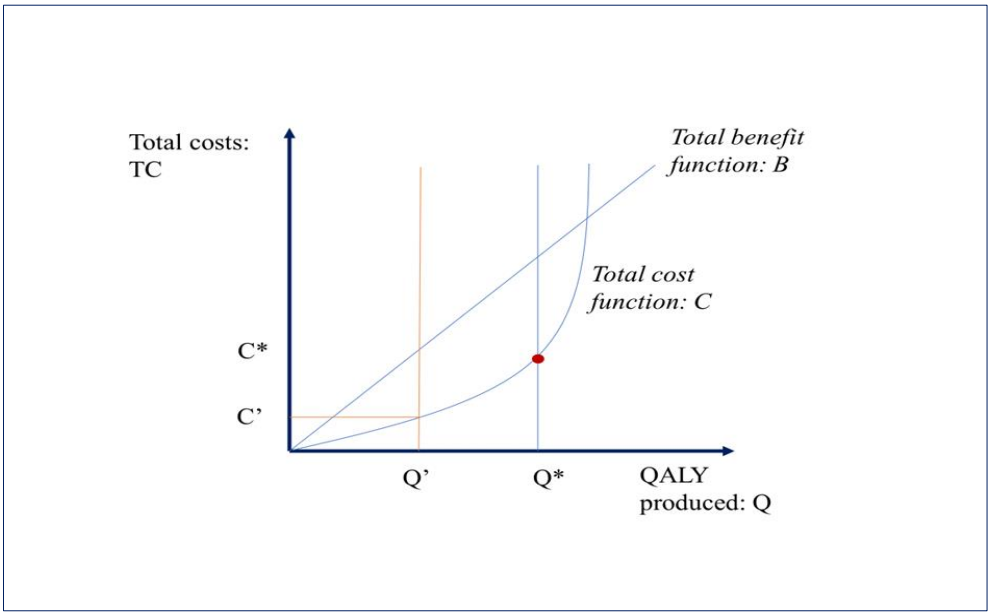

Figure 2 - Optimal QALY production in welfarist and extra welfarist approach

Note: Own editing based on Baker et al. [29]

A classical economic approach would assume, that aggregating the per unit value of the total health gain (that refers to the amount of health production, e.g. expressed as QALYs) is defined by the total benefits (given as B=TC/Q). The assumption of diminishing returns says that the marginal cost (MC) of additional health gain (Q) production is increasing. Presuming ceteris paribus, the optimal level of health gain production is (Q*), where the marginal cost and marginal benefits are equal (the gradient is illustrated with the red dot). Suppose that a public expenditure constraint is imposed (for example set by a fixed healthcare budget/ fixed financing

threshold), where the C' limits the QALY production and the optimum level shifts to Q'. The extra welfarist approach allows to provide health intervention if the QALY unit gain is below or equal to the marginal cost at Q'. The welfarist approach requires from public health service providers to supply only if the marginal cost-benefit ratio exceeds 1. (Figure 2)

2.2. Quality of life

Quality of life has an extend concept, its interpretation varies across scientific disciplines. Consequently, there is no straightforward guidance which instruments measure clearly QoL. More definitions exist parallelly, the subjective factors include individuals' satisfaction with one's life in the context of their value system [30]. Objective factors include physical, emotional and social well-being. Defining health as a "state of complete physical, mental and social well-being" made huge influence on development of health status measurement tools [31].

2.2.1. Health-related quality of life

In medicine and health economic fields HRQoL measurements as narrower concept had expanded as new quantitative methods were developed to assess health over the past 40 years. The measurement instruments were genuinely used to assess the impact of diseases on physiological, psychological and social domains of health [32].

The multidimensionality of HRQoL measures make it difficult to operationalize the term, though certain common domains commonly appear in measures: physical and mental health, social functioning, well-being. Though health utility methods rather capture physiological and psychological aspects of health, well-being is rather evaluated by capability measures. HRQoL measures have two approaches – preference and non-preference based methods (also referred as profile and decision theory) - to assess health state [1, 33].

Preference-based methods are built on a decision making theory, where people can explicitly address values to certain health states, expressed in a single score. The measurement attempts to generate a single index by weighting different dimensions of health. Health utility methods became the greatest part of preference-based HRQoL measurements. *Non-preference based* evaluation methods follow the multi-attribute utility (MAU) technique, where people rate specified domains of health using questionnaire items with a defined scoring system. This psychometric measurement creates an individual profile about the evaluated dimensions that can be summarized into a health status score.

<table>
<tr><td colspan="5">HRQoL measurement type</td></tr>
<tr><td colspan="2">preference-based (choice task)</td><td colspan="3">non-preference based (MAU instrument)</td></tr>
<tr><td>Direct utility assessment</td><td>Contingent valuation</td><td colspan="3">Indirect utility assessments</td></tr>
<tr><td>time-trade off method (TTO)</td><td rowspan="2">willingness to pay</td><td>generic HRQoL instruments</td><td>disease-group specific instrument</td><td>disease specific instruments</td></tr>
<tr><td>standard gamble (SG)</td><td rowspan="3">e.g.: EQ-5D, Health Utility Index (HUI), Short Form-6D (SF-6D)</td><td rowspan="3">e.g.: Patients Health Questionnaire-9 (PHQ-9), Dermatology Life Quality Index (DLQI)</td><td rowspan="3">e.g.: Beck Depression Inventory (BDI), Hamilton Depression Scale (HADS)</td></tr>
<tr><td>discrete choice experiment (DCE)</td><td rowspan="2">willingness to accept</td></tr>
<tr><td>best-worst scaling (BWS)</td></tr>
</table>

Table 6 - Typology of HRQoL measurements

Note: The typology is based on [34] and does not list all instruments.

2.2.2. Quality-adjusted life years

QALY is a standard health outcome measure, very frequently used in health economic evaluations to quantify health gain. QALY combines in a single measure life years (mortality) adjusted by health quality weights (mortality), enabling direct comparisons across disease states and intervention programs. The quality weight is measured by a preference-based health utility elicitation method. One QALY is equivalent to one year spend in full health [4].

If an individual lives 10 years in full health, plus 10 years in an imperfect health state with a utility of 0.4 than dies, that results in 10*1.0 + 10*0.4 = 14 QALYs. Suppose that the individual gets a treatment, where 10 years in full health is followed by 7 years spend in a utility of 0.8 that is followed by 5 years valued at 0.6 utility. The treatment created two additional life years, but 4.6 additional QALYs = 10*1.0 + 7*0.8 + 5*0.6 = 18.6 QALYs.

The increase in QALY can be seen differently due the combined two components: (1) quality of life could be improved; (2) quantity of life could be extended; (3) the quantity improved but quality decreased, (4) the quality improved but quantity reduced, though the society values most the intervention that (5) increases both quality and quantity of life.

Consequently, society faces a trade-offs between quality or quantity of life when choosing treatment. For example a cohort of 100 people has a life expectancy of 75.1 years which consist of 50 ill people living at the age of 66, 49 people makes beyond 85-years-old, and one baby dies at the age of 1. Evaluators inform the decision maker that according to the current

healthcare capacities two alternatives are offered having the same health outcome (both extend life, resulting 50 QALYs gained):

- Scenario A: Treatment for a rare disease, that saves the child, who can live up to 50 more years.
- Scenario B: Treatment for a common illness, that extends the life of 50 people by one year.

The QALY approach is the preferred outcome measure in country-specific guidelines for economic evaluations. Due it captures changes quality and quantity of life the measure is applicable in all population groups and diseases, and has strong theoretical foundation with the health utility theory QALY measurement is preferred in CUA and clinical outcome studies [35]. The QALY concept is in line with the rational economic theory assumptions, thus does not consider the societal distribution of accumulated health gain, namely which groups life is more valuable. The discussion (8.1) lists the limitations of the QALY concept in more details.

2.2.3. Well-being and capability measures

Health economic evaluations rely on HRQoL measurements, since these instruments accurately capture quality of health states. Capability approach was developed by Amartya Sen to provide an alternative evaluation framework of well-being beyond the standard utilitarian health assessment [36]. The framework emphasizes the importance of individuals ability and free will to achieve objectives they value; thus capability is defined as a non-health domain of well-being [37].

Capability instruments are developed to measure different dimensions of well-being, such as ability to progress in life, attachment to others, enjoyment safety, self-fulfilment, independence of decisions, etc. Preference and non-preference weighted forms of capability outcome measures are being extensively used in economic evaluations. A growing number of capability questionnaires and empirical studies also indicates the increasing interest towards capability assessment. In the past 15 years fourteen capability instruments were designed, some widely expanded internationally [38]:

- ICECpop CAPability for Adults/Elderly (ICECAP-A/O)
- Adult Social Care Outcomes Toolkit (ASCOT)
- Oxford Capabilities questionnaire – Mental Health (OxCAP-MH)

The ICECAP instruments are recommended for economic evaluations applied 22 times in the past decade [39, 40]. ASCOT shows excellent psychometric properties, more countries developed preference (direct utility) based valuation studies recently [41-43]. OxCAP-MH was specifically designed to assess capabilities in current mental health state. More language versions – among them Hungarian - are validated, though so far only disease specific studies were published [44-47], no large-sample general population study is available so far.

The capability-based well-being measurement is relevant from context of health economic evaluation that focuses of health production function to account for the benefits of an intervention. The results of such assessments can be misleading, when intervention effects go beyond physiological and psychological domains of health (like in long-term care), therefor more health-experts suggest capability-measurement outcomes to be considered in economic evaluations (for instance in social care) [48-50].

2.2.4. Patient-reported outcomes

PRO is defined as patient's self-reported health condition, where the outcomes cover a wide range of QoL/HRQoL assessments. The PRO measurement instruments require patients to rate their own-current health status, using a validated questionnaire. In practice, PROs are used in clinical research to compare health outcomes of interventions [51]. There is no recommended gold-standard PRO measure, therefor researchers use different instruments that seem to fulfil the criteria of responsiveness, validity and reliability [52].

2.3. Health state utility measurement

HSU measurement is a common HRQoL assessment method. Utility refers on quality of the health state, measured on a $>0 -1$ interval scale, where 1.0 means full health, 0 is equivalent to dead, while negative values represents WTD health states (when living any time in the given condition is considered to be worse than dying immediately) [5]. The lower limit of negative utilities is not -1, but dependents on the elicitation method, though, it is usually scaled symmetrically to >0. Health utility assessment methods can be divided into direct and indirect measurement categories [53].

2.3.1. Direct utility measurements

Utility is directly obtained if a preference-based valuation task is used. The direct HSU methods – most commonly time trade-off (TTO) [12, 54], standard gamble (SG) [55, 56], discrete choice experiment (DCE) [57, 58], best-worst scaling (BWS) [59] – are based on rational decision making theory assuming that people can clearly choose between two alternatives according to their preferences. Respondents (patients or general population) valuate self-experienced (own) or vignette-based (hypothetical condition described in the task) health states.

2.3.2. Indirect utility measurements

Indirect HSU measures use self-completed HRQoL questionnaires as a health state descriptive system, where items are rated with a scoring system (often a Likert-scale). The instruments item sum score represents the generic or disease(group)-specific HRQoL score. The item scores can be transformed into utilities using a value set that is based on direct utility measurement [60].

The value sets, that reflect the societal preferences towards specific health domains (e.g. self-care, pain, bodily integrity), are generated by performing the direct utility elicitation task using the descriptive system of the indirect HSU instrument and associate the direct utility values with the non-preference based HRQoL instrument item scores. The “tariff” or “weight” obtain with the direct task from the population is used to transform HRQoL item scores into a single utility score [61]. The most common generic preference-accompanied HRQoL instrument by far is the EQ-5D, studies often apply Health Utility Index (HUI), Short Form-6 Dimension (SF-6D) and Quality of Well-being Index (QWB) measurement instruments as well [62]. The most frequently used dermatology-disease group specific instrument is the Dermatology Life Quality Index (DLQI), while a common disease specific tool in major depression is Beck Depression Inventory (BDI).

Value sets can represent the preferences of patients or the general population. Usually, country specific value sets are used for economic evaluations, but it is often “borrowed” from a culturally similar society if national in not available.

Utility estimates yield systematically different results [63-65]. Generally direct methods output higher scores, but there are exceptions for instance in dermatological diseases [66-68]. The measurement tools methodological attributes tend to influence the utility scores as well [69-73]. Patients have different perceptions about heath states than general population, especially when valuing self-experienced vs. vignette-based conditions [74-76].

Sociodemographic characteristics – like age, education, gender, marital status, religious believes – also largely impact utilities [77-79].

2.4. Time trade-off method

The most commonly used direct utility elicitation method, strongly suggested for determining preference-weights for indirect utilities is the time trade-off (TTO) method. Due to its explicit relationship with QALY; taking into account its relative simplicity (as compared to SG), and better compliance with the theoretical axioms of economic evaluations (as compared to a visual analogue scale), TTO has become very popular among direct health state preference elicitation techniques over the past 30 years. Respondents opinion also support the application of TTO task [80, 81]. The number of empirical TTO studies exceed all other methods of direct utility elicitations by 2-3 times (nn PubMed; using the search: "time trade off" OR "time trade-off" OR "time tradeoff" OR TTO [all fields]; had n=2400+ search results in May 2023).

TTO task is designed to make the respondent choose between quality and quantity of life. The task offers to options: (A) to live for a 't' period of time in a "full" (better) health state or (B) or to live 'x' period of time in a particular "imperfect" health state. Traded years 't-x' represent the value that the respondent is willing to sacrifice for quality over quantity of life. Where the respondent's preferences are equal with respect to the two alternative A vs B health states, it is called the point of indifference (t=x). Utility is calculated directly from the indifference point, examples for the better and worse than dead utilities are presented in the next section [12, 54].

As a result of extended use, TTO has diverse methodological attributes [14]. The attribute of methods include: *TTO task type, mode of data administration, timeframe, iteration process, health state description, self-experienced or vignette-based health state, number* and *order of evaluated health states, task presentation* [82]. There is no defined guideline, because the method choices depend on purpose and circumstances of study, defined by the researchers. The considerable differences in methods - evidently impacting TTO utility results – have started a methodology-harmonization effort [83]. The EuroQol group who developed the most recommended and used – EQ-5D-3L/5L – generic HRQoL questionnaires for eliciting indirect utility, advised the new EQ-VT valuation protocol in 2012, replacing the previous Measurement and Valuation of Health (MVH) study quality check guide [84].

Protocol	Task type	TTO framework	Iteration process	Anchor state	Data collection	Respondent training
MVH	conventional	10 years	up/downward titration with 1 year increments	perfect health state, described with EQ-5D-3L	Face-to-face interview with paper administration	one warm-up valuation task
EQ-VT	composite	10 and 20 years	up/downward titration with 6 month increments	full health	Face-to-face interview with computer administration and visual aid	one example task and three practices

Table 7 - TTO method attributes by valuation protocols

Note: MVH=measurement and Valuation of Health (protocol), EQ-VT=EuroQol Valuation Technology (protocol), TTO=time trade-off (the guide is based on Oppe et al. [84])

Researchers must compromise between adjusting TTO method to the specific attributes of their studies and/or following a standard protocol to ensure the comparability of utility results. Using conventional or composite task type, timeframe setting, valuation of self-experienced or vignette-based health states, responding population, iteration process, anchor health state can all alter TTO utility outputs. TTO is challenged when evaluation acute (short therm) health conditions like a bone-fracture or seasonal flu. It is difficult to choose the right timeframe, while age and time preference of respondents makes it complicated (elderly people's life expectancy is unlikely to be 20-30 years, their discount rate regarding quality of life is potentially unlike than young adults).

2.4.1. Conventional and composite task

Conventional TTO type evaluates better than dead (BTD) health states, where the theoretical utility range is 0-1 [85]. In the TTO exercise, the respondent must choose between living 'x' years with self-experienced or vignette-based imperfect health state (such as moderate depression) or living 't' years with full health (the better health state) immediately followed by death. The task starts with the iteration, for example using a top down approach offering 10 years in full health vs 10 years in moderate depression. If the respondent accepts this, then the utility of moderate depression is equivalent to full health (1.0). If not, the step-by-step titration continues with decreasing the number of life years in the full health by standard time increments (usually 1 year) until the respondent can no longer choose between the two health states. For example, if the respondent is indifferent between spending 8 years in full health and 10 years

in moderate depression, the utility of moderate depression is calculated from the indifference point as the ratio of equivalent health states length of lifetime period: U= x/t = 8/10 resulting 0.8 (Figure 3). While the traded years 't-x = 10-8 = 2' represent the price (opportunity cost) that a respondent is willing to sacrifice for quality over quantity of life. To demonstrate relationship between TTO utility and QALY, the example shows that the alternative health states are equal in terms of QALY outcome:

8 years × utility of full health (1.0) = 8 QALY = 10 years × utility of moderate depression (0.8)

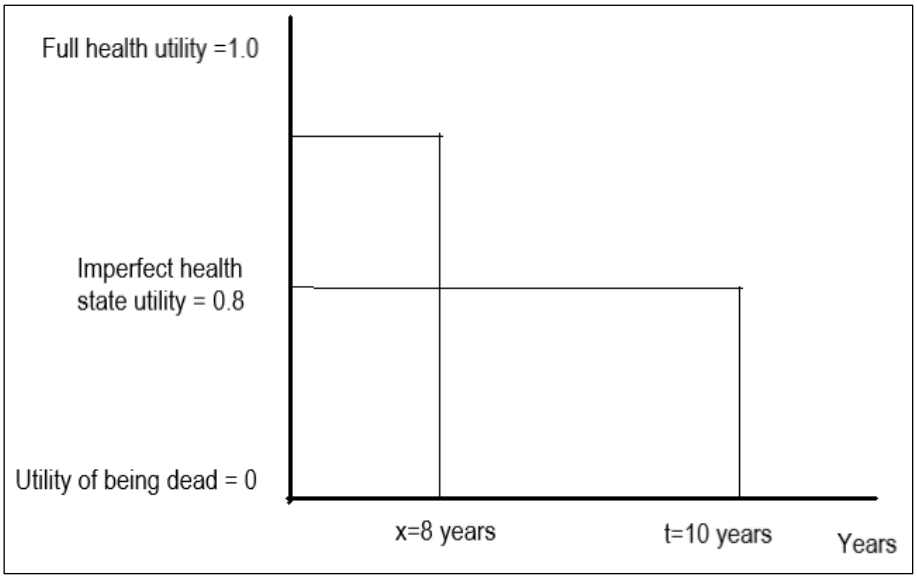

Figure 3 - Utility calculation of conventional TTO

Note: own editing, based on: [86]

In case of the conventional format the task ends if the respondent answers that 0 years (immediate death) in full health is equivalent to 10 years in the described 'moderate depression' health state, resulting 0 utility.

The *composite* (also referred as lead-time/lag-time) *time trade-off (cTTO)* presents another task if the respondent chooses not to live any years in the imperfect health state, assuming that there are worse than dead health states possible. Composite TTO is an advanced version of the conventional TTO type, meaningfully different in besides valuing better than dead states it offers valuation of worse than dead health states [15]. Composite task also starts with valuing better than dead format, however, resumes if the respondent choose immediate

death (0 years in full health = 10 years in moderate depression) [87, 88]. To continue the previous example task, in this WTD format task initially offers 10 years in full health has to be compared to 10 years in full health followed by 10 years in moderate depression. Now the respondent would sacrifice three years in this format. The health utility of 'moderate depression' in WTD format is calculated as: U = (x – t) / (2t-t) = (x-t)/t = (7-10)/10 = -0.3 (so the individual was willing to sacrifice 13 out of 20 years of full health and indifferent between living 7 years in full health or 10+10 years in the combined health, as presented on Figure 4).

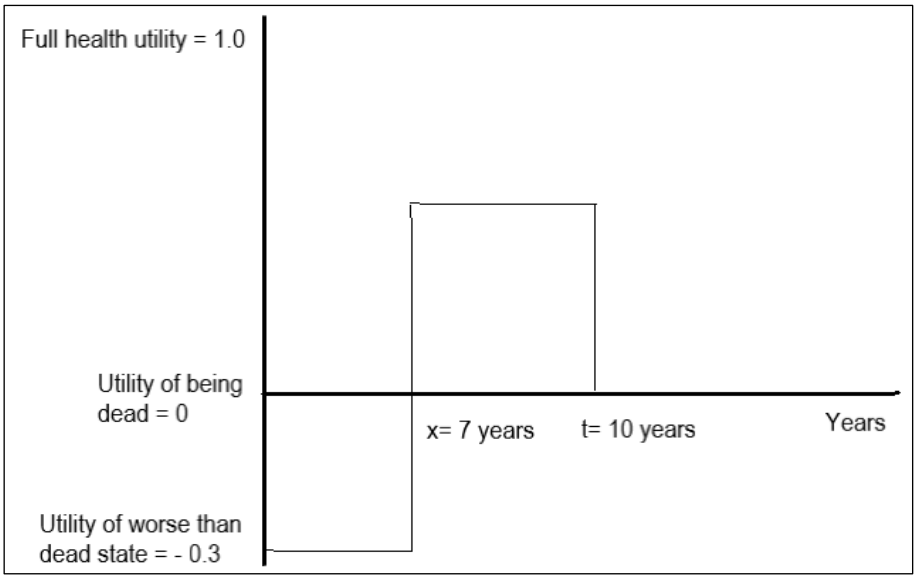

Figure 4 - Utility calculation of composite TTO

Note: the graph is based on: [15]

Using the QALY notation, the health outcome of the two health states are equally 7 QALYs and can be expressed as:

$$7\ years \times\ utility\ of\ full\ health\ (1.0) =$$
$$= 10\ years\ in\ full\ health$$
$$+ 10\ years \times (-0.3)\ utility\ of\ moderate\ depression\ (\frac{7-10}{10} years)$$

The lead/lag time typically uses equal duration for full and imperfect health, usually 10+10 years, but not necessarily (see Figure 5) [89]. Theoretically, the cTTO utility falls between minus infinity and 0 (while the imperfect state in WTD format plus the timeframe life spent in full health add up to zero) [90].

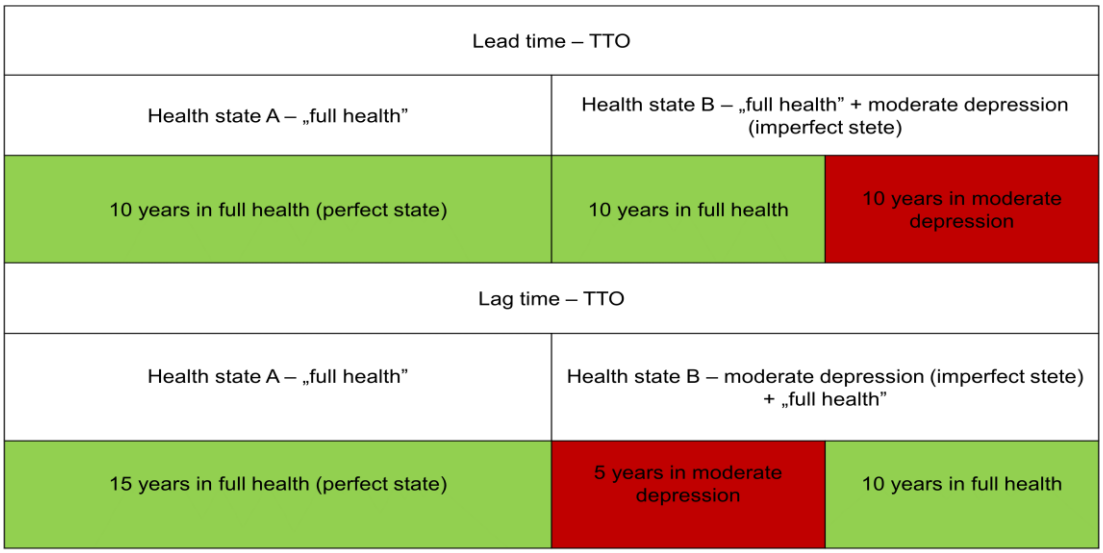

Figure 5 - Lead and lag time composite TTO variants

2.4.2. Vignette-based and self-experienced assessments

TTO task may elicit HSU score of the self-experienced health state or a vignette-based (often referred as “hypothetical”) health state [70]. Self-experienced state is usually valuated as the current state of patients, but they can evaluate vignette-based states too. General population principally values vignette-based health states that they have to imagine based on the health state describing vignette.

Health state vignette is the tool applied to describe the given health state in the TTO task. The aim of the vignette is to depict the disease (or health condition) as precisely as possible [91]. Vignettes include a description of disease specific attributes as health domains while simultaneously differentiating the stages of disease severity [92]. Development has three main practices: (1) literature review/scoping, (2) consultation with health professionals, or (3) the combination of the two processes. Differences in similar health state descriptions, like ‘mild depression’ are possibly influence the imagination of health state, leading to systematic shifts in evaluation [93].

McSad-based vignette		**Hamilton-based vignette**		**PHQ-9 based vignette**	
Domain	Description	Domain	Description	Domain	Description
Role function	Frank finds little pleasure	Depressive mood	Feels sad frequently.	Happiness: little interest in doing things	not at all

Sleeping	in doing things. He often wakes up at night, and has less appetite than usually. Frank considers himself average, nowadays there are fast shifts in his mood. There are only a few things that he enjoys. He rarely meets with friends. He does not like to think on the future.	Feeling guilt	Sometime feels letting down on people.	Mood: feeling down or depressed	several days
Eating		Suicidal attempt	Thinking that there is nothing to live for.	Sleeping: trouble falling or staying asleep	several days
Self-appraisal		Sleeping disorder	It is hard to fall asleep all night.	Agility: feeling tired	several days
Emotions/reactions		Agility	Often being feeble and unhappy.	Eating: poor appetite or overeating	not at all
Happiness		Anxiety	Feeling stressful and irritated	Self-appraisal: feeling bad about yourself	more than half of the days
Cognition		Physiological symptoms	Having no appetite many times.	Foci: trouble concentrating on things	not at all
Behaviour/relations		Somatic symptoms	Usually feeling symptoms of a disease.	Role function: moving more slowly or fidgety than usual	several days
-				Suicidal attempt: thoughts that you would be better off dead	not at all

Table 8 – Different health state descriptions of 'mild depression' by vignettes.

Description vignettes may differ in terms of *covered disease domains/dimensions; number of domains, presentation of the description, origin/base of the description tool*. Health state description is extremely important to elicit accurate HSU scores, several studies - controversially - report that presentation of the valuation task has an impact on the utility score [94, 95]. (Table 8)

2.4.3. Patients and general populations valuations

Countries practices differ in considering patients or general populations perspective in the context of health economic evaluations. Increasing number of analysis and guidelines suggest that health technology assessment (HTA) should include utilities based on both patients and general populations perspective. The literature argues that patients persistently report higher

HSU scores regarding their self-experienced health state, practically impacting the cost-effectiveness analysis by reducing the QALY gain associated with the intervention [96-98].

The discrepancy between patients and general populations health perception opens an other research gap. Since the MAU-based indirect utility instrument scores are transformed using the weight of a preference-based direct utility assessment, general populations values are used to obtain patients utility scores [99]. When valuing vignette-based (hypothetical) health states patient vs general population utility scores differ according TTO, SG, DCE measures [75]. There are a number of reasons of difference in perceptions to mention [100, 101]:

a. understanding the health state description differently
b. patients adapt to health states, while it is hard to imagine for non-ill people
c. healthy respondents do not consider all aspects of never experienced health states
d. patients forget what it is like to live in full health

According to current reviews of HTA guidelines most countries (e.g. Australia, England, France, Hungary, Spain) prefer preference weights of the general population, for (1) they fund the healthcare systems, (2) they have no interest in valuing imperfect health states and (3) death and full health anchor in tasks raises concerns in case of patients who actually experience the health state [81].

2.4.4. Mapping of measurement scores

Many studies use different measurement instruments to assess HRQoL, therefor researchers often use statistical tools, like mapping algorithms to convert one measurement's score into a more commonly used HSU value (like EQ-5D) [102]. If respondent's item level HRQoL instrument scores are available, these can mapped to health utility scores, often referred as 'crosswalk' HRQoL score into utilities [103].

There is huge demand for mapping algorithms, since many researchers use non-preference accompanied measures of HRQoL. Mapping algorithms are regression based prediction models (e.g.: ordinary least square, tobit, two-part, generalized linear or Bayesian models), setting the target measurement score as dependent and base instrument item responses as explanatory variables [104].

3. Research objective and hypothesis

Healthcare resource allocations rely on health economic evaluations, to aid decision-making based on efficiency. Health outcome most frequently is expressed with QALY measure, that requires HSU assessment to quantify health gain by adjusting life years by quality of life. Direct and indirect utility measurement methods yield systematically different utility results, that influence the evaluations.

Differences in methodological attributes of HSU measurement instruments (like health state description; responding population, timeframe, task type, etc.) and individual factors (such as sociodemographic characteristics, disease-specific variables) potentially yield different results. Recent thesis aims to explore the impact of methodological and individual attributes on health state utility results. Beyond the evaluation of health quality, the well-being is further assessed with capability approach as an alternative of the QALY concept. The research hypothesis states that:

- H1: Methodological attributes (task type, health state description, timeframe, responding population) make impact on TTO utilities.
- H2: Individual characteristics (age, gender, level of education, employment status, marital status, disease severity) have effect on utility estimates.
- H3: Direct (TTO) and indirect (EQ-5D) HSU measures yield systematically different results.
- H4: Mental health capability measure (OxCAP-MH) effectively capture non-health domains of well-being.

Note: The following four chapters (4-7) present the methods and results of four of Our previously published two articles [105, 106] and two handed-in manuscripts [107, 108] are presented. As main author (Péter Balázs) in the four publications, several parts of the studies were used in this thesis.

4. Systematic review of time trade-off measurements in Hungary

4.1. Introduction of the review study in Hungary

The chronic diseases cause long lasting societal and patient level impairments, posing burden on the healthcare system. Chronic diseases may reduce both HRQoL and life expectancy, utility measurement is suitable to examine the magnitude of the disease burden. Hungary finances the healthcare system out of public social insurance contribution, financing decisions therefore should reflect domestic societal preferences. Indirect utility measurements have been reviewed previously, but directly measured utility values are not summarized in Hungary. The most often used direct HSU measurement − that most Hungarian health evaluation submissions also applied − is the TTO method [109]. Therefore, the objective of Our previously published review was to collect all direct HSU measurement studies by TTO in Hungary [105].

4.2. Methodology of systematic literature review in Hungary

A keyword-based literature search related to time trade-off method following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline was conducted in January 2020 [110]. Two international (PubMed, Web of Science) and one domestic (Matarka) online databases were searched.

Study characteristics related information of sample size, responding population, mean age, gender proportion and TTO method attributes as type, health state description, timeframe, number and order of health states, mode of data administration were retrieved from all studies. Utility – mean, standard deviation, median, inter quartile range, proportion of ‘1’ and ‘0’ responses – were extracted uniformly if available. Proportion of non-traders was examined, defined as those respondents who rated all health states equivalent to full health (not willing to sacrifice any life years for exchange of life quality) in the TTO study [111].

4.2.1. Search terms and strategy

The search strategy using no language restrictions was developed based on combination of search terms and constructing a "search filter" as suggested for HSU studies [112]:

#	Search terms

1	“time trade-off” OR “time trade off” OR “time tradeoff” OR TTO
2	“utilities” OR (“utility” AND (health OR index OR indices OR measure* OR weight* OR score* OR value* OR gain*))
3	“quality-adjusted life year” OR QALY
4	“illness state*” OR “health state”
5	“health gain*”
6	#1 OR #2 OR #3 OR #4 OR #5
7	“Hungary” OR “Hungarian”
8	#6 AND #7

Table 9 - Search strategy of TTO studies in Hungary

4.2.2. Study selection criteria

The review selected studies that met three inclusion criteria: (1) original studies, which (2) measure utilities with TTO method, (3) using a Hungarian sample. Publications that are non-original (reviews, abstracts), or not measured TTO utilities empirically, or had no Hungarian sample results available independently were excluded. The search hits were independently screened by two reviewers, first by title and abstract, then the included studies were read in full-length.

4.3. Results of systematic literature review in Hungary

Overall, n= 643 hits were found in the three online databases, n=83 were removed as duplicates, n=551 were screened by title and abstract. Exclusion criteria sorted out n=520 publications, n=31 were full-text assessed. Finally, n=9 articles were included, according to the three inclusion criteria [113-121]. (Figure 6) [105]

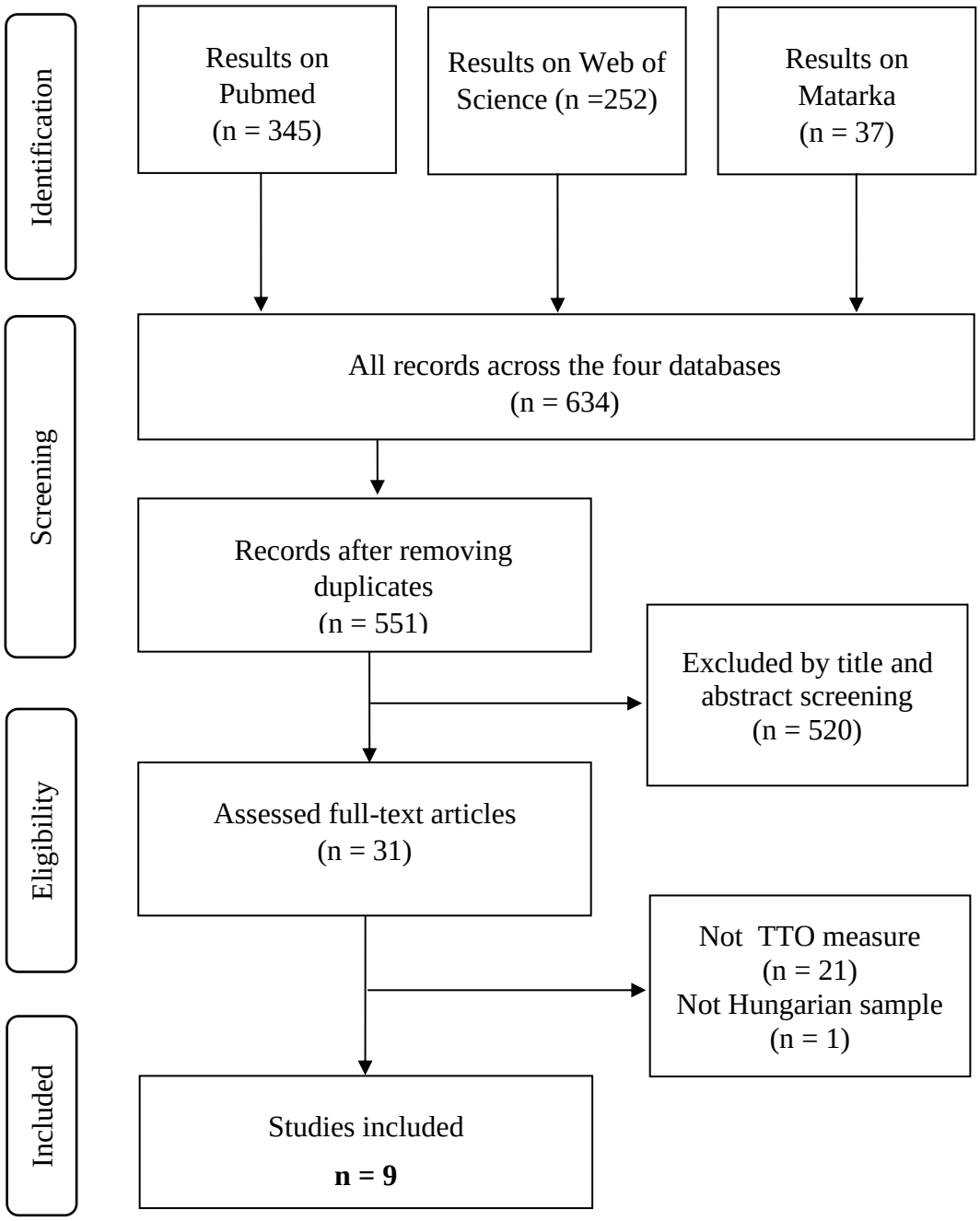

Figure 6 - PRISMA flowchart of the systematic review of Hungarian TTO studies

4.3.1. Results of selected studies

In total, the nine included articles publish the results of seven original TTO studies. The studies published between 2012-2019 investigate HRQoL and HSU of seven chronic diseases: rheumatoid arthritis, chronic migraine, pemphigus, psoriasis, primer dysmenorrhea, age-related macular degeneration and Crohn's disease. One study investigated dermatological health states described with vignettes based on DLQI instrument among patients and general population.

All studies were observational cross-sectional, sample size varied between 108 and 1996. Mean age of samples widely ranged from 25.6 to 75.2, likewise the proportion of women 32.1-100%. Two studies had general population sample, two had patient sample, three measured utility among both general population and patients. The method of data collection was paper-based in five cases (n=5; 71%), once online (n=1; 14%) and one study collected data online and paper based combined. All studies used self-administrated mode, in one case complemented with focus group interviews. (Table 10)

Study	Year	Disease	Population	Sample size	Women's proportion (%)	Data collection method
Inotai et al. [113]	2012	Rheumatoid arthritis	patients	168	85.5	paper based self-administrated
Rencz et al. [114]	2015	Chronic migraine	general population (self-reported patients among them)	180	71.1	paper based self-administrated
Rencz et al. and Hajdu et al. [115, 120]	2015 and 2019	Pemphigus (and disease described with DLQI)	general population	108	58.3	paper based self-administrated (and group interviews)
			patients	108	64.8	
Rencz et al. and Poór et al. [116, 117]	2016 and 2017	Psoriasis	general population	308	68.6	paper and online based self-administrated
			patients	238	27.4	
Péntek et al. [119]	2017	Macular degeneration	patients	122	62.3	paper based self-administrated
Rencz et al. [118]	2017	Primer dysmenorrhea	general population (self-reported patients among them)	1996	100	online self-administrated
	2019		patients	206	45.1	

Rencz et al. [121]		Crohn's disease	general population	221	32.1	paper based self-administrated

Table 10 - Study characteristics of TTO studies in Hungary

4.3.2. Results of TTO utility measurement in Hungary

Conventional TTO type was applied three times, three times the simplified – indifference in one answer – version of conventional TTO was used, one study evaluated both BTD and WTD formats with composite cTTO type. Two studies measured vignette-based HSU in primer dysmenorrhea and chronic migraine of general population. Two studies evaluated only patients self-experienced health states in rheumatoid arthritis and age-related macular degeneration. Three studies elicited both vignette-based and self-experienced TTO utilities among patients and general population in pemphigus, psoriasis and Crohn's disease. The timeframe was set to the common 10-years in three studies, subjective life expectancy with not exactly defined time period was applied in two cases, while the in the cTTO task the standard 10-years in BTD 10+10 years in WTD format was followed. Most studies used BTD format (n=6; 86%), one applied composite TTO task with BTD+WTD health states possible. The smallest tradable unit of the iteration was either 6 month or 1 year (but not reported in three studies). The number of TTO task health states presented in the surveys varied between 1 and 7, most studies (n=5; 71%) fixed, while two randomized the presentation order of the health states. (Table 11)

Study	TTO type	Health state description	Timeframe	Smallest tradable unit	Format	Number of examined health states	Presenting order of health states
Inotai et al. [113]	indifference in one answer	self-experienced	subjective life expectancy	n/a	BTD	1	fixed
Rencz et al. [114]	indifference in one answer	vignette-based	20 and 80 years	1-year	BTD	2	fixed
Rencz et al. and Hajdu et al. [115, 120]	composite	vignette-based and self-experienced	10 years BTD and 10+10 years WTD	6-month	BTD + WTD	3 vignette-based + 1 self-experienced	fixed
Rencz et al. and Poór et al. [116, 117]	conventio-nal	vignette-based and self-experienced	10 years	6-month	BTD	6 vignette based + 1 self-experienced	randomized
Péntek et al. [119]	indifference in one answer	self-experienced	subjective life expectancy	n/a	BTD	1	fixed

Rencz et al. [118]	conventional	vignette-based	10 years	n/a	BTD	2	fixed
Rencz et al. [121]	conventional	vignette-based and self-experienced	10 years	6-month	BTD	4 vignette-based + 1 self-experienced	randomized

Table 11 - Attributes of TTO method among Hungarian studies

Note: BTW/WTD refer to better/worse than dead health states

4.3.3. Health utility results of TTO studies in Hungary

Overall, 35 health states – five self-experienced disease states and thirty vignette-based hypothetical health states – were evaluated in the seven included studies. Altogether, 45 TTO utilities of patients and general population were extracted. The mean utilities ranged from 0.34 (uncontrolled pemphigus vulgaris, vignette-based evaluation of general population) to 0.94 (mild primer dysmenorrhea, vignette-based evaluation of affected general population). The rate of non-traders among those who completed the TTO task was 0-29%. (Table 12)

Study and year	Health state	Description and population	n	mean	SD	median	IQR	non-traders
Inotai et al. 2012 [113]	rheumatoid arthritis	self-experienced (patients)	168	0.77	0.21	n/a	n/a	n/a
Rencz et al. 2015 [114]	migraine 2x weekly-for 20 years	vignette-based (affected population)	110	0.80	0.26	0.80	n/a	11.0%
		vignette-based (healthy population)	70	0.77	0.27			
	migraine 2x weekly-for 80 years	vignette-based (affected population)	110	0.82	0.20	0.86		
		vignette-based (healthy population)	70	0.85	0.12			
	migraine 2x monthly-for 20 years	vignette-based (affected population)	110	0.84	0.27	0.90		
		vignette-based (healthy population)	70	0.84	0.26			
	migraine 2x monthly-for 80 years	vignette-based (affected population)	110	0.90	0.15	0.92		
		vignette-based (healthy population)	70	0.88	0.12			
Rencz et al. 2016 [115]	uncontrolled pemphigus vulgaris	vignette-based (general population)	108	0.34	0.38	0.40	0.27-0.41	0%
	uncontrolled pemphigus foliaceus		108	0.51	0.32	0.50	0.45-0.57	
	controlled pemphigus		108	0.75	0.31	0.80	0.69-0.81	
	pemphigus (own-current)	self-experienced (patients)	90	0.76	0.30	0.90	n/a	7.4%

Hajdu et al. 2019 [120]	uncontrolled pemphigus vulgaris	vignette-based (general population)	86	0.41	0.45	0.40		
	uncontrolled pemphigus foliaceus		89	0.52	0.42	0.50		
	controlled pemphigus		91	0.66	0.36	0.80		
Rencz et al. 2016 [116]	mild-DLQI state A	vignette-based (general population)	125	0.64	0.32	0.80	n/a	14.1%
	mild-DLQI state B		130	0.75	0.27	0.85		
	mild-DLQI state C		126	0.62	0.30	0.70		
	moderate-DLQI state A		124	0.66	0.31	0.80		
	moderate-DLQI state B		127	0.64	0.28	0.70		
	moderate-DLQI state C		125	0.59	0.29	0.60		
	severe-DLQI state		125	0.56	0.29	0.60		
Poór et al. 2017 [117]	psoriasis (own-current)	self-experienced (patients)	238	0.93	0.16	n/a	n/a	29.0%
	mild-DLQI state A	vignette-based (general population)	103	0.89	0.19	0.95	0.90-1	
	mild-DLQI state B		102	0.91	0.15	0.95	0.90-1	
	mild-DLQI state C		99	0.85	0.20	0.95	0.80-1	
	moderate-DLQI state A		101	0.83	0.22	0.90	0.75-1	
	moderate-DLQI state B		100	0.85	0.22	0.85	0.80-1	
	moderate-DLQI state C		96	0.84	0.20	0.93	0.70-1	
	severe-DLQI state		98	0.84	0.20	0.90	0.80-1	
Péntek et al. 2017 [119]	age-related macular degeneration	self-experienced (patients)	122	0.72	0.30	n/a	n/a	n/a
Rencz et al. 2017 [118]	mild primer dysmenorrhea	vignette-based (healthy population)	159	0.93	0.20	1.00	n/a	24.0%
		vignette-based (affected population)	1808	0.94	0.18	1.00		
	severe primer dysmenorrhea	vignette-based (healthy population)	159	0.83	0.27	0.90		
		vignette-based (affected population)	1802	0.85	0.26	0.95		
Rencz et al. 2019 [121]	Crohn's disease (own-current)	self-experienced (patients)	201	0.83	0.28	1.00	0.80-1	24.3%
	mild Crohn's disease	vignette-based (general population)	221	0.88	0.25	1.00	0.90-1	11.8%
		vignette-based (patients)	199	0.87	0.26	1.00	0.90-1	
	mild Crohn's disease-with fistulas	vignette-based (general population)	221	0.80	0.26	0.90	0.80-1	
		vignette-based (patients)	193	0.80	0.29	0.90	0.70-1	

	severe Crohn's disease	vignette-based (general population)	221	0.65	0.29	0.70	0.50-0.90	
		vignette-based (patients)	199	0.73	0.31	0.90	0.50-1	
	severe Crohn's disease-with fistulas	vignette-based (general population)	221	0.59	0.31	0.65	0.40-0.85	
		vignette-based (patients)	197	0.69	0.33	0.80	0.50-1	

Table 12 - TTO utility catalogue of Hungarian studies

The proportion of respondents who were not willing to sacrifice any life years (answered to each impaired health state as equivalent to full health '1') varied between 1-58% among general population and 14-51% among patients. The rate of "being dead or worse than dead" equivalent utilities (who answered being indifferent between death '=0' or would rather die than live in the given impaired health state '<0' ranged 0-12%. Out of 13 comparable vignette-based health states, patients TTO utility was higher in seven cases, equal in two and lower in two cases compared to general population values. (Figure 7)

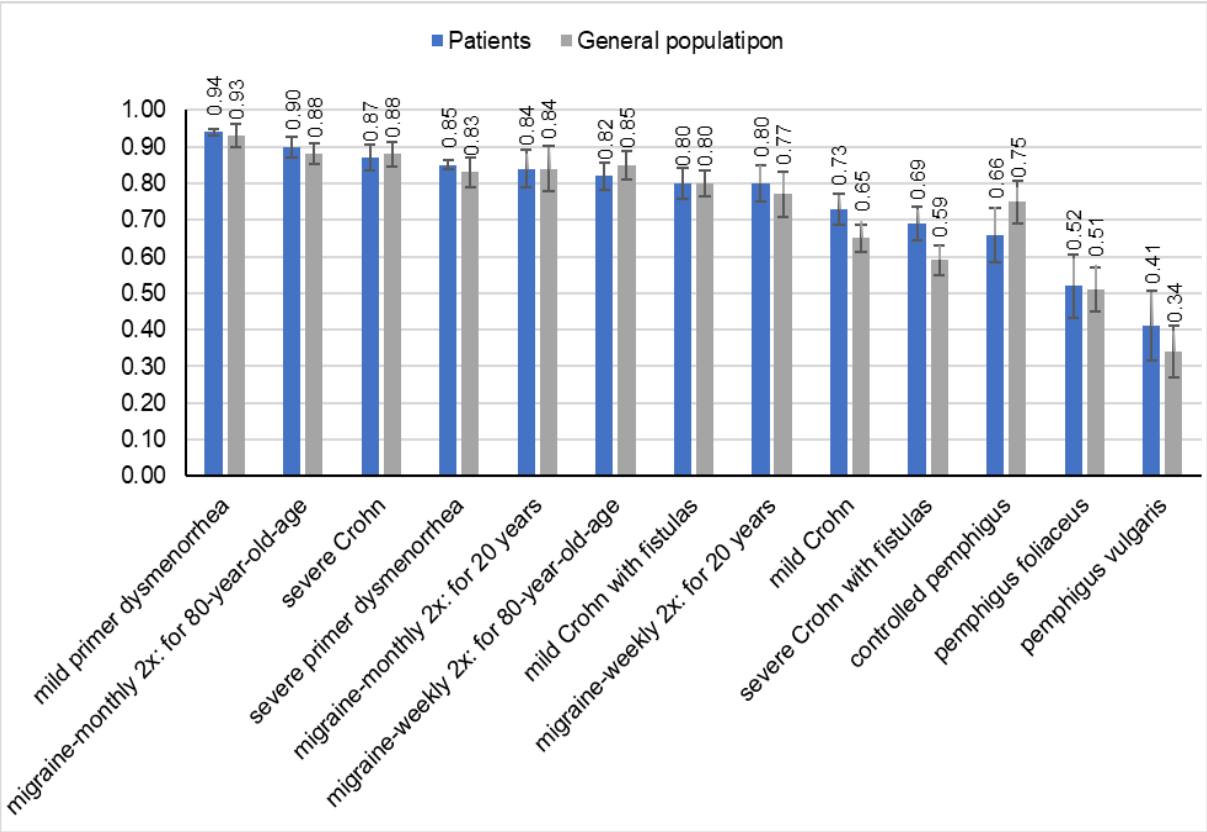

Figure 7 - Comparison of vignette-based TTO utilities between patients and general population

Note: mean utility and 95% confidence interval are illustrated

5. Review and meta-analysis of time trade-off measurements in depression

5.1. Introduction of the depression review

Depression is a common mental disorder, affecting 260+ million people worldwide. It has a broad range of disease severity levels. The degree of depression intensity and form varies (e.g., major depression, bipolar disorder, affective disorder), and it is characterized by typical physiological and mental symptoms [122, 123].

Depression can disrupt role functions, causes sleeping and eating dysfunctions, impair emotional and cognitive functions, harm individuals' self-assessments. Depression is considered as a chronic health condition that may lead to deterioration in HRQoL or decline in subjective perceptions of social, occupational and individual well-being [124, 125].

There are multiple disease specific HRQoL instruments – besides the direct and indirect HSU assessments – to measure the impact of depression on people's health. Hamilton Depression Rating Scale (HADS), Beck Depression Inventory (BDI), Patient Health Questionnaire-9 item (PHQ-9) are all valid and reliable tools, broadly used in clinical research to score depression severity [126-128].

Although there are several direct HSU studies in depression, the literature lacks a review to synthetise findings on utility values. The primary aim of Our previously published review therefore was to catalogue TTO utility estimates in depression-related health states [106].

5.2. Methodology of systematic review in depression

A systematic literature search was conducted in November 2020 and updated in March 2022 following the principles of PRISMA protocol [110]. Four online databases were searched: PubMed, Web of Science, PsycINFO and Cochrane Database of Systematic Reviews. No language or publication date restrictions were applied.

5.2.1. Search strategy

The search strategy was developed as a combination of TTO and depression search terms. The following terminological variants were used: 'time trade-off' or 'time tradeoff' or 'time trade off' or 'TTO' and 'depression'. Manual search among the reference list of included studies was

employed to track all potential eligible studies. The original search conducted in November 2020 was updated in May 2022, to find potentially relevant new studies. (Table 13)

#	Search terms	PubMed	Web of Science	Cochrane	PsycInfo
#1	"time tradeoff"[All Fields] OR "time trade-off"[All Fields] OR "time trade off"[All Fields] OR "TTO"[All Fields]	2,223	3,427	292	461
#2	"depressed"[All Fields] OR "depression"[MeSH Terms] OR "depression"[All Fields] OR "depressions"[All Fields] OR "depression's"[All Fields] OR "depressive disorder"[MeSH Terms] OR ("depressive"[All Fields] AND "disorder"[All Fields]) OR "depressive disorder"[All Fields] OR "depressivity"[All Fields] OR "depressive"[All Fields] OR "depressively"[All Fields] OR "depressiveness"[All Fields] OR "depressives"[All Fields]	567,587	581,299	87,719	376,874
#3	#1 AND #2 (original search: November of 2020)	104	107	21	32
updated search (March or 2022)		121	124	22	39

Table 13 - Search strategy of TTO studies in depression

5.2.2. Method of study selection

The results of four databases were summarized in Excel. After removing duplicate studies, two independent reviewers screened the titles and abstracts of selected articles. Eligibility according to three inclusion and eight exclusion criteria was based on full-text assessments of articles. The included articles must be (1) empirical studies, (2) TTO studies valuing depression or depressed states and (3) TTO utility obtained from samples of patients, health professionals, general population or combined population. The exclusion criteria were: (i) no English abstract, (ii) English abstract of non-English full text paper, (iii) not a journal article (e.g., abstract, editorial letter, review), (iv) animal/in vitro/in silico or other preclinical study, (v) abstract non-related to the relevant disease, (vi) abstract was a review/secondary search, (vii) EQ-5D valuation studies, (viii) use of TTO to elicit utility for health states described by EQ-5D descriptive system.

5.2.3. Data extraction

General information regarding the articles and study were summarized in terms of (1) authors, (2) years of publication, (3) study settings, (4) countries, (5) main study objectives and (1) study population; (2) sample size; (3) proportion of women; (4) sample age; (5) data collection method. Further data was extracted related to the TTO method and health state vignettes: (1) TTO type; (2) time frame; (3) state description; (4) iteration process; (5) format of health state evaluation and (1) presentation of description used; (2) number of health domains in the description; (3) origin of the health state vignette; (4) number of health states assessed.

A TTO utility catalogue in depression was assembled collecting mean (and SD) utility values of self-experienced or vignette-based depression states.

5.2.4. Study quality assessment

Quality assessment of TTO measurements was performed by monitoring the TTO task-related information using seven aspects of the EuroQol's earlier Measurement and Valuation of Health (MVH) valuation protocol that regards the TTO task [84]: *framework, time horizon, anchor state, iteration algorithm, mode of administration, method of data collection and respondent training*. A scoring system was set to evaluate the studies quality, +1 point if met the criteria, -1 if not, 0 if the information was not reported. The reason of using the former MVH protocol is that most studies were conducted before the launch of the current EQ-VT protocol in 2012.

5.2.5. Method of meta-analysis

Meta-analysis was conducted using random effect (REML) model to evaluate mean utility estimates. Two eligibility criteria were established: (1) the pooled utility of mild, moderate, and severe depression (2) described by vignettes was included in the meta-analysis, ensuring the consistency of comparisons (done only between vignette-based health states). The missing standard deviation (SD) data were replaced by the sample size-weighted average $\bar{u} = \frac{\sum u_i \times n_i}{\Sigma n_i}$ of reported SDs [129].

The mean TTO utility of vignette-based mild, moderate, and severe depression - were compared across depressed and nondepressed populations. Meta-regression was used to estimate the effects of four binary coded variables on utility values: (1) vignette type (McSad or other); (2) population group (depressed or nondepressed); (3) method of data collection (self-

completed or interviewer-administered); and (4) depression severity (mild or severe). Heterogeneity was tested using I2, which measures the proportion of observed variation between studies, where the differences were expressed in terms of utility estimates (0-100%). The dispersion between studies (variance of utilities) was estimated by computing T^2 and Tau [130]. The meta-analysis and the meta-regression were performed using Stata 16.0 (StataCorp LLC).

5.3. Results of systematic review in depression

Overall, 306 records were found in the four databases in the updated search (May 2022): PubMed: 121; Web of Science: 124; PsycINFO: 39; Cochrane: 22. Duplicates (n=142) were removed, the abstracts and titles of 164 publications were screened, of which 146 were excluded. Accordingly, 18 articles were included for full text analysis, where 4 articles were based on the inclusion criteria (not in English = 1; not focused on depression = 2; design for an unrealized trial = 1). The updated search found 42 new records compared to the previous in November 2020 (PubMed: 104; Web of Science: 107; PsycINFO: 32; Cochrane: 21), but no additional eligible studies completing the inclusion criteria were found. [131-134].

In total, 14 articles met the inclusion criteria, but two publications were merged, while reporting the results of the same study[106]. An additional empirical TTO study was found by hand searching the reference lists of included articles, resulting in 14 included individual studies [135-149]. (Figure 8, based on [110])

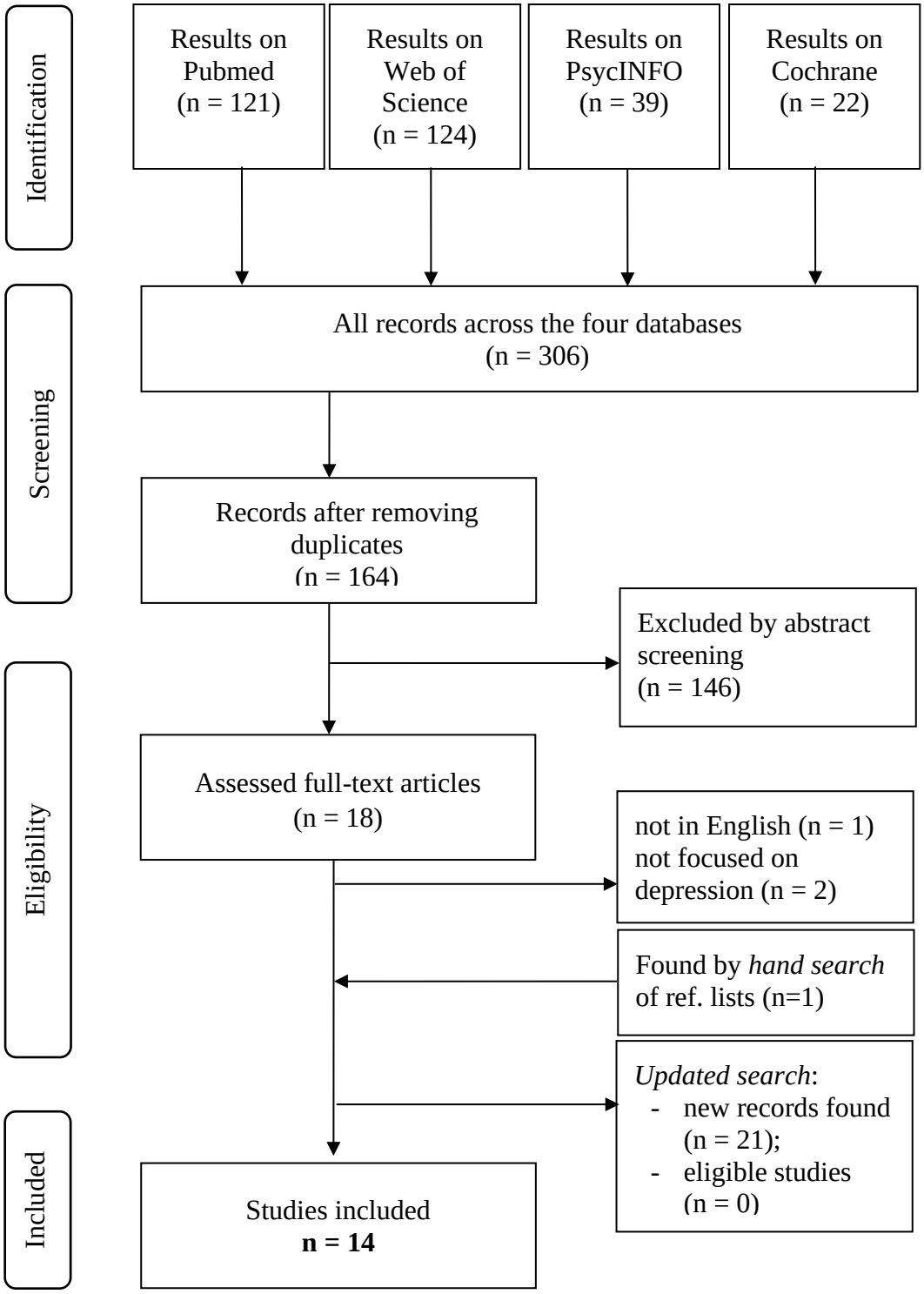

Figure 8 - PRISMA flowchart of systematic review of TTO studies in depression

5.3.1. Results of studies characteristics in depression

The study settings, patient characteristics, depressive health status, data collection, and description of vignettes were heterogeneous across TTO studies. Studies were published between 1991 and 2020, covering 9 countries: 4 in USA, 2 in Canada, 2 in the Netherlands and one each in the United Kingdom, Thailand, Australia, Sweden, Spain and Germany. Majority of the studies examined patient samples with depression (n=9; 64%). Three studies (n=3; 21%) included the general population, one study focused on health professionals and one study covered a mixed populations of patients, general population and health professionals. The study designs were mostly cross-sectional (n=11, 79%), randomized controlled trials (n=2, 14%), and case control study (n=1, 7%). The most frequently applied data collection methods were semi-structured interviews (n=8, 57%), paper-based self-completion questionnaires (n=4, 29%), and in two cases online self-completed questionnaires (n=2, 14%). The sample sizes varied widely between 32 and 3,986; similarly, the mean age of respondents varied between 32.0 and 52.8 years (two studies reported only the age range: 20-64 years). The proportions of women ranged 11.5-93.5%. Nine studies (n=9; 64%) reported the utilities of depression patients (physician or self-diagnosed), two studies (n=2; 14%) focused on the nondepressed population (general population, health professionals), and three investigated (n=3; 21%) both the depressed and the non-depressed population. (Table 14)

Study	Study setting	Main objective	Country	Study population	Data collection	Sample size	Women %	Age (mean/ range)
self-experienced health valuing TTO studies								
Oldridge et al. 1991, 1993 [135, 136]	Randomized controlled trial	Evaluating clinical effectiveness of rehabilitation	Canada	Mildly/moderately depressed patients diagnosed with myocardial infarction	Semi-structured interviewer administrated	165	11.5%	52.8
Wells et al. 1999 [137]	Cross-sectional	Compare both HRQoL and utility for current health	USA	Self-reported (on DSM-IV) depression outpatients	Semi-structured interviewer administrated	750	62.0%	n/a
Tsevat at al. 2000 [138]	Cross-sectional	Measure health values of patients with bipolar disorder	USA	Treatment receiving bipolar disorder patients	Semi-structured interviewer administrated	53	62.0%	43.0

Voruganti et al. 2000 [139]	Case-control	Explore the feasibility of traditional utility evaluation techniques	Canada	Diagnosed major depression patients (on DSM-	Semi-structured interviewer administrated	32	53.3%	43.6
Sherbourne et al. 2001 [140]	Randomized controlled trial	Evaluating the responsiveness of seven HRQoL measures	USA	Clinician assessed patients with depressive symptoms	Paper based - self completed	1136	71.0%	44.3
Isacson et al. 2005 [141]	Cross-sectional	Measuring the impact of depression on HRQoL	Sweden	General population with self-reported depression	Paper based - self completed	3986	53.7%	20-64
König et al. 2009 [142]	Cross-sectional	Analyse the TTO method properties in patients with mental disorders	Germany	Patients diagnosed with affective disorder (ICD-10)	Semi-structured interviewer administrated	153	66.7%	46.8
Leykin et al. 2017 [143]	Cross-sectional	Compare preferences of depression patients with comorbid patient groups	USA	Major depressive disorder patients diagnosed via DSM-IV	Semi-structured interviewer administrated	61 depre. patients	54.1% depre. patients	40.1 depre. patients
						58 depre.+ pain comorbid patient	60.3% depre.+ pain comorbid patient	50.9 depre.+ pain comorbid patient
vignette based depression states valuing TTO studies								
Sanderson et al. 2004 [144]	Cross-sectional pilot study	Modelling changes in health status	Australia	Health professionals: general practitioners	Paper based - self completed	42	63.0%	35-54
Montejo et al. 2011 [145]	Cross-sectional	Multi-attribute utility (MAU) tool development	Spain	Bipolar disorder and schizophrenia patients (according to DSM-IV)	Paper based - self completed	70	36.2%	41.9
Papageorgiu et al. 2014 [146]	Cross-sectional pilot study	Pilot test TTO valuation task and vignettes	Netherlands	Quota sampled volunteers of general population	Semi-structured interviewer administrated	60	50.0%	35.0
Papageorgiu et al. 2015 [147]	Cross-sectional	Direct utility elicitation with TTO in vignette based depression	Netherlands	Stratified sampling of general population	Online - self completed	n/a	51.1%	46.7

		health states						
Flood et al. 2018 [148]	Cross-sectional	Develop and assess health state vignettes and TTO task	United Kingdom	Volunteers among mental service users, health professionals and caregivers	Online - self completed	46 patients	73.9%	32.0
						31 general population	93.5%	36.0
						28 health profession.	67.9%	39.0
Nontarak et al. 2020 [149]	Cross-sectional	Determine disability weights for depression	Thailand	Diagnosed major depressive disorder patients (ICD-11: F32)	Semi-structured interviewer administrated	75	69.3%	47.9

Table 14 - TTO study characteristics in depression

5.3.2. Results of applied TTO methods in depression

Majority of studies employed the conventional time trade-off method (n=9; 64%), three studies (21%) used the simplified indifference in one answer TTO task, one study (7%) applied lead-time/lag-time TTO, and one paper did not clarify the TTO type. Most studies used a 10-year timeframe (n=8; 57%), and other six studies used various time frames: 10 years + y years lead/lag time, 20 years, 50 years, 80 year-old-age, subjective life expectancy and alternating periods of time (of 20, 25, 30, 35 and 40 years). The iteration process was seldom reported, missing in six cases (n=6; 43%). Three studies used a single question, two studies used the incremental bottom-up method, another two studies used the ping-pong method, and one study employed top-down steps. (Table 15)

Self-experienced health – of affective disorder/bipolar disorder/depression/depressive symptoms/depression following infarction – was assessed by seven studies. Vignette-based health states was assessed by six studies focusing on no, mild, moderate, and severe depression. One study evaluated imagined perfect health vs. self-experienced health vs. hypothetical mild depression. All but one studies evaluated BTD health states, König et al complemented it with a lag-time TTO task, which could yield utilities ranging from negative values to one. (Table 14)

Publication	**TTO type**	**Timeframe**	**Iteration process**	**Health state description**	**Format**
self-experienced health states valuing TTO studies					

Oldridge et al. 1991 and 1993 [135, 136]	conventional	10 years	n/a	current health	BTD
Wells et al. 1999 [137]	conventional	10 years	n/a	current health	BTD
Tsevat et al. 2000 [138]	conventional	subjective life expectancy	n/a	current mental health	BTD
Voruganti et al. 2000 [139]	conventional	50 years	n/a	current mental health, worst state mental health, desirable mental health	BTD
Sherbourne et al. 2001 [140]	indifference in one answer	10 years	single question	current health	BTD
Isacson et al. 2005 [141]	conventional	20 years	n/a	current health	BTD
König et al. 2009 [142]	conventional	10 years +10 years waiting	top-down steps	current health	BTD+waiting trade off
Leykin et al. 2017 [143]	conventional	10 years	ping-pong method (7 years basepoint)	perfect health vs current health and current health vs mild depression	BTD
vignette-based depression states valuing TTO studies					
Sanderson et al. 2004 [144]	indifference in one answer	10 years	single question	vignettes: remitted, few symptom, some symptom, many symptom depression	BTD
Montejo et al. 2011 [145]	n/a	20, 25, 30, 35, 40 years	n/a	vignette: single statement	BTD
Papageorgiou et al. 2014 [146]	conventional	80-year-old age	ping-pong method	vignettes: mild and severe depression + co-occurring with cancer, diabetes, heart disease (2+6)	BTD
Papageorgiou et al. 2015 [147]	indifference in one answer	10 years	single question (ping-pong method as warm up)	vignettes: no, mild, moderate, severe depression	BTD
Flood et al. 2018 [148]	conventional	10 years	incremental (bottom-up)	vignettes: severe depression	BTD
Nontarak et al. 2020 [149]	conventional	10 years	top-down steps	vignettes: mild, moderate, severe depression	BTD

Table 15 - TTO attributes of depression studies

5.3.3. Results of health state vignette comparison

Overall, six out of fourteen studies (43%) employed a vignette-based TTO task describing remitted/no, mild, moderate, and severe hypothetical states of depression [144-149]. The health

state descriptions notably differed across studies; no identical vignettes were used. Altogether, the vignettes covered 11 dimensions: *anxiety*, *behaviour*, *cognition*, *emotions*, *functioning mood, physiology, role function, self-appraisal, social relations* and *usual activities*. The number of dimensions used by the studies ranged from 1 to 6, with a mode of six. Minimum one maximum eight health states were evaluated per study. Two studies evaluated mild, moderate and severe depression; another two studies investigated only severe state of depression; one study evaluated no/remitted, mild, moderate and severe states of depression. One study evaluated mild and severe depression separately and alongside with three cooccurring diseases (cancer, diabetes, and heart disease).

Studies designed one to thirty vignettes to describe different states of depression. Almost all studies (n=5; 83%) used one vignette for each different level of depression severity, while one study designed 4 mild, 17 moderate and 9 severe vignettes to differentiate depression-related health states. Half of studies used scenarios (which were interpreted from a third-person perspective) as a presentation of health states, other half applied statements (first-person perspective sentences) as descriptions. Disease domains were covered similarly: items focusing on emotions, physiological functioning, and cognition appeared in five out of six descriptions. Three descriptions were based on the McSad depression scale, which was originally developed for direct utility measurement, although the method of presentation differed across cases (statements vs. scenario). One study used 6 items of the Short Form-12 questionnaire, which were presented as statements, and altered the original response options to describe the 4 levels of depression severity. One study applied a single dimension, single statement description, using the first item of the mental health-specific Tolerability and Quality of Life (TooL) questionnaire. Vignette development and origin was not disclosed by one study, though the description closely resembled the scenario-based McSad vignette. (Table 16)

Publication	**Dimensions (covered)**	**Health states (assessed)**	**Origin of description**	**Number of vignettes used**	**Presentation**
Sanderson et al. 2004 [144]	emotions, physiology, mood, social relations (4)	remitted, few symptoms, some symptoms, many symptoms (4)	SF-12 questionnaire	16	scenario
Montejo et al. 2011 [145]	anxiety/depression (1)	severe depression (1)	TooL questionnaire	1	single statement

Papageorgiou et al. 2014 [146]	emotions, self-appraisal, cognition, physiology, behaviour, role function (6)	mild/severe depression + mild depression co-occurring with moderate or severe cancer/diabetes/heart disease (2+6)	McSad depression scale	8	statements
Papageorgiou et al. 2015 [147]	emotion, self-appraisal, cognition, physiology, behaviour, role function (6)	no, mild, moderate, severe depression (4)	McSad depression scale	30	statements
Flood et al. 2018 [148]	self-appraisal, physiology, functioning, emotions, social-relations, usual activities (6)	severe depression (1)	McSad depression scale	1	scenario
Nontarak et al. 2020 [149]	emotion, usual activities, physiology, cognition (4)	mild, moderate, severe	not disclosed	3	scenario

Table 16 - Comparison of health state vignettes in depression studies

5.3.4. Results of quality assessment based on MVH

The applied TTO task of the 13 included studies were quality check based on the MVH protocol. The TTO study quality met the requirements of MVH protocol mostly in attributes of framework (79%) and timeframe (64%). The major shortcomings regarded the detailing of iteration process and respondent training, where these aspects were rather poorly reported (n/a= 50% and 57% missing). Three studies achieved four scores out of the maximum of seven, other three were below zero, indicating insufficient TTO study description. (Table 17)

Study	Protocol	TTO frame work	Time horizon	Anchor state	Iteration process	Mode of administration	Method of data collection	Respondent training	Score
Oldridge et al. 1991, 1993 [135, 136]	unclear	✓	✓	✓	n/a	✓	n/a	n/a	4
Wells et al. 1999 [137]	unclear	✓	✓	-	n/a	✓	n/a	n/a	2
Tsevat at al. 2000 [138]	unclear	✓	-	-	n/a	✓	✓	n/a	1
Voruganti et al. 2000 [139]	unclear	✓	-	✓	n/a	✓	✓	✓	4
Sherbourne et al .2001 [140]	unclear	✓	✓	-	n/a	-	-	n/a	-1
Sanderson et al. 2004 [144]	unclear	-	✓	-	-	-	✓	✓	-1

Isacson et al. 2005 [141]	unclear	✓	-	✓	n/a	-	n/a	n/a	0
König et al. 2009 [142]	MVH	✓	✓	-	-	✓	✓	nr.	2
Montejo et al. 2011 [145]	unclear	n/a	-	n/a	n/a	-	n/a	n/a	-2
Papageorgiu et al. 2014 [146]	MVH	✓	-	-	✓	✓	✓	✓	3
Papageorgiu et al. 2015 [147]	MVH	-	✓	✓	-	✓	✓	✓	3
Leykin et al. 2017 [143]	unclear	✓	✓	-	✓	✓	-	✓	3
Flood et al. 2018 [148]	unclear	✓	✓	✓	-	✓	✓	✓	5
Nontarak et al. 2020 [149]	unclear	✓	✓	✓	-	✓	✓	n/a	4

Table 17 - Quality check of TTO protocols in depression studies

Note: checklist based on MVH criteria, n/a= information not available

5.3.5. Results on meta-analysis and regression in depression

The REML meta-analysis examined 3 different vignette-based depression health states (mild: n=5; moderate: n=4; severe: n=9) derived from patients (n=3), general population of depressed/nondepressed persons (n = 4) and health professionals (n=2). Utility estimates were pooled and compared between the study populations with (n=7) and without (n = 11) depression (n=7).

The overall pooled mean TTO utilities were 0.75 in mild depression, 0.66 in moderate depression and 0.50 in severe depression. The estimated mean utilities of vignette-based mild, moderate, and severe depression values (with upper-lower confidence level of 95%) were: 0.82 (0.72-0.91), 0.73 (0.66-0.81), and 0.53 (0.46-0.61) among the non-depressed population. Among depressed population the mean utilities (95% CI) of mild, moderate and severe depression were: 0.68 (0.64-0.72), 0.57 (0.42-0.73), and 0.46 (0.30-0.63), respectively. (forest plot Figures 9-11) [106, 144-149].

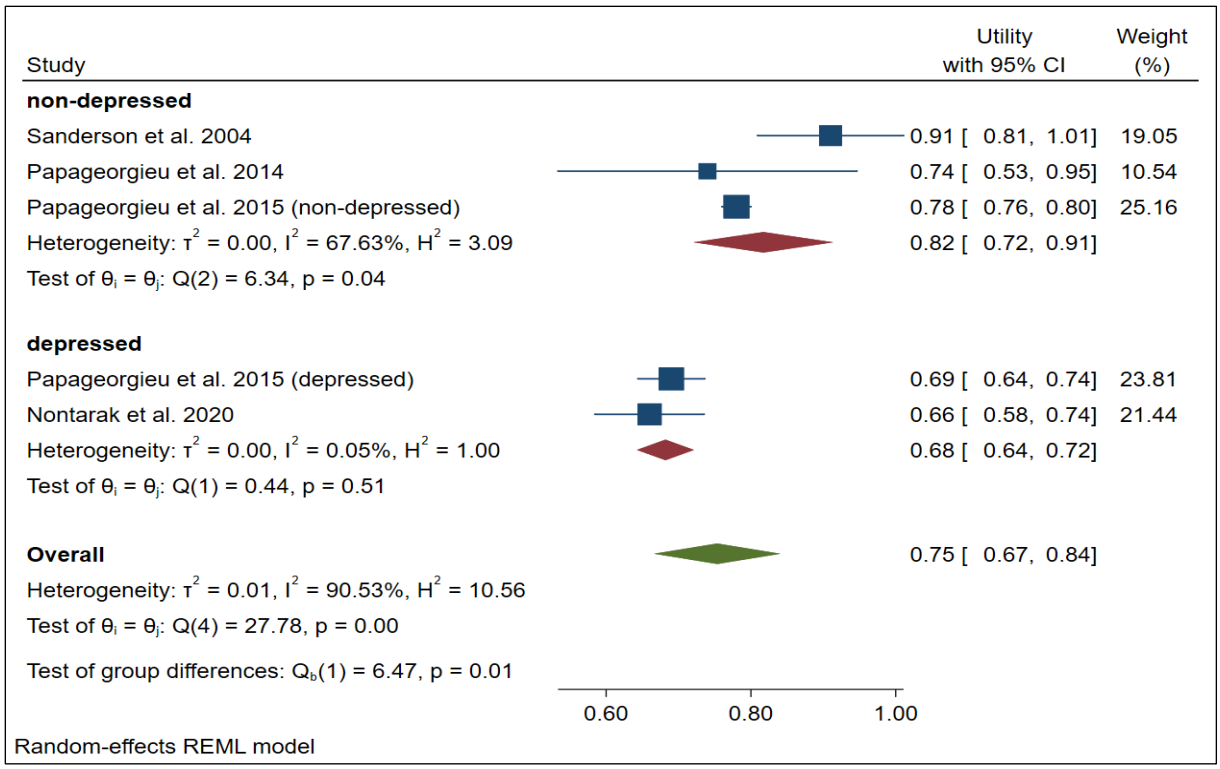

Figure 9 - Meta analysis of vignette-based mean utilities in mild depression

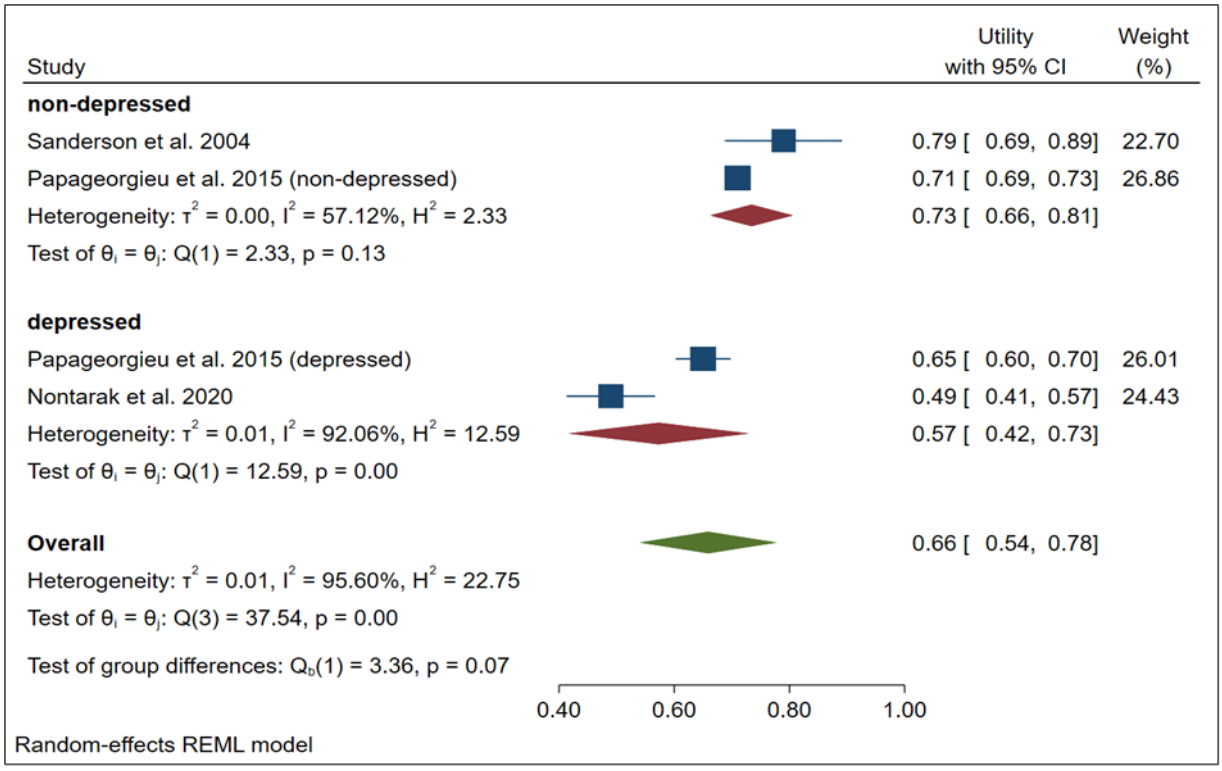

Figure 10 - Meta analysis of vignette-based mean utilities in moderate depression

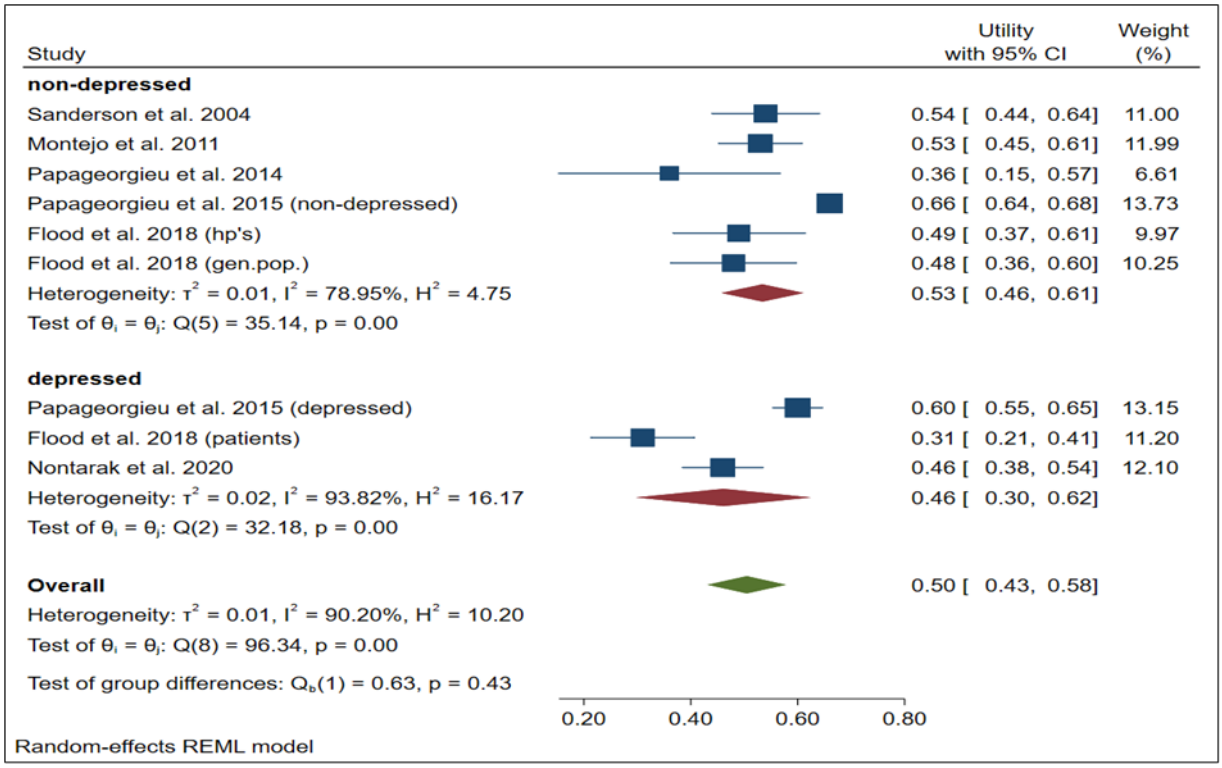

Figure 11 - Meta analysis of vignette-based mean utilities in severe depression

Meta-regression shows that evaluating severe depression state (β= -0.155) and having depressed population sample (β= -0.128) had significant (p<0.02) small negative impact on the vignette-based TTO utility estimates. A large proportion of I^2 was observed among the pooled utilities of the moderate (92.1%) and severe (93.8%) depression subgroups. All subgroups had low standard deviation of utilities across studies (T^2=0.000-0.020). The high proportion of heterogeneity found across studies suggests the existence of additional subgroup or moderator effects, especially in cases of severe and mild depression. (Table 18)

Moderator		**Coefficient**	**S.E.**	**p-level**
depression (reference coded: moderate depression)	mild	0.088	0.065	0.178
	severe	*-0.155*	0.062	*0.012*
vignette type	McSad based (ref. category: other)	0.016	0.053	0.756
data collection	interview (ref. category: self-completed)	0.037	0.063	0.552
population sample	depressed population (ref. category nondepressed)	*-0.128*	0.054	*0.017*
residual I^2	88.80%			
I^2	97.50%			
residual T^2	0.0008			
T^2	0.021			

Table 18 - Meta-regression results of effects on vignette-based utility mean values

Note: T refers to Tau (amount of true variance), SE= standard error

5.3.6. TTO utility catalogue in depression

Overall, 61 HSU values were extracted from 14 original studies, with cataloguing the responding population, the exact health state, utility mean and SD, along with the number of respondents.

Eight studies (57%) reported 36 utilities pertaining to 33 different self-experienced depression states among respondents, ranging from 0.89 (self-experienced health state of US depression patients) to 0.24 (worst self-experienced own health state of Canadian depression patients). Six studies (43%) calculated 25 vignette-based utilities for no, mild, moderate and severe depression states as well as six comorbid conditions co-occurring with depression. These values ranged from 0.96 (remitted depression as evaluated by health professionals) to 0.31 (patient perceptions of severe depression resulting from mental illness). (Table 19)

Publication	Study population	Health state	n	Utility mean (SD)
Oldridge et al. 1991 and 1993 [135, 136]	depression after myocardial infarction patients-rehabilitation	baseline health state	78	0.72 (0.24)
		2 month follow up	78	0.80 (n/a)
		4 month follow up	78	0.82 (n/a)
		8 month follow up	78	0.87 (n/a)
		12 month follow up	78	0.87 (n/a)
	depression after myocardial infarction patients-conventional therapy	baseline health state	87	0.77 (0.25)
		2 month follow up	87	0.82 (n/a)
		4 month follow up	87	0.82 (n/a)
		8 month follow up	87	0.83 (n/a)
		12 month follow up	87	0.86 (n/a)
Wells et al. 1999 [137]	primary care depression patients	current health state (probable depression)	750	0.89 (0.14)
		current health state (depression + comorbid medical condition)	2146	0.87 (0.13)
Tsevat et al. 2000 [138]	patients with bipolar disorders	current health state (depression)	53	0.71 (0.37)
Voruganti et al. 2000 [139]	depression patients	desired health state in the future	30	1.0 (n/a)
		current health state	30	0.73 (0.19)
		worst health state experienced	30	0.24 (0.02)
Sherbourne et al. 2001 [140]	patients with depressive symptoms	(baseline) current health state (depression)	1018	0.83 (0.31)
	Usual care received patients	6 month follow up	372	0.80 (n/a)
		12 month follow up		0.81 (n/a)
		18 month follow up		0.84 (n/a)
		24 month follow up		0.86 (n/a)
	Quality improved management received patients	6 month follow up	361	0.79 (n/a)
		12 month follow up		0.81 (n/a)
		18 month follow up		0.84 (n/a)
		24 month follow up		0.86 (n/a)
	Quality improved psychotherapy received patients	6 month follow up	403	0.76 (n/a)
		12 month follow up		0.80 (n/a)
		18 month follow up		0.83 (n/a)
		24 month follow up		0.87 (n/a)
Sanderson et al. 2004 [144]	health professionals (GP's)	remitted depression	42	0.96 (n/a)
		few symptom-depression	42	0.91 (n/a)
		some symptom-depression	42	0.79 (n/a)
		many symptom-depression	42	0.54 (n/a)
Isacson et al. 2005 [141]	general population	current health state	3835	0.93 (0.15)
	general population (self-reported depression)		151	0.80 (0.25)
König et al. 2009 [142]	patients	affective disorder	153	0.66 (0.44)

Montejo et al. 2011 [145]	schyzophrenia and bipolar disorder patients	severe depression /anxiety	70	0.53 (n/a)
Papageorgiu et al. 2014 [146]	general population	mild depression	10	0.74 (n/a)
		severe depression	10	0.36 (n/a)
		mild depression-cooccurring with moderate cancer	7	0.65 (n/a)
		mild depression-cooccurring with moderate diabetes	6	0.63 (n/a)
		mild depression-cooccurring with moderate heart disease	7	0.63 (n/a)
		mild depression-cooccurring with severe cancer	7	0.45 (n/a)
		mild depression-cooccurring with severe diabetes	6	0.47 (n/a)
		mild depression-cooccurring with severe heart disease	7	0.59 (n/a)
Papageorgiu et al. 2015 [147]	general population - depressed	mild depression	200	0.69 (n/a)
		moderate depression		0.65 (n/a)
		severe depression		0.60 (n/a)
	general population - non depressed	mild depression	1068	0.78 (n/a)
		moderate depression		0.71 (n/a)
		severe depression		0.66 (n/a)
Leykin et al. 2017 [143]	depression patients	own depression-compared to perfect health	61	0.46 (0.31)
		own depression-compared to mild depression	61	0.79 (0.34)
	comorbid patients (depression+pain)	own depression-compared to perfect health	58	0.52 (0.34)
		own depression-compared to mild depression	58	0.81 (0.30)
Flood et al. 2018 [148]	mental illness patients	severe depression	46	0.31 (0.35)
	general population		31	0.48 (0.30)
	health professionals		28	0.49 (0.34)
Nontarak et al. 2020 [149]	major depressive disorder patients	mild depression	75	0.66 (n/a)
		moderate depression	75	0.49 (n/a)
		severe depression	75	0.46 (n/a)

Table 19 - TTO utility catalogue in depression

6. Measurement agreement between direct and indirect utility measures

6.1. Introduction to the agreement study

Over the past three decades the researchers found consensus that different health state utility measures yield systematically different scores. The applied measurement tool is beyond doubt impacts the estimated health outcome. Discrepancies of HSU scores by different measurement methods raises questions regarding interpretation and comparison of utility scores, affecting both aspects of economic evaluations (incremental health gain of therapies and value for money). For instance, direct measurement methods often output significantly higher HSU values than indirect instruments, but there are exceptions especially in many dermatological conditions [64, 68]. More concerns are raised by TTO-DCE agreement studies, founding poor agreement between the valuations. Argument facilitated around the divergence between head to head quality of life comparisons (DCE: comparing two similar imperfect states) versus comparing an imperfect health state to death/full health (TTO task) [150, 151]. Several observations record also poor agreement between TTO and EQ-5D with wide limits of agreement (LOA) [152, 153]. The other well studied effects on utility results are sociodemographic factors (being married, having children, religiousness, older age, higher education level upwards the utility values) [77, 78, 154].

The primary aim of the study is to examine measurement agreement between direct and indirect HSU measures. To assess the impact of sociodemographic variables on direct and indirect HSU is considered as a secondary goal.

Chronic dermatological diseases are associated with impaired physical and social functioning, sexual life, pain and lowered self-esteem, that cause substantial decrease in HRQoL [155]. In adult populations, psoriasis and atopic dermatitis are among the most common chronic dermatological conditions, while hidradenitis suppurativa and pemphigus are one of the most severe conditions impacting HRQoL. Atopic dermatitis is characterized by skin lesions and itching, with 15-25% prevalence among all population groups worldwide [156]. Hidradenitis suppurativa is a painful immune-mediated skin disease, causing recurrent abscesses, fistulas and scars [157]. Psoriasis, besides red and painfully itching skin is associated with psychological effects on self-esteem and depression, while affects approximately 1-9% of world's population [158]. Pemphigus is a rare disease of skin and mucosa; without treatment the health outcome is lethal [159].

6.2. Study methodology

Data of four multi-centre cross sectional surveys conducted between 2015-2021 was collected from atopic dermatitis, hidradenitis suppurativa, pemphigus and psoriasis adult (18≤) patients in Hungary. Paper-based self-completed questionnaires requested HRQoL-related, sociodemographic (age, sex, level of education, employment status) and clinical characteristics (disease duration, skin severity, outpatient care visits) information. Descriptive statistics were computed for the pooled sample according to disease diagnoses. The detailed study methods were published in the previous articles using the same sample data [115, 117, 160, 161]. Database was built in Excel (Microsoft Office 365) RStudio (version 4.1.2 'devtools' package) was used for statistical analysis.

6.2.1. Direct and indirect utility measurement tools

Direct HSU measures − most frequently TTO, less frequently SG and DCE − all rely on rational decision-making theory, assuming that people know their constant preferences and they are plainly able to choose between two alternatives. Indirect HSU measures such as EQ-5D, HUI or SF-6D follow MAU theory [162]. The generic indirect instruments are questionnaires with a standard health state descriptive system, where the items are rated on a scale, that enable to summarize HRQoL in a single score. The similarly functioning disease-specific HRQoL instruments, like DLQI or Skindex-16, use dermatology related items (e.g. dermatology-specific domains of emotions, functioning, symptoms, personal relations) to describe and rate skin condition. Preference-accompanied (indirect) measurement instrument item sum scores can be converted into health utility, using a value set or mapping algorithms [61, 163].

Recent study includes the comparison of four HSU measures – (1) *conventional TTO*, (2) *EQ-5D-5L*, (3) *mapping-based DLQI utility* and (4) *value set based DLQI utility* – in four dermatological diseases (atopic dermatitis, hidradenitis suppurativa, psoriasis, pemphigus). Mean, standard deviation (SD), median, interquartile range (IQR) were calculated for TTO, EQ-5D-5L, mapping-based DLQI and value set based DLQI utilities.

TTO was used for direct HSU measures in all four studies, valuing patient's current health vs. full health. In three cases conventional TTO method with 10-year timeframe, using top-down iteration process with fixed 1/0.5-year steps were used. Composite TTO was employed in case of hidradenitis suppurativa, though only the answers on BTD health format were included to ensure methodological synergy, thus responses on lead-time valuations with negative utilities were excluded.

EQ-5D-5L was applied as a generic preference-accompanied HRQoL measure to elicit indirect utilities using the Hungarian value set. The five dimensions (5D): Mobility, Self-care, Usual activities, Pain/discomfort, and Anxiety/Depression are rated on five levels (5L): no problems, slight problems, moderate problems, severe problems, and extreme problems/unable to do [164]. The items cover 3125 possible health states, the best health state is represented by 11111, the worst by 55555. The Hungarian value set transforms the item scores to utilities ranging from -0.848 to 1, but responses of health states with negative values were truncated to save the harmonization on utility range of 0-1.

Further two indirect utilities were estimated using patient's responses on the Dermatology Life Quality Index (DLQI) [165, 166]. The 10-item DLQI is the most commonly used dermatology-specific HRQoL measure, covering six health domains: *symptoms and feelings, daily activities, leisure, work and school, personal relationships, treatment.* Items are rated on a 0-3 severity scale (ranging between 0-30), the higher the score the worse the HRQoL.

The first approach was mapping the DLQI item responses into EQ-5D-3L utilities that rely on the UK value set. Mapping based DLQI utility **(**in the following: **mDLQI)** was estimated by transforming the DLQI scores into EQ-5D-3L utilities by using a OLS regression-based mapping algorithm [167]:

$$EQ5D3L = Predicted\ EQ5D(1) - \left(1 - Predicted\ EQ5D3L(1)\right) * \ imperfect\ health\ utility(x)$$

The best mDLQI estimation was achieved if inputting item level responses, age and sex information of Hungarian patient sample.

The second approach to generate indirect utilities was applying the recently published Hungarian DLQI value set. The value set based DLQI utility (**vDLQI** in the following) was established with the censored regression that assigns disutility to DLQI item level responses (level score L1, L2, L3) [168]:

$$y = 0.873 + (-disutility(L1, L2, L3) of\ Item\ 1 + \ldots disutility\ of\ Item\ 10)$$

The DLQI value set is based on conventional TTO method valuations from members of the Hungarian adult general population assessing vignette-based health states. Table 15 describes the categorization of HRQoL measurements according to HSU type:

Utility measurement attribute	Direct	Indirect		
Measurement tool	TTO	EQ-5D-5L	mDLQI	vDLQI
Instrument specificity (theory)	preference-based (rational choice)	generic (multi-attribute utility)	dermatology specific (multi-attribute utility)	
Preferences	patients	societal		
Country of the preference weight	Hungarian patient's current health	Hungary	United Kingdom	Hungary
Theoretical utility range	0-1	0-1	0.25-0.98	0.57-0.87

Table 20 - Categorization of compared direct and indirect utility measurements

For utility measurement comparison, convergent validity of utilities was analysed by Pearson's correlation (r). Known-group validity was assessed by comparing patient subgroups (according to sociodemographic) applying Kruskal-Wallis or Mann-Whitney nonparametric tests.

6.2.2. Methods of agreement investigation

Bland-Altman (B-A) plots and intraclass correlation coefficient (ICC) were used to discriminate agreement parameters from measurement error. B-A plots are used to investigate relationship of discrepancies between two measurements and the true value, so it detects systematic differences between two measurement instruments (proportional bias practically meaning that one method gives higher/lower values than the other). Bland-Altman visualization is useful, when studying the agreement between different measurement techniques that measure the same construct (e.g. health state utility) to search for potential proportional bias across the four measures. The difference of measurement methods was regressed on the average of the two methods to evaluate the relationship of discrepancies [169].

To calculate ICC values, two-way random model with absolute agreement was used that randomizes both subject and instrument effects. ICC reflects the reliability of measurements, representing the proportion of total variance due the variations between cluster members. It indicates poor, moderate and strong agreement if ICC>0.50, 0.51-0.74 and 0.75-1 [170].

6.2.3. Regression analysis of sociodemographic factors

Four multivariate Tobit regression models were conducted to analyse the effect of sociodemographic and disease-related variables on TTO, EQ-5D-5L, mDLQI and vDLQI utility values. Censored regression was selected to treat probable – often reported left-sided – skewedness in utility data distribution, while assuming linear relationship between the dependent and independent variables [171]. The three continuous and five dummy coded explanatory variables were: age, sex (male reference coded), level of education (primary reference coded), employment status (unemployed reference coded), disease duration in years, outpatient care visit during the past 3 month (yes or no), skin condition severity (0-1 normalized score), type of skin disease (pemphigus reference coded). Skin severity variable was created by normalizing ABSIS, PASI, SARTORIUS, SCORAD disease-specific clinical measures to a range of 0-1 score (with 1 referring to the worst possible severity) in order to make skin disease severity rating comparable. Right-censor was set as upper limit at one, because utility estimates are typically peaked at 1. The regression equation is the following in all four models:

$$(health\ state)\ utility_i = \beta_0 + \beta_i Age + \beta_i Gen + \beta_i Educ + \beta_i Emp + \beta_i Disdur + \beta_i Outp + \beta_i Skincond + \gamma_i Diseasetype + \epsilon_i = utility_i^* \text{ if } utility_i^* > 0 \text{ otherwise, } utility_i = 0 \text{ if } utility_i^* \leq 0$$

6.3. Results of measurement agreement

Overall, N=765 dermatology patient responses have been analysed (atopic dermatitis n=218; hidradenitis suppurativa n=200; pemphigus n=109; psoriasis n=238). Proportion of females was 47.3%, the mean age of the sample was 41.5 (SD=16.2). Most of the sample population was secondary educated (56.3%) and full-time employed (50.1%). Average disease duration was 12.8 (SD=12.6) years. Altogether 435 (56.9%) patients used outpatient care service in the past 12 month. Normalized skin severity scores in atopic dermatitis, hidradenitis suppurativa, psoriasis and pemphigus were 0.51, 0.22, 0.18 and 0.14, respectively. Mean total DLQI score was 9.9 (SD=8.4), the highest, referring to the worst condition in atopic dermatitis (13.4), the lowest in pemphigus (5.4). (Table 21)

Variables	**n (%) or mean (SD)**				
	Total	Pemphigus	Psoriasis	Hidradenitis suppurativa	Atopic dermatitis
Characteristics					
Total	765 (100)	109 (14.2)	238 (31.1)	200 (26.1)	218 (28.5)
Age in years	41.5 (16.2)	57.1 (14.8)	47.4 (15.2)	37.1 (12.4)	31.3 (11.7)

Sex					
female	362 (47.3)	70 (64.2)	89 (37.4)	77 (38.5)	126 (57.8)
male	403 (52.7)	39 (35.8)	149 (62.6)	123 (61.5)	92 (42.2)
Education (missing n=3)					
primary	93 (12.2)	22 (20.2)	19 (8.0)	40 (20.0)	12 (5.5)
secondary	431 (56.3)	58 (53.2)	132 (55.5)	129 (64.5)	112 (51.4)
tertiary	238 (31.1)	29 (26.6)	87 (36.6)	30 (15.0)	92 (42.2)
Employment status					
full-time employed	383 (50.1)	41 (37.6)	121 (50.8)	117 (58.5)	104 (47.7)
part-time employed	37 (4.8)	4 (3.7)	13 (5.5)	9 (4.5)	11 (5.0)
student	83 (10.8)	1 (0.9)	7 (2.9)	21 (10.5)	54 (24.8)
retired	97 (12.7)	38 (34.9)	44 (18.5)	4 (2.0)	11 (5.0)
disability pensioner	63 (8.2)	15 (13.8)	26 (10.9)	14 (7.0)	8 (3.7)
unemployed	53 (6.9)	6 (5.5)	13 (5.5)	25 (12.5)	9 (4.1)
other (e.g. housewife, caregiver)	49 (6.4)	4 (3.7)	14 (5.9)	10 (5.0)	21 (9.6)
Outpatient care use (missing n=6)					
yes	435 (56.9)	54 (49.5)	211 (88.7)	71 (35.5)	99 (45.4)
no	324 (42.7)	55 (50.5)	27 (11.3)	126 (63.0)	116 (53.2)
Disease duration in years (missing n=3)	12.8 (12.6)	3.8 (4.9)	18.1 (12.3)	4.8 (6.7)	19.0 (12.9)
Disease severity*	0.28 (0.24)	0.14 (0.20)	0.18 (0.19)	0.22 (0.18)	0.51 (0.21)
DLQI score (0-30) (missing n=4)	9.9 (8.4)	5.4 (6.8)	7.1 (7.4)	11.7 (8.1)	13.4 (8.5)

Table 21 - Sample characteristics of dermatology patients in the agreement study

6.3.1. Comparison of utility scores

Total mean (SD) TTO, EQ-5D-5L, mDLQI and vDLQI utilities were: 0.83 (0.24); 0.81 (0.24), 0.77 (0.14) and 0.81 (0.08), respectively. TTO values showed the highest median (IQR) utility score of 0.95 (0.74-1), followed by EQ-5D-5L with 0.89 (0.75-0.97), vDLQI with 0.84 (0.76-0.87) and mDLQI with 0.79 (0.70-0.86). Utility data were skewed to the left according to all utility measurement methods, TTO and EQ-5D-5L peaked at 1. Converted mDLQI utility was closer to normal distribution, scores ranged between 0.25-0.98 (only n=3 respondents had maximum utility). The Hungarian vDLQI utility had a range of 0.57 (worst health state) to 0.87 (best health state), where n=289 patients had maximum utility. (Figure 12)

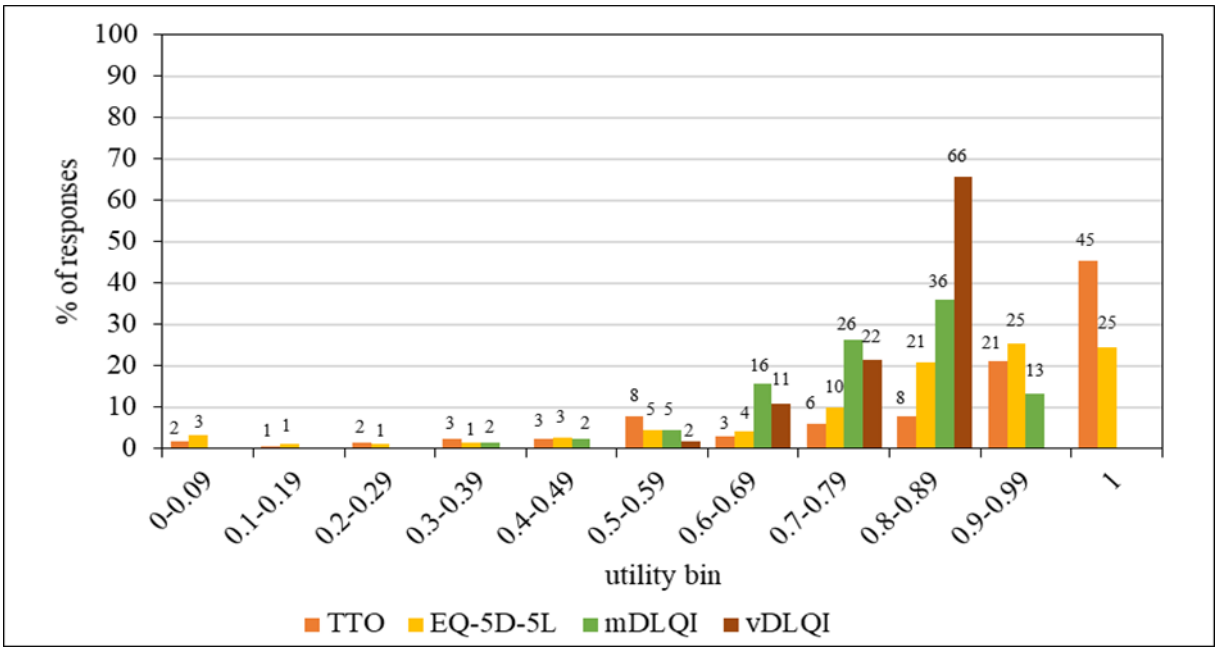

Figure 12 - Distribution of utility values according to TTO, EQ5D5L, mDLQI and vDLQI measures

Note: mDLQI/vDLQI stand for mapping based and value set based dermatology utilities

Altogether in five grouping variables 18 utility scores per each four utility measures were categorizing were computed, whereby TTO produced the highest mean utility in 10 (56%) cases. In five cases (28%) the vDLQI and 3 (17%) cases the EQ-5D-5L resulted the highest mean utility.

Psoriasis patients had the highest, whereas pemphigus the lowest mean utility according to TTO (0.91 vs 0.72). Similarly, the highest mean EQ-5D-5L utility was observed in psoriasis, but the lowest in hidradenitis suppurativa. Both m/vDLQI utility estimations produced the same order: highest mean utility in pemphigus the lowest in atopic dermatitis. Among all sample characteristic variables, the range of mean utilities was the biggest according to TTO (0.19 in disease type), somewhat smaller in EQ-5D-5L (0.18 in employment status). The mDLQI and vDLQI mean utility had smaller scatter in subgroups. (Table 22)

Weak to strong significant correlations were observed between the utility estimates. The strongest correlation was between the mDLQI and vDLQI utilities ($r=0.900$; $p<0.01$). EQ-5D-5L moderately correlated with mDLQI and vDLQI utility ($r=0.598$ and 0.556; $p<0.01$), and weakly with TTO ($r=0.287$; $p<0.01$). TTO showed weak correlation with both vDLQI ($r=0.257$, $p<0.01$) and mDLQI ($r=0.257$ and 0.244; $p<0.01$).

Subgroup comparison of mean utility values across all utility measurement methods showed statistically significant differences in disease type, educational level and employment

status ($p<0.05$). Patients that were higher (tertiary) educated and full- or part-time employed had significantly higher mean utility scores according to all assessment methods. Both m/vDLQI utility means significantly differed in sec, showing higher values for males. Mean utilities depending on the use of outpatient care did not differ significantly according to any of the utility assessments. (Table 22)

Variable	TTO utility				EQ-5D-5L utility				mapping-based DLQI utility				value set-based DLQI utility			
	n	*Mean (SD)*	*Median*	*p level*	*n*	*Mean (SD)*	*Median*	*p level*	*n*	*Mean (SD)*	*Median*	*p level*	*n*	*Mean*	*Median*	p level
Total	730	0.832 (0.24)	0.95	-	761	0.814 (0.24)	0.89	-	761	0.769 (0.13)	0.79	-	764	0.805 (0.08)	0.84	-
Disease																
Psoriasis	238	0.91 (0.16)	1	< 0.001	237	0.84 (0.22)	0.92	< 0.001	238	0.80 (0.11)	0.81	< 0.001	238	0.83 (0.07)	0.87	< 0.001
Atopic dermatitis	193	0.85 (0.22)	0.95		218	0.83 (0.21)	0.89		218	0.74 (0.14)	0.76		218	0.78 (0.09)	0.81	
Hidradenitis suppurativa	194	0.78 (0.27)	0.9		198	0.76 (0.28)	0.86		198	0.75 (0.13)	0.78		200	0.79 (0.09)	0.82	
Pemphigus	105	0.72 (0.32)	0.9		108	0.82 (0.25)	0.91		107	0.80 (0.10)	0.83		108	0.84 (0.06)	0.87	
Gender																
female	347	0.84 (0.24)	0.95	0.490	360	0.81 (0.23)	0.89	0.112	360	0.75 (0.13)	0.77	< 0.001	361	0.79 (0.09)	0.83	< 0.001
male	383	0.83 (0.24)	0.95		401	0.82 (0.25)	0.92		401	0.79 (0.12)	0.81		403	0.82 (0.07)	0.85	
Education																
primary	89	0.75 (0.30)	0.9	< 0.001	91	0.69 (0.31)	0.79	< 0.001	91	0.74 (0.12)	0.75	0.003	92	0.78 (0.09)	0.80	< 0.001
secondary	409	0.81 (0.25)	0.95		430	0.81 (0.24)	0.89		430	0.77 (0.12)	0.78		431	0.80 (0.08)	0.84	
tertiary	230	0.90 (0.18)	1		237	0.86 (0.18)	0.92		237	0.78 (0.12)	0.81		238	0.82 (0.08)	0.85	
Employment status																
full-time employed	366	0.85 (0.23)	0.95	0.025	382	0.86 (0.20)	0.92	< 0.001	381	0.79 (0.12)	0.81	< 0.001	383	0.81 (0.08)	0.85	< 0.001
part-time employed	35	0.89 (0.12)	0.95		37	0.81 (0.24)	0.85		37	0.77 (0.13)	0.79		37	0.80 (0.09)	0.84	
student (university)	82	0.80 (0.25)	0.9		83	0.84 (0.21)	0.91		83	0.78 (0.13)	0.79		83	0.79 (0.08)	0.82	
retired	94	0.83 (0.27)	1		96	0.77 (0.25)	0.88		96	0.74 (0.09)	0.74		96	0.83 (0.06)	0.85	
disability pensioner	60	0.76 (0.30)	0.9		62	0.68 (0.29)	0.76		63	0.72 (0.15)	0.75		63	0.78 (0.10)	0.81	

unemployed	49	0.78 (0.25)	0.9		52	0.69 (0.32)	0.83		52	0.74 (0.12)	0.78		53	0.78 (0.09)	0.81	
other	44	0.90 (0.19)	1		49	0.79 (0.26)	0.88		49	0.73 (0.15)	0.77		49	0.78 (0.10)	0.84	
Outpatient care visiting																
yes	416	0.84 (0.25)	0.95	0.130	433	0.81 (0.25)	0.88	0.805	433	0.77 (0.13)	0.79	0.872	435	0.81 (0.08)	0.85	0.137
no	309	0.83 (0.23)	0.95		322	0.83 (0.22)	0.91		322	0.77 (0.11)	0.79		323	0.81 (0.08)	0.84	

Table 22 - Comparison of four utility measurement scores in four dermatological diseases

6.3.2. Agreement results of direct and indirect utility measures

Measurement agreement was found only between TTO and EQ-5D-5L (mean difference: 0.016; SD= 0.287; p=0.124), with moderate agreement in between individuals (ICC=0.445; 95% CI: 0.36-0.52; p<0.001). Agreement was present between EQ-5D-5L and vDLQI, though linear regressing of the bias and measurement revealed high level of proportional bias. TTO and mDLQI/vDLQI as well as EQ-5D and mDLQI/vDLQI measures showed presence of systemic proportional bias. Bland-Altman plots further supported the extent of disagreements between the TTO/EQ-5D-5L and two DLQI utility estimates, while the scores misfit limits of agreement especially at the lower end of the utility scale. (Table 22)

Absolute agreement between individuals in mDLQI and vDLQI measures was strong (ICC=0.872; p<0.001), moderate in EQ-5D-5L and mDLQI/vDLQI measures (ICC=0.646 and 0.505; p<0.001), close to moderate between TTO and EQ-5D-5L (ICC= 0.445; p<0.001) but rather poor between TTO and the two DLQI utility measures (ICC= 0.314 and 0.263; p<0.001). (Table 23)

measure	**TTO-ED5D**	**TTO-mDLQI**	**TTO-vDLQI**	**EQ5D-mDLQI**	**EQ5D-vDLQI**	**mDLQI-vDLQI**
Bland-Altman plot results						
Mean difference (t test p-level)	*0.016 (0.124)*	0.062 (<0.001)	0.025 (0.004)	0.045 (<0.001)	*0.008 (0.317)*	0.037 (<0.001)
Standard deviation of differences	0.287	0.243	0.235	0.193	0.206	0.063
Limits of agreement (95% lower-upper CI)	-1.125	-0.953	-0.922	-0.756	-0.807	-0.247
B-A regression β coeffic. (p-level)	*0.033 (0.546)*	0.999 (<0.001)	1.397 (<0.001)	0.770 (<0.001)	1.187 (<0.001)	-0.446 (<0.001)
ICC results						
ICC	0.445	0.314	0.263	0.646	0.505	0.872
95% lower-upper CI	0.358-0.520	0.205-0.408	0.148-0.362	0.582-0.700	0.430-0.571	0.728-0.927
p-level	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table 23 - Measurement agreement results between direct and indirect utility measures

Note: ICC= intraclass correlation coefficient, CI=confidence interval, B-A=Bland-Altman

The six B-A plots visualize the agreement between the four estimates, with mean vs mean difference and LOA (as 95% confidence interval) highlighted. (Figure 13-18)

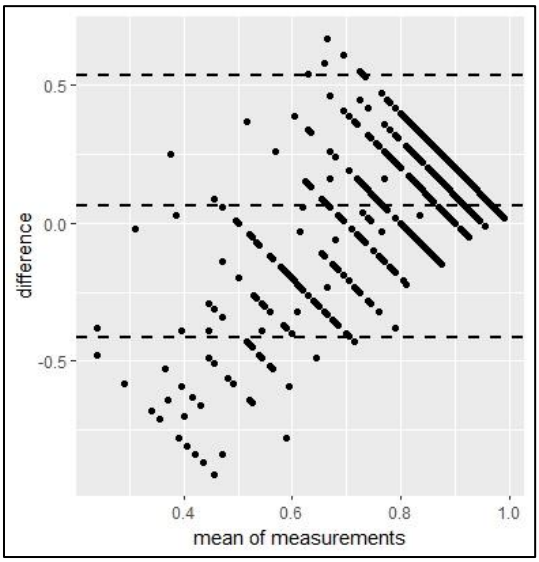

Figure 13 – TTO-mDLQI agreement plot

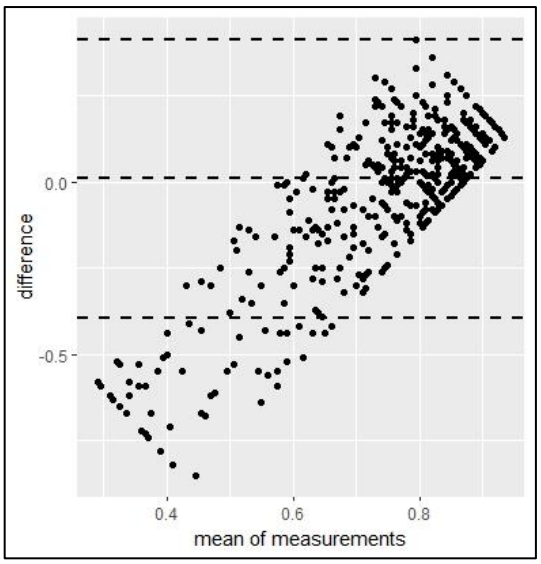

Figure 18 – TTO-EQ5D5L agreement plot

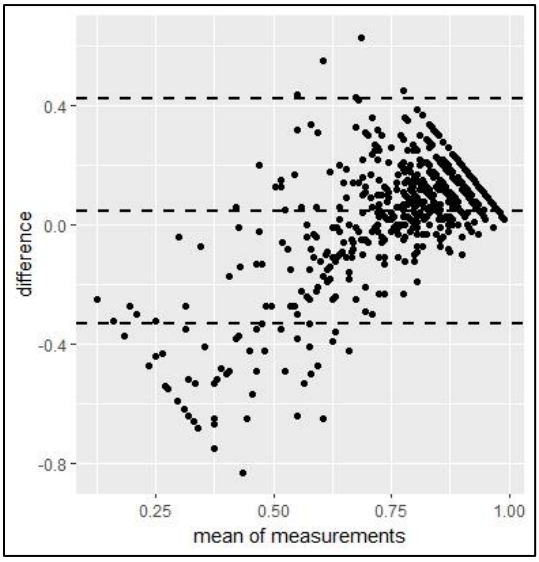

Figure 16 – TTO-vDLQI agreement plot

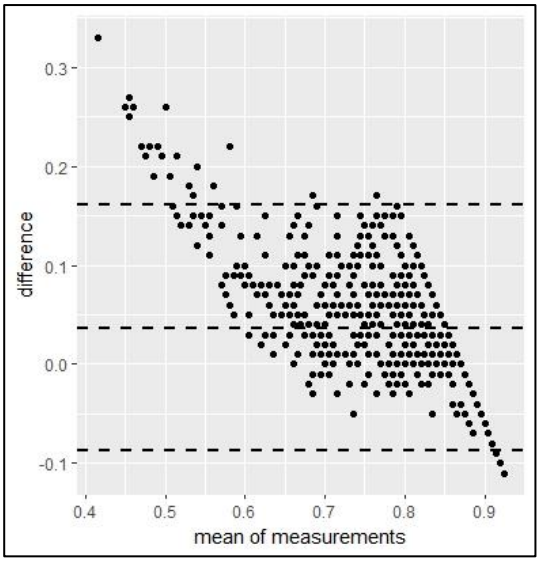

Figure 17 – EQ5D5L-vDLQI agreement plot

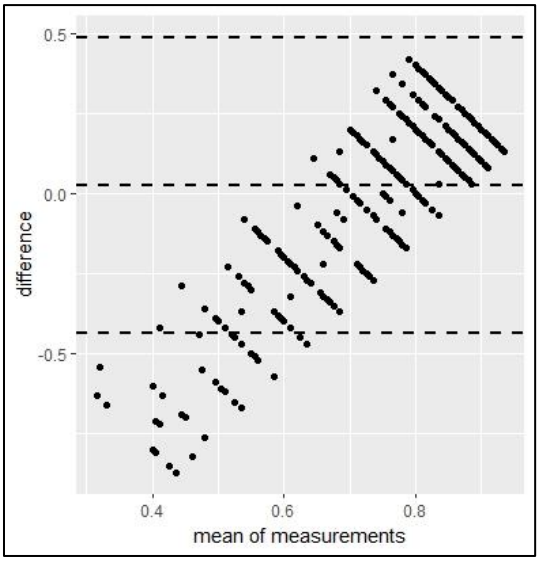

Figure 15 – EQ5D5L-mDLQI agreement plot

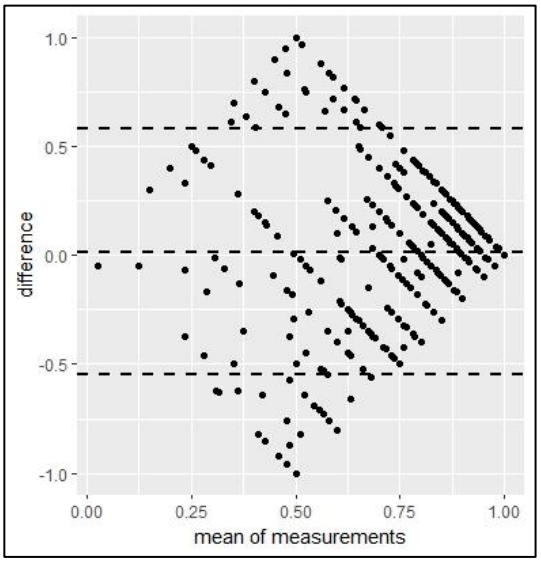

Figure 14 – mDLQI-vDLQI agreement plot

6.3.3. Regression results of impacts on utilities

Out of the four tobit models, the greatest variance was explained in EQ-5D-5L utility by the observed sociodemographic and clinical characteristics (R^2=35.2%; σ=0.07). EQ-5D-5L utility was higher in patients with secondary or higher education background (β=0.08 and 0.13) and in those being full-time employed (β=0.09) or in those having more severe skin disease (β=-0.51). TTO utilities were significantly higher in higher educated patients (β=0.15) and those having either atopic dermatitis (β=0.35), psoriasis (β=0.29) or hidradenitis suppurativa (β=0.13). Disease severity was negatively associated with TTO utilities (β=-0.45). Older age, female gender, outpatient care visiting, living in hidradenitis suppurativa slightly but significantly decreased mDLQI utilities. Higher vDLQI utilities were observed in older and highly educated, and lower in female, those having worse disease severity or hidradenitis suppurativa. (Table 24)

Variables/ coefficients	**TTO utility**		**EQ-5D-5L utility**		**mDLQI utility**		**vDLQI utility**	
	β (SE)	*p-value*	*β (SE)*	*p-value*	*β (SE)*	*p-value*	*β (SE)*	*p-value*
Age	0.002 (0.00)	0.118	-0.001 (0.00)	0.287	**-0.002 (0.00)**	**<0.001**	**0.001 (0.00)**	**0.039**
Gender	0.022 (0.03)	0.46	-0.038 (0.04)	0.065	**-0.050 (0.01)**	**<0.001**	**-0.024 (0.01)**	**<0.001**
Secondary education	0.036 (0.04)	0.424	**0.079 (0.03)**	**0.011**	0.009 (0.01)	0.435	0.009 (0.01)	0.268
Higher education	**0.152 (0.05)**	**0.003**	**0.129 (0.04)**	**<0.001**	0.021 (0.01)	0.112	**0.019 (0.01)**	**0.033**
Full-time employed	0.036 (0.05)	0.499	**0.094 (0.04)**	**0.010**	0.012 (0.01)	0.388	0.009 (0.01)	0.349
Part-time employed	0.053 (0.08)	0.491	0.040 (0.05)	0.451	-0.007 (0.02)	0.734	0.002 (0.01)	0.855
Retired	-0.032 (0.08)	0.681	-0.041 (0.05)	0.441	-0.013 (0.02)	0.495	-0.002 (0.01)	0.891
Disability pensioner	-0.080 (0.08)	0.275	-0.064 (0.05)	0.209	-0.029 (0.02)	0.130	-0.020 (0.01)	0.110
Student	-0.014 (0.07)	0.830	**0.091 (0.05)**	**0.045**	0.015 (0.02)	0.371	0.012 (0.01)	0.266
Other employment	0.126 (0.08)	0.115	0.033 (0.05)	0.531	-0.007 (0.02)	0.737	-0.001 (0.01)	0.929
Disease duration(years)	0.000 (0.00)	0.890	-0.001 (0.00)	0.416	0.000 (0.00)	0.649	0.000 (0.00)	0.319
Outpatient care use (y/n)	-0.020 (0.03)	0.525	-0.025 (0.02)	0.257	-0.013 (0.01)	0.104	-0.010 (0.01)	0.087
Skin disease severity score	**-0.447 (0.07)**	**<0.001**	**-0.486 (0.05)**	**<0.001**	**-0.292 (0.02)**	**<0.001**	**-0.172 (0.01)**	**<0.001**
Psoriasis	**0.293 (0.06)**	**<0.001**	0.017 (0.04)	0.661	-0.015 (0.01)	0.291	0.004 (0.01)	0.657
Hidradenitis suppurativa	**0.130 (0.05)**	**0.013**	**-0.119 (0.04)**	**0.001**	**-0.070** (0.01)	**<0.001**	**-0.030 (0.01)**	**0.001**
Atopic dermatitis	**0.352 (0.07)**	**<0.001**	0.073 (0.05)	0.135	-0.005 (0.02)	0.787	0.014 (0.01)	0.241

Constant	0.682 (0.10)	<0.001	0.952 (0.07)	<0.001	0.961 (0.03)	<0.001	0.839 (0.02)	<0.001
Regression model indices								
R2	0.140		0.352		0.346		0.174	
observations(n)	724		753		753		756	
uncensored	394		569		753		756	
right censored	330		184		0		0	
log likelihood	-362.9		-176.6		682.5		979.1	
prob > (chi)	<0.001		<0.001		<0.001		<0.001	
variance (σ)	0.12		0.066		0.01		0.004	

Table 24 - Tobit regression results of sociodemographic variables impact on utility estimates

Four tobit regression models showed that disease type and skin condition severity along with socioeconomic variables (age, gender, level of education and employment status) had small significant effect on at least one of the HSU measures. More severe skin disease significantly downwards TTO (β=-0.45), EQ-5D-5L (β=-0.49), mDLQI (β=-0.29) and vDLQI (β=-0.17) utility scores. Similar effects of sociodemographic factors on utility values are well investigated [172-174].

7. Capability measurement in mental health states

7.1. Introduction to the capability study in Hungary

Most often health economic evaluations − following the utilitarian theory − use the QALY concept to express health outcome. HRQoL instruments are used as golden standards, to calculate indirect HSU, since these questionnaires accurately capture health state quality [3]. However, new measures that go beyond HRQoL and cover wider concept of well-being are indeed needed. Especially when pandemic/epidemic periods (like the COVID-19 in 2019-2022) caused circumstances increase demand both on healthcare and social welfare systems [175]. Episodic appearance of pandemics elevate stress with a variety of traumas, pose burden on social care systems [176]. While utility approach focuses on the *needs of individuals* and to satisfy physical and psychological health domains, capability approach emphasizes the *human desires* where well-being is the (cap)ability to achieve objectives that people value [36]. Capability approach – as a viable alternative/complement to the utilitarian health assessment approach – rapidly expanded for presenting practical relevance in health economic evaluations [177].

Apart from occasional unexpected events, many countries report long-term growth in the prevalence of chronic mental diseases [178]. In the absence of proper counterfactuals, the reason of increase in mental disease conditions cannot be clearly determined. It may be attributable to more factors, like (i) the better sensitivity of measurement tools, (ii) particular health policy focus on health states, or (iii) true increase in proportions. The population normative data permits to compare individuals scores to a mean of larger group who are similar in terms of age, sex, level of education, residence, employment status, etc. Normative data of mental capabilities reflecting on current performance status of healthy population enables to track the change in capabilities as well as frequency of impaired social groups [179].

Oxford CAPabilities-Mental Health questionnaire (OxCAP-MH) was particularly designed to be used among groups with mental health impairment [36]. Similarly, to the frequently used ICEpop CAPability (ICECAP-A) measure for adults the OxCAP-MH was developed in the UK, currently has 3 further language validations: German, Hungarian and Luganda version [44-46, 180]. Previous studies report good psychometric properties and stable dimensionality of the, but all studies had disease specific (HIV, schizophrenia) or genuinely small (N=9-172) samples. No OxCAP population norms have been established anywhere before.

The objective of the study would be to establish the first OxCAP-MH population norm, to set current mental health related capabilities of the Hungarian general population. The complementary aim is psychometric assessment of OxCAP-MH to explore test-retest reliability, internal consistency, construct and convergent validity of the instrument.

7.2. Methods of mental capability study

A large sample (N=2000) online survey was conducted among Hungarian general population in 2021 August. Respondents were recruited from a panel database of a professional survey company (NRC). Population composition quotas were set to obtain sample representativeness on Hungarian population with respect of age, gender, level of education and residence, although stratification is only applicable with sample weights (that was not used for the recent analysis). Online self-administrated data collection was followed. The questionnaire consisted of validated HRQoL instruments, capability measurement tools and sociodemographic questions.

7.2.1. Psychometric evaluation of the OxCap-MH

The psychometric properties of the OxCap-MH were evaluated with classical test theory methods. Internal consistency and test-retest reliability was explored with Cronbach's alpha (good if >0.8) and ICC using two-way mixed model of absolute agreement (poor agreement if <0.5, moderate if 0.5-0.75, good if >0.75) [181]. Item reliability was further assessed with corrected item-total correlation. Construct validity was examined with nonparametric t-tests among known-groups. Pearson's correlation coefficients (r) were calculated between OxCap-MH standardized scores and ICECAP-A, PHQ-9 and GAD-7 instruments to assess convergent validity.

7.2.2. Mental health measurement instruments

The OxCAP-MH consists of 16 items, each scored on a 1-5 Likert scale, where five refers to worst domain and one to the best capability in case of items 2,4,5,6,9-16 while all other items (1,3,7,8) are reverse coded. The change in interpretation was built in to avoid pattern answering, though complicates interpretation. Mental capability is expressed on a easy to read standardized 0 to 100 score, where 100 represents the best capabilities [182]. The standardized capability score is calculated as:

$$100 * (OxCAP - MH\ item\ total\ score - minimum\ possible\ score)/(max - min\ score)$$

The measurement tool covers several essential but "non-physiological health-related" life domains such as *social activities, recreation, influence on decisions, freedom of expression* that determine capabilities of individuals. The survey used the validated Hungarian language adaptation.

Patient Health Questionnaire-9 item (PHQ-9) was developed to measure self-experienced severity of depression [183]. The nine questions requests information from the past two weeks. Items are rated on a 4-point Likert scale, where 0 means "not at all" 3 denotes "nearly every day". Item sum score ranges between 0-27, originally four severity categories were established: no (0-4), mild (5-9), moderate (10-14), severe (15-19), extremely severe depression (20-27). Many studies investigated the cut-off score to signify severity of depression, most clinical studies set the sensitivity of the instrument as: non-depressed >10 score≤ depressed.

Generalized Anxiety Disorder-7 item (GAD-7) scale is a self-reported measurement tool that assesses anxiety symptom severity [184]. The instrument is also on a 4-point Likert scale describing frequency of experienced symptoms (0=not at all and 3=nearly every day) during the last two weeks. Item responses are summed resulting in score ranging between 0 and 21. The individual results are interpreted as no (>5), mild (5-9), moderate (10-14), severe (15≤) anxiety.

The ICECAP-A is a preference-based (non-mental specific) instrument designed to measure the capabilities of general population aged 18-64. It covers five domains of well-being: (1) attachment; (2) stability; (3) achievement; (4) enjoyment; (5) autonomy each rated on a 1-4 level scale [185]. Higher score represents better capability at current state (maximum total score = 20). Since no tariff was established for OxCAP-MH/PHQ-9/GAD-7 measures in Hungary, therefor we reported the total item sum scores of ICECAP-A (∑Item 1-5) because keeping the generality appears more suitable to compare the measurement scores than applying unit-weights (external tariffs) of dissimilar population. The Hungarian version proved to be valid and reliable, though it revealed little difference among population subgroups.

7.2.3. Population norm

Population normative data of OxCap-MH score was presented according to age groups and sex in groups of education level, residence, employment status, marital status, PHQ-9 and GAD-7

severity categories. No sample weight was used to achieve representativeness. The cells report the OxCAP-MH standardized score. All analysis was carried out using STATA 16.0 (StataCorp LLC).

Characteristics	**Categories**
Age group	18-24; 25-34; 35-44; 45-54; 55-64; 65≤
Sex	male; female
Level of education	primary; secondary; tertiary
Residence type	Budapest; Town; Village (less than 10,000 inhabitants)
Employment status	full-time employed or entrepreneur; part-time employed; unemployed; student; retired; inactive or disability pensioner; other such as housewife of caregiver
Marital status	married or in permanent relationship, single; divorced or widowed
PHQ-9 severity level	no; mild; moderate; severe; extremely severe
GAD-7 severity level	no; mild; moderate; severe

Table 25 - Population characteristic categories used for normative data presentation

To examine the determinants of the OxCAP-MH score an ordinary least square (OLS) multivariate regression was used including the eight variables of population characteristics as explanatory variables:

$$OxCAPMH_{\mathrm{i}} = \beta^0 + \beta_{\mathrm{i}} Age + \beta_{\mathrm{i}} Sex + \beta_{\mathrm{i}} Educ + \beta_{\mathrm{i}} Resid + \beta_{\mathrm{i}} Employ + \beta_{\mathrm{i}} Maritals + \beta_{\mathrm{i}} Depr + \gamma_{\mathrm{i}} Anx + \epsilon_{\mathrm{i}}$$

7.3. Mental capability results of the Hungarian general population

The sample consists of N=2000 responses (response rate: 79%) of Hungarian adult general population, with mean age of 46.3%, majority being female (57.3%). Most respondents completed secondary education (45.5%), worked full time or as entrepreneurs (48.7%), lived in bigger cities of Hungary (48.9%). Majority of the respondents was married or in a permanent relationship (62.1%) and fall into the non-depressed (53.1%) and non-anxious (57.0%) category. Sociodemographic characteristics (sex, age, level of education, employment status, residence, marital status, mental condition expressed with PHQ-9 and GAD-7 severity categories) are summarized − for the total sample and (for two subgroups) those who are <10 and ≥10 PHQ scores to distinguish respondents with and without depressed mood − as descriptive statistics. (Table 26)

Variable	groups	**Total**		**PHQ-9 score <10 group**		**PHQ-9 score ≥10 group**	
		n	%	n	%	n	%
Sex	male	855	42.8	701	45.0	154	34.8
	female	1145	57.3	857	55.0	288	65.2
Age (year)	18-24	202	10.1	136	8.7	66	14.9
	25-34	441	22.1	324	20.8	117	26.5
	35-44	337	16.9	249	16.0	88	19.9
	45-54	285	14.2	228	14.6	57	12.9
	55-64	337	16.9	275	17.7	62	14.0
	65<	398	19.9	346	22.2	52	11.8
Education	primary	544	27.2	390	25.0	154	34.8
	secondary	909	45.5	700	44.9	209	47.3
	tertiary	547	27.4	468	30.0	79	17.9
Residence	Budapest	390	19.5	307	19.7	83	18.8
	Town	979	48.9	769	49.4	210	47.5
	Countryside	631	31.6	482	30.9	149	33.7
Employment	full-time employed/entrepreneur	973	48.7	769	49.4	204	46.2
	part-time employed	101	5.1	71	4.6	30	6.8
	unemployed	91	4.6	62	4.0	29	6.6
	student	68	3.4	51	3.3	17	3.8
	retired	502	25.1	432	27.7	70	15.8
	other (homemaker)	189	9.5	128	8.2	61	13.8
	disability pensioner/inactive	76	3.8	45	2.9	31	7.0
Marital status	single	472	23.6	331	21.2	141	31.9
	married/in relationship	1242	62.1	1002	64.3	240	54.3
	divorced/widowed	286	14.3	225	14.4	61	13.8
PHQ-9 category	no	1061	53.1	1061	68.1	-	-
	mild	497	24.9	497	31.9	-	-
	moderate	238	11.9	-	-	238	53.8
	severe	132	6.6	-	-	132	29.9
	extremely severe	72	3.6	-	-	72	16.3
GAD-7 category	no	1140	57.0	1116	71.6	24	5.4
	mild	528	26.4	384	24.6	144	32.6
	moderate	220	11.0	54	3.5	166	37.6
	severe	112	5.6	4	0.3	108	24.4

Table 26 - Sample characteristics of the mental capability (OxCAP-MH) study

7.3.1. OxCap-MH psychometric properties

The total reliability of the OxCAP-MH instrument was good (Cronbach α=0.85), no items deleted would improve internal consistency, corrected item-total correlations of items ranged between 0.29-0.64. ICC showed strong agreement, indicating high level of test-retest reliability ($p<0.001$; 0.817). Out of 16 items, 15 were significantly correlating ($p<0.01$) on slight to moderate level (r=-0.032-0.624), except item 9 (influence on local decisions) with reverse coded items of 7 and 8. Known-group (construct) validity of the item was established

convincingly, total scores significantly ($p<0.005$) differed according to age group, residence, employment status, marital status, PHQ-9 and GAD-7 mental health state severity groups. Significant ($p<0.001$) moderately strong correlations between OxCAP-MH and two mental health state HRQoL measures (PHQ-9: r=-0.610; GAD-7: r=-0.580) and a generic capability measure (ICECAP-A: r=0.620) demonstrated fine convergent validity.

The central tendency of items suggest that most people agreed with statement 9 (15%) and 8 (unlikely to be discriminated=8%) positively influencing the capabilities. Most disagreed with statements 11-13 (appreciating nature=47%; respecting people=39%; enjoying friendly support=39%), derogating mental capability immensely. People were neutral rather neutral when responding on statement 9 and 10 (freedom of expression=34%). Item responses are presented on Figure 19, the labels record the score of the response, the interpretation is different according to instrument coding.

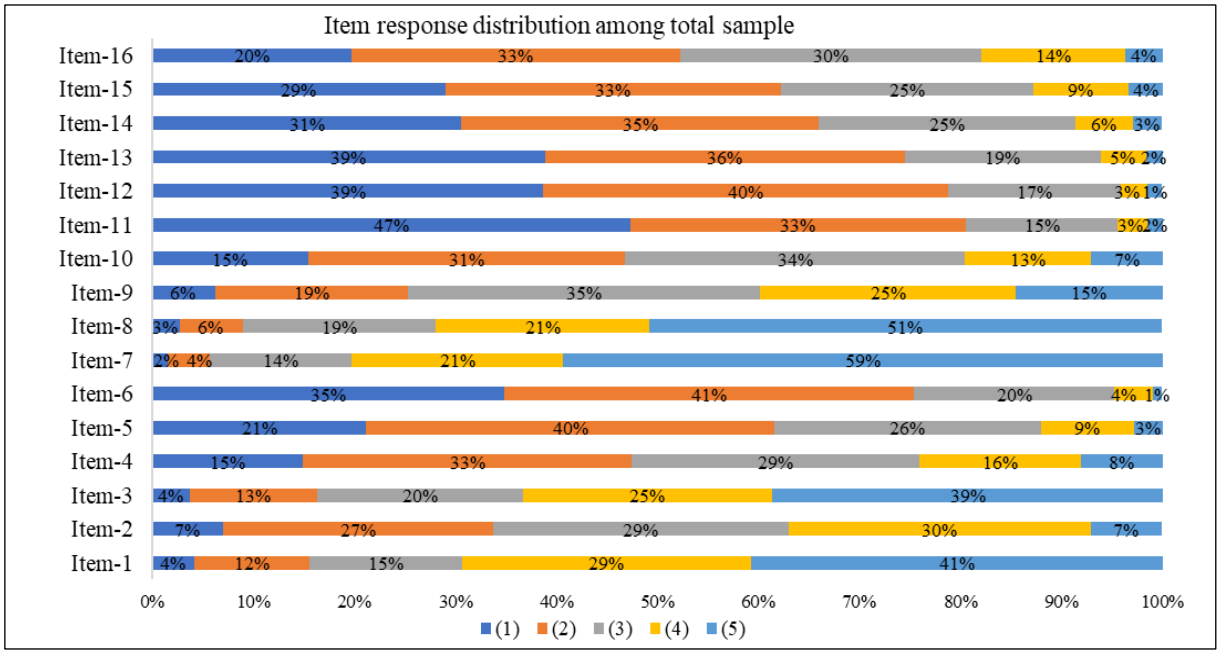

Figure 19 - Item response distribution on OxCAP-MH instrument

7.3.2. Comparison of mental health measurement scores

The mean OxCAP-MH score in the total sample was 68.5 (SD=14.4), with women having slightly − but not significantly − lower score (68.2, SD=14.1) then men (68.9, SD=14.7). Respondents with older age, higher educational level, living in Budapest, being students/retired, living as married/divorced and having no depression and anxiety symptoms had significantly ($p<0.005$) higher average OxCAP-MH score. This tendency of scores in subgroups was

consistent according to PHQ-9, GAD-7 and ICECAP-A measures as well. Significant differences were found in mean PHQ-9 and GAD-7 score in subgroups of sex, age, education level, employment, marital status, depression and anxiety severity. Average ICECAP-A total score significantly differed according to level of education, residency, employment status, marital status, depression and anxiety severity. Both capability measurement scores were the highest in no depression subgroup, and the lowest in people with extremely severe depression. (Table 27)

Sociodemographic variables	**Subgroups**	**PHQ-9** total score (0-27)	**GAD-7** total score (0-21)	**ICECAP** total score (0-20)	**OxCap-MH** standardized score (0-100)	**t-test** p-level
Total	all respondents (N=2000)	**5.9 (5.9)**	**4.8 (5.0)**	**13.5 (2.9)**	**68.5 (14.4)**	-
Sex	male	*5.2 (5.7)*	*4.0 (4.7)*	13.6 (2.9)	68.9 (14.7)	0.283
	female	*6.5 (5.9)*	*5.4 (5.1)*	13.5 (2.8)	68.2 (14.1)	
Age group	18-24	*7.8 (6.5)*	*6.3 (5.0)*	13.6 (2.8)	*66.0 (15.1)*	*<0.001*
	25-34	*6.5 (5.8)*	*5.7 (5.0)*	13.6 (2.9)	*66.3 (13.7)*	
	35-44	*6.3 (6.3)*	*5.3 (5.2)*	13.3 (3.0)	*66.5 (14.7)*	
	45-54	*5.7 (5.7)*	*4.4 (4.6)*	13.3 (2.9)	*69.9 (14.6)*	
	55-64	*5.3 (5.6)*	*4.4 (4.9)*	13.6 (3.0)	*69.5 (14.5)*	
	65<	*4.6 (5.0)*	*3.4 (4.5)*	13.8 (2.7)	*72.0 (13.3)*	
Education level	primary	*7.1 (6.9)*	*5.7 (5.6)*	*12.9 (3.1)*	*64.3 (15.0)*	*<0.001*
	secondary	*6.0 (5.6)*	*4.8 (4.8)*	*13.5 (2.8)*	*68.7 (14.2)*	
	tertiary	*4.7 (4.8)*	*3.9 (4.3)*	*14.2 (2.8)*	*72.4 (12.9)*	
Residence type	Budapest	5.7 (5.5)	4.7 (4.8)	*13.9 (2.8)*	*69.3 (13.9)*	*0.048*
	Town	5.8 (5.8)	4.7 (4.9)	*13.7 (2.9)*	*69.1 (14.3)*	
	Countryside	6.3 (6.1)	5.1 (5.2)	*13.1 (2.9)*	*67.1 (14.7)*	
Employment status	full-time employed/entrepren.	*5.6 (5.5)*	*4.6 (4.7)*	*13.9 (2.8)*	*69.3 (14.0)*	*<0.001*
	part-time employed	*7.6 (6.4)*	*6.0 (5.3)*	*12.4 (2.6)*	*62.7 (14.3)*	
	unemployed	*7.9 (6.8)*	*6.7 (5.6)*	*12.4 (3.1)*	*63.3 (14.7)*	
	student	*6.4 (5.6)*	*5.4 (4.5)*	*13.7 (2.5)*	*71.3 (14.1)*	
	retired	*4.7 (5.2)*	*3.7 (4.7)*	*13.8 (2.8)*	*71.3 (14.1)*	
	other (homemaker, caregiver)	*7.6 (6.6)*	*6.7 (5.2)*	*12.9 (3.0)*	*64.7 (13.6)*	
	disability pensioner/inactive	*9.0 (7.1)*	*6.7 (5.8)*	*12.5 (3.1)*	*60.4 (15.1)*	
Marital status	single	*7.2 (6.3)*	*5.6 (5.1)*	*13.0 (3.0)*	*65.2 (15.1)*	*<0.001*
	married/in relationship	*5.4 (5.7)*	*4.6 (4.9)*	*13.8 (2.9)*	*69.6 (14.2)*	
	divorced/widowed	*5.9 (5.5)*	*4.6 (4.8)*	*13.3 (2.5)*	*69.5 (12.9)*	
PHQ-9 category	no (0-4)	*1.6 (1.4)*	*1.7 (2.3)*	*14.8 (2.6)*	*75.5 (12.0)*	*<0.001*
	mild (5-9)	*7.0 (1.5)*	*5.7 (3.1)*	*12.9 (2.5)*	*65.6 (11.3)*	
	moderate (10-14)	*11.7 (1.5)*	*9.3 (3.6)*	*12.0 (2.2)*	*59.0 (10.3)*	
	severe (15-19)	*16.7 (1.4)*	*12.2 (4.4)*	*11.1 (2.6)*	*53.1 (10.7)*	
	extremely severe (20+)	*22.9 (2.3)*	*16.3 (4.4)*	*9.9 (2.6)*	*45.2 (12.3)*	
GAD-7 category	no (0-4)	*2.5 (2.8)*	*1.3 (1.5)*	*14.6 (2.6)*	*74.6 (12.0)*	*<0.001*
	mild (5-9)	*7.8 (4.0)*	*6.7 (1.3)*	*12.8 (2.5)*	*64.6 (12.0)*	
	moderate (10-14)	*12.5 (4.6)*	*11.9 (1.5)*	*11.8 (2.3)*	*56.5 (11.4)*	

	severe (15+)	*18.8 (5.1)*	*17.5 (2.2)*	*10.2 (2.5)*	*48.6 (12.5)*	

Table 27 - Comparison of capability (OxCAP-MH, ICECAP-A) and mental health status measurement (PHQ-9, GAD-7) mean scores

7.3.3. OxCap-MH population norm

The highest OxCAP-MH scores (75.5) were observed among no depression (PHQ-9) subgroup, while the lowest among extremely severe depression group (45.2) in the total sample. It was so in the male and female subsamples as well. The detailed OxCAP-MH population normative results as mean and SD are presented in Table 28 (the cells labelled with one*/two** asterisk represent subgroups that had n<5/n=1 respondents).

Variables	Groups	Total							Female							Male						
		total	**18-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65<**	**total**	**18-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65<**	**total**	**18-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65<**
Education	primary	64.3 (15.0)	56.7 (15.4)	63.0 (14.2)	61.9 (14.4)	66.6 (14.3)	63.8 (15.7)	69.0 (15.0)	63.8 (14.4)	58.7 (14.6)	62.9 (13.7)	63.5 (10.8)	65.1 (13.8)	61.4 (16.5)	69.1 (15.7)	65.0 (15.9)	54.0 (16.5)	63.5 (15.6)	60.0 (17.4)	69.0 (14.9)	66.8 (14.2)	68.8 (14.3)
	secondary	68.7 (14.1)	67.9 (14.3)	65.6 (13.6)	66.9 (15.5)	72.5 (13.4)	70.3 (14.1)	71.2 (12.7)	68.9 (14.0)	69.1 (13.9)	65.7 (13.1)	65.2 (16.4)	73.9 (13.1)	70.9 (13.8)	71.3 (12.9)	68.4 (14.3)	66.7 (14.8)	65.4 (14.5)	68.5 (14.6)	70.9 (13.7)	69.4 (14.7)	71.0 (12.5)
	tertiary	72.4 (12.9)	69.9 (13.5)	69.3 (13.2)	70.9 (12.3)	73.0 (15.5)	75.4 (10.5)	74.8 (12.3)	71.7 (12.7)	69.6 (14.1)	70.0 (13.2)	71.0 (11.3)	74.1 (14.1)	75.9 (10.8)	72.3 (13.0)	73.1 (13.1)	72.9 (6.5)	67.9 (13.4)	70.8 (13.7)	72.2 (16.7)	75.6 (10.3)	76.2 (11.7)
Residence	Budapest	69.3 (14.0)	63.0 (11.9)	66.0 (13.6)	69.2 (13.1)	70.5 (13.0)	72.8 (13.8)	73.4 (14.8)	69.4 (13.6)	64.4 (12.1)	66.2 (13.3)	70.8 (11.9)	70.5 (12.7)	68.7 (14.6)	74.9 (14.3)	69.3 (14.3)	60.9 (11.5)	65.6 (14.1)	67.9 (14.0)	70.5 (13.6)	77.7 (11.3)	71.8 (15.2)
	Town	69.1 (14.3)	68.1 (14.9)	67.5 (13.8)	67.2 (15.3)	70.0 (15.5)	69.8 (14.3)	71.8 (12.5)	68.8 (14.0)	69.6 (14.8)	67.5 (13.1)	66.0 (14.7)	68.3 (15.5)	71.4 (14.9)	70.2 (12.5)	69.5 (14.6)	65.6 (15.0)	67.6 (15.1)	68.2 (15.8)	71.8 (15.5)	68.1 (13.5)	73.7 (12.4)
	Countryside	67.1 (14.7)	63.6 (17.0)	64.1 (13.7)	64.1 (14.7)	69.5 (14.1)	67.8 (15.0)	71.4 (13.2)	66.8 (14.4)	62.9 (15.1)	64.8 (14.2)	65.0 (13.5)	70.1 (13.7)	66.5 (15.3)	69.8 (14.4)	67.6 (15.2)	64.3 (18.8)	62.2 (12.2)	62.4 (16.8)	68.7 (14.8)	70.3 (14.3)	73.4 (11.6)
Employment	full-time employed/ente.	69.3 (14.0)	67.3 (14.9)	67.5 (13.8)	68.3 (14.3)	72.2 (14.0)	71.1 (12.6)	79.3 (10.2)	68.9 (13.9)	70.7 (13.9)	67.3 (14.0)	68.7 (14.0)	72.0 (13.8)	67.3 (12.8)	77.7 (13.5)*	69.7 (14.1)	63.1 (15.3)	67.8 (13.6)	68.0 (14.6)	72.3 (14.2)	74.1 (11.7)	80.0 (9.2)
	part-time employ.	62.7 (14.3)	61.9 (13.8)	59.2 (13.2)	59.2 (17.9)	64.9 (12.5)	64.9 (11.9)	72.9 (14.2)	62.7 (12.8)	58.4 (8.8)	60.4 (14.1)	61.6 (13.8)	63.3 (12.7)	65.9 (13.3)	73.4 (12.4)*	62.7 (16.2)	64.9 (17.1)	56.8 (12.1)	56.1 (22.9)	67.0 (12.7)	63.4 (10.1)	72.4 (18.6)*
	unemployed	63.3 (14.7)	59.1 (12.8)	62.4 (14.8)	66.4 (13.8)	63.3 (16.6)	63.1 (15.1)	84.4 (-)**	63.2 (13.9)	55.0 (13.3)	62.4 (13.3)	64.4 (13.9)	64.4 (15.8)	66.8 (15.1)	-	63.6 (16.2)	62.5 (12.5)	62.5 (19.5)	73.4 (12.3)*	60.9 (20.2)	58.9 (15.0)	84.4 (-)**
	student	71.3 (14.1)	72.8 (12.9)	56.5 (18.8)	-	-	64.1 (-)**	-	72.1 (13.1)	72.7 (12.8)	64.6 (16.6)*	-	-	-	-	70.0 (15.7)	73.0 (13.3)	48.4 (20.3)*	-	-	64.1 (.)**	-
	retired	71.3 (14.1)	-	52.3 (3.3)*	48.8 (20.4)*	61.7 (16.5)	72.3 (14.6)	71.9 (13.3)	71.6 (14.0)	37.5 (-)**	-	-	58.6 (16.5)	74.2 (13.6)	71.1 (13.6)	71.0 (14.3)	-	52.3 (3.3)*	48.8 (20.4)*	63.7 (17.2)	67.4 (16.2)	72.8 (12.9)
	other (hom./care.)	64.7 (13.6)	56.3 (15.0)	67.4 (11.6)	62.8 (13.3)	68.6 (12.1)	66.4 (16.3)	67.8 (14.4)	65.2 (13.2)	58.1 (12.5)	67.6 (11.7)	63.3 (13.0)	68.3 (12.5)	66.0 (18.5)	61.5 (12.7)*	60.5 (16.3)	50.9 (18.8)	61.7 (3.3)*	55.2 (18.3)*	71.9 (-)**	67.5 (9.7)	77.3 (14.4)*
	disability pensioner/inac.	60.4 (15.1)	58.3 (10.9)	54.5 (8.7)	62.8 (12.9)	63.6 (16.1)	59.9 (17.9)	65.2 (10.2)*	60.9 (15.8)	65.6 (11.3)*	56.6 (10.3)	62.0 (5.9)*	65.3 (15.5)	56.8 (18.8)	71.9 (8.8)*	59.7 (14.4)	54.7 (9.5)	52.0 (6.4)*	63.5 (19.5)*	58.2 (19.5)	64.2 (16.2)	58.6 (7.7)*
Marital status	single	65.2 (15.1)	66.5 (15.7)	62.9 (14.3)	64.7 (15.2)	65.8 (16.1)	68.7 (15.2)	67.3 (13.2)	64.4 (14.6)	66.1 (15.8)	62.4 (14.6)	64.6 (13.8)	64.3 (13.7)	64.9 (16.0)	67.6 (13.4)	65.9 (15.6)	66.9 (15.8)	63.6 (14.0)	64.7 (16.0)	66.7 (17.6)	71.6 (14.3)	66.5 (13.5)
	married/in relatio.	69.6 (14.2)	66.1 (14.5)	67.9 (13.1)	67.7 (14.6)	72.0 (13.8)	69.8 (15.2)	72.6 (13.9)	69.1 (14.1)	68.3 (13.6)	68.2 (12.5)	67.3 (14.0)	71.6 (13.9)	69.3 (16.3)	70.8 (14.9)	70.2 (14.4)	61.0 (15.6)	67.3 (14.4)	68.2 (15.4)	72.4 (13.8)	70.3 (13.9)	73.8 (13.0)
	divorced/ widow.	69.5 (12.9)	54.4 (7.7)	64.1 (17.2)	64.2 (13.3)	66.1 (13.9)	69.2 (12.2)	72.1 (12.1)	69.7 (12.8)	60.9 (8.8)*	64.8 (21.1)*	63.9 (10.1)	65.5 (14.8)	69.5 (12.0)	72.0 (12.2)	68.9 (13.2)	51.2 (5.5)*	63.0 (14.5)	64.6 (18.2)	67.6 (11.9)	68.4 (13.3)	72.4 (12.1)
PHQ-9 category	no	75.5 (12.0)	74.5 (13.5)	73.3 (12.6)	74.1 (12.8)	76.6 (12.2)	76.0 (11.3)	77.4 (10.3)	75.6 (11.7)	74.1 (14.4)	73.3 (12.4)	74.6 (11.9)	76.7 (11.8)	77.3 (10.8)	77.2 (10.1)	75.4 (12.3)	75.0 (12.3)	73.3 (13.1)	73.6 (13.7)	76.5 (12.6)	74.8 (11.8)	77.6 (10.5)
	mild	65.6 (11.3)	66.3 (13.0)	64.5 (11.2)	64.4 (10.8)	65.9 (12.4)	65.1 (9.9)	67.8 (10.6)	66.9 (10.8)	70.2 (11.1)	65.3 (11.1)	64.5 (9.1)	68.0 (12.0)	65.4 (10.4)	69.8 (9.7)	63.6 (11.8)	61.2 (13.8)	62.5 (11.4)	64.4 (12.8)	63.0 (12.6)	64.6 (9.1)	65.2 (11.3)
	moderate	59.0 (10.3)	60.2 (11.0)	58.1 (10.2)	58.8 (9.7)	60.5 (9.1)	58.6 (13.3)	59.0 (8.8)	59.4 (10.9)	61.6 (11.3)	58.4 (11.1)	58.6 (11.1)	60.5 (8.3)	58.8 (13.7)	60.0 (9.5)	58.3 (9.3)	58.2 (10.6)	57.7 (8.5)	59.0 (8.3)	60.5 (11.1)	57.8 (13.1)	56.5 (6.7)
	severe	53.1 (10.7)	53.8 (8.4)	54.9 (8.9)	50.7 (10.8)	53.4 (13.2)	52.8 (13.6)	52.3 (9.9)	54.5 (10.7)	57.4 (7.8)	55.7 (9.8)	53.2 (8.8)	54.8 (14.3)	52.9 (14.4)	53.3 (9.0)	49.9 (9.9)	50.2 (7.8)	52.6 (4.6)	45.8 (13.0)	49.6 (10.2)*	52.6 (12.5)	47.9 (14.5)*
	extremely severe	45.2 (12.3)	46.0 (13.1)	47.0 (9.3)	44.1 (9.2)	49.4 (13.0)	43.2 (19.5)	39.5 (11.0)	44.3 (10.1)	48.2 (9.9)	48.4 (8.5)	45.1 (6.4)	47.1 (12.6)	36.3 (8.6)	35.2 (8.3)	46.6 (15.1)	38.0 (22.6)*	45.5 (10.6)	43.2 (11.8)	52.2 (14.4)	52.0 (27.0)*	52.3 (7.7)*
GAD-7 category	no	74.6 (12.1)	74.1 (12.5)	72.6 (12.7)	72.8 (12.9)	75.9 (12.8)	75.1 (11.1)	76.2 (10.8)	75.2 (11.4)	74.8 (12.7)	73.3 (12.3)	73.4 (12.1)	76.5 (11.7)	76.7 (10.1)	76.4 (10.2)	74.0 (12.7)	73.3 (12.4)	71.5 (13.3)	72.3 (13.6)	75.2 (13.7)	73.6 (12.0)	76.0 (11.4)
	mild	64.6 (12.0)	64.8 (13.0)	64.0 (11.8)	63.9 (12.2)	64.9 (12.0)	65.4 (12.5)	65.2 (10.6)	65.3 (11.8)	67.0 (12.6	64.0 (11.7)	64.7 (10.8)	66.2 (11.6)	65.4 (13.6)	66.7 (10.6)	63.2 (12.2)	60.5 (12.8)	64.1 (12.3)	63.0 (13.7)	63.0 (12.5)	65.5 (10.4)	61.6 (9.8)
	moderate	56.5 (11.4)	56.2 (12.3)	56.7 (8.4)	54.9 (13.3)	56.8 (11.7)	56.3 (11.5)	59.1 (12.0)	58.3 (11.4)	59.7 (13.1)	57.9 (8.8)	57.1 (11.5)	56.9 (12.2)	57.1 (12.0)	61.4 (11.7)	53.1 (10.7)	51.0 (9.2)	54.1 (7.0)	51.8 (15.2)	56.6 (10.9)	53.8 (10.3)	52.9 (11.0)
	severe	48.6 (12.5)	47.8 (15.0)	50.4 (11.5)	50.4 (10.7)	50.4 (11.5)	43.2 (15.5)	46.8 (12.4)	47.5 (10.0)	49.2 (11.0)	49.6 (8.0)	50.7 (9.2)	47.9 (8.8)	42.1 (10.6)	43.1 (11.6)	50.8 (16.6)	44.9 (23.0)	51.7 (16.2)	49.7 (14.2)	55.2 (16.7)*	46.5 (27.9)*	57.8 (7.8)

Table 28 – OxCAP-MH population normative results

According to the results of the OLS multivariate linear regression, respondents with older age (β=0.09), living in towns as compared to villages (β=1.32), employed full time/entrepreneurs and students (β=4.30 and β=9.11), being married/in permanent relationship (β=1.48) had significantly higher OxCAP-MH capability scores. The marginal effect on increasing the capability was the highest of being student compared to disability pensioners/inactive. Also, females (β=-1.96); mild-moderate-severe-extremely severe (β=-6.97, -10.78, -14.75, -19.59) depression; and mild-moderate-severe (β=-4.01, -7.01, -9.82) anxiety groups in contrast to asymptotic population had significantly lower OxCAP-MH capability scores, respectively. The seven explanatory sociodemographic variables included into the model explained 41% of the variation in the mental-capability score, the overall model was significant (R^2= 40.9, p<0.001). (Table 29)

Variables/subgroups		**β coefficient**	**standard error**	**p-value**
Age	years	***0.091***	***0.02***	***<0.001***
Sex	female	***-1.961***	***0.53***	***<0.001***
Education level	secondary	***2.458***	***0.62***	***<0.001***
	tertiary	***4.164***	***0.70***	***<0.001***
Residence	Budapest	1.267	0.73	0.082
	town	***1.317***	***0.57***	***0.022***
Employment	full-time/entrepreneur	***4.302***	***1.35***	***0.002***
	part-time	0.684	1.70	0.688
	unemployed	1.910	1.74	0.273
	student	***9.111***	***1.96***	***<0.001***
	retired	1.836	1.46	0.208
	homemaker/other	2.865	1.56	0.066
Marital status	married/in relationship	***1.481***	***0.65***	***0.022***
	widowed/divorced	1.314	0.94	0.164
PHQ-9	mild	***-6.969***	***0.72***	***<0.001***
	moderate	***-10.775***	***1.04***	***<0.001***
	severe	***-14.745***	***1.33***	***<0.001***
	extremely severe	***-19.592***	***1.87***	***<0.001***
GAD-7	mild	***-4.011***	***0.74***	***<0.001***
	moderate	***-7.011***	***1.11***	***<0.001***
	severe	***-9.820***	***1.62***	***<0.001***
Regression indices	Constant	***64.700***	***1.86***	***<0.001***
	N observation	2000		
	R^2	40.9%		
	RMSE	11.1		
	significance level	p<0.001		

Table 29 - Regression results of sociodemographic factors impacting the OxCAP-MH capability score

8. Discussion

The thesis investigated direct and indirect health state utility measurement methods in different diseases and general population. The thesis consists of eight parts, first, the introduction covers the broad theoretical background and practical application of HSU measurement methods. Chapters 4-7 report the findings of Our four previously published studies compacting the research methods and results. Finally, the main points of the thesis are discussed here and summarized.

The introduction overviews health economic evaluations models, that apply direct/indirect HSU assessment for HRQoL measurement to express health outcome (as QALY gain). The novel capability approach was employed to examine non-health domains of well-being as an alternative approach beyond QALY. Two qualitative analytical studies (Chapter 4 and 5 answer Hypothesis 1) report the results of two systematic reviews of (i) TTO measurements in seven chronic diseases among Hungarian population, and (ii) TTO studies in depression with a meta-analysis of vignette-based utility results. Two applied parts (Chapter 6 and 7 answering on Hypothesis 2-4) report the results of "in-process" quantitative studies of (iii) measurement agreement between direct (TTO) and indirect (EQ-5D, DLQI) utility assessments is four common dermatological diseases, and (iv) evaluating the Hungarian population's mental capability by OxCAP-MH instrument. The discussion (Chapter 8) unfolds theoretical limitations of the economic evaluations, followed by the discussion of studies and their limitations/conclusions one by one.

8.1. Challenges of health economic evaluations

The major advantage of the QALY measurement is its generalizability, that is often referred as "a QALY is a QALY, that is a QALY" [186]. Although such generalizability conceals weaknesses. The limitations of the health outcome measurement is twofold.

i. Firstly, the QALY has conceptual shortcomings. Measurement across diverse disease categories is not settled, while in practice for example health states in palliative care and dermatology are hardly comparable. The core element of the QALY is given by different utility measures, albeit evidence shows that utility measurement instruments yield systematically different results. The source of quality weights is also unsolved, on the one hand the taxpaying general population is the financer of the healthcare system, on the other hand the

patient's preferences are more adequate related to the disease. Further tense debate is about discounting the costs and health outcomes, while time is an important factor for instance in regenerative medicine where cost reduce but quality of life may improve over time. Majority states that the net present value of both cost and benefits shall be discounted at the same rate, contrarily some pharmaco-economists say that outcomes should not be discounted at all [187, 188].

ii. Secondly, there are a number of technical constraints due to its quantitative nature. QALY is indifferent whether the improvement comes from a small or large number of people. The QALY is non-sensitive for clinically meaningful small changes in health status as well as to contextual (individual are societal) factors.

Utility assessments are either *direct* that are preference-based valuations based on rational choice theory or *indirect* multi-attribute utility instruments rated on a scale. Due to the difference in cognitive evaluation the measurements yield different results. In the introduction the deviation because of TTO method attributes and EQ-5D item-weights (tariffs) were described. However, systematic distortions may also influence utility estimates [189]:

(1) People value recent health gains more due to their time preference, but tasks do not discount future states.
(2) Framing effect creates an overly optimistic perceptions, when people over/under value the effect of events.
(3) Behavioural economic assumptions state that individuals value gains less than avoiding losses.

Therefor the QALY concept built quality weighted quantity of life has several technical shortfalls. Also, QALY maximization models disregard resource allocation issues (egalitarian arguments), like whom the society prioritizes (priority to young, more severe patients, caregivers, marginalized groups). The quality of life may correlate with the income level, that is double counted when calculating the incremental cost-effectiveness ratio. Further concern is that QALY gain is not adjusted for the initial severity of the health state, while a small 0.05 QALY improvement is not the same for a worse basis (0.1) compared to a relatively good health state (0.9). The combination of quality and quantity of life could be misleading, as mentioned in the introduction, improving just life expectancy or HRQoL is extreme. In brief, the QALY concept and the determination of its components is still under development. Evaluations that

apply QALY for more effective financial decisions can be as biased as the ones that ignore economic outcomes if the societal setting is neglected [190].

8.2. Systematic literature review of TTO in Hungary study

The systematic literature review of TTO measurement in Hungary found nine papers reporting seven original studies evaluating seven chronic diseases in total. In Central and Eastern Europe, Hungary leads the number of TTO studies [191]. However, almost all countries in the region (Hungary [61], Poland [192], Slovenia [193], and Romania [194]) have EQ-5D value sets based on TTO, suitable for indirect HSU elicitation.

The majority of the selected studies focused on two areas of chronic diseases, on the one hand, diseases for which effective but high-cost biological therapies have appeared in the last two decades (Crohn's disease, age-related macular degeneration, pemphigus, psoriasis, and rheumatoid arthritis), on the other hand, diseases associated with chronic pain (migraine and primary dysmenorrhea). Internationally, TTO has already been used in determining the quality of life in many chronic conditions with a high disease burden (cardiovascular diseases, diabetes, asthma, cancer and mental illnesses) but no such research has yet been conducted in Hungary [195].

Several evidences were reported on errors related to the use of TTO, mostly as incomplete communication of methodology and results [196, 197]. In only two cases (rheumatoid arthritis and macular degeneration) were found deficiencies in the detailed description of methodology and interpretation of results. Researchers tried to adapt those elements of TTO method that fit to the nature of the examined health state and population. The results of the review suggest that majority of the TTO studies in Hungary are in line with international quality standards.

In Hungary, the cost-effectiveness analysis of healthcare technologies is mandatory [198]. The current health economics directive recommends using direct HSU measurement if indirect utility measurement (EQ-5D) is not available. Health state utility values based on preferences of the Hungarian society vastly contribute to more accurate health economic evaluations (such as health technology assessment of new medicines, diagnostic procedures and medical aids [199]) by enabling not only cost data, but health quality data input from domestic sources. National HSU measurement promotes effective resource allocation of the healthcare financer. There is EQ-5D utility review [191], but no previous study systematized TTO studies in Hungary.

The limitation of the study is that (1) from HSU measures only TTO assessment were selected. Due the various chronic diseases and heterogeneity of applied TTO methods, the (2) utility outcomes cannot be compared, thus meta-analysis was not conducted. Besides the TTO, (3) other direct utility measurement methods (like SG, DCE, BWS) have also appeared in, that was not introduced in this study. There is no official TTO methodological harmonization, (3) though no quality check of task was performed based on the current recommendations and guides (MVH, Paris or EQ-VT protocol) [84].

The systematic review of TTO studies conducted in Hungary summarizes the results of nine publications, reporting results in seven chronic diseases. Studies covered a broad-range of TTO methodological attributes, regarding the type, timeframe, and evaluated health states. Economic evaluations using domestic cost and quality of life data facilitate the opportunity for more effective decision-making in healthcare.

8.3. Systematic literature review and meta-analysis in depression study

In the review of TTO studies, HSU of vignette-based or self-experienced depression-related health states were catalogued. The analysis compared the health description vignettes used for the hypothetical assessments. The pooled mean utilities of mild, moderate and severe depression elicited in depressed and non-depressed groups were estimated in a REML meta-analysis.

Overall, 14 articles reporting 36 self-experienced and 25 vignette-based utilities for depression-related health states were identified. The mean utility of self-experienced depression in patient groups ranged between 0.89 and 0.24, while the vignette-based mean utility of mild, moderate and severe depression (in self-reported depressed and healthy groups) ranged between 0.91-0.66, 0.79-0.49 and 0.66-0.31, respectively. A previous review reported somewhat lower SG and EQ-5D utilities in the context of unipolar depression, that ranged from 0.92-0.09 and 0.90-0.14, respectively [200].

Little evidence is reported on description vignettes impact on direct utility [92]. The review compared six vignettes in terms of origin, covered health domains, mode of presentation, that concluded in revealing five different methodological approaches to vignette development. Although methodological recommendations are provided [201], the nonstandardized application of vignettes resulted in different health state description, even if the same scale (McSad) was employed. These finding haste the use of a common approach to vignette development.

The meta-analysis of vignette-based TTO utilities for mild, moderate and severe depression state results were: 0.75, 0.66 and 0.50, respectively. Meta-regression revealed that evaluating severe depression (β =-0.16) and having depressed population (β =-0.13) has a significant negative impact on the pooled utility estimates. The findings of the review support the claim that patients report systematically different health utilities than the general population [74, 76]. However, the analysis contradicts the assumption that patients report higher utilities, especially in self-experienced states, as proven in many previous studies [202]. In all comparable (n=18) vignette-based health states, depressed groups had lower pooled mean utility than non-depressed in mild, moderate and severe depression. The heterogeneity of patients self-experienced depression states prevented to further support the evidence of discrepancy in patient's vs healthy population's HSU results.

Nine of the included studies used other methods besides TTO to elicit utilities pertaining to the same health state. Three studies reported consistent order in self-experienced utilities: rating scale (RS)<TTO<SG [138, 139, 148]. One research group also reported that RS<TTO utility in a vignette-based assessment [144]. Another three valuations revealed that TTO<SG utility [137, 140, 143]. Vignette-based evaluation of mild, moderate & severe depression comparison resulted in the order of EQ-5D<TTO<VAS, with the three approaches presenting considerable differences [149].

Limitations of the study conceal (i) the substantial heterogeneity of the included studies regarding the depression states and TTO task. Heterogeneous studies using different vignettes (ii) were included in the meta-analysis as well. The quality evaluation of TTO studies (iii) was based on the former MVH protocol, as being closer to the publication dates of the studies, than the EuroQol's EQ-VT valuation protocol that launched in 2012.

As a conclusion the results of the review suggest that the utility generation method may significantly impact the utility value associated with depression. To our knowledge, this is the first study to provide a comprehensive utility catalogue in depression and compare description vignettes. The meta-regression showed decreased pooled mean utility when severe level of depression and depressed sample was evaluated. Interestingly, patients' perceptions of hypothetical depression health states were worse than those of healthy respondents.

8.4. Measurement agreement between direct and indirect utility measures study

The study aimed to explore the agreement between direct (TTO), one indirect general (EQ-5D-5L) and two indirect dermatology-specific (mDLQI and vDLQI) HSU measures using a pooled

dataset of four multi-centre, cross-sectional, patient reported outcome studies in atopic dermatitis, hidradenitis suppurativa, pemphigus and psoriasis. Bland-Altman analysis revealed measurement agreement without proportional bias was only found between TTO and EQ-5D-5L, while no agreement was found between TTO-mDLQI/vDLQI, and EQ-5D-5L – mDLQI/vDLQI, showing considerable differences in all cases. Despite the proportional bias in measurement scores, EQ-5D index and vDLQI may present a certain level of in-between individual agreement. Although the source of the observed agreement may be attributable to the common national value set, representing the preferences of the Hungarian general population.

Total sample mean and median scores in descending order was as follows: TTO utility scores (0.83 and 0.95), EQ-5D-5L (0.81 and 0.89), vDLQI utility (0.81 and 0.84) and mDLQI (0.77 and 0.79). Similar tendency was observed in mean utility values in subgroups, where TTO was the highest in ten out of eighteen cases (psoriasis, atopic dermatitis, both genders, higher educated, part time and other employment status, outpatient care visitors), vDLQI utility in five and EQ-5D-5L utility was on top in three subgroups. In fourteen subgroups mDLQI converted utility, in three cases the EQ-5D-5L, once the TTO (pemphigus) had the lowest score among HSU measures. Thus, the findings support the claim that higher mean utilities are elicited with direct measurement method, though there are exceptions reported in the literature [66, 203, 204].

Considering the data distribution, the ceiling results of TTO affirmed it is more likely to result perfect health state, while DLQI utility was closer to normal distribution. The high rate of non-trading respondents imply that trading of life years is dependent on more latent factors (e.g. personal believes, intrafamilial considerations). The four tobit regression models showed that disease type and skin condition severity had significant effect on all HSU estimates. Socioeconomic variables of age, gender, level of education and employment status made impact on certain HSU measures.

Direct utility measures that follow rational choice theory output significantly different results than indirect utility measurements based on MAU theory. Beyond the often mentioned shortfalls of the measures − (1) the TTO being too demanding, (2) the EQ-5D-5L being too narrow/oversimplified, (3) converting HRQoL instrument scores with mapping into utility may yield inadequate values − more pitfalls may distinguish direct and indirect HSU measures. The primary source of the problem is the difference in cognitive process required from the respondents: preference-base evaluations present a demanding task that encourage using

heuristics/framing effect/relativization/mutable preferences, while item-rating MAU tasks face with pattern answers/non-revealed health domains/ex-post weighting the items with population tariffs [205-208]. This is the first study that investigated measurement agreement between TTO, EQ-5D-5L and two DLQI-based utility estimates, also pinpoint that except DLQI currently there is no other skin-specific MAU instrument offering mapping algorithm or value set to calculate indirect HSU.

The study is limited in not considering that (1) the mapping algorithm for DLQI utility was optimized for the British EQ-5D-3L value set. Further issue is (2) the dissimilar nature of diseases, that yields differences in HRQoL domains. Converting the DLQI item scores into utilities in four fairly different conditions limit the comparability of utility results (the DLQI total score in pemphigus was 5.4, but 13.4 in atopic dermatitis, thus the converted mean vDLQI utility was 0.84 in pemphigus, while 0.78 in atopic dermatitis. As opposed the TTO utility was the lowest in pemphigus (0.72), the highest in atopic dermatitis (0.85). This phenomenon highlights, that DLQI items describe a certain set of skin-condition related problems (e.g. itching, appearance, social discernment), while TTO rather focuses on respondents overall health state associated with a disease, but there is no necessarily a match between the two [209].

To conclude, measurement agreement was established only between TTO and EQ-5D-5L measures, also both showed disagreement with m/vDLQI measures. In between person agreement was solid in case of EQ-5D-5L – DLQI. The mean utility results in subgroups contribute to the evidence that direct measures yield higher utility values. The abyss found between TTO and DLQI converted utility in atopic dermatitis and pemphigus warns to compare DLQI-based utilities in different dermatological conditions with extreme caution.

8.5. Mental capability measurement and OxCAP-MH population norm study

The primary objective of the study was to develop the fist OxCAp-MH mental capability population normative data and assess the psychometric properties of the instrument with CTT. Large-sample, cross-sectional study among the Hungarian general population was conducted in 2021 August. In total N=2000 respondents answered the online, self-administrated questionnaire. The OxCAP-MH instrument showed good internal consistency (Cronbach's α=0.85), high test-retest reliability (ICC=0.817), fine construct validity (all but one items significantly correlated on a slight to moderate level: r=-0.032-0.624), and robust convergent validity (moderately strong significant correlations were observed between capability and HRQoL tools: r=-0.580-0.620) confirming solid validity and reliability [45].

The total sample mean OxCAP-MH score was 68.5 (SD=14.4), the highest in people with no depression symptoms (75.5), the lowest among people with extremely severe depression (45.2). The average standardized mental-capability score reported as normative data, was significantly higher among older people (β=0.09), living in towns (β=1.32), full-time employed/entrepreneurs (β=4.30) or students (β=9.11) and married (β=1.48).

Nevertheless, interesting controversies were found between the capability items. The *influence on local decisions* (item 9) was the only non-correlating item with the *likelihood of assault* and *likelihood of future discrimination* (item 7 and 8). The lack of relationship may be attributable to the scoring system specificity, while items 7 and 8 are reverse coded (1=strong agreement with a negatively framed statement) and item 9 is the first question that is rated on a regular 1-5 Likert-scale, that could be due disturbance in understanding. Also, item 9 has the weakest correlations with more reverse coded statements regarding *daily activities* and *worries* (r=0.032 and 0.088, item 1 and 3). Further explanation could be that concerns about the future (only asked in question 7 and 8) are poorly manifested in respondents' current expectations.

OxCAP-MH measure was specifically designed to consider mental-health related capabilities, that appeared to work well, while the measure was more sensitive in subgroups according to depression and anxiety severity [36]. The convergence analysis pointed out that the OxCAP-MH in this sample had slightly better discriminatory power than ICECAP-A in age groups. The commonly recommended ICECAP-A for health economic evaluations might quantify the benefit of interventions in terms of capability accurately among working age general population [38]. Although, OxCAP-MH can meaningfully contribute to assess well-being, when the mental health state of the population matter. None of the two instruments cover health domains separately, however, change in mental-state is likely to be better captured with mental-capabilities, while if general population is assessed the change in ability may be caused by either physical/mental-state impact [210, 211].

Limitations of the study should be mentioned. Firstly, the respondents were recruited from a panel database, during the end on the third wave of the COVID-19 pandemic. Secondly, the responsiveness of the instrument was not assessed, thus causality examination between statements is inconclusive. Thirdly, general adult population sample was analysed, that disregards the uniqueness of the mental-capability instrument.

To summarise the findings, validity and reliability of the OxCAP-MH was confirmed. The first mental-capability population norm provides a viable alternative to express current well-being of the Hungarian general population. Further research can investigate the choice of

instruments linked to sample population, and the potential role of mental-capability measurement in health economic evaluations.

9. Summary

9.1 Answering the research question

In the following, the results of the above introduced applied studies will be used to fill four research gaps:

H1: Methodological attributes of the TTO task, make impact on utility estimates. The role of task type and timeframe was evident from the findings of the review of Hungarian studies, while meta-regression proved that health state description and responding population affect mean TTO utility. (Chapter 4.2 and 5.2)

H2: Sociodemographic (age, sex, education level, employment and marital status) and disease specific characteristics (disease type and severity) make impact on utility estimates, as found by the tobit regression analysis in four dermatological diseases. (Chapter 6.3)

H3: Direct (TTO) and indirect (EQ-5D-5L) HSU measures do yield systematically different results, but agreement was found between the two methods. (Chapter 6.3)

H4: OxCAP-MH mental-capability measure effectively captures non-health domains of well-being, clearly discriminates capability scores along depression and anxiety severity groups. (Chapter 7.3)

9.2 Conclusion

The thesis contributed to analyse health state utility measurement – as a crucial element of providing quality of life component for economic evaluations – with direct and indirect methods. The review of TTO studies in Hungary overviewed HRQoL in seven chronic diseases and implicitly found difference in valuation of patients and general population. The systematic review of TTO studies in depression catalogued 61 depression-related health states, while compared six health description vignettes. The meta-analysis estimated pooled mean utilities for vignette-based mild (0.75), moderate (0.66) and severe (0.50) depression and found that evaluating severe health state and having depressed population decreases TTO utility. Empirical analysis of disease-related and individual factors in four dermatological conditions shows systematic difference in direct and indirect utility results, although measurement agreement was found between TTO and EQ-5D-5L assessments. Mental-capability of the Hungarian general population was assessed by OxCAP-MH instrument, where the normative data indicated better

capabilities with older age, higher education, living in town, being employed/student and living as married [105-108].

References

1. Ferrans, C.E., et al., *Conceptual model of health-related quality of life.* Journal of nursing scholarship, 2005. **37**(4): p. 336-342.
2. Folland, S., A.C. Goodman, and M. Stano, *The economics of health and health care: Pearson new international edition*. 2016: Routledge.
3. Brazier, J., et al., *Measuring and valuing health benefits for economic evaluation*. 2017: OXFORD university press.
4. Weinstein, M.C., G. Torrance, and A. McGuire, *QALYs: the basics.* Value Health, 2009. **12 Suppl 1**: p. S5-9.
5. Tsevat, J., *What do utilities measure?* Med Care, 2000. **38**(9 Suppl): p. Ii160-4.
6. Gandjour, A., *Drug pricing and control of health expenditures: a comparison between a proportional decision rule and a cost-per-QALY rule.* Int J Health Plann Manage, 2015. **30**(4): p. 395-402.
7. Banta, D., *The development of health technology assessment.* Health policy, 2003. **63**(2): p. 121-132.
8. Sirgy, M., *A Quality-of-Life Theory Derived from Maslow's Developmental Perspective.* American Journal of Economics and Sociology, 2006. **45**: p. 329-342.
9. McSweeny, A.J. and T.L. Creer, *Health-related quality-of-life assessment in medical care.* Disease-a-Month, 1995. **41**(1): p. 6-71.
10. Muldoon, M.F., et al., *What are quality of life measurements measuring?* Bmj, 1998. **316**(7130): p. 542.
11. Fishburn, P.C. and G.A. Kochenberger, *TWO-PIECE VON NEUMANN-MORGENSTERN UTILITY FUNCTIONS*.* Decision Sciences, 1979. **10**(4): p. 503-518.
12. Torrance, G.W., *Measurement of health state utilities for economic appraisal.* J Health Econ, 1986. **5**(1): p. 1-30.
13. Torrance, G.W., *Utility approach to measuring health-related quality of life.* Journal of chronic diseases, 1987. **40**(6): p. 593-600.
14. Attema, A.E., et al., *Time trade-off: one methodology, different methods.* Eur J Health Econ, 2013. **14 Suppl 1**(Suppl 1): p. S53-64.
15. Janssen, B.M., et al., *Introducing the composite time trade-off: a test of feasibility and face validity.* The European Journal of Health Economics, 2013. **14**(1): p. 5-13.
16. Sen, A., *Capability and well-being73.* The quality of life, 1993. **30**: p. 270-293.
17. Briggs, A., M. Sculpher, and K. Claxton, *Decision modelling for health economic evaluation*. 2006: Oup Oxford.
18. Briggs, A.H. and B.J. O'Brien, *The death of cost-minimization analysis?* Health Econ, 2001. **10**(2): p. 179-84.
19. Engström, A., et al., *A case study of cost-benefit analysis in occupational radiological protection within the healthcare system of Sweden.* J Appl Clin Med Phys, 2021. **22**(10): p. 295-304.

20. Brouwer, W., et al., *When is it too expensive? Cost-effectiveness thresholds and health care decision-making*. 2019, Springer. p. 175-180.

21. Thavorncharoensap, M., et al., *A systematic review of demand-side methods of estimating the societal monetary value of health gain.* Value in Health, 2021. **24**(10): p. 1423-1434.

22. Cameron, D., J. Ubels, and F. Norström, *On what basis are medical cost-effectiveness thresholds set? Clashing opinions and an absence of data: a systematic review.* Global health action, 2018. **11**(1): p. 1447828.

23. Iino, H., M. Hashiguchi, and S. Hori, *Estimating the range of incremental cost-effectiveness thresholds for healthcare based on willingness to pay and GDP per capita: A systematic review.* PLoS One, 2022. **17**(4): p. e0266934.

24. Shiroiwa, T., et al., *International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness?* Health Econ, 2010. **19**(4): p. 422-37.

25. Nimdet, K., et al., *A systematic review of studies eliciting willingness-to-pay per quality-adjusted life year: does it justify CE threshold?* PLoS One, 2015. **10**(4): p. e0122760.

26. Gloria, M.A.J., et al., *A Systematic Review of Demand-Side Methods of Estimating the Societal Monetary Value of Health Gain.* Value Health, 2021. **24**(10): p. 1423-1434.

27. Kouakou, C.R.C. and T.G. Poder, *Willingness to pay for a quality-adjusted life year: a systematic review with meta-regression.* Eur J Health Econ, 2022. **23**(2): p. 277-299.

28. Brouwer, W.B., et al., *Welfarism vs. extra-welfarism.* Journal of health economics, 2008. **27**(2): p. 325-338.

29. Baker, R., et al., *Searchers vs surveyors in estimating the monetary value of a QALY: resolving a nasty dilemma for NICE.* Health Economics, Policy and Law, 2011. **6**(4): p. 435-447.

30. Haraldstad, K., et al., *A systematic review of quality of life research in medicine and health sciences.* Qual Life Res, 2019. **28**(10): p. 2641-2650.

31. Karimi, M. and J. Brazier, *Health, Health-Related Quality of Life, and Quality of Life: What is the Difference?* PharmacoEconomics, 2016. **34**(7): p. 645-649.

32. Kaplan, R.M. and A.L. Ries, *Quality of Life: Concept and Definition.* COPD: Journal of Chronic Obstructive Pulmonary Disease, 2007. **4**(3): p. 263-271.

33. Taillefer, M.-C., et al., *Health-Related Quality of Life Models: Systematic Review of the Literature.* Social Indicators Research, 2003. **64**(2): p. 293-323.

34. Drummond, M.F., et al., *Methods for the economic evaluation of health care programmes*. 2015: Oxford university press.

35. Husereau, D., et al., *Consolidated Health Economic Evaluation Reporting Standards (CHEERS)—Explanation and Elaboration: A Report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force.* Value in Health, 2013. **16**(2): p. 231-250.

36. Simon, J., et al., *Operationalising the capability approach for outcome measurement in mental health research.* Social Science & Medicine, 2013. **98**: p. 187-196.

37. Helter, T.M., et al., *Does the relative importance of the OxCAP-MH's capability items differ according to mental ill-health experience?* Health Qual Life Outcomes, 2022. **20**(1): p. 99.

38. Baji, P., et al., *Capability of well-being: validation of the Hungarian version of the ICECAP-A and ICECAP-O questionnaires and population normative data.* Quality of Life Research, 2020. **29**(10): p. 2863-2874.

39. Proud, L., C. McLoughlin, and P. Kinghorn, *ICECAP-O, the current state of play: a systematic review of studies reporting the psychometric properties and use of the instrument over the decade since its publication.* Qual Life Res, 2019. **28**(6): p. 1429-1439.

40. Afentou, N. and P. Kinghorn, *A Systematic Review of the Feasibility and Psychometric Properties of the ICEpop CAPability Measure for Adults and Its Use So Far in Economic Evaluation.* Value Health, 2020. **23**(4): p. 515-526.

41. Shiroiwa, T., et al., *Development of Japanese utility weights for the Adult Social Care Outcomes Toolkit (ASCOT) SCT4.* Qual Life Res, 2020. **29**(1): p. 253-263.

42. Trukeschitz, B., et al., *What's important when caring for a loved one? Population-based preference weights for the Adult Social Care Outcomes Toolkit for informal carers (ASCOT-Carer) for Austria.* Qual Life Res, 2021. **30**(7): p. 1975-1984.

43. Rand, S., et al., *Validity and test-retest reliability of the self-completion adult social care outcomes toolkit (ASCOT-SCT4) with adults with long-term physical, sensory and mental health conditions in England.* Health Qual Life Outcomes, 2017. **15**(1): p. 163.

44. Łaszewska, A., et al., *Measuring broader wellbeing in mental health services: validity of the German language OxCAP-MH capability instrument.* Qual Life Res, 2019. **28**(8): p. 2311-2323.

45. Helter, T.M., et al., *Internal and external aspects of freedom of choice in mental health: cultural and linguistic adaptation of the Hungarian version of the Oxford CAPabilities questionnaire-Mental Health (OxCAP-MH).* BMC Psychol, 2021. **9**(1): p. 161.

46. Simon, J., et al., *Cultural and linguistic transferability of the multi-dimensional OxCAP-MH capability instrument for outcome measurement in mental health: the German language version.* BMC Psychiatry, 2018. **18**(1): p. 173.

47. Vergunst, F., et al., *Psychometric validation of a multi-dimensional capability instrument for outcome measurement in mental health research (OxCAP-MH).* Health Qual Life Outcomes, 2017. **15**(1): p. 250.

48. Nederland, Z., *Guideline for economic evaluations in healthcare.* Diemen: ZIN, 2016.

49. Proud, L., C. McLoughlin, and P. Kinghorn, *ICECAP-O, the current state of play: A systematic review of studies reporting the psychometric properties and use of the instrument over the decade since its publication.* Quality of Life Research, 2019. **28**: p. 1429-1439.

50. Dworzynski, K., et al., *Rehabilitation after stroke: summary of NICE guidance.* Bmj, 2013. **346**.

51. Alaqeel, S., et al., *Patient-reported outcome measures in Arabic-speaking populations: a systematic review.* Quality of Life Research, 2022. **31**(5): p. 1309-1320.

52. Pattinson, R.L., et al., *Patient-Reported Outcome Measures in Dermatology: A Systematic Review.* Acta Derm Venereol, 2021. **101**(9): p. adv00559.

53. Culyer, A.J., *Encyclopedia of health economics*. 2014: Newnes.

54. Torrance, G.W., *Utility approach to measuring health-related quality of life.* J Chronic Dis, 1987. **40**(6): p. 593-603.

55. Torrance, G.W., *Social preferences for health states: an empirical evaluation of three measurement techniques.* Socio-economic planning sciences, 1976. **10**(3): p. 129-136.

56. Torrance, G.W., W.H. Thomas, and D.L. Sackett, *A utility maximization model for evaluation of health care programs.* Health services research, 1972. **7**(2): p. 118.

57. Ryan, M., *Discrete choice experiments in health care*. 2004, British Medical Journal Publishing Group. p. 360-361.

58. Ryan, M., K. Gerard, and M. Amaya-Amaya, *Using discrete choice experiments to value health and health care*. Vol. 11. 2007: Springer Science & Business Media.

59. Flynn, T.N. and A.A. Marley, *Best-worst scaling: theory and methods*, in *Handbook of choice modelling*. 2014, Edward Elgar Publishing. p. 178-201.

60. Poudel, N., et al., *Methodological similarities and variations among EQ-5D-5L value set studies: a systematic review.* J Med Econ, 2022. **25**(1): p. 571-582.

61. Rencz, F., et al., *Parallel valuation of the EQ-5D-3L and EQ-5D-5L by time trade-off in Hungary.* Value in Health, 2020. **23**(9): p. 1235-1245.

62. Richardson, J., et al., *Comparing and explaining differences in the magnitude, content, and sensitivity of utilities predicted by the EQ-5D, SF-6D, HUI 3, 15D, QWB, and AQoL-8D multiattribute utility instruments.* Medical Decision Making, 2015. **35**(3): p. 276-291.

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

64. Arnold, D., et al., *Comparison of direct and indirect methods of estimating health state utilities for resource allocation: review and empirical analysis.* Bmj, 2009. **339**.

65. Doctor, J.N., H. Bleichrodt, and H.J. Lin, *Health utility bias: a systematic review and meta-analytic evaluation.* Medical Decision Making, 2010. **30**(1): p. 58-67.

66. Blieden Betts, M., et al., *Differences in utility elicitation methods in cardiovascular disease: a systematic review.* Journal of Medical Economics, 2018. **21**(1): p. 74-84.

67. Meregaglia, M. and J. Cairns, *A systematic literature review of health state utility values in head and neck cancer.* Health and quality of life outcomes, 2017. **15**(1): p. 1-13.

68. Liu, L., et al., *Health state utilities and subjective well-being among psoriasis vulgaris patients in mainland China.* Quality of Life Research, 2018. **27**(5): p. 1323-1333.

69. Matza, L.S., et al., *The time horizon matters: results of an exploratory study varying the timeframe in time trade-off and standard gamble utility elicitation.* The European Journal of Health Economics, 2016. **17**(8): p. 979-990.

70. Hinz, A., et al., *The relationship between perceived own health state and health assessments of anchoring vignettes.* International Journal of Clinical and Health Psychology, 2016. **16**(2): p. 128-136.

71. Attema, A.E. and W.B.F. Brouwer, *In search of a preferred preference elicitation method: A test of the internal consistency of choice and matching tasks.* Journal of Economic Psychology, 2013. **39**: p. 126-140.

72. Luo, N., et al., *The effects of lead time and visual aids in TTO valuation: a study of the EQ-VT framework.* The European Journal of Health Economics, 2013. **14**(1): p. 15-24.

73. King Jr, J.T., et al., *"Perfect health" versus "disease free": the impact of anchor point choice on the measurement of preferences and the calculation of disease-specific disutilities.* Medical decision making, 2003. **23**(3): p. 212-225.

74. Helgesson, G., et al., *Whom should we ask? A systematic literature review of the arguments regarding the most accurate source of information for valuation of health states.* Quality of Life Research, 2020. **29**(6): p. 1465-1482.

75. McTaggart-Cowan, H., *Elicitation of informed general population health state utility values: a review of the literature.* Value in Health, 2011. **14**(8): p. 1153-1157.

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

77. van Nooten, F.E., et al., *"Back to the future": Influence of beliefs regarding the future on TTO answers.* Health and quality of life outcomes, 2016. **14**(1): p. 1-7.

78. van Nooten, F., et al., *What should we know about the person behind a TTO?* 2018, Springer. p. 1207-1211.

79. Hao, Y., V. Wolfram, and J. Cook, *A structured review of health utility measures and elicitation in advanced/metastatic breast cancer.* Clinicoecon Outcomes Res, 2016. **8**: p. 293-303.

80. Lipman, S., W. Brouwer, and A. Attema, *What is it going to be, TTO or SG?* Health Economics, 2020.

81. Kennedy-Martin, M., et al., *Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines.* The European Journal of Health Economics, 2020. **21**(8): p. 1245-1257.

82. Boye, K.S., et al., *Challenges to time trade-off utility assessment methods: when should you consider alternative approaches?* Expert Rev Pharmacoecon Outcomes Res, 2014. **14**(3): p. 437-50.

83. Arnesen, T. and M. Trommald, *Are QALYs based on time trade-off comparable?–A systematic review of TTO methodologies.* Health economics, 2005. **14**(1): p. 39-53.

84. Oppe, M., et al., *EuroQol protocols for time trade-off valuation of health outcomes.* Pharmacoeconomics, 2016. **34**(10): p. 993-1004.

85. Dolan, P., et al., *The time trade-off method: results from a general population study.* Health economics, 1996. **5**(2): p. 141-154.

86. Lee, H.-J., et al., *Review of valuation methods eliciting preference for health states.* J Health Tech Assess, 2017. **5**(2): p. 69-76.

87. Devlin, N.J., et al., *A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach.* Health Econ, 2011. **20**(3): p. 348-61.

88. Lipman, S.A., A.E. Attema, and M.M. Versteegh, *Correcting for discounting and loss aversion in composite time trade-off.* Health Economics, 2022. **31**(8): p. 1633-1648.

89. Augustovski, F., et al., *Lead versus lag-time trade-off variants: does it make any difference?* The European Journal of Health Economics, 2013. **14**(1): p. 25-31.

90. Lugnér, A.K. and P.F.M. Krabbe, *An overview of the time trade-off method: concept, foundation, and the evaluation of distorting factors in putting a value on health.* Expert Review of Pharmacoeconomics & Outcomes Research, 2020. **20**(4): p. 331-342.

91. Grol-Prokopczyk, H., et al., *Promises and pitfalls of anchoring vignettes in health survey research.* Demography, 2015. **52**(5): p. 1703-1728.

92. Zigler, C.K., et al., *The importance of development standards for anchoring vignettes: an illustrative example from pediatric localized scleroderma quality of life.* Quality of Life Research, 2020. **29**(12): p. 3263-3272.

93. Hansen, T.M., K. Stavem, and K. Rand, *Time trade-off with someone to live for: impact of having significant others on time trade-off valuations of hypothetical health states.* Quality of Life Research, 2022. **31**(4): p. 1199-1207.

94. Angelini, V., et al., *Do Danes and Italians rate life satisfaction in the same way? Using vignettes to correct for individual-specific scale biases.* Oxford bulletin of Economics and Statistics, 2014. **76**(5): p. 643-666.

95. Au, N. and P.K. Lorgelly, *Anchoring vignettes for health comparisons: an analysis of response consistency.* Quality of Life Research, 2014. **23**(6): p. 1721-1731.

96. Wale, J., et al., *Why patients should be involved in health technology assessment.* International Journal of Technology Assessment in Health Care, 2017. **33**(1): p. 1-4.

97. Versteegh, M. and W. Brouwer, *Patient and general public preferences for health states: a call to reconsider current guidelines.* Social Science & Medicine, 2016. **165**: p. 66-74.

98. Mott, D.J., *Incorporating quantitative patient preference data into healthcare decision making processes: is HTA falling behind?* 2018, Springer. p. 249-252.

99. Rowen, D., et al., *International Regulations and Recommendations for Utility Data for Health Technology Assessment.* PharmacoEconomics, 2017. **35**(1): p. 11-19.

100. Ubel, P.A., G. Loewenstein, and C. Jepson, *Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public.* Quality of life Research, 2003. **12**(6): p. 599-607.

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

102. Dakin, H., *Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database.* Health and Quality of Life Outcomes, 2013. **11**(1): p. 151.

103. Gheorghe, A., et al., *Health Utility Values Associated with Surgical Site Infection: A Systematic Review.* Value Health, 2015. **18**(8): p. 1126-37.

104. Mortimer, D. and L. Segal, *Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY-weights.* Medical decision making, 2008. **28**(1): p. 66-89.

105. Balázs, P.G., V. Brodszky, and F. Rencz, *Health utility measurement by time trade-off method in Hungary.* Orvosi hetilap, 2021. **162**(14): p. 542-554.

106. Balázs, P.G., et al., *Time trade-off health state utility values for depression: a systematic review and meta-analysis.* Quality of Life Research, 2022: p. 1-15.

107. Balazs, P. and V. Brodszky, *138 Comparative analysis of patient's direct indirect and dermatology specific utility values in four dermatological diseases.* Journal of Investigative Dermatology, 2022. **142**(12): p. S203.

108. Balázs, P. and V. Brodszky, *EPH195 Psychometric Evaluation of the Hungarian Version of Oxford Capabilities-Mental Health Questionnaire.* Value in Health, 2022. **25**(12): p. S228.

109. Lajos, B., *Egészségügyi gazdaságtani elemzések (Health Economic Analysis)*. 2019: JATEPress Kiadó.

110. Moher, D., et al., *Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement.* Systematic reviews, 2015. **4**(1): p. 1-9.

111. Jakubczyk, M., D. Golicki, and M. Niewada, *The impact of a belief in life after death on health-state preferences: True difference or artifact?* Quality of Life Research, 2016. **25**(12): p. 2997-3008.

112. Arber, M., et al., *Performance of ovid medline search filters to identify health state utility studies.* International Journal of Technology Assessment in Health Care, 2017. **33**(4): p. 472-480.

113. Inotai, A., et al., *Health-related quality of life and utility in patients receiving biological and non-biological treatments in rheumatoid arthritis.* Rheumatology international, 2012. **32**(4): p. 963-969.

114. Rencz, F., et al., *Health state utilities for migraine based on attack frequency: a time trade-off study.* Neurological Sciences, 2015. **36**(2): p. 197-202.

115. Rencz, F., et al., *Valuation of pemphigus vulgaris and pemphigus foliaceus health states: a convenience sample experiment.* British Journal of Dermatology, 2016. **175**(3): p. 593-599.

116. Rencz, F., et al., *Discrepancies between the Dermatology Life Quality Index and utility scores.* Quality of Life Research, 2016. **25**(7): p. 1687-1696.

117. Poór, A.K., et al., *Is the DLQI appropriate for medical decision-making in psoriasis patients?* Archives of Dermatological Research, 2018. **310**(1): p. 47-55.

118. Rencz, F., et al., *Bleeding out the quality-adjusted life years: evaluating the burden of primary dysmenorrhea using time trade-off and willingness-to-pay methods.* Pain, 2017. **158**(11): p. 2259-2267.

119. Péntek, M., et al., *Subjective health expectations of patients with age-related macular degeneration treated with antiVEGF drugs.* BMC geriatrics, 2017. **17**(1): p. 1-9.

120. Hajdu, K., et al., *Patient-assigned health utility values for controlled and uncontrolled pemphigus vulgaris and foliaceus.* Journal of the European Academy of Dermatology and Venereology, 2019. **33**(11): p. 2106-2113.

121. Rencz, F., et al., *Patient and general population values for luminal and perianal fistulising Crohn's disease health states.* The European Journal of Health Economics, 2019. **20**(1): p. 91-100.

122. Longyear, R.L. and K. Kushlev, *Can mental health apps be effective for depression, anxiety, and stress during a pandemic?* Practice Innovations, 2021.

123. Bodden, D.H.M., Y. Stikkelbroek, and C.D. Dirksen, *Societal burden of adolescent depression, an overview and cost-of-illness study.* Journal of affective disorders, 2018. **241**: p. 256-262.

124. Cuijpers, P., *The challenges of improving treatments for depression.* Jama, 2018. **320**(24): p. 2529-2530.

125. Abdin, E., et al., *A comparison of the reliability and validity of SF-6D, EQ-5D and HUI3 utility measures in patients with schizophrenia and patients with depression in Singapore.* Psychiatry research, 2019. **274**: p. 400-408.

126. Kroenke, K. and R.L. Spitzer, *The PHQ-9: a new depression diagnostic and severity measure*. 2002, SLACK Incorporated Thorofare, NJ. p. 509-515.

127. Richter, P., et al., *On the validity of the Beck Depression Inventory. A review.* Psychopathology, 1998. **31**(3): p. 160-8.

128. Maier, W., *The Hamilton Depression Scale and its alternatives: a comparison of their reliability and validity.* Psychopharmacol Ser, 1990. **9**: p. 64-71.

129. Furukawa, T.A., et al., *Imputing missing standard deviations in meta-analyses can provide accurate results.* Journal of clinical epidemiology, 2006. **59**(1): p. 7-10.

130. Borenstein, M., et al., *Introduction to meta-analysis*. 2021: John Wiley & Sons.

131. Banihashemian, M., et al., *Health state utilities for patient's current health from bipolar type I disorder.* J Ment Health Policy Econ, 2018. **21**(1): p. 3-10.

132. Group, C.E.S., *Association between recombinant human erythropoietin and quality of life and exercise capacity of patients receiving haemodialysis.* BMJ: British Medical Journal, 1990: p. 573-578.

133. Doctor, J.N., L.A. Zoellner, and N.C. Feeny, *Predictors of health-related quality-of-life utilities among persons with posttraumatic stress disorder.* Psychiatric Services, 2011. **62**(3): p. 272-277.

134. Center, L.U.M., *The (Cost-) Effectiveness of Mindfulness-training and Cognitive Behavioural Therapy in Adolescents and Young Adults With Deliberate Self Harm (DSH).* 2009.

135. Oldridge, N., et al., *Effects on quality of life with comprehensive rehabilitation after acute myocardial infarction.* The American journal of cardiology, 1991. **67**(13): p. 1084-1089.

136. Oldridge, N., et al., *Economic evaluation of cardiac rehabilitation soon after acute myocardial infarction.* The American journal of cardiology, 1993. **72**(2): p. 154-161.

137. Wells, K.B. and C.D. Sherbourne, *Functioning and utility for current health of patients with depression or chronic medical conditions in managed, primary care practices.* Archives of general psychiatry, 1999. **56**(10): p. 897-904.

138. Tsevat, J., et al., *Health values of patients with bipolar disorder.* Quality of Life Research, 2000. **9**(5): p. 579-586.

139. Voruganti, L.N., et al., *Assessing health utilities in schizophrenia.* Pharmacoeconomics, 2000. **17**(3): p. 273-286.

140. Sherbourne, C.D., et al., *Can utility-weighted health-related quality-of-life estimates capture health effects of quality improvement for depression?* Medical Care, 2001: p. 1246-1259.

141. Isacson, D., K. Bingefors, and L. von Knorring, *The impact of depression is unevenly distributed in the population.* European Psychiatry, 2005. **20**(3): p. 205-212.

142. König, H.-H., et al., *Utility assessment in patients with mental disorders.* Pharmacoeconomics, 2009. **27**(5): p. 405-419.

143. Leykin, Y., L.B. Dunn, and R.F. Muñoz, *The effect of depression on the decision to join a clinical trial.* Journal of consulting and clinical psychology, 2017. **85**(7): p. 751.

144. Sanderson, K., et al., *Using the effect size to model change in preference values from descriptive health status.* Quality of Life Research, 2004. **13**(7): p. 1255-1264.

145. Montejo, Á.L., et al., *Estimation of a multiattribute utility function for the Spanish version of the TooL questionnaire.* Value in Health, 2011. **14**(4): p. 564-570.

146. Papageorgiou, K., et al., *Valuation of depression co-occurring with a somatic condition: feasibility of the time trade-off task.* Health Expectations, 2014. **18**(6): p. 3147-3159.

147. Papageorgiou, K., et al., *Do individuals with and without depression value depression differently? And if so, why?* Quality of Life Research, 2015. **24**(11): p. 2565-2575.

148. Flood, C., et al., *What utility scores do mental health service users, healthcare professionals and members of the general public attribute to different health states? A co-produced mixed methods online survey.* Plos one, 2018. **13**(10): p. e0205223.

149. Nontarak, J., S. Assanangkornchai, and S. Callinan, *Patients' self-reported disability weights of top-ranking diseases in Thailand: Do they differ by socio-demographic and illness characteristics?* International Journal of Environmental Research and Public Health, 2020. **17**(5): p. 1595.

150. Spencer, A., et al., *Do time trade-off values fully capture attitudes that are relevant to health-related choices?* The European Journal of Health Economics, 2019. **20**: p. 559-568.

151. Pullenayegum, E.M., A.S. Pickard, and F. Xie, *Latent class models reveal poor agreement between discrete-choice and time tradeoff preferences.* Medical Decision Making, 2019. **39**(4): p. 421-436.

152. Ariza-Ariza, R., et al., *Assessing utility values in rheumatoid arthritis: a comparison between time trade-off and the EuroQol.* Arthritis Care & Research: Official Journal of the American College of Rheumatology, 2006. **55**(5): p. 751-756.

153. Katanyoo, K., et al., *A comparison of six approaches for measuring utility values among patients with locally advanced cervical cancer.* Expert Review of Pharmacoeconomics & Outcomes Research, 2022. **22**(1): p. 107-117.

154. Van Nooten, F., et al., *"Married with children" the influence of significant others in TTO exercises.* Health and Quality of Life Outcomes, 2015. **13**(1): p. 1-7.

155. Chen, S.C., et al. *A catalog of dermatology utilities: a measure of the burden of skin diseases*. in *Journal of Investigative Dermatology Symposium Proceedings*. 2004. Elsevier.

156. Song, H.J., et al., *Estimation of Health Utilities Based on the Response to Treatment in Atopic Dermatitis: a Population-based Study.* Clinical Therapeutics, 2019. **41**(4): p. 700-713.

157. Shih, T., et al., *Hyperhidrosis treatments in hidradenitis suppurativa: A systematic review.* Dermatol Ther, 2022. **35**(1): p. e15210.

158. Grajdeanu, I.A., et al., *Imaging techniques in the diagnosis and monitoring of psoriasis.* Exp Ther Med, 2019. **18**(6): p. 4974-4980.

159. Martin, L.K., et al., *A systematic review of randomized controlled trials for pemphigus vulgaris and pemphigus foliaceus.* J Am Acad Dermatol, 2011. **64**(5): p. 903-8.

160. Gergely, L.H., et al., *Validity of EQ-5D-5L, Skindex-16, DLQI and DLQI-R in patients with hidradenitis suppurativa.* J Eur Acad Dermatol Venereol, 2020. **34**(11): p. 2584-2592.

161. Koszorú, K., et al., *General and skin-specific health-related quality of life in patients with atopic dermatitis before and during the COVID-19 pandemic.* Dermatitis, 2022. **33**(6S1): p. S92-S103.

162. Richardson, J.R.J., J.R. Mckie, and E.J. Bariola, *Multiattribute utility instruments and their use*, in *Encylopedia of Health Economics, Volume 2*. 2014, Elsevier. p. 341-357.

163. Mukuria, C., et al., *An Updated Systematic Review of Studies Mapping (or Cross-Walking) Measures of Health-Related Quality of Life to Generic Preference-Based Measures to Generate Utility Values.* Appl Health Econ Health Policy, 2019. **17**(3): p. 295-313.

164. Herdman, M., et al., *Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L).* Quality of life research, 2011. **20**: p. 1727-1736.

165. Finlay, A.Y. and G. Khan, *Dermatology Life Quality Index (DLQI)—a simple practical measure for routine clinical use.* Clinical and experimental dermatology, 1994. **19**(3): p. 210-216.

166. Basra, M., et al., *The Dermatology Life Quality Index 1994–2007: a comprehensive review of validation data and clinical results.* British Journal of Dermatology, 2008. **159**(5): p. 997-1035.

167. Davison, N.J., et al., *Generating EQ-5D-3L Utility Scores from the Dermatology Life Quality Index: A Mapping Study in Patients with Psoriasis.* Value Health, 2018. **21**(8): p. 1010-1018.

168. Ruzsa, G., F. Rencz, and V. Brodszky, *Assessment of health state utilities in dermatology: an experimental time trade-off value set for the dermatology life quality index.* Health and Quality of Life Outcomes, 2022. **20**(1): p. 1-19.

169. Bland, J.M. and D. Altman, *Statistical methods for assessing agreement between two methods of clinical measurement.* The lancet, 1986. **327**(8476): p. 307-310.

170. Koo, T.K. and M.Y. Li, *A guideline of selecting and reporting intraclass correlation coefficients for reliability research.* Journal of chiropractic medicine, 2016. **15**(2): p. 155-163.

171. Dolan, P. and J. Roberts, *To what extent can we explain time trade-off values from other information about respondents?* Social Science & Medicine, 2002. **54**(6): p. 919-929.

172. Sayah, F.A., et al., *Determinants of time trade-off valuations for EQ-5D-5L health states: data from the Canadian EQ-5D-5L valuation study.* Qual Life Res, 2016. **25**(7): p. 1679-85.

173. Stefanidou, M., et al., *Willingness to pay and quality of life in patients with pruritic skin disorders.* Archives of Dermatological Research, 2019. **311**(3): p. 221-230.

174. Alinia, C., et al., *The health-related quality of life in Iranian patients with COVID-19.* BMC Infect Dis, 2021. **21**(1): p. 459.

175. de Andrade, L.H., et al., *The impact of social welfare and COVID-19 stringency on the perceived utility of food apps: A hybrid MCDM approach.* Socio-economic planning sciences, 2022. **82**: p. 101299.

176. Gloster, A.T., et al., *Impact of COVID-19 pandemic on mental health: An international study.* PLoS One, 2020. **15**(12): p. e0244809.

177. Helter, T.M., et al., *Does Assessing Outcomes in Terms of Capability for Schizophrenic Patients with Depression Provide more Information Than use of the NICE Recommended QALY?-An Empirical Comparison of the OxCAP-MH, ICECAP-A and EQ-5D-5L Instruments.* 2020.

178. OECD and E. Union, *Health at a Glance: Europe 2022*. 2022.

179. Abbott, R.A., et al., *Normative data from linear and nonlinear quantile regression in CANTAB: Cognition in mid-to-late life in an epidemiological sample.* Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 2019. **11**: p. 36-44.

180. Katumba, K.R., et al., *Cultural and linguistic adaptation of the multi-dimensional OXCAP-MH for outcome measurement of mental health among people living with HIV/AIDS in Uganda: the Luganda version.* Journal of patient-reported outcomes, 2021. **5**(1): p. 1-9.

181. DeVellis, R.F., *Classical test theory.* Medical care, 2006: p. S50-S59.

182. Vergunst, F., et al., *Application of Sen's capability approach to outcome measurement in mental health research: psychometric validation of a novel multi-dimensional instrument (OxCAP-MH).* Human Welfare, 2014. **3**(1).

183. Kroenke, K., R.L. Spitzer, and J.B. Williams, *The PHQ-9: validity of a brief depression severity measure.* Journal of general internal medicine, 2001. **16**(9): p. 606-613.

184. Spitzer, R.L., et al., *A brief measure for assessing generalized anxiety disorder: the GAD-7.* Archives of internal medicine, 2006. **166**(10): p. 1092-1097.

185. Al-Janabi, H., T. N Flynn, and J. Coast, *Development of a self-report measure of capability wellbeing for adults: the ICECAP-A.* Quality of life research, 2012. **21**: p. 167-176.

186. Weinstein, M.C., *A QALY is a QALY is a QALY—or is it?* 1988, North-Holland. p. 289-290.

187. Moore, M.A., A.E. Boardman, and A.R. Vining, *More appropriate discounting: the rate of social time preference and the value of the social discount rate.* Journal of Benefit-Cost Analysis, 2013. **4**(1): p. 1-16.

188. Attema, A.E., W.B. Brouwer, and K. Claxton, *Discounting in economic evaluations.* Pharmacoeconomics, 2018. **36**(7): p. 745-758.

189. Morris, S., et al., *Economic analysis in healthcare*. 2012: John Wiley & Sons.

190. Pettitt, D., et al., *The limitations of QALY: a literature review.* Journal of Stem Cell Research and Therapy, 2016. **6**(4).

191. Rencz, F., et al., *EQ-5D in central and Eastern Europe: 2000–2015.* Quality of Life Research, 2016. **25**: p. 2693-2710.

192. Golicki, D., et al., *Valuation of EQ-5D-5L health states in Poland: the first EQ-VT-based study in Central and Eastern Europe.* Pharmacoeconomics, 2019. **37**: p. 1165-1176.

193. Rupel, V.P., A. Srakar, and K. Rand, *Valuation of EQ-5D-3l health states in Slovenia: VAS based and TTO based value sets.* Slovenian Journal of Public Health, 2020. **59**(1): p. 8-17.

194. Paveliu, M.S., et al., *Estimating an EQ-5D-3L value set for Romania using time trade-off.* International Journal of Environmental Research and Public Health, 2021. **18**(14): p. 7415.

195. Paracha, N., A. Abdulla, and K.S. MacGilchrist, *Systematic review of health state utility values in metastatic non-small cell lung cancer with a focus on previously treated patients.* Health and quality of life outcomes, 2018. **16**(1): p. 1-30.

196. Wright, D.R., et al., *Methods for measuring temporary health states for cost-utility analyses.* Pharmacoeconomics, 2009. **27**: p. 713-723.

197. Attema, A.E., et al., *Time-tradeoff sequences for analyzing discounting and time inconsistency.* Management Science, 2010. **56**(11): p. 2015-2030.

198. Kovács, S., et al., *Should Hungary Pay More for a QALY Gain than Higher-Income Western European Countries?* Applied Health Economics and Health Policy, 2022. **20**(3): p. 291-303.

199. Gulácsi, L., et al., *History of health technology assessment in Hungary.* International journal of technology assessment in health care, 2009. **25**(S1): p. 120-126.

200. Mohiuddin, S. and K. Payne, *Utility values for adults with unipolar depression: systematic review and meta-analysis.* Medical Decision Making, 2014. **34**(5): p. 666-685.

201. Matza, L.S., et al., *Vignette-based utilities: usefulness, limitations, and methodological recommendations.* Value in Health, 2021. **24**(6): p. 812-821.

202. Brazier, J., et al., *Experience-based utility and own health state valuation for a health state classification system: why and how to do it.* The European Journal of Health Economics, 2018. **19**: p. 881-891.

203. Yousefi, M., et al., *Assessing the performance of direct and indirect utility eliciting methods in patients with colorectal cancer: EQ-5D-5L versus C-TTO.* Health Services and Outcomes Research Methodology, 2019. **19**(4): p. 259-270.

204. Liem, Y.S., J.L. Bosch, and M.M. Hunink, *Preference-based quality of life of patients on renal replacement therapy: a systematic review and meta-analysis.* Value in Health, 2008. **11**(4): p. 733-741.

205. Kroneberg, C. and F. Kalter, *Rational choice theory and empirical research: Methodological and theoretical contributions in Europe.* Annual review of sociology, 2012. **38**: p. 73-92.

206. Furnham, A. and H.C. Boo, *A literature review of the anchoring effect.* The journal of socio-economics, 2011. **40**(1): p. 35-42.

207. Kahneman, D., *A perspective on judgment and choice: mapping bounded rationality.* American psychologist, 2003. **58**(9): p. 697.

208. Feng, Y.-S., et al., *Combining EQ-5D-5L items into a level summary score: demonstrating feasibility using non-parametric item response theory using an international dataset.* Quality of Life Research, 2021: p. 1-13.

209. Vilsbøll, A.W., et al., *Mapping Dermatology Life Quality Index (DLQI) scores to EQ-5D utility scores using data of patients with atopic dermatitis from the National Health and Wellness Study.* Quality of Life Research, 2020. **29**: p. 2529-2539.

210. Helter, T.M., et al., *Comparison of capability and health-related quality of life instruments in capturing aspects of mental well-being in people with schizophrenia and depression.* BJPsych Open, 2022. **8**(4): p. e117.

211. Helter, T.M., et al., *Capability instruments in economic evaluations of health-related interventions: a comparative review of the literature.* Quality of Life Research, 2020. **29**: p. 1433-1464.