ANGOL NYELVŰ TÉZISGYŰJTEMÉNY

Balázs Péter

Egészség-hasznosságmérési módszerek összehasonlító elemzése
című PhD értekezéshez

Témavezető:
Prof. Brodszky Valentin

Budapest, 2023

Corvinus University of Budapest
Department of Health Policy

BOOKLET

Comparative analysis of health state utility measurement methods

Candidate:
Péter Balázs

Supervisor:
Prof. Valentin Brodszky

CONTENT

1. BACKGROUND

1.1. Introduction

Health is defined as an infinite demand, which is supplied by the finite health care system. Any health improvement is considered as a societal public good. 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 to inform decision makers − allocative decisions require rich information [1].

Health economics evaluations hold on a primary aim: application of effective redistribution of scarce resources, by quantifying costs and benefits of an intervention to maximize welfare.

Health economic evaluations cover a broad range of systematic analytical methods that are used to compare the effectiveness of various interventions by comparing their cost and outcomes (the four common types are: *cost minimization, cost-benefit, cost-effectiveness* and *cost-utility analysis*). The costs are usually defined as total costs summing direct (associated with the intervention) and indirect (associated with reduced productivity due to health state) costs, while outcomes are measured by health gain (incremental benefit of the intervention) [2].

1.2. Research framework

1.2.1. QALY

Most frequently health economic evaluations apply a standardized health outcome measure called quality-adjusted life year (QALY). The QALY combines two components, (1) *quality of life* – expressed in terms of life years gained/saved and (2) *quality of life* – measured by a health-related quality of life (HRQoL), usually a health state utility (HSU) assessment. One year spent in full health equals one QALY [3, 4].

1.2.2. HRQoL and Capability

To measure the quality of a health status in medicine and economic fields several HRQoL instruments have evolved the past 30 years. HRQoL measures have a particular focus on physical and psychological components 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* domains [5, 6]; while (3) ability to function, well-being and personal relationships are captured by *capability measurements* [7, 8]. Measuring domains beyond health show increasing importance, while HRQoL measures are rather designed to capture physical and psychological domains of health. The capability approach emphasizes the individual's ability to achieve their valued objectives, therefore defined as capable to do things that are important in life.

1.2.3. Direct and indirect HSU measurement

A specific type of HRQoL measurements are the widely used HSU elicitation methods, that solely focus on physical and psychological domains of health. 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 [9].

In *direct HSU* valuation tasks are preference-based measurements that apply rational decision making theory and offer choices between two alternatives for respondents. The tasks requires from respondents to make a trade-off between quality/quantity of life. Assessed health states can be

hypothetical, where vignettes describe the health states, or self-experienced current health state. The *indirect HSU* elicitations are generic or disease specific HRQoL questionnaires that apply multi-attribute utility theory. The multidimensional instrument items cover health domains that are rated on a scale, where total scores are converted into HSU using value sets (country specific weights) [10, 11].

Table 1 - HRQoL measurement typology

<table>
<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</td><td rowspan="2">willingness to pay (WTP)</td><td>generic HRQoL instruments</td><td>disease-group specific instrument</td><td>disease specific instruments</td></tr>
<tr><td>standard gamble</td><td rowspan="3">e.g.: EQ-5D, Health Utility Index (HUI), Short Form-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</td><td rowspan="2">willingness to accept</td></tr>
<tr><td>best-worst scaling</td></tr>
</table>

1.2.4. TTO

The most frequently used direct HSU measurement is the time trade-off (TTO) method [12]. In the conventional 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 fixed time increments (usually 1 year) until the respondent cannot choose between the two health alternatives. 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 HSU. To demonstrate the close relationship between TTO utility and QALY, the example shows that the alternative health states are equal in terms of QALY outcome:

$$8 \text{ years} \times \text{utility of full health } (1.0) = 8 \text{ QALY} = 10 \text{ years} \times \text{utility of moderate depression } (0.8)$$

1.3. Research objective

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. Health economic evaluations use QALY to quantify health gain, that rely on quality of life assessments. Quality components are measured by direct and indirect HSU measurement methods that yield systematically different results, influencing the evaluations.

Differences in *methodological attributes* of HSU measurement instruments (like health state description; responding population, timeframe, task type, iteration process, etc.) and *individual factors* (such as sociodemographic characteristics, disease-specific and clinical characteristics) potentially impact utility results. Recent thesis aims to explore the impact of methodological and individual attributes on HSU. 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.

2. SYSTEMATIC REVIEW OF TTO IN HUNGARY

2.1. Methods of TTO review in Hungary

2.1.1. Study selection

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 [13]. Two international (PubMed, Web of Science) and one domestic (Matarka) online databases were searched. The review selected studies that met three inclusion criteria: (1) original studies, which (2) measure utilities with TTO method, (3) using a Hungarian sample.

Data extraction regarded (i) study characteristics related information of sample size, responding population, mean age, gender proportion and (ii) TTO method attributes as task type, health state description, timeframe, number and order of health states, mode of data administration. HSU as mean, standard deviation, median, inter quartile range, proportion of '1' and '0' responses – were retrieved.

2.1.2. Search strategy

Search terms were combined to construct a "search filter" as suggested for HSU studies [14].

2.2. Results of TTO review in Hungary

2.2.1. Selected studies

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.

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 [15-23]*. 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. All studies used self-administrated data collection mode, the method was paper-based in five cases (71%), once online and one study collected data online and paper based combined.

2.2.2. TTO results of studies

The applied TTO methods were heterogeneous. Conventional TTO type was applied three times, three times indifference in one answer TTO was used, one study choose 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 was applied in two cases, while the in the cTTO task the standard 10-years and 10+10 years format was followed. The smallest tradable unit of the iteration was either 6 month or 1 year. The number of TTO task health states presented in the surveys varied between 1 and 7, most studies fixed, while two randomized the presentation order of the health states.

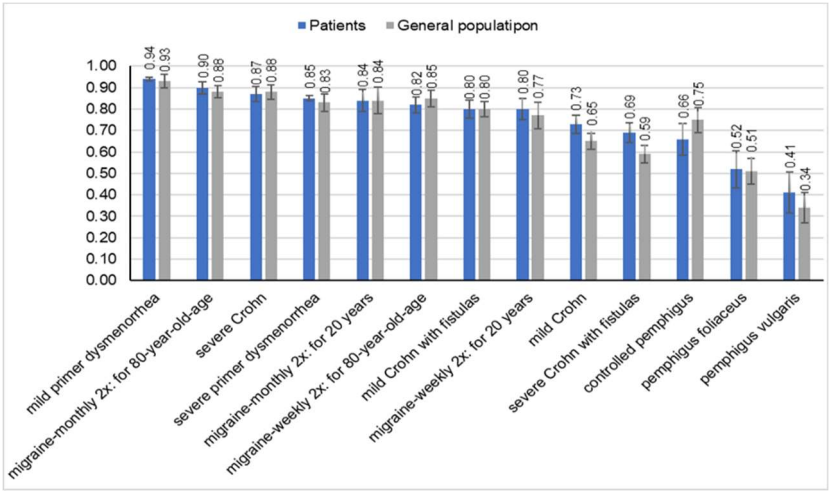
Table 2 - Studies application of TTO methodology in Hungary

Study	TTO type	Health state description	Timeframe	Smallest tradable unit	Order of health states
Inotai et al.[16]	indifferen-ce in one answer	self-experienced	subjective life expectancy	n/a	fixed
Rencz et al.[20]	indifferen-ce in one answer	vignette-based	20 and 80 years	1-year	fixed
Rencz et al. and Hajdu et al.[15, 21]	composite	vignette-based and self-experienced	10 years BTD and 10+10 years WTD	6-month	fixed
Rencz et al. and Poór et al.[18, 19]	conventio-nal	vignette-based and self-experienced	10 years	6-month	random
Péntek et al.[17]	indifference in one answer	self-experienced	subjective life expectancy	n/a	fixed
Rencz et al. [22]	convention-nal	vignette-based	10 years	n/a	fixed
Rencz et al.[23]	convention-nal	vignette-based and self-experienced	10 years	6-month	random

2.2.3. TTO utilities in Hungary

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%. 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 1 - TTO utilities of patients' vs general population



3. REVIEW AND META-ANALYSIS IN DEPRESSION

3.1. Methods of systematic review of TTO studies in depression

3.1.1. Study selection

A systematic literature search was conducted in March 2022 following the principles of PRISMA protocol. Four online databases were searched: PubMed, Web of Science, PsycINFO and Cochrane Database of Systematic Reviews. No language or publication date restrictions were applied. Articles that are (1) empirical studies, (2) TTO studies valuing depression or depressed states and (3) report TTO utility estimates of any sample population were included

3.1.2. Search strategy

The strategy was developed as a combination of TTO and depression search terms: '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 (the original search was conducted in November 2020 and updated in May 2022).

3.1.3. Quality assessment

Measurement and Valuation of Health (MVH) valuation protocol's attributes that regards the TTO task were used for quality assessment of TTO studies: framework, time horizon, anchor state, iteration algorithm, mode of administration, method of data collection and respondent training [24]. 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.

3.1.4. Meta-analysis methods

Random effect (REML) model was used 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). Heterogeneity was tested using T^2 and I^2 and moderator effect of sample, vignettes and health state by meta-regression [25].

3.2. Results

3.2.1. Search results 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 excluded. An additional empirical TTO study was found by hand searching the reference lists of included articles, resulting in 15 included individual studies, but two publications were merged due reporting the same study results. In total, *14 articles* met the inclusion criteria, six fit for the meta-analysis [26-41].

3.2.2. Depression study characteristics

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 (64%). Three studies (21%) included the general population. Study designs were mostly cross-sectional (79%) and the

most frequent data collection methods were semi-structured interviews (57%) or paper-based self-completion questionnaires (29%). The sample sizes varied widely between 32 and 3,986; the mean age of respondents ranged between 32.0 and 52.8 years, the proportions of women was 11.5-93.5%.

3.2.3. TTO quality assessment results

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 (50% and 57% missing).

3.2.4. TTO tasks in depression studies

Majority of studies employed the conventional time trade-off method (64%), three studies (21%) used the simplified indifference in one answer TTO task, one study applied lead-time/lag-time TTO, and one paper did not clarify the TTO type. Most studies used a 10-year timeframe (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. The iteration process was seldom reported, missing in seven cases. Three studies used a single question iteration, two studies used the incremental bottom-up method, another two studies used the ping-pong method, and one study employed top-down steps.

Overall, six out of fourteen studies (43%) employed a vignette-based TTO task describing remitted/no, mild, moderate, and severe hypothetical states of depression. 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. 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. 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. Three descriptions were based on the McSad depression scale, one study used 6 items of the Short Form-12 questionnaire, another one applied a single dimension, single statement description, using the first item of the mental health-specific Tolerability and Quality of Life (TooL) questionnaire.

Table 3 - Depression health state vignette attributes

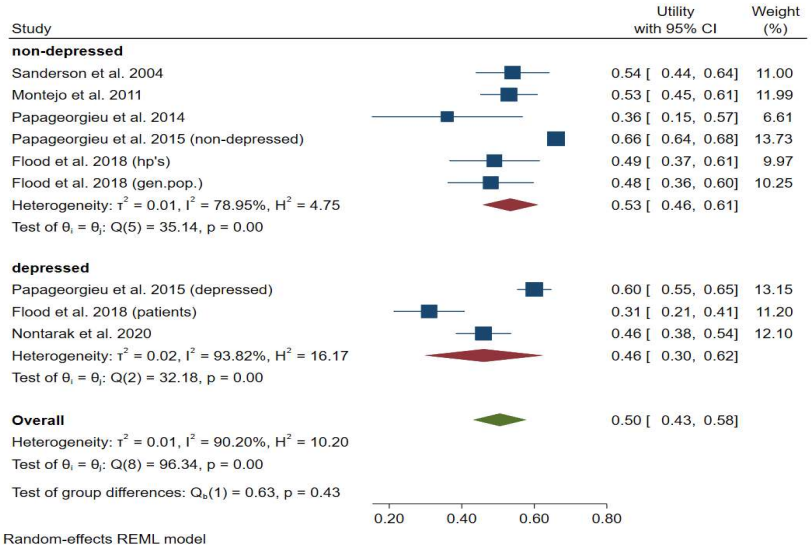
Publication	Dimensions (covered)	Health states (assessed)	Origin of description	Presenting
Sanderson et al. 2004 [37]	emotions, physiology, mood, social relations (4)	remitted, few symptoms, some symptoms, many symptoms (4)	SF-12 questionnaire	scenario
Montejo et al. 2011 [31]	anxiety/depression (1)	severe depression (1)	TooL questionnaire	single statement
Papageorgiou et al. 2014 [35]	emotions, self-appraisal, cognition, physiology, behaviour, role function	mild/severe depression + mild depression co-occurring with moderate or severe cancer/diabetes/heart disease (2+6)	McSad depression scale	statements

	(6)			
Papageorgiou et al. 2015 [36]	emotion, self-appraisal, cognition, physiology, behaviour, role function (6)	no, mild, moderate, severe depression (4)	McSad depression scale	statements
Flood et al. 2018 [27]	self-appraisal, physiology, functioning, emotions, social-relations, usual activities (6)	severe depression (1)	McSad depression scale	scenario
Nontarak et al. 2020 [32]	emotion, usual activities, physiology, cognition (4)	mild, moderate, severe	not disclosed	scenario

3.2.5. Results of the meta-analysis in depression

The REML meta-analysis examined 3 different vignette-based depression health states (mild: n=5; moderate: n=4; severe: n=9). Utility estimates were pooled and compared between the study populations with (n=7) and without (n = 11) depression. The overall pooled mean TTO utilities were 0.75 in mild depression, 0.66 in moderate depression and 0.50 in severe depression. Meta-regression shows that evaluating severe depression state (β= -0.16) and having depressed population sample (β= -0.13) 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 4 - Forest plot of TTO utilities in severe depression



3.2.6. Utility catalogue in depression

Overall, 61 HSU values were extracted from 14 original studies, 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 calculated 25 *vignette-based utilities* for no, mild, moderate and severe depression states, where the values ranged from 0.96 (remitted depression as evaluated by health professionals) to 0.31 (patient perceptions of severe depression).

4. MEASUREMENT AGREEMENT OF UTILITY MEASURES

4.1. Methodology of agreement measurement between direct and indirect HSU measures

4.1.1. Study design

Data of four multi-centre cross sectional surveys conducted between 2015-2021 was collected from atopic dermatitis, hidradenitis suppurativa, pemphigus and psoriasis adult patients in Hungary using paper-based self-completed questionnaires, that 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 comparison of four HSU measures – (i) conventional TTO, (ii) EQ-5D-5L, (iii) mapping-based mDLQI utility and (iv) value set based vDLQI utility – was performed to examine measurement agreement in four dermatological diseases [42-44]. Mean, SD, median and IQR were calculated.

4.1.2. Agreement measure

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 detects proportional bias (practically meaning that one method gives higher/lower values than the other) [45, 46]. 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. ICC indicates poor, moderate and strong agreement if ICC>0.50, 0.51-0.74 and 0.75.

4.1.3. Regression analysis

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. Right-censor was set as upper limit at one, because utility estimates are typically peaked at 1 [47]. The regression equation is the following in all four models:

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

4.2. Agreement results between direct and indirect HSU measurements

4.2.1. Sample characteristics

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 patient sample population was secondary educated (56.3%) and full-time employed (50.1%). Average disease duration was 12.8 (SD=12.6) years, the mean total DLQI score was 9.9 (SD=8.4).

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 utility score of 0.95, followed by EQ-5D-5L with 0.89, vDLQI with 0.84 and mDLQI with 0.79.

4.2.2. Agreement results between TTO, EQ5D and DLQI

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; p<0.001). TTO and mDLQI/vDLQI as well as EQ-5D and mDLQI/vDLQI measures showed presence of systemic proportional bias. Bland-Altman plots supported the presence of disagreements, scores misfit limits of agreement especially at the lower end of the utility scale.

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), but rather poor between TTO and the two DLQI utility measures (ICC= 0.314 and 0.263; p<0.001).

4.2.3. Regression results of factors impacting utilities

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 in literature [48-51].

Table 5 - Regression analysis of factors impacting HSU estimates

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	
uncensored	394		569		753		756	

5. MENTAL CAPABILITY MEASUREMENT

5.1. Methods of mental capability measurement

5.1.1. Study design

A large sample (N=2000) online self-administrated survey was conducted among Hungarian general population in 2021 August. The questionnaire consisted of validated HRQoL instruments (PHQ-9, GAD-7), capability measurement tools (OxCAP-MH, ICECAP-A/O) and sociodemographic questions. The ***OxCAP-MH*** is a validated mental-capability measurement, that consists of 16 items, each scored on a 1-5 Likert scale, where mental capability is expressed on a easy to read standardized 0 to 100 score, where 100 represents the best capabilities, standardized score calculated as [52, 53]:

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

5.1.2. Analysis methods

The psychometric properties of the OxCap-MH were evaluated. Item *reliability* was 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 and other instruments to assess *convergent* validity. Population normative data of OxCap-MH standardized mean 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 [54].

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:

$$OxCAPMH_i = \beta 0 + \beta_i Age + \beta_i Sex + \beta_i Educ + \beta_i Resid + \beta_i Employ + \beta_i Maritals + \beta_i Depr + \gamma_i Anx + \epsilon_i$$

5.2. Results of OxCAP-MH capability measure

5.2.1. Study results

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.

The OxCAP-MH instrument had good reliability (Cronbach α=0.85), no items deleted would improve internal consistency, corrected item-total correlations of items ranged between 0.29-0.64. Out of 16 items, 15 were significantly correlating, except item 9 (influence on local decisions) with item 7 & 8. Construct validity of the item was convincing, total scores significantly ($p<0.005$) differed according to age, education, residence, employment status, marital status, PHQ-9 and GAD-7 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 measure (ICECAP-A: r=0.620) shows fair convergent validity.

5.2.2. Population normative results

The mean OxCAP-MH score in the total sample was 68.5 (SD=14.4), the highest mean scores (75.5) were observed among no depression 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.

Table 6 - HRQoL and capability instrument mean (SD) scores

Sociodemographics	**Subgroups**	**PHQ-9** total score (0-27)	**GAD-7** total score (0-21)	**ICECAP** total score (0-20)	**OxCap-MH** standardized score (0-100)
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)
	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)*
	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)*
	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)*
	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)*
Employ-ment status	full-time employed/entrepren.	*5.6 (5.5)*	*4.6 (4.7)*	*13.9 (2.8)*	*69.3 (14.0)*
	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)*
	married/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)*
	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)*
	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)*

The seven observed explanatory sociodemographic variables included into the model explained 41% of the variation in the mental-capability score, the overall OLS multivariate linear regression model was significant (R^2= 40.9, $p<0.001$). According to the regression result, respondents with older age (β=0.09), living in towns (β=1.32) as compared to villages, 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. Being female (β=-1.96); mild-moderate-severe-extremely severe depression (β=-6.97, -10.78, -14.75, -19.59); and mild-moderate-severe anxiety (β=-4.01, -7.01, -9.82) groups in contrast to asymptotic population had significantly lower OxCAP-MH capability scores, respectively.

6. DISCUSSION

6.1. Answer on research questions

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 estimates in depression.

H2: Sociodemographic characteristics (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.

H3: Direct (TTO) and indirect (EQ-5D-5L, DLQI) HSU measures do yield systematically different results, agreement was only found between TTO and EQ-5D methods.

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

6.2. 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 the measurement agreement found between TTO and EQ-5D-5L assessments are good news. 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 towns, being employed/student and living as married.

The practical implications of the thesis can be considered when applying health economic evaluations to guide decisions on effective allocation of resources. The diversity of HSU measurement can yield systematically different results, that warns to consider alternative (supplementary) approaches or revise previous.

7. REFERENCES

1. Folland, S., A.C. Goodman, and M. Stano, *The economics of health and health care: Pearson new international edition*. 2016: Routledge.
2. Drummond, M.F., et al., *Methods for the economic evaluation of health care programmes*. 2015: Oxford university press.
3. Tsevat, J., *What do utilities measure?* Med Care, 2000. **38**(9 Suppl): p. Ii160-4.
4. Weinstein, M.C., G. Torrance, and A. McGuire, *QALYs: the basics.* Value Health, 2009. **12 Suppl 1**: p. S5-9.
5. 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.
6. Ferrans, C.E., et al., *Conceptual model of health-related quality of life.* Journal of nursing scholarship, 2005. **37**(4): p. 336-342.
7. 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.
8. Simon, J., et al., *Operationalising the capability approach for outcome measurement in mental health research.* Social Science & Medicine, 2013. **98**: p. 187-196.
9. Torrance, G.W., *Utility approach to measuring health-related quality of life.* Journal of chronic diseases, 1987. **40**(6): p. 593-600.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. Rencz, F., et al., *Discrepancies between the Dermatology Life Quality Index and utility scores.* Quality of Life Research, 2016. **25**(7): p. 1687-1696.
20. 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.
21. 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.
22. 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.
23. 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.
24. Oppe, M., et al., *EuroQol protocols for time trade-off valuation of health outcomes.* Pharmacoeconomics, 2016. **34**(10): p. 993-1004.
25. Borenstein, M., et al., *Introduction to meta-analysis*. 2021: John Wiley & Sons.
26. 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.

27. 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.
28. 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.
29. König, H.-H., et al., *Utility assessment in patients with mental disorders.* Pharmacoeconomics, 2009. **27**(5): p. 405-419.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. 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.
37. 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.
38. 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.
39. Tsevat, J., et al., *Health values of patients with bipolar disorder.* Quality of Life Research, 2000. **9**(5): p. 579-586.
40. Voruganti, L.N., et al., *Assessing health utilities in schizophrenia.* Pharmacoeconomics, 2000. **17**(3): p. 273-286.
41. 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.
42. 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.
43. 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.
44. 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.
45. 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.
46. 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.
47. 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.
48. 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.
49. 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.
50. 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.
51. van Nooten, F., et al., *What should we know about the person behind a TTO?* 2018, Springer. p. 1207-1211.

52. 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.
53. Sen, A., *Capability and well-being73.* The quality of life, 1993. **30**: p. 270-293.
54. 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.

8. RELATED OWN PUBLICATIONS

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.

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.

Balázs, P.G. 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.

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