Dissertation
zum Erwerb des Doctor of Philosophy (Ph.D.)
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

From outcome measurement to outcome prediction in patient management: The case for individually valued health-related quality of life

erstellt am
Helmholtz Zentrum München - Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH),
Institut für Gesundheitsökonomie und Management im Gesundheitswesen

vorgelegt von
Manuel Bernhard Huber

aus
Dachau, Germany

am
18. September 2019

Supervisor:	Prof. Dr. Reiner Leidl
Second expert:	Prof. Dr. Rolf Holle
Dean:	Prof. Dr. med. dent. Reinhard Hickel
Date of oral defense:	15.01.2020

Dedicated to my parents

Table of contents

List of abbreviations

AI	Artificial intelligence
EMA	European Medicines Agency
EQ-5D-3L	3-level version of the EQ-5D
EQ-5D-5L	5-level version of the EQ-5D
FDA	U. S. Food and Drug Administration
HRQoL	Health-related quality of life
NHS	National Health Service
OHS	Oxford Hip Score
OKS	Oxford Knee Score
PROMs	Patient-reported outcome measures
SG	Standard gamble
TTO	Time-trade-off
VAS	Visual analog scale

List of publications included in this thesis

Article 1: Huber MB, Felix J, Vogelmann M, et al. Health-Related Quality of Life of the General German Population in 2015: Results from the EQ-5D-5L. *International journal of environmental research and public health*. 2017; 14(4):426.

Article 2: Huber M, Vogelmann M, Leidl R. Valuing health-related quality of life: systematic variation in health perception. *Health and quality of life outcomes*. 2018; 16(1): 156.

Article 3: Huber M, Kurz C, Leidl R. Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning. *BMC Medical Informatics and Decision Making*. 2019; 19(1):3.

1 Introductory Summary

1.1 Relevance

Health-related quality of life (HRQoL) is a multidimensional construct to measure and evaluate the health perception of individuals or groups [1]. Compared to hard outcomes such as survival, HRQoL is more versatile and accounts for the multi-layered nature of health states. Both the European Medicines Agency (EMA) and the U.S. Food and Drug Administration (FDA) recognize the importance of HRQoL for clinical trials [2-4]. Moreover, registries for elective surgery in England and Sweden gather HRQoL on a routine basis [5, 6], while a wide variety of studies include or focus on HRQoL as an outcome of interest [7-13]. As shown in Figure 1, a key word count on PubMed illustrates the increasing importance of HRQoL in the scientific literature since 1994.

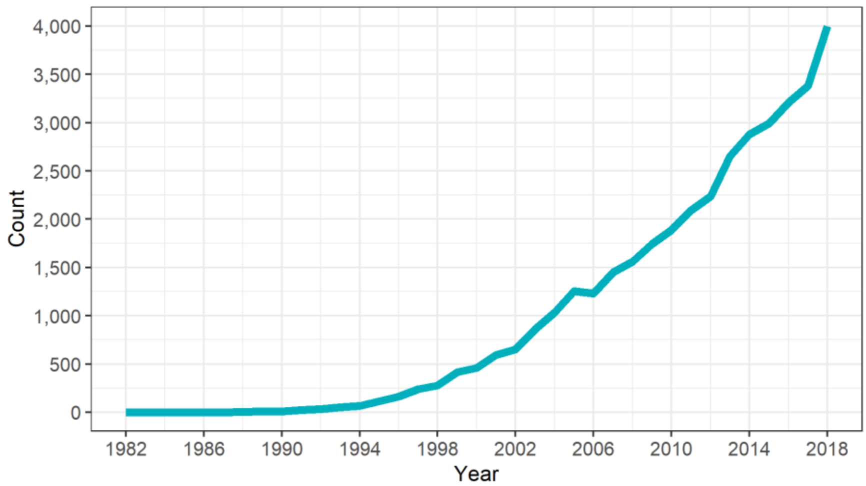

Figure 1 Key word count for 'health-related quality of life' on PubMed

Despite these developments, barriers to implement HRQoL for primary and secondary care exist, and effective improvement of quality of care from the patient perspective is still needed [14].

1.2 Aims

To improve the understanding of health perception and to foster the implementation of HRQoL in clinical practice, this thesis evaluates three aspects of HRQoL research. This is done using three articles, moving from the macro (country-level) to the micro (patient-level) perspective.

By evaluating the results of a cross-sectional survey in Germany from 2015, the aim of the first article is to supply an accurate description of individuals' perceptions of their own health (henceforth, health perception) in a representative sample of the general German population. The article complements previous population studies [15-17]. Component analysis is used to evaluate the influence of valuation and description on health perception changes over the evaluated years. Sociodemographic variables, comorbidities, and other variables are evaluated to identify possible drivers of HRQoL.

The second article connects to the first one by focusing on those survey participants who reported being problem-free. Despite not indicating any problems in the descriptive parts of one HRQoL instrument, many older participants stated significantly lower values for their aggregated health perception [17]. The association between increasing age and reduced health perception is well documented [13–15], but the decline in aggregated health perception among people reporting being otherwise problem-free nevertheless deserves attention, because it reflects discrepancy between what people perceive and what values are reported for further evaluation.

The third article supplements the first two by testing new methods to predict patient-reported outcomes (PROMs) that are based on HRQoL, following hip and knee replacement surgery. It incorporates linear and machine learning based prediction models. The aim is to find the best performing model, measured by respective metrics, and to identify the most important variables. Practical aspects and clinical relevance of PROMs prediction are discussed. The overall goal of the third article is to improve shared decision-making between doctors and patients by enabling better estimation of outcomes following hip and knee replacement surgery.

1.3 Methods

1.3.1 Data

Article 1 and 2 evaluate data from an annual survey in Germany that gathers information about health trends and health status. To include a representative sample of the German population, the survey uses a random-route procedure to select participating households. The survey is of high quality and allows to spot annual changes of overall health status. Article 3 uses routinely gathered registry data for hip and knee replacement surgery from the National Health Service (NHS). Due to its significant number of yearly observations and the included variables, this provider-based dataset is an excellent source to train prediction models. No comparable dataset could be accessed for Germany. Both datasets incorporate validated and widely accepted instruments to measure HRQoL.

1.3.2 Instruments to measure health-related quality of life

Instruments to measure HRQoL can be divided into two main categories: generic and disease-specific [18]. As implied by their names, disease-specific instruments are developed for specific health conditions, to address specific symptoms associated with disease. For example, the Oxford Hip Score (OHS) [19], which is used in Article 3, consists of 12 questions highly related to problems with the hip (e.g. hip pain or the ability to walk). On the other hand, generic instruments are applicable across different populations as well as interventions. They incorporate more general questions, for instance concerning overall pain or the ability to complete daily tasks. A widely accepted and validated generic instrument is the EQ-5D-3L/5L [20], which is used in all three articles in this thesis. The EQ-5D-5L is a newer version of the EQ-5D-3L questionnaire; the former includes only three answer levels for each dimension, while the latter includes five. Both versions consist of five questions, called dimensions, and the visual analog scale (VAS). The five dimensions of the EQ-5D-3L/5L are mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Based on their answers, patients report a health state that is transformed into a five-digit number. For example, 11111 indicates no problems in all five dimensions, while 55555 is the worst possible health state, translating into extreme problems. Survey takers are also asked to mark their current health state on the VAS, which is an overall measure of patient health perception on the day of the survey. It ranges from 0 (worst state) to 100 (best state). The psychometric properties of the EQ-5D-3L/5L have been evaluated across different diseases [21-24], and the questionnaire is used by a wide variety of arthroplasty registries worldwide [25].

1.3.3 Valuing health states

To support allocation decisions in health care and to evaluate HRQoL, health states often need to be valued via value sets. Value sets provide one corresponding index value for every possible health state (e.g. 1 for the problem-free health state of 11111 in the EQ-5D-5L). Otherwise, it would be difficult to differentiate between the same answer levels across different dimensions. For example, are slight problems with walking as inconvenient as slight pain? Value sets can mainly be divided into two categories. To create population-based value sets, people from the general public value hypothetical health states, while to form experience-based value sets, study participants who actually experience respective health states express their perception. Different approaches, such as time-trade-off (TTO) or standard gamble (SG), are used in valuation studies. With TTO, participants waive time for perfect health; with SG, they make probability-based decisions for health status. Anchoring scales [26] and response shift [27] are also important methodological aspects but go beyond the scope of this thesis. Furthermore, since health perception differs across countries, country-specific value sets exist, such as for England [28], Sweden [29], and Germany [26, 30, 31].

There is some discussion concerning whether population perception or patient experience should be used for value sets [32-35]. For example, population-based value sets may not be fully informed, since participants do not experience respective health states [36]. Moreover, convincing arguments exist, to use VAS valuation for cost-utility analysis [37]. Article 2 accounts for this issue by comparing the VAS results of the most prevalent health state – namely being problem-free – with the most commonly assigned values for this health state. Articles 1 and 3 do not incorporate value sets. In addition to the EQ-5D-3L, Article 3 also uses the OHS and Oxford Knee Score (OKS). All three instruments are well accepted, and they are incorporated by respective registries for elective surgery as PROMs [38].

1.3.4 Patient-reported outcome measures

According to the FDA, one inherent feature of PROMs is that they directly measure patients' health perception, without interference from anyone else [39]. PROMs capture the health perception of patients at one or more points in time, such as before and following surgery. PROMs improve quality of care by enabling comparison of service providers, but they also make it possible, among others, to determine efficacy, cost effectiveness, and triggers for surgery [38]. Furthermore, the information provided by PROMs can be used to support shared decision-making between doctors and patients to estimate likely intervention outcomes. By preventing unnecessary or harmful procedures, PROMs can improve health and reduce societal costs [40]. The NHS PROMs Programme and the Swedish quality registers successfully implement routine PROMs collection for elective surgery on a national scale [6]. Both registries use generic as well as disease-specific instruments, and they gather PROMs shortly before and at least 6 or 12 months following surgery. In contrast, despite being among the top four OECD countries where hip and knee replacement surgeries are performed [41], German quality registers do not gather PROMs on a routine basis, nor do they make their data publicly available. Therefore, public NHS PROMs data is used in Article 3. The quality of this data is good but restrictive in some ways (e.g. age groups instead of actual patient age are reported), to conform to data privacy protection standards.

1.3.5 Minimal important differences

Since changes in HRQoL between pre and post intervention can be very small, it is important to know what is relevant for patients. Minimal important differences (MIDs) have been developed to this end. MIDs define changes in HRQoL that are relevant or meaningful to patients [42, 43]. For example, the MID for patients with chronic obstructive pulmonary disease undergoing pulmonary rehabilitation has been estimated to be around 8 points for the EQ-5D-5L VAS [44]. For the EQ-5D-5L index score, MIDS were estimated to range from 0.037 to 00.69, depending on country [45]. MIDs differ by patient population, disease, and instrument. A change that is relevant or meaningful for one patient may not be relevant or meaningful for another. Thus, MIDs must be case specific. Two of the most commonly used methods to calculate MIDs are anchor-based (external indicators to measure change) and distribution-based (statistical distribution of HRQoL in a sample) approaches [46]. The former uses external indicators and the latter uses statistical distributions of observed HRQoL scores to measure relevant change [46]. When suitable, MIDs can also be incorporated from the available literature. In all three articles in this thesis, results are evaluated in reference to respective MIDs. Article 3 also uses MIDs to define the binary states of 'improvement' and 'no improvement' following surgery.

1.3.6 Statistical evaluation

The three articles in this thesis use different evaluation methods, including descriptive analysis, linear approaches, and machine learning. Descriptive methods – used in all three articles of this thesis – describe characteristics or summarize overall findings such as age distributions, while linear regression evaluates the relationship between one or more variables and a response variable. Linear models are very fast and can be interpreted easily, but machine learning sometimes offers better predictive performance, mostly at the cost of explanatory insight [47]. Interestingly, despite being highly popular lately, many machine learning algorithms can be dated back several decades [48, 49]. Machine learning is part of artificial intelligence (AI), which covers algorithms that enable computers to learn from experience [50]. Article 3 benchmarks linear against machine learning models. Because there is no free lunch in optimization, no model works best for all problems and comparison becomes necessary to reach maximal preditive performance [51]. To better understand what is referred to by machine learning in this context, it should be noted that machine learning can be divided into three main paradigms: supervised machine learning, unsupervised machine learning, and reinforcement learning [52]. Of these three categories, supervised machine learning is the most prevalent. It works by mapping predictors to pre-defined outcomes, for example, when preoperative variables are used to predict postoperative outcomes. Unsupervised machine learning, on the other hand, is mainly used for pattern detection, when data is not labeled and no pre-defined outcomes exist. The third paradigm, reinforcement learning, tries to maximize rewards by optimizing sequential behavior. Recent popular examples of AI incorporating reinforcement learning are AlphaGo [53], its successor AlphaGo Zero [54], and the follower Alpha Zero [55]. These AIs have not only beat the world's best human players in specific games, but they have also become increasingly efficient at beating other world-leading AIs in decreasing amounts of time [54]. However, given the research aims and data, only supervised machine learning is considered in this thesis.

1.4 Results summary

Article 1 reports stability in the health perception of the general German population over the last years. Age, depression, heart disease, and diabetes contribute significantly to reductions in perceived health. A slight increase in health perception compared with 2012 can be traced back to a higher number of participants in the problem-free health state. Compared with other countries, like Spain, England, and Canada, health perception in Germany (mean VAS of 85.1) was and is on a high level.

Next, Article 2 identifies systematic variation in health perception among the EQ-5D-5L study population, and especially among older participants with comorbidities. Diabetes and obesity are significantly associated with reporting lower VAS scores in the problem-free health state. The basic principle of current value sets – assigning one value for one health state – partly fails to reflect this patient/participant heterogeneity, since perceived health status may go beyond what is captured in the descriptive system of the EQ-5D-5L.

Finally, Article 3 delivers benchmark results for several machine learning and linear patient outcome prediction models following hip and knee replacement surgery. Overall, some machine learning classifiers, such as extreme gradient boosting, mostly outperform linear alternatives, but the margin is small. Predicting generic improvement (VAS) is easier than predicting disease-specific improvement (OHS/OKS). Preoperative generic and disease-specific HRQoL are the most important predictors for the corresponding postoperative outcome. Benchmarking of different models is needed to maximize predictive performance and to find the optimal prediction model.

Detailed results are provided in the corresponding section of each article.

1.5 Conclusion and outlook

Moving from the macro to the micro (patient-level) perspective, this dissertation illustrates the importance of measuring and evaluating HRQoL for patient management. Routine gathering of HRQoL for representative samples of the general German population is necessary to identify trends and to measure the effects of health policy decisions. Reductions in perceived health should be counteracted. Moreover, with an increasing patient focus, health perception and individual valuation decisions should be discussed more openly. Finally, to foster shared decision-making between doctors and patients, accurate prediction models are needed, and machine learning offers new and interesting ways to build them. Recent technological and methodological advances as well as significant cost increases make it important to evaluate what patients want – namely, to improve their individual health state.

2 Article 1: Health-Related Quality of Life of the General German Population in 2015: Results from the EQ-5D-5L

Reference: Huber MB, Felix J, Vogelmann M, et al. Health-Related Quality of Life of the General German Population in 2015: Results from the EQ-5D-5L. *International journal of environmental research and public health*. 2017; 14(4):426.

DOI: 10.3390/ijerph14040426

3 Article 2: Valuing health-related quality of life: systematic variation in health perception

Reference: Huber M, Vogelmann M, Leidl R. Valuing health-related quality of life: systematic variation in health perception. *Health and quality of life outcomes*. 2018; 16(1): 156.

DOI: 10.1186/s12955-018-0986-8

4 Article 3: Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning

Reference: Huber M, Kurz C, Leidl R. Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning. *BMC Medical Informatics and Decision Making*. 2019; 19(1):3.

DOI: 10.1186/s12911-018-0731-6

5 References

1. Testa MA, Simonson DC: **Assessment of Quality-of-Life Outcomes**. *New England Journal of Medicine* 1996, **334**(13):835-840.
2. European Medicines Agency: **Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQoL) measures in the evaluation of medicinal products**. In., edn; 2005.
3. U. S. Department of Health: **Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance**. *Health and quality of life outcomes* 2006, **4**(1):79.
4. Hines PA, Guy RH, Humphreys AJ, Papaluca-Amati M: **The European Medicines Agency's goals for regulatory science to 2025**. *Nature reviews Drug discovery* 2019, **18**(6):403-404.
5. Devlin NJ, Parkin D, Browne J: **Patient-reported outcome measures in the NHS: new methods for analysing and reporting EQ-5D data**. *Health economics* 2010, **19**(8):886-905.
6. Prodinger B, Taylor P: **Improving quality of care through patient-reported outcome measures (PROMs): expert interviews using the NHS PROMs Programme and the Swedish quality registers for knee and hip arthroplasty as examples**. *BMC health services research* 2018, **18**(1):87.
7. Wick W, Gorlia T, Bendszus M, Taphoorn M, Sahm F, Harting I, Brandes AA, Taal W, Domont J, Idbaih A *et al*: **Lomustine and Bevacizumab in Progressive Glioblastoma**. *New England Journal of Medicine* 2017, **377**(20):1954-1963.
8. Donovan JL, Hamdy FC, Lane JA, Mason M, Metcalfe C, Walsh E, Blazeby JM, Peters TJ, Holding P, Bonnington S *et al*: **Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer**. *New England Journal of Medicine* 2016, **375**(15):1425-1437.
9. von Minckwitz G, Procter M, de Azambuja E, Zardavas D, Benyunes M, Viale G, Suter T, Arahmani A, Rouchet N, Clark E *et al*: **Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer**. *New England Journal of Medicine* 2017, **377**(2):122-131.
10. Hennessy CH, Moriarty DG, Zack MM, Scherr PA, Brackbill R: **Measuring health-related quality of life for public health surveillance**. *Public Health Rep* 1994, **109**(5):665-672.
11. Burström K, Johannesson M, Diderichsen F: **Swedish population health-related quality of life results using the EQ-5D**. *Quality of Life Research* 2001, **10**(7):621-635.
12. Leow MK-S, Griva K, Choo R, Wee H-L, Thumboo J, Tai ES, Newman S: **Determinants of Health-Related Quality of Life (HRQoL) in the Multiethnic Singapore Population – A National Cohort Study**. *PLOS ONE* 2013, **8**(6):e67138.
13. Sonneveld P, Verelst SG, Lewis P, Gray-Schopfer V, Hutchings A, Nixon A, Petrucci MT: **Review of health-related quality of life data in multiple myeloma patients treated with novel agents**. *Leukemia* 2013, **27**:1959.
14. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S: **Patient reported outcome measures in practice**. *BMJ : British Medical Journal* 2015, **350**:g7818.
15. Mielck A, Vogelmann M, Schweikert B, Leidl R: **[Health status of adults in Germany: results from a representative survey using the EuroQol 5D (EQ-5D)]**. *Gesundheitswesen* 2010, **72**(8-9):476-486.
16. Mielck A, Vogelmann M, Leidl R: **Health-related quality of life and socioeconomic status: inequalities among adults with a chronic disease**. *Health and quality of life outcomes* 2014, **12**:58.
17. Huber MB, Reitmeir P, Vogelmann M, Leidl R: **EQ-5D-5L in the General German Population: Comparison and Evaluation of Three Yearly Cross-Section Surveys**. *International journal of environmental research and public health* 2016, **13**(3).

18. Fayers PM, Hays R, Hays RD: **Assessing Quality of Life in Clinical Trials: Methods and Practice**: Oxford University Press; 2005.
19. Dawson J, Fitzpatrick R, Carr A, Murray D: **Questionnaire on the perceptions of patients about total hip replacement**. *J Bone Joint Surg Br* 1996, **78**(2):185-190.
20. **EuroQol--a new facility for the measurement of health-related quality of life**. *Health policy (Amsterdam, Netherlands)* 1990, **16**(3):199-208.
21. Stark RG, Reitmeir P, Leidl R, Konig HH: **Validity, reliability, and responsiveness of the EQ-5D in inflammatory bowel disease in Germany**. *Inflamm Bowel Dis* 2010, **16**(1):42-51.
22. Konerding U, Elkhuizen SG, Faubel R, Forte P, Malmstrom T, Pavi E, Janssen MF: **The validity of the EQ-5D-3L items: an investigation with type 2 diabetes patients from six European countries**. *Health and quality of life outcomes* 2014, **12**:181.
23. Obradovic M, Lal A, Liedgens H: **Validity and responsiveness of EuroQol-5 dimension (EQ-5D) versus Short Form-6 dimension (SF-6D) questionnaire in chronic pain**. *Health and quality of life outcomes* 2013, **11**:110.
24. Hunger M, Sabariego C, Stollenwerk B, Cieza A, Leidl R: **Validity, reliability and responsiveness of the EQ-5D in German stroke patients undergoing rehabilitation**. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2012, **21**(7):1205-1216.
25. Rolfson O, Eresian Chenok K, Bohm E, Lübbeke A, Denissen G, Dunn J, Lyman S, Franklin P, Dunbar M, Overgaard S *et al*: **Patient-reported outcome measures in arthroplasty registries**. *Acta Orthop* 2016, **87 Suppl 1**:3-8.
26. Leidl R, Reitmeir P: **An Experience-Based Value Set for the EQ-5D-5L in Germany**. *Value in Health* 2017(In Press, Corrected Proof).
27. Rapkin BD, Schwartz CE: **Advancing quality-of-life research by deepening our understanding of response shift: a unifying theory of appraisal**. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2019.
28. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B: **Valuing health-related quality of life: An EQ-5D-5L value set for England**. *Health economics* 2018, **27**(1):7-22.
29. Burstrom K, Sun S, Gerdtham UG, Henriksson M, Johannesson M, Levin LA, Zethraeus N: **Swedish experience-based value sets for EQ-5D health states**. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2014, **23**(2):431-442.
30. Ludwig K, Graf von der Schulenburg JM, Greiner W: **German Value Set for the EQ-5D-5L**. *PharmacoEconomics* 2018, **36**(6):663-674.
31. Leidl R, Reitmeir P: **A value set for the EQ-5D based on experienced health states: development and testing for the German population**. *PharmacoEconomics* 2011, **29**(6):521-534.
32. Dolan P: **Thinking about it: thoughts about health and valuing QALYs**. *Health economics* 2011, **20**(12):1407-1416.
33. Brazier J, Akehurst R, Brennan A, Dolan P, Claxton K, McCabe C, Sculpher M, Tsuchyia A: **Should patients have a greater role in valuing health states?** *Applied health economics and health policy* 2005, **4**(4):201-208.
34. Brazier J, Rowen D, Karimi M, Peasgood T, Tsuchiya A, Ratcliffe J: **Experience-based utility and own health state valuation for a health state classification system: why and how to do it**. *Eur J Health Econ* 2018, **19**(6):881-891.
35. Dolan P: **NICE should value real experiences over hypothetical opinions**. *Nature* 2009, **462**(7269):35-35.
36. Karimi M, Brazier J, Paisley S: **Are preferences over health states informed?** *Health and quality of life outcomes* 2017, **15**:105.
37. Parkin D, Devlin N: **Is there a case for using visual analogue scale valuations in cost-utility analysis?** *Health economics* 2006, **15**(7):653-664.

38. Kingsley C, Patel S: **Patient-reported outcome measures and patient-reported experience measures**. *BJA Education* 2017, **17**(4):137-144.
39. **Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance**. *Health and quality of life outcomes* 2006, **4**:79.
40. Hansson-Hedblom A, Jonsson E, Fritzell P, Hagg O, Borgstrom F: **The Association Between Patient Reported Outcomes of Spinal Surgery and Societal Costs: A Register Based Study**. *Spine* 2019.
41. OECD: **Health at a Glance 2017**; 2017.
42. Jaeschke R, Singer J, Guyatt GH: **Measurement of health status. Ascertaining the minimal clinically important difference**. *Controlled clinical trials* 1989, **10**(4):407-415.
43. McGlothlin AE, Lewis RJ: **Minimal Clinically Important Difference: Defining What Really Matters to PatientsMinimal Clinically Important DifferenceJAMA Guide to Statistics and Methods**. *JAMA* 2014, **312**(13):1342-1343.
44. Della Patrona S, Zanini A, Aiello M, Adamo D, Casale S, Cherubino F, Raimondi E, Zampogna E, Chetta A, Spanevello A: **Estimation of minimum clinically important difference in EQ-VAS score after pulmonary rehabilitation in COPD patients**. *European Respiratory Journal* 2014, **44**(Suppl 58).
45. McClure NS, Sayah FA, Xie F, Luo N, Johnson JA: **Instrument-Defined Estimates of the Minimally Important Difference for EQ-5D-5L Index Scores**. *Value in Health* 2017, **20**(4):644-650.
46. Revicki D, Hays RD, Cella D, Sloan J: **Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes**. *Journal of clinical epidemiology* 2008, **61**(2):102-109.
47. Shmueli G: **To Explain or to Predict?** *Statist Sci* 2010, **25**(3):289-310.
48. Cortes C, Vapnik V: **Support-vector networks**. *Machine Learning* 1995, **20**(3):273-297.
49. McCulloch WS, Pitts W: **A logical calculus of the ideas immanent in nervous activity**. *The bulletin of mathematical biophysics* 1943, **5**(4):115-133.
50. Mitchell TM: **Machine Learning**: McGraw-Hill; 1997.
51. Wolpert DH, Macready WG: **No free lunch theorems for optimization**. *IEEE transactions on evolutionary computation* 1997, **1**(1):67-82.
52. Jordan MI, Mitchell TM: **Machine learning: Trends, perspectives, and prospects**. *Science* 2015, **349**(6245):255-260.
53. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M *et al*: **Mastering the game of Go with deep neural networks and tree search**. *Nature* 2016, **529**:484.
54. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A *et al*: **Mastering the game of Go without human knowledge**. *Nature* 2017, **550**:354.
55. Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, Lanctot M, Sifre L, Kumaran D, Graepel T *et al*: **A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play**. *Science* 2018, **362**(6419):1140-1144.

6 Acknowledgements

My sincere gratitude to my supervisor, Prof. Dr. Reiner Leidl, who supported me on this long journey and who made this dissertation possible. His substantial experience in health economics and health care management greatly contributed to the quality of this research. I cannot imagine a better advisor and mentor.

Prof. Dr. Rolf Holle and Prof. Dr. Eva Grill, my other two thesis advisory committee members, also deserve extensive gratitude. They initiated relevant and constructive discussion and gave highly viable methodological advice. Their support goes beyond the articles of this dissertation.

Moreover, I would like to thank all my co-authors and colleagues at IGM, especially Christoph, Christian, Boglárka and Julia.

7 Affidavit

I hereby declare, that the submitted thesis entitled:

From outcome measurement to outcome prediction in patient management: The case for individually valued health-related quality of life

is my own work. I have only used the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

Munich, 15.01.2020

Place, date

Manuel Huber

Signature

8 Confirmation of congruency

I hereby declare, that the electronic version of the submitted thesis entitled:

From outcome measurement to outcome prediction in patient management:

The case for individually valued health-related quality of life

is congruent with the printed version both in content and format.

Munich, 15.01.2020 Manuel Huber

Place, date Signature